Statistical Extreme Value Analysis of JFK Taxiway Centerline Deviations for 747 Aircraft

Prepared under a

Cooperative Research and Development Agreement Between
The Federal Aviation Administration (FAA) and The Boeing Company

Statistical Extreme Value Analysis of JFK Taxiway Centerline Deviations for 747 Aircraft*

Summary

This report describes the analysis of 747 taxiway centerline deviation data that were collected from 6/24/1999 to 2/17/2000 at New York's John F. Kennedy International Airport (JFK). Deviations were measured for nose and main gear at two laser locations for each of two parallel 75 ft straight taxiway segments with shoulder called ALPHA and BRAVO, respectively.

The discrimination of 747 aircraft was mainly based on the landing gear geometry, i.e., the 76.54 ft longitudinal distance between nose and main gear and the outer to outer main gear tire width of 41.33 ft. The raw data was processed and filtered using various consistency criteria after which 2518 cases remained, giving us a centerline deviation for nose and main gear at each of the two lasers in each case.

The purpose of the data collection was to provide a basis for understanding the extreme behavior of such centerline deviations. This behavior as observed from the roughly 2,500 deviation events is extrapolated to more extreme deviation levels as they could be encountered maybe once during much higher numbers of event exposures, e.g., during 10^6 or 10^7 such events.

This somewhat vague statement concerning the number of occurrences of extreme and rare events will benefit from some further explanation. This should also clarify the kind of precision one can expect in statements about such rare events. If the extreme event has chance $p = 10^{-6}$ of occurring and if we focus on a sequence of $n = 1/p = 10^6$ events we can ask how likely it is that we see $k = 0, 1, 2, 3, 4, 5, 6, \ldots$ extreme events among them. The respective probabilities for these counts are .3679, .3679, .1839, .0613, .0153, .0031, .0005, with the remaining counts having chance $< .000084$. Although one would expect 1 such rare event in n such trial events the probabilities themselves make it clear that there could be quite some variability in such a count, with most likely and equally probable counts of 0 and 1 and half that chance of seeing 2 extreme events.

*Data gathered and provided by the FAA. Analysis and report prepared by Fritz Scholz, The Boeing Company, October 13, 2003.

The calculation of the above probabilities for extreme event counts is relatively insensitive to how small p is, as long as it is small, say $p < .01$. All that is additionally required is the link between p and n through $n = 1/p$ so that $np = 1$.

The methodology [9] that extrapolates the extreme deviation behavior observed during 2518 events to much higher levels of event exposures (say $n = 10^6$) had previously been applied by Andrew Booker [1] to similar data collected from Schipol Airport at Amsterdam and more recently in [10] to such data collected at Anchorage International Airport (ANC).

The cleaned data were subjected to various consistency checks, such as plotting nose gear versus main gear deviations at the same laser for each event and plotting deviations of same gears at the two lasers against each other. These plots showed great consistency between deviations for the same events. The main gear deviations at Laser 1 from each taxiway centerline became the focus for further extrapolation analysis since Laser 1 was farther away from any turn-offs. However, no obvious effects of turning maneuvers were evident, in contrast to the experience observed with the ANC data. The main gear deviations were chosen because that is where the exposure to collision and leaving the taxiway is greatest and most relevant. The reason for not combining nose and main gear deviations at the same laser or main gear deviations at the two lasers is that they are highly correlated and "doubling" or "quadrupling" the sample size this way would be quite misleading. Furthermore, the methodology in [9] and in much of the extreme value literature is based on a sample of independent measurements.

The distributions of the selected main gear deviations were then examined by heading and taxiway for symmetry around zero and for possible pooling into one data set. It turned out that there were biases, consistently away from the lasers, but more so for one heading than for the other. As in the ANC study [10] we isolated two forms of bias, a parallax bias, which would change sign with the heading, and another bias that would not change sign with the heading. The parallax bias, presumably resulting from the pilot sitting to one side of the aircraft centerline (also the longitidinal aircraft axis), was estimated to be .04 ft at ALPHA and as .10 ft at BRAVO. These should be seen as average offsets over the 1428 and 1090 events, respectively.

The other bias, was estimated as 0.66 ft at ALPHA and 0.24 ft at BRAVO. The latter biases were much less consistent with each other than observed during the ANC study. Possible reasons for this different bahavior was the different offset of centerlights from the centerline at JFK (21 inches) as opposed to that at ANC (12 inches) and the fact that the centerlights were on the far side from the lasers at ALPHA and on the near side at BRAVO. In the ANC study [10] the centerlights were on the far side for either taxiway. However, a clear resolution of this difference in bias behavior is not apparent.

After subtracting the biases from the deviation data they were again examined for symmetry around zero and to see whether data for different headings and different taxiways could be combined. Having resolved these questions in the affirmative, mostly based on a visual graphical basis, the data were pooled into one bias corrected sample of 2518 deviations. The range of this set of deviations is $[-8.63, 7.53]$ ft.

This sample was then examined to determine how many (k) of the extremes should reasonably be used for extrapolation purposes. This determination is a subtle process and requires a fair amount of judgment. It is not yet automated and will be the focus of a future research review so that it can become an automated part of the extrapolation process given in [9]. After that has been done the data will be reanalyzed some time in the future.

This determination of k was done for the positive, negative, and absolute deviations separately and the extrapolations were carried out for each case, although the results for the absolute deviations were used for the final risk/deviation threshold answers. The analyses for separate positive and negative deviations were carried out to examine consistency, as it was done in the ANC study.

The extrapolation deviation thresholds for various risk levels were then back-corrected for the previously subtracted biases, the parallax and the other bias. Rather than having different back-corrections for different taxiways we averaged the parallax biases over the two taxiways and for the other bias we chose to take the larger of the two biases, namely 0.66 ft, to represent that bias back-correction conservativley. As a result the overall bias back-correction consists of the sum of .07 ft and 0.66 ft which was rounded up to .75 ft to agree with the result from the ANC study.

This bias back-correction will usually be a conservative step since adding the maximum bias correction of .75 ft to the derived thresholds will mainly be relevant for a particular aircraft heading and a particular side of the taxiway. These back-adjusted thresholds are given in Table A for various risk levels. These are deviation thresholds for 747 centerline deviations at the main gear location from the taxiway centerline.

Table A: Back-Adjusted Thresholds by Exceedance Risk
for 747 Centerline Deviations from Taxiway Centerline

From JFK Data	two-sided exceedance risk				
$n = 2518$	$2 \cdot 10^{-7}$	$2 \cdot 10^{-6}$	$2 \cdot 10^{-5}$	$2 \cdot 10^{-4}$	$2 \cdot 10^{-3}$
estimate	15.49 ft	13.70 ft	11.80 ft	9.78 ft	7.62 ft
95% upper bounds	16.81 ft	14.85 ft	12.75 ft	10.52 ft	8.14 ft

The reverse process of establishing risk levels for given deviation thresholds is also presented. It was used to calculate the estimated risk levels for the outer main gear tire edge to exceed the edge of the taxiway for various taxiway widths recommended by the FAA and ICAO design standards. These estimated risks and corresponding 95% upper confidence bounds are presented in Table B. A distance of 41.33 ft (47.03 ft, 52.49 ft) for the outer-to-outer main gear tire edge was assumed for the 747 (A380, NLA) main gear geometry. Using the

A380 and NLA (New Large Airplane) main gear dimension is meant to show the change in risk if a 747 had these main gear widths. This assumes that this change in main gear width would not affect the behavior of the 747 centerline deviations from the taxiway centerline. This could be the case if the deviation path is a parallel offset from the taxiway centerline. However, if the aircraft path is of sinusoidal form then the greater distance ($L_2 = 93.34$ ft) between nose and main gear centroids for the A380 gear geometry (as compared to the corresponding distance of $L_1 = 78$ ft for the 747) has a 20% greater dampening effect on the main gear oscillations. Not knowing how much of the deviation path is made up of a parallel offset and how much of it is sinusoidal variation and with what amplitude makes a clear comparison difficult. No such assessment is possible for the NLA since the corresponding L_3 is not available.

The extremely low risks (of order 10^{-13} and smaller) in the last column of Table B should be interpreted with caution. Most likely these small risks are understated as a result of using an extreme value index that is negative. The extreme value index plays a crucial role in describing the extreme tail behavior of a distribution. A negative value for this index implies that there is a hard threshold beyond which deviations would not be seen. This negative extreme value index may result from the relatively small sample size of $n = 2518$ and the fact that the taxiway edge may have induced some taxiway deviation behavior in the data that appears limiting. However, it would be a stretch to call the taxiway edge a hard limit. A more convincing hard limit might be a wall.

Table B: Two-Sided Exceedance Risks for 747

Risk of Outer Main Gear Tire Edge Exceeding the Taxiway Edge
for Different Taxiway Width Standards, Using 747, A380 & NLA Main Gear Dimensions.
Does Not Compensate for the Different Nose to Main Gear Distance for the A380 or NLA.

Main Gear Dimensions	Taxiway Width	75 ft	82 ft	100 ft
from 747	Estimated Risk	$3.17\ 10^{-8}$	$1.62\ 10^{-10}$	$1.17\ 10^{-18}$
	95% Upper Bound on Risk	$1.94\ 10^{-7}$	$2.13\ 10^{-9}$	$5.73\ 10^{-16}$
from A380	Estimated Risk	$1.41\ 10^{-6}$	$1.26\ 10^{-8}$	$1.34\ 10^{-15}$
	95% Upper Bound on Risk	$5.24\ 10^{-6}$	$8.74\ 10^{-8}$	$1.42\ 10^{-13}$
from NLA	Estimated Risk	$3.79\ 10^{-5}$	$5.25\ 10^{-7}$	$3.88\ 10^{-13}$
	95% Upper Bound on Risk	$9.50\ 10^{-5}$	$2.21\ 10^{-6}$	$1.39\ 10^{-11}$

These exceedance risks are pointwise, i.e., they refer to the exceedance at one particular prespecified point along a straight taxiway. Such pointwise assessments may be relevant when dealing with the risk of collision with some stationary structure, but the risk would have to be recalculated to take into account the distance of the structure from the taxiway centerline and the wingspan of the 747.

This pointwise exceedance risk is not the same as the lengthwise risk of the outside main gear tire edge exceeding the taxiway width somewhere along the length of the taxiway. This risk is bound to be higher than the pointwise risk. This lengthwise risk is currently out of reach since data were collected at just two points along the taxiway. The measurement design used at Frankfurt airport does address this point to some extent and data from this collection effort may be useful in answering this broader and more relevant question of lengthwise risk.

Having analyzed both the data underlying the ANC study and those of the current study a comparison of the deviation experiences offers itself as a natural step. It turned out that the bias corrected deviations from either study showed the same distributional character except for the fact that the JFK deviations appeared to spread out more by a factor of 1.1 (1.097 to be precise). This allowed pooling both data sets into one after dividing the JFK bias corrected deviations by 1.1. The resulting pooled data set then comprised $12314 = 9796 + 2518$ deviations which again were subjected to the extrapolation methodology.

Even though a larger sample was available than for either study and should lead to more trustworthy results we have to present these in differentiated form for JFK and ANC since for JFK the deviation thresholds need to be scaled up by the factor 1.1. The results corresponding to Table A and B are given in Tables C and D, respectively. Note that the estimates from the first risk column in Table C relate to each other as follows: $17.89 = (16.37 - .75) \times 1.097 + .75$.

Since the results reported here are based on data collected for 747s at JFK on 75 ft straight taxiway segments with shoulders the question arises whether the findings carry over to other aircraft and airports, to other taxiway widths, and to taxiways without shoulders. The comparison with the ANC data showed that the distributional character was the same except for a scaling factor. This did not prevent us from pooling the data and thus getting more assured results out of them. However, the final results had to be presented in differentiated form. We expect similar experiences with data from other airports. One needs to have several studies of this form to understand to what extent differences play out. All of the mentioned factors could reasonably affect the deviation behavior in some way. The importance of the airport may be less obvious, so we make it clear. At ANC the 747 traffic consisted mainly of freighters while at JFK that may not be the case. Piloting freighters may be different from piloting passenger planes. Also, at JFK the centerlights are offset from the taxiway centerline by about 18 to 21 inches while at ANC the offset is about 12 inches.

The usual cautionary reminder is given with respect to the risk extrapolations presented here. There may be additional extreme value behavior that has not yet had a chance to manifest itself and that may exhibit more extreme behavior than indicated by the observed linear pattern used for extrapolation. On the other hand there may be a natural feedback

Table C: Back-Adjusted Thresholds by Exceedance Risk
for 747 Centerline Deviations from Taxiway Centerline

Based on Analysis of Combined ANC/JFK Data

$n = 12314$	two-sided exceedance risk				
	$2 \cdot 10^{-7}$	$2 \cdot 10^{-6}$	$2 \cdot 10^{-5}$	$2 \cdot 10^{-4}$	$2 \cdot 10^{-3}$
Applied to ANC Context	Thresholds				
estimate	16.37 ft	13.72 ft	11.24 ft	8.94 ft	6.79 ft
95% upper bounds	17.10 ft	14.31 ft	11.70 ft	9.27 ft	7.01 ft
Applied to JFK Context	Thresholds				
estimate	17.89 ft	14.97 ft	12.26 ft	9.73 ft	7.38 ft
95% upper bounds	18.69 ft	15.62 ft	12.76 ft	10.10 ft	7.61 ft

loop through the pilot's increased awareness of the approaching taxiway edge that would inhibit much larger extreme observations than already observed.

As far as the confidence bounds are concerned, they do not take into account the uncertainty in the choice k of the number of extremes to be used and the estimation uncertainty of the extreme value index c that is used to make the extrapolation pattern look linear. These uncertainties arise from the inherent sampling variation in the data. This means that a different study at JFK with a different collection of $\approx 2,500$ deviations (different due to sampling variations) might have led to a somewhat different choice of k and a different estimate of the extreme value index c. This might have led to different estimates and confidence bounds. To some extent this deficiency will be counteracted by the conservative nature of the nonparametric confidence bounds. They do not rely on the assumptions underlying extreme value analysis. It is hoped that future research will allow taking all these concerns into account and adjust for them.

Table D: Two-Sided Exceedance Risks for 747 Based on ANC/JFK Data

Risk of Outer Main Gear Tire Edge Exceeding the Taxiway Edge
for Different Taxiway Width Standards, Using 747, A380 & NLA Main Gear Dimensions.
Does Not Compensate for the Different Nose to Main Gear Distance for the A380 or NLA.

Main Gear Dimensions	Taxiway Width	75 ft	82 ft	100 ft
	as applied to ANC context			
from 747	Estimated Risk	$1.36\ 10^{-7}$	$8.46\ 10^{-9}$	$1.70\ 10^{-11}$
	95% Upper Bound on Risk	$2.48\ 10^{-7}$	$1.69\ 10^{-8}$	$4.07\ 10^{-11}$
from A380	Estimated Risk	$1.57\ 10^{-6}$	$7.97\ 10^{-8}$	$1.07\ 10^{-10}$
	95% Upper Bound on Risk	$2.64\ 10^{-6}$	$1.48\ 10^{-7}$	$2.44\ 10^{-10}$
from NLA	Estimated Risk	$1.98\ 10^{-5}$	$7.97\ 10^{-7}$	$6.94\ 10^{-10}$
	95% Upper Bound on Risk	$3.01\ 10^{-5}$	$1.37\ 10^{-6}$	$1.50\ 10^{-9}$
	as applied to JFK context			
from 747	Estimated Risk	$4.51\ 10^{-7}$	$3.25\ 10^{-8}$	$8.66\ 10^{-11}$
	95% Upper Bound on Risk	$7.90\ 10^{-7}$	$6.22\ 10^{-8}$	$1.98\ 10^{-10}$
from A380	Estimated Risk	$4.55\ 10^{-6}$	$2.72\ 10^{-7}$	$5.05\ 10^{-10}$
	95% Upper Bound on Risk	$7.33\ 10^{-6}$	$4.86\ 10^{-7}$	$1.10\ 10^{-9}$
from NLA	Estimated Risk	$4.91\ 10^{-5}$	$2.40\ 10^{-6}$	$3.00\ 10^{-9}$
	95% Upper Bound on Risk	$7.21\ 10^{-5}$	$3.95\ 10^{-6}$	$6.21\ 10^{-9}$

Acknowledgements

I thank George Legarreta of the FAA and Jerry Robinson of Boeing for their untiring efforts to get this project (COOPERATIVE RESEARCH AND DEVELOPMENT AGREEMENT 01-CRDA-0164 between THE FEDERAL AVIATION ADMINISTRATION WILLIAM J. HUGHES TECHNICAL CENTER and THE BOEING COMPANY) going, and again George Legarreta for his careful review of drafts of this report. I thank Ryan King, Jim White and Pete Sparacino of the FAA William J. Hughes Technical Center for their data collection effort, for the steady support during the analysis and for much of the material and illustrations used in this report. I thank Joe MacDonald, Boeing 747 Chief Project Pilot, for his perspectives and inputs on taxiing 747s and David Steiner for his solution relating main gear oscillations to nose gear oscillations. I further thank Kaz Konya, David Nielson, Ed Gervais, Cory Gibbs, and Brad Bachtel of Boeing for various feedback and support, and again especially Jerry Robinson for his cheerful and engaged support throughout. Finally, I would like to thank Dan Cohen-Nir and Richard Marchi of Airports Council International-North America for making available their draft report on the analysis of similar data. This report caused us to look at the data quality more critically.

1 Data Description

At New York's John F. Kennedy International Airport (JFK) 747 taxiway deviation data were collected from 6/24/1999 to 2/17/2000 at two laser locations for each of two parallel taxiway segments called ALPHA and BRAVO, respectively. The schematic layout of the lasers is shown in Figure 1 while photographic views of ALPHA and one of the lasers are shown in Figures 2 and 3. In particular, Figure 2 shows the centerlights positioned on the eastside of the taxiway centerline, i.e., away from the lasers. However, for BRAVO the centerlights are positioned on the westside of the taxiway centerline, as indicated in the schematic view of Figure 1. The aerial view in Figure 4 also indicates distances to any taxiway crossings.

The laser heights were set up to capture the distances to nose and forward main gears as they crossed in front of the laser beam. The measurement accuracy of the lasers was given as 1-2 inches as stated in [11] which is our main source for details of the data collection process.

Each of the two lasers recorded distance data (also called hits) from anything that crosses its beam on either taxiway. Usually five such hits would be recorded per object (nose or forward main gear wheel or whatever it is that crosses the laser beam). When one of the five hits was not properly recorded it was registered as a zero. The average of all non-zero hits was reported as the final distance to the object recorded, together with the time and date.

Initially it was these averages that were presented for analysis of centerline deviations of 747 aircraft. The determination of a 747 was made based on the landing gear geometry [11], see also Figure 5, namely the combination of 76.54 ft longitudinal distance between nose and main gear and 20.665 ft lateral distance between the aicraft longitudinal centerline and the outside tire edge of the forward main gear (assuming a 19 in wide main gear tire). This combination is unique to the B-747 aircraft series, except for the 747SP. The lateral distance of 20.665 ft translates to a lateral distance of 18.46 ft between out side tire edges of nose gear and forward main gear (assuming a 17 in wide nose gear tire).

For the initially presented data all the lateral distances between outside tire edges of nose and forward main gear fell within 1.5 ft of 18.5 ft. Thus it appears that an allowance of ± 1.5 ft was made in checking for this lateral distance. What allowance was made for the longitudinal distance is not clear.

The determination of the longitudinal distances requires some knowledge of aircraft (or object) speed. According to [11] the time between lasers was taken as average speed. That alone does not allow the calculation of the longitudinal distance between nose and forward main gear unless we also have the time between wheels at each laser. Initially the average distance data for each laser were only given with one date and time stamp for each set of nose and main gear hits. Thus there must have been some tracking internal to the data collection process that would provide this time between wheels. Multiplying the time between wheels with this average speed between lasers would then yield an estimate of the longitudinal distance between gears.

After encountering integrity problems with the originally provided paired distance data,

Figure 1: Schematic Layout of Lasers at Taxiways ALPHA & BRAVO at JFK

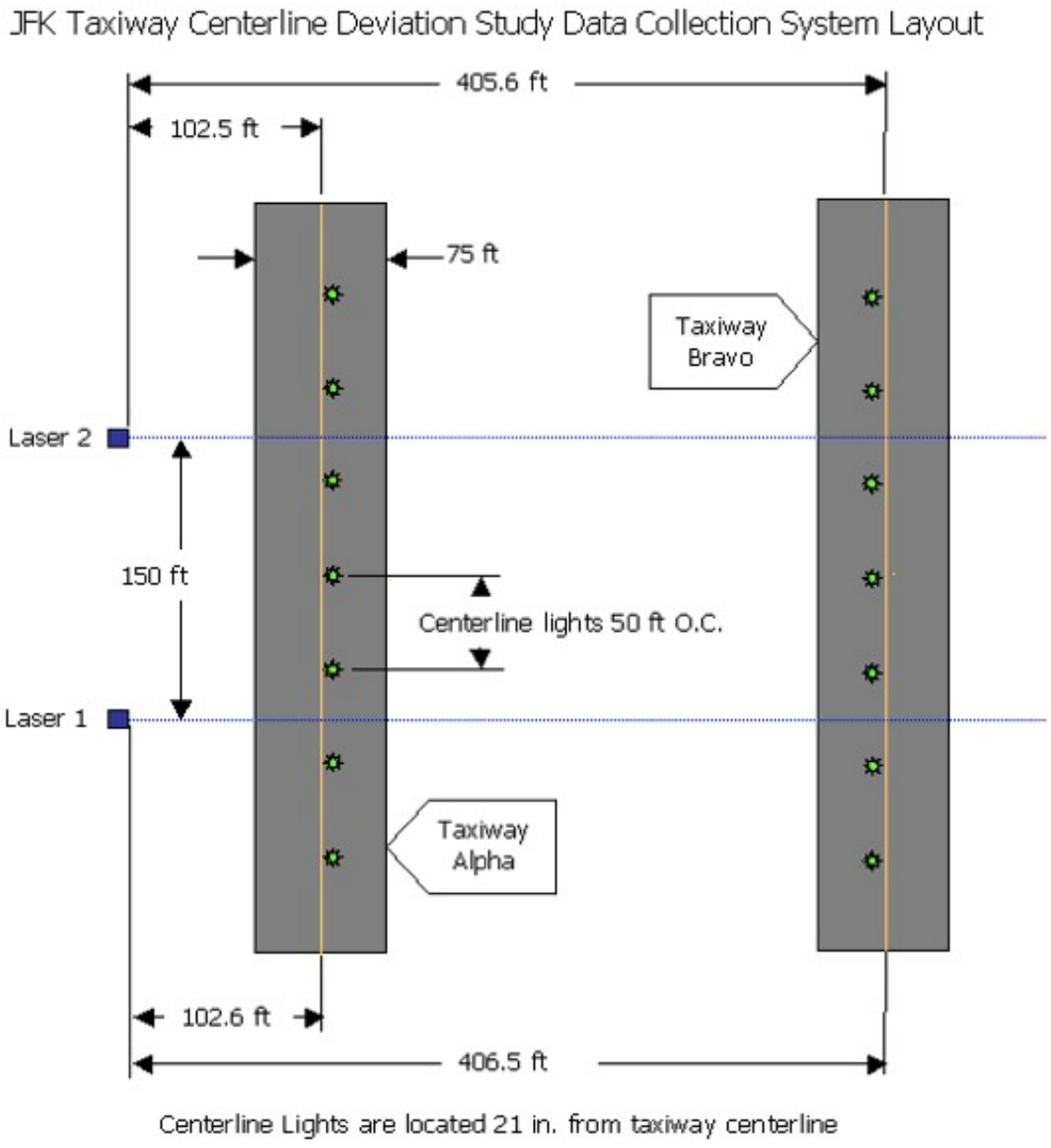


Figure 2: Centerline, Centerlight, and Laser for Taxiway ALPHA at JFK, Looking Southwest

Figure 3: Centerline, Edge, and Laser for Taxiway ALPHA at JFK, Looking South

Figure 4: Aerial View of Lasers for Taxiways ALPHA and BRAVO at JFK

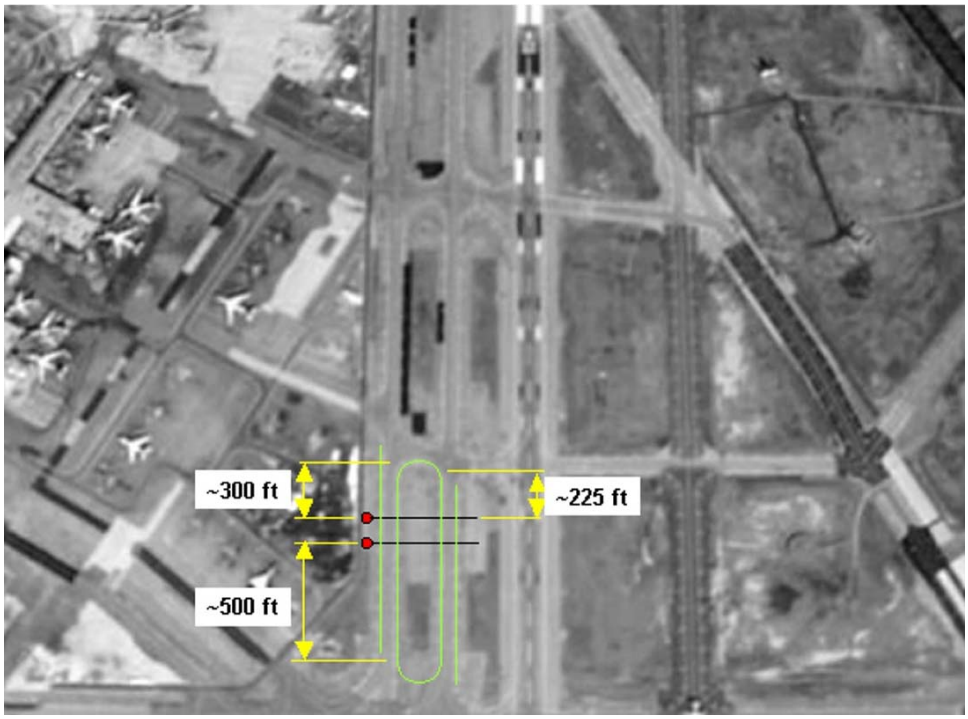
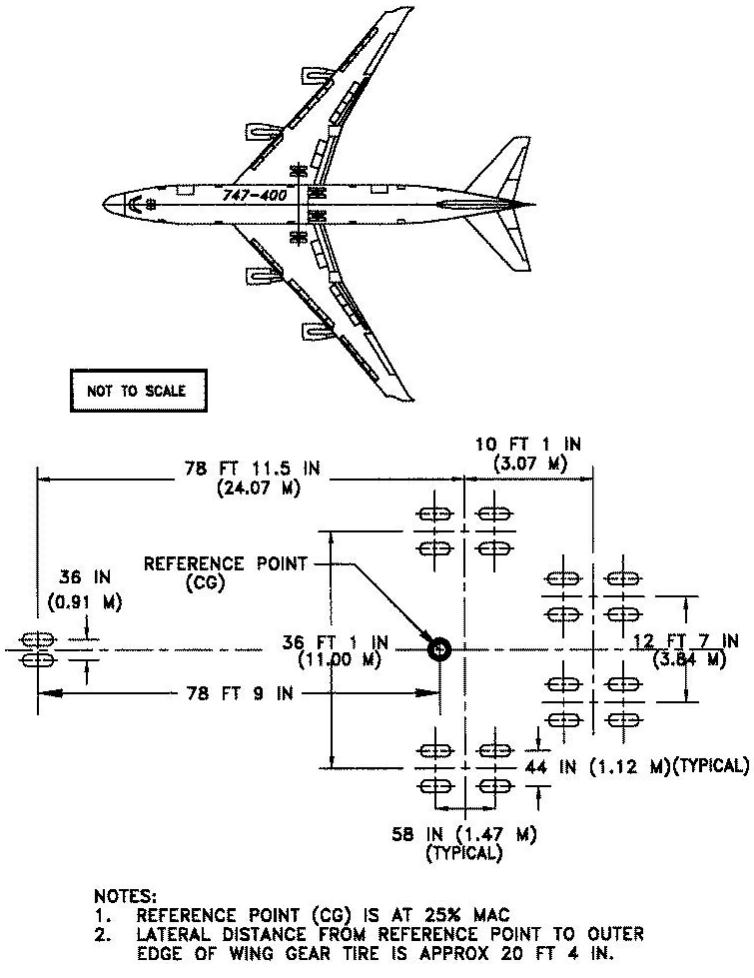


Figure 5: 747 Aircraft Dimensions



i.e., average distances to nose and main gear at each of the two lasers, and getting no clear resolution on how the raw data were originally filtered for quality and classified as 747 events it was decided to start from the very beginning of the raw data. An Excel spreadsheet was provided by Ryan King to serve as baseline master data set. It gave the event data in date and time sorted form. In addition to the date/time stamp for each set of laser hits this raw data set also contained a data value labeled "time between wheels" and the original 5 hits for nose and main gear or whatever was hit.

Although the data were given in time ordered sequence it was not clear which records could be paired as recordings on the same aircraft at the two lasers. To arrive at such a pairing we had to preprocess the raw data.

2 Data Preprocessing and Pairing

We now describe the preprocessing and the appropriate pairing of the raw data. In the course of this a significant amount of data is discarded, but never with any apparent bias influencing the extreme taxiway centerline deviation behavior of the involved aircraft. Many of the discards represent non-747 aircraft.

The provided spreadsheet was organized in rows/records which gave the measurements for either laser 1 or 2, a date and time, the time between wheels and the five hits for nose and main gear at either ALPHA or BRAVO, organized by separate columns. Only rows 2539-41883 appeared to come from alternating lasers. All data before row 2539 corresponded to recordings for laser 2 only and those that came after row 41883 corresponded to recordings for laser 1 only. The latter cases are explained by the failure of Laser 2 on January 13, 2000, after which no more data came from that laser.

In order to sort out 747 aircraft (based on landing gear geometry and speed between lasers) and to check for data quality and consistency readings from both lasers are required. The time ordering of the rows suggests that only adjacent rows could possibly be paired as belonging to the same event (aircraft crossing both lasers). Thus we worked only with the data in rows 2539-41883, i.e., with a possible 19672 paired rows or events. We mention here that in the originally provided data there were 4739 paired events. Thus a significant culling must have taken place to arrive at that number, presumably eliminating non-747 aircraft and other events.

2.1 Splitting Data into ALPHA and BRAVO

As a first step we checked whether there were cases with hits on both ALPHA and BRAVO. There were 193 such cases. They were eliminated after which the data could be separated unambiguously into data from ALPHA and data from BRAVO. However, each case kept its original case number (determined by the date/time sequence) so that later only cases with adjacent case numbers could be paired. This left 26923 cases for ALPHA and 12229 for BRAVO.

2.2 Consistency, Number of Nonzero Hits, and Lateral Distances

Part of the previously mentioned data integrity problems concerned the discovery that the average distance of the nonzero hits was calculated without regard to consistency of the laser hits. Some of these hits differed by more than 34 ft which clearly makes no sense. This discrepancy was discovered during the analysis of the original data when following up on a few cases where the average distances for nose and main gear did not correlate well. The examination of the raw data for those cases revealed these inconsistencies and ultimately led to the reprocessing of the raw data.

Given that the accuracy of the laser is about 1-2 inches and given the profile of the wheel (tire and hub well) one would expect some variation among the five hits on a wheel. Thus we looked at the ranges (maximum minus minimum distance) of the five or fewer nonzero hits and examined their distributions. The results are shown in Figure 6 for nose and main gear broken down by taxiway. Clearly the vast majority of the cases had very consistent ranges of half a foot or less. We decided to drop any case for which the range exceeded 1 ft. For ALPHA and BRAVO there were 5849 and 1595 such cases, respectively. Note that the number of these cases for each taxiway does not equal the sum of such exceedances for nose gear and main gear (as indicated in Figure 6), e.g., $5849 \neq 2897 + 4754$. This discrepancy arises from the fact that ranges in excess of 1 ft can occur for both nose and main gear or for just one gear, i.e., there can be overlap in the counts.

We also checked the lateral distance between nose and main gear to see whether it deviated by more than 1.5 ft from 18.5 ft. Recall that in the originally provided data all cases had a lateral distance within 1.5 ft of 18.5 ft. We found 19424 and 8450 cases which deviated by more than 1.5 ft for ALPHA and BRAVO, respectively. These were also removed.

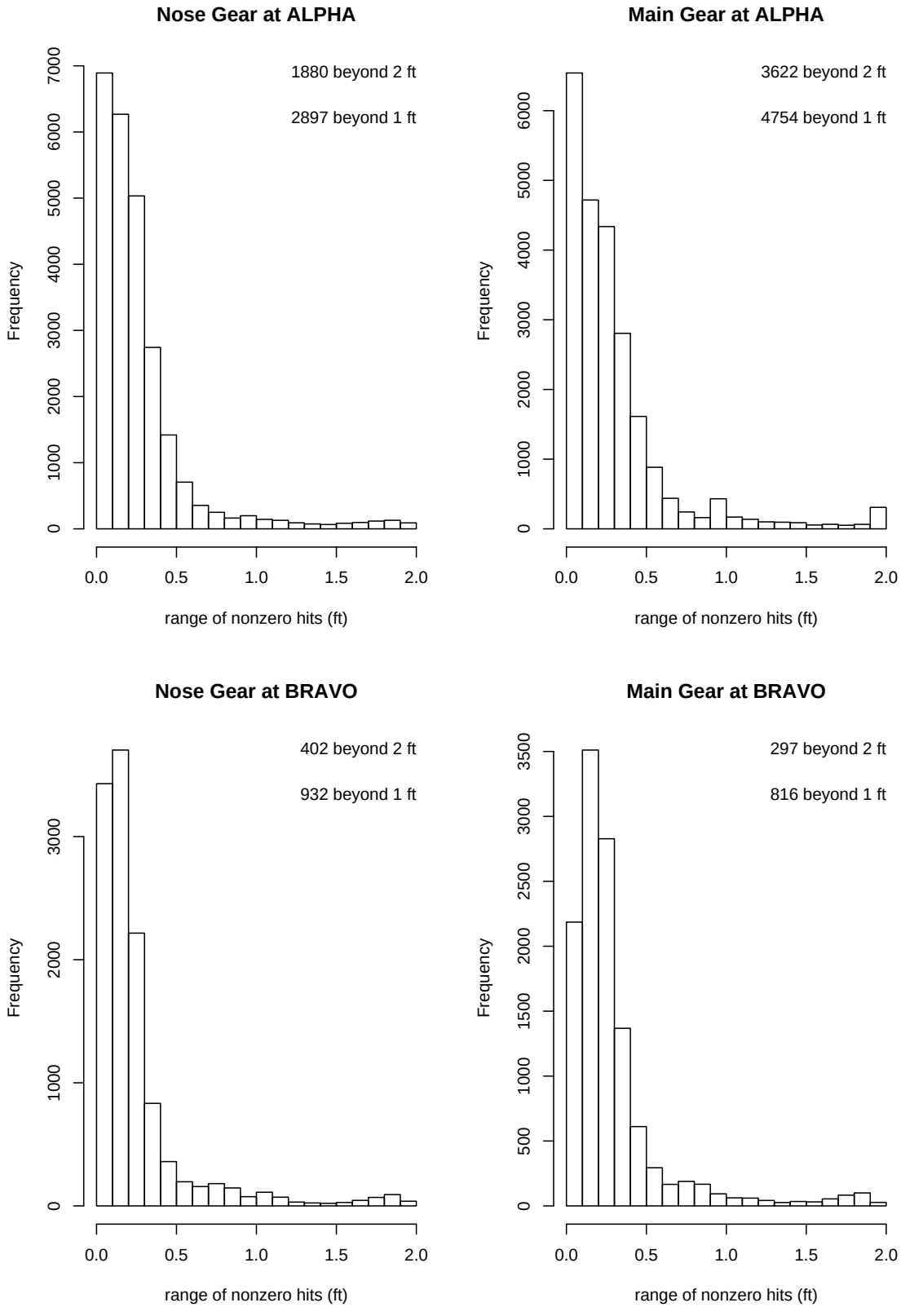
Since the range of nonzero hits degenerates to zero when there is only one nonzero hit there is no way to check the consistency of the hits in that case and we decided to remove such cases. There were 4628 and 706 of these cases for ALPHA and BRAVO, respectively.

When eliminating all cases with high ranges, or lateral nose to main gear distances outside the range 18.5 ± 1.5 ft, or with just one nonzero hit on at least one of the gears we removed 21624 cases for ALPHA and 9120 cases for BRAVO, leaving us with 5299 and 3109 viable cases for ALPHA and BRAVO, respectively.

2.3 Time Gap Processing

These remaining cases were then examined for time gaps between them. These gaps were marked as "short" when they were 60 seconds or less and otherwise they were marked as "long." Since the cases were time ordered to begin with one would expect to see an alternating pattern of short and long gaps or at least no adjacent short gaps. Mostly that was the case, but occasionally one would see several short gaps in a row. This may be possible if the aircraft are queueing. When we encounter an even number of short gaps in a row, say two short gaps, then it is not clear which two of the three involved cases should be

Figure 6: Ranges of Nonzero Hits



paired. Should we pair the first two or the last two? As it turned out there were no situations with an even number of short gaps in a row where that number was greater than two. For the 5299 and 3109 viable cases from ALPHA and BRAVO we respectively encountered 15 and 18 situations with two short gaps in a row. These were examined on a case by case basis and resolved. Typically one of the short gaps would be quite short (a few seconds) while the other would be 30 or more seconds long. The odd case (different from the two that define the shorter time gap) was then removed. This left us with 5284 and 3091 viable cases for pairing.

Next we processed these viable sets in an iterative fashion starting at the beginning of each set and proceeding sequentially through all remaining cases. Each pair of cases adjacent in time order was treated as a potential pair for serving as the pair of measurements on the same aircraft at both lasers. We accepted it as such a pair when the following three pairing criteria were satisfied. The time gap for the pair was short, the involved lasers were different, and the original sequence numbers differed by exactly one. This last requirement ruled out pairings which originally were separated by in between measurements that had since been removed by prior processing.

Finding a pair satisfying all three pairing criteria we would allocate both cases to the collection of paired data cases. If a pair did not satisfy all three pairing criteria we would remove the first case in that pair and treat the second case as the first case of the next pair to be examined for the three pairing criteria. This process was continued until all cases were either paired or removed. In the process of pairing we also assigned the direction of travel based on timing and laser id. This process resulted in 1483 pairs (aircraft events) for ALPHA and 1098 pairs for BRAVO for a total of 2581 events for both taxiways.

2.4 Checks for Velocity and Longitudinal Nose to Main Gear Distance

As a next step we examined the velocities of the aircraft (or whatever was captured) for each event when calculating it from the time between lasers and from the time between wheels. For the velocity between lasers one takes the 150 ft and divides it by the time between lasers. For the velocity at the lasers one would take the longitudinal distance between nose and main gear and divide that by the time between wheels at that laser. This calculation assumes some knowledge concerning that longitudinal distance ℓ. If we deal with 747s then we would take $\ell = L = 76.54$ ft. But if it is not a 747 one should use another value. One can calculate four velocities, namely the velocities of nose and main gears between lasers and the velocities between gears at lasers 1 and 2.

If the remaining paired cases all represent 747s then these velocities should be nearly the same. When plotted against each other these velocities should fall near the main diagonal. However, an intitial plot of velocities at a laser against the velocity between lasers showed quite a few points that strayed far from the main diagonal. These can be seen in the middle two plots of Figures 7 and 8. The significance of the heavy dots (not all are stray points) will be explained later.

Figure 7: Velocity & Landing Gear Geometry Checks for ALPHA

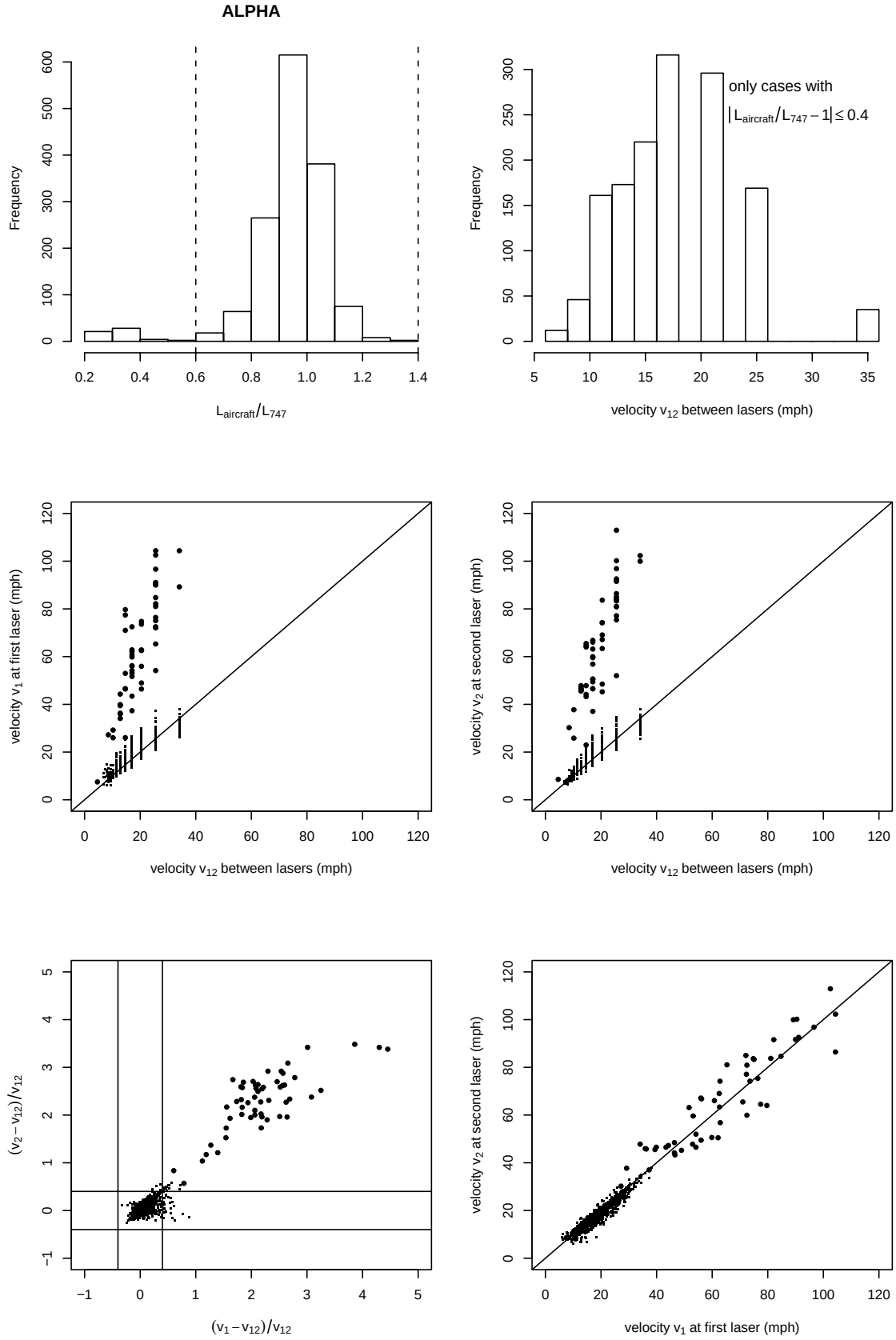
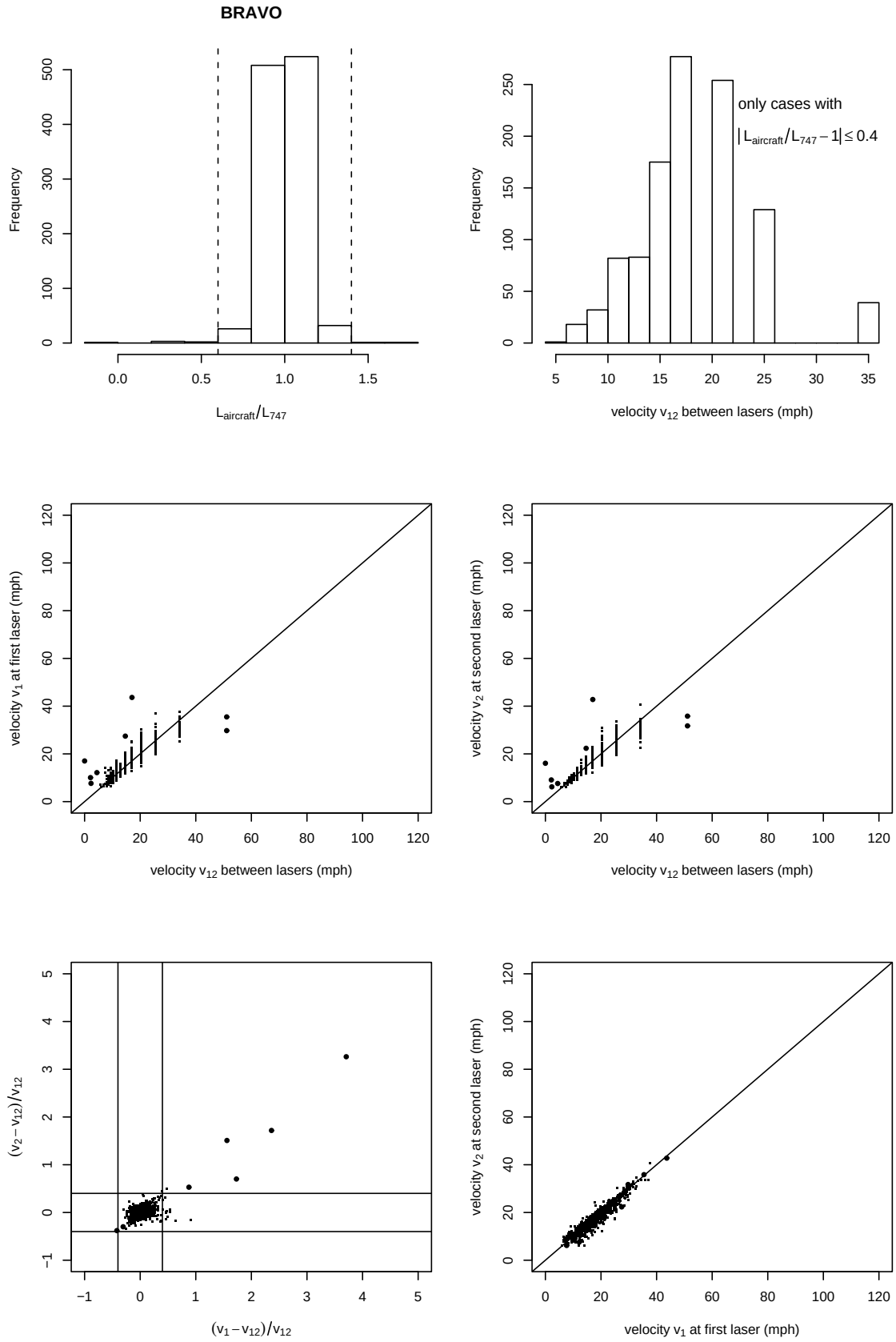


Figure 8: Velocity & Landing Gear Geometry Checks for BRAVO



The bottom left plot shows the relative differences of the velocities at lasers 1 and 2 (v_1 and v_2) when compared to the velocity between lasers (v_{12}). If we were dealing only with 747s we should see small relative differences. The box around the main point cloud has limits $[-.4, .4]$ on either dimension. The bottom right plot shows comparisons of velocities at laser 1 and 2. There the points all fall nicely along the main diagonal. This just expresses the fact that whatever longitudinal distance between nose and main gears applies at one laser also applies at the other laser and that the times between wheels are more or less consistent.

Clearly some of these velocities are very high and could well be the result of using the wrong longitudinal distance between nose and main gear, namely dividing the large $L = 76.54$ by the shorter time that would typically accompany a shorter true longitudinal distance ℓ between nose and main gear.

This led us to consider the following model for backing out the actual distance ℓ or the ratio ℓ/L. Assume that the motion of the aircraft or vehicle that is captured by the lasers has constant acceleration a, which includes constant velocity as special case ($a = 0$). Then the distance $s(t)$ (ft) traveled by the aircaft and its velocity $v(t)$ can be described at a function of elapsed time t (sec) by the following equation:

$$s(t) = s_0 + v_0 t + \frac{1}{2} a t^2 \qquad \text{and} \qquad v(t) = v_0 + at \ ,$$

where $s_0 = 0$ and v_0 are the distance and velocity at the time $t = 0$. When calculating an average velocity by dividing the distance $s(t_2) - s(t_1)$ by the time difference $t_2 - t_1$ we get from the above equations

$$v(t_1, t_2) = \frac{s(t_2) - s(t_1)}{t_2 - t_1} = v_0 \frac{t_2 - t_1}{t_2 - t_1} + a \frac{t_2^2 - t_1^2}{t_2 - t_1} = v_0 + a \frac{t_1 + t_2}{2} \ .$$

If $t_1 < t_2$ are the two times available for the first laser ($t_2 - t_1$ being the time between wheels) and if $t_3 < t_4$ are the two times available for the second laser then one could view $v_1 = v(t_1, t_2)$ and $v_2 = v(t_3, t_4)$ as the velocities at the two lasers.

If we assume $s(t_2) - s(t_1) = s(t_4) - s(t_3) = L = 76.54$ ft we would have the velocity of a 747. However, at this point we will make no such assumption and use $s(t_2) - s(t_1) = s(t_4) - s(t_3) = \ell$ where ℓ is unknown. Denote these velocities by $\tilde{v}_1 = \tilde{v}(t_1, t_2)$ and $\tilde{v}_2 = \tilde{v}(t_3, t_4)$ when using ℓ and denote them by $v_1 = v(t_1, t_2)$ and $v_2 = v(t_3, t_4)$ when using L. Then we have the following relationships: $\tilde{v}_1 = v_1 \ell / L$ and $\tilde{v}_2 = v_2 \ell / L$. Here v_1 and v_2 can be computed because of the known quantity L. Taking the difference of $\tilde{v}_1$ and $\tilde{v}_2$ we get on the one hand

$$\tilde{v}_2 - \tilde{v}_1 = a \frac{t_3 + t_4}{2} - a \frac{t_1 + t_2}{2} = a \left(\frac{t_3 + t_4}{2} - \frac{t_1 + t_2}{2} \right)$$

and on the other hand

$$\tilde{v}_2 - \tilde{v}_1 = \left(\frac{\ell}{t_4 - t_3} - \frac{\ell}{t_2 - t_1} \right) = \left(\frac{L}{t_4 - t_3} - \frac{L}{t_2 - t_1} \right) \frac{\ell}{L} \ .$$

Combining this gives us an equation in the two unknowns ℓ/L and a:

$$a\left(\frac{t_3+t_4}{2}-\frac{t_1+t_2}{2}\right)=\left(\frac{L}{t_4-t_3}-\frac{L}{t_2-t_1}\right)\frac{\ell}{L}\,.$$

Similarly we can get another equation by taking the difference of the two velocities between lasers. These give us

$$v(t_2,t_4)-v(t_1,t_3)=a\frac{t_2+t_4}{2}-a\frac{t_1+t_3}{2}=a\left(\frac{t_2+t_4}{2}-\frac{t_1+t_3}{2}\right)$$

and

$$v(t_2,t_4)-v(t_1,t_3)=\left(\frac{150}{t_4-t_2}-\frac{150}{t_3-t_1}\right)\,.$$

Combining these we get

$$a\left(\frac{t_2+t_4}{2}-\frac{t_1+t_3}{2}\right)=\left(\frac{150}{t_4-t_2}-\frac{150}{t_3-t_1}\right)\,.$$

This equation can be solved for the acceleration a. That value can then be used in solving the first equation for ℓ/L. This process has one minor exception, namely when $t_2-t_1=t_4-t_3=\Delta>0$, i.e., when both times between wheels are the same. Then the second equation degenerates to $a\Delta=0$ and implies $a=0$ while the first becomes $0=0$ ℓ/L and ℓ/L cannot be determined this way. However, when $a=0$, as is implied from this case, one can go directly to the motion equations, namely get

$$v(t)=v_0=\frac{150}{t_3-t_1}=\frac{\ell}{t_2-t_1}=\frac{\ell}{L}\,\frac{L}{t_2-t_1}$$

and from here get

$$\frac{\ell}{L}=\frac{150}{L}\frac{t_2-t_1}{t_3-t_1}=\frac{150}{76.54}\frac{t_2-t_1}{t_3-t_1}\,.$$

When we are dealing with a 747 we would expect to obtain a value of ℓ/L near one, allowing for some measurement error and mild deviations from the constant acceleration model. However, computed values of ℓ/L far from one would throw doubt on the assumption that we deal with a 747 and one should discard such cases.

The top left plots of Figures 7 and 8 show the histograms of ℓ/L computed in the above fashion. Most of the cases gave values of ℓ/L distributed closely around one. Also shown are two vertical lines at $1\pm.4$ beyond which the value of ℓ/L was judged to be excessive. Those points are also marked by heavy dots in the remaining plots in these Figures. Clearly the previously noted bad behavior in most cases falls in line with the interpretation that a length ℓ different from L is at work in those cases. We thus decided to remove those event cases, 55 at ALPHA and 8 at BRAVO, leaving us with 1428 and 1090 events at ALPHA and BRAVO, respectively.

2.5 Lateral Deviations Between Nose and Main Gear

Recall that the deviation data had been filtered using the restriction that only those cases would be accepted for which the difference in the distances from laser to nose gear and to forward main gear is within ± 1.5 ft of 18.5 ft. The plots in Figure 9 show the scatter of these lateral differences when plotting those from Laser 2 against those from Laser 1, separated by direction. It appears that the scatter clouds are reasonably centered at $(18.5, 18.5)$ showing no turning effects as they were seen in similar plots for the ANC data.

2.6 Velocity and Acceleration Checks

The screening of the ANC deviation data employed velocity and acceleration checks while there were no checks for the longitudinal distance between nose and main gear as it was employed here. We point out that these checks are somewhat equivalent since the velocities for the ANC data assume a 747 wheel base. If this latter assumption is wrong and a much shorter wheel base is appropriate then the velocities at the lasers will come out much higher than the velocities between the lasers. This in turn will lead to some unusual accelerations. For comparison purposes we also provide the velocity and acceleration plots for the filtered JFK data in Figures 10-12. All histograms show well behaved patterns, actually better than those seen for the ANC data. This is not surprising in view of the larger distance of the lasers from any possible turns to and from ALPHA and BRAVO.

Figure 9: Lateral Distances Between Nose Gear and Main Gear
after cleaning of data

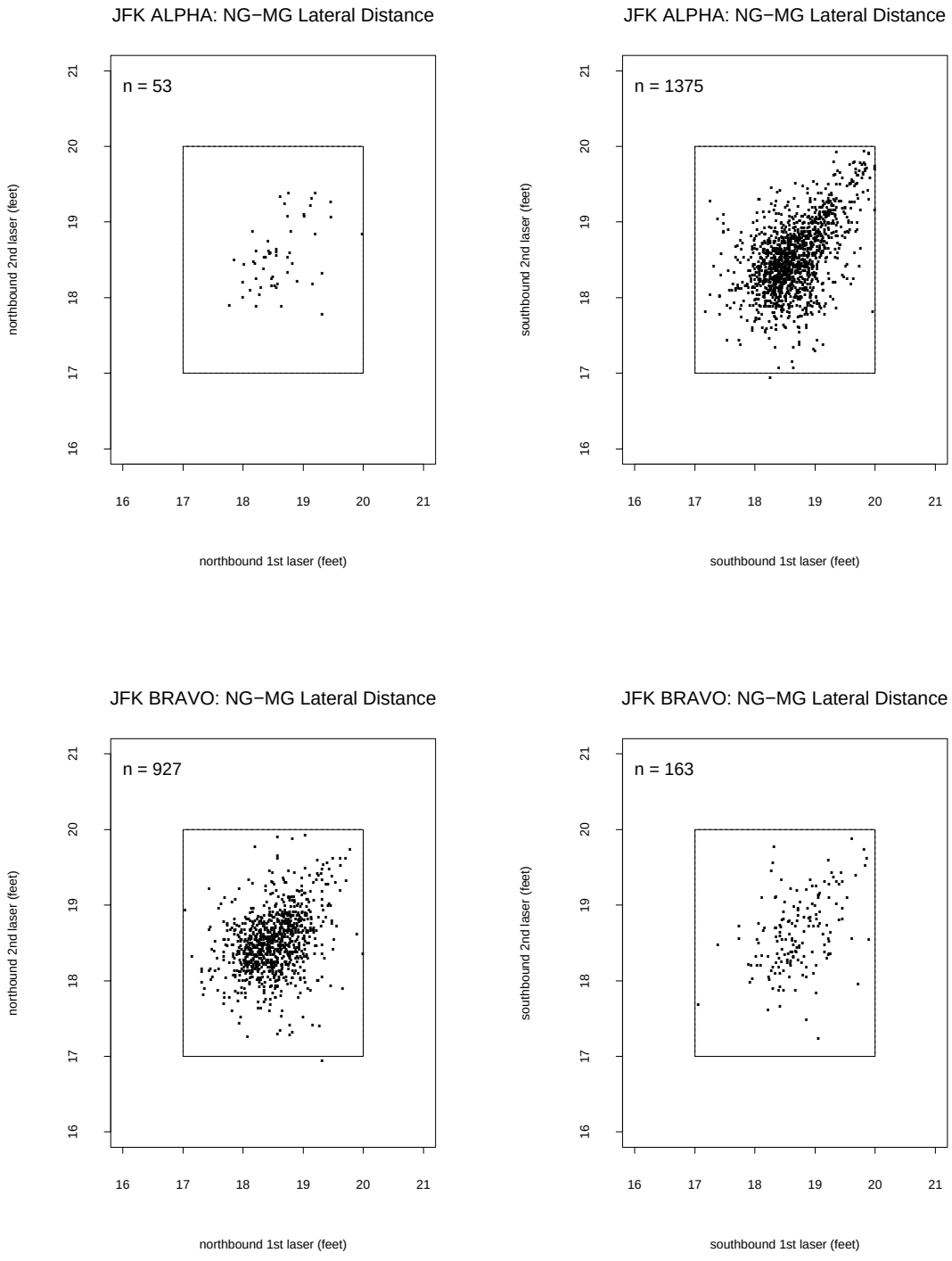


Figure 10: Maximum Speeds at Each Laser & From Laser to Laser
after cleaning of data

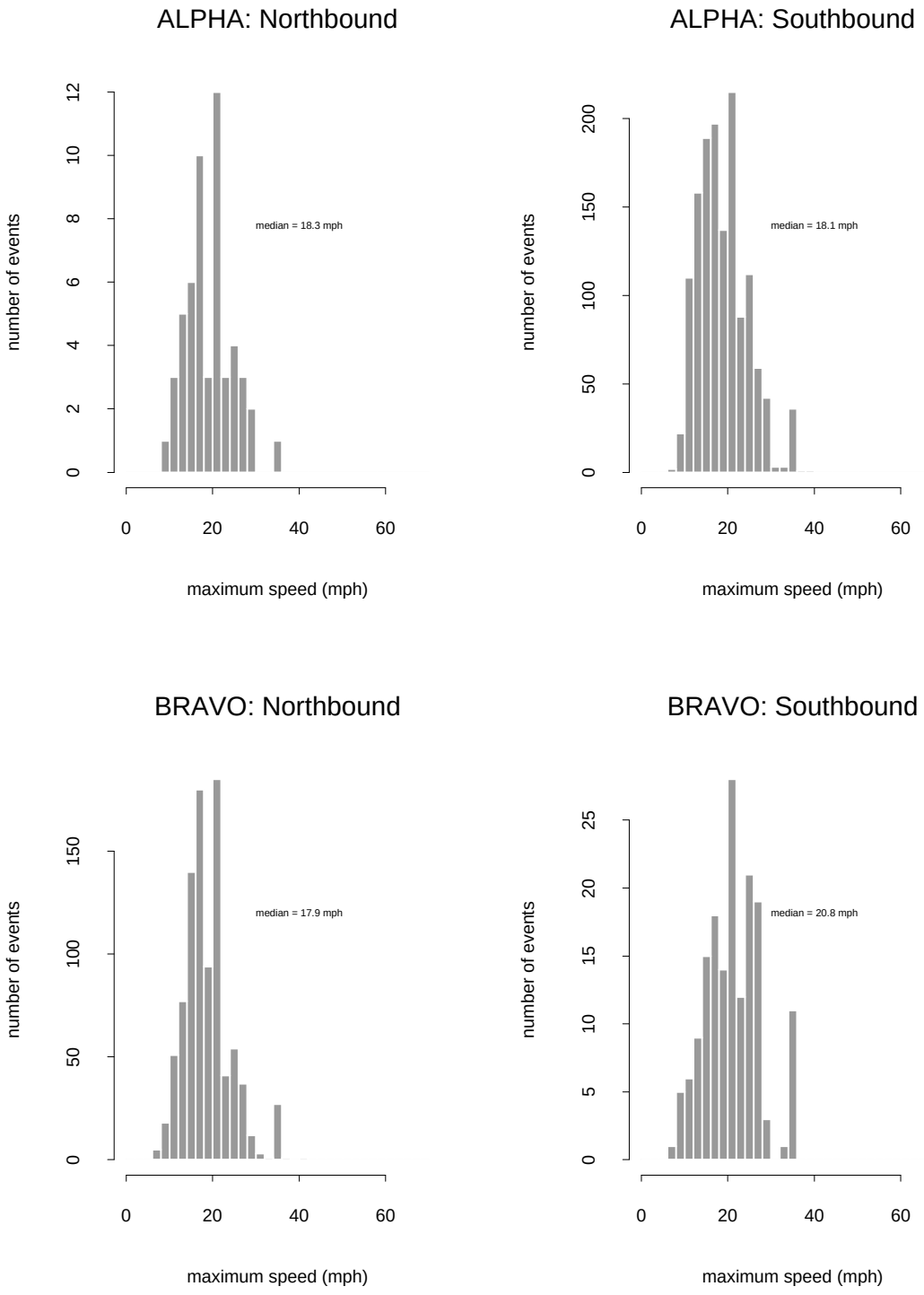
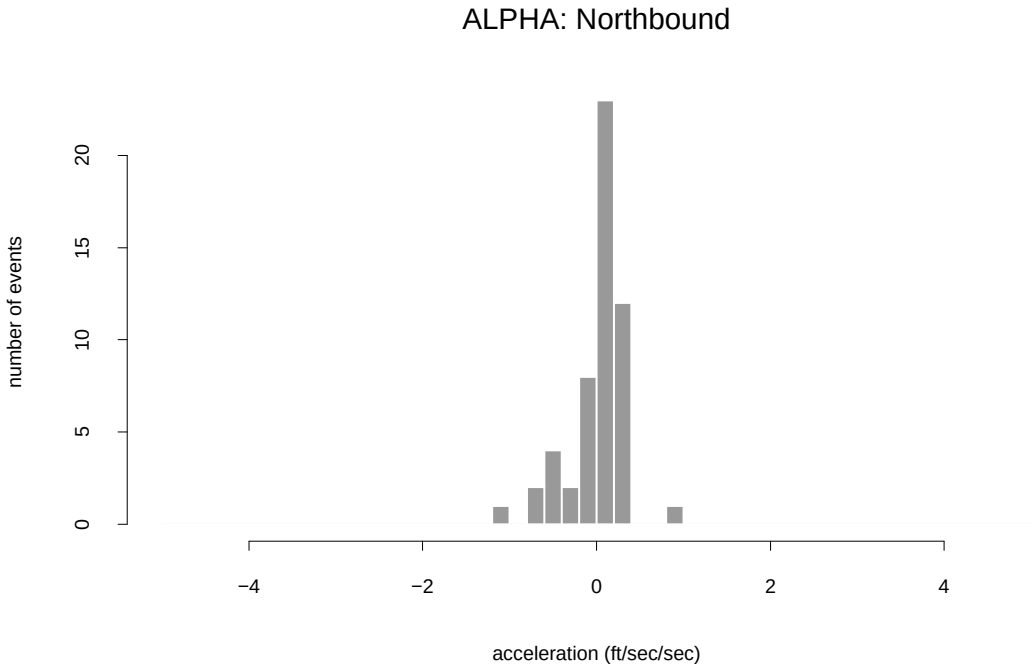


Figure 11: Acceleration Check After Cleaning of Data



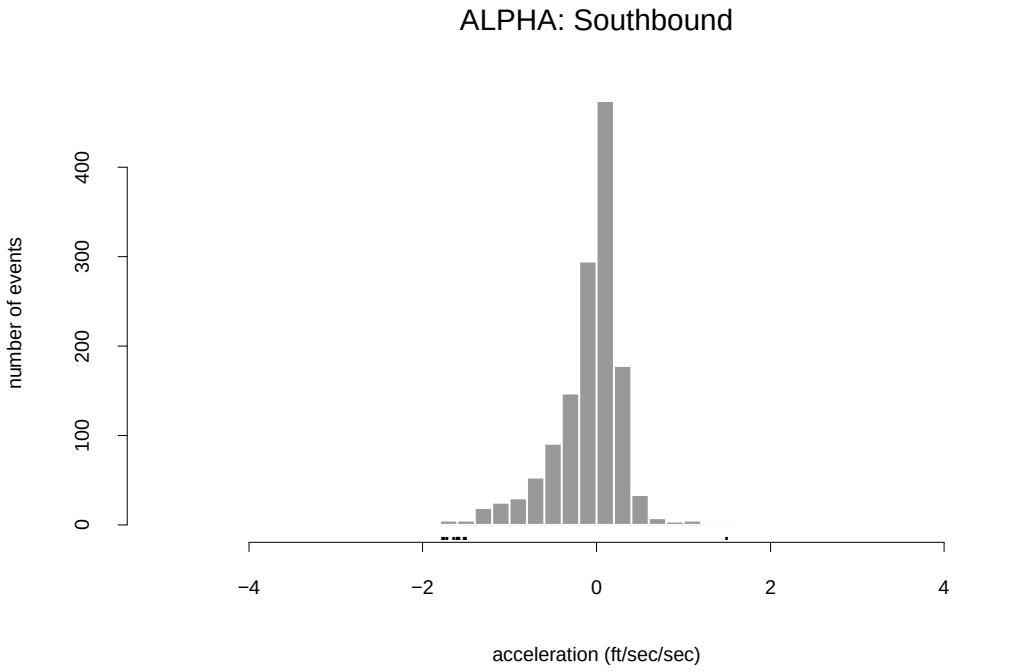
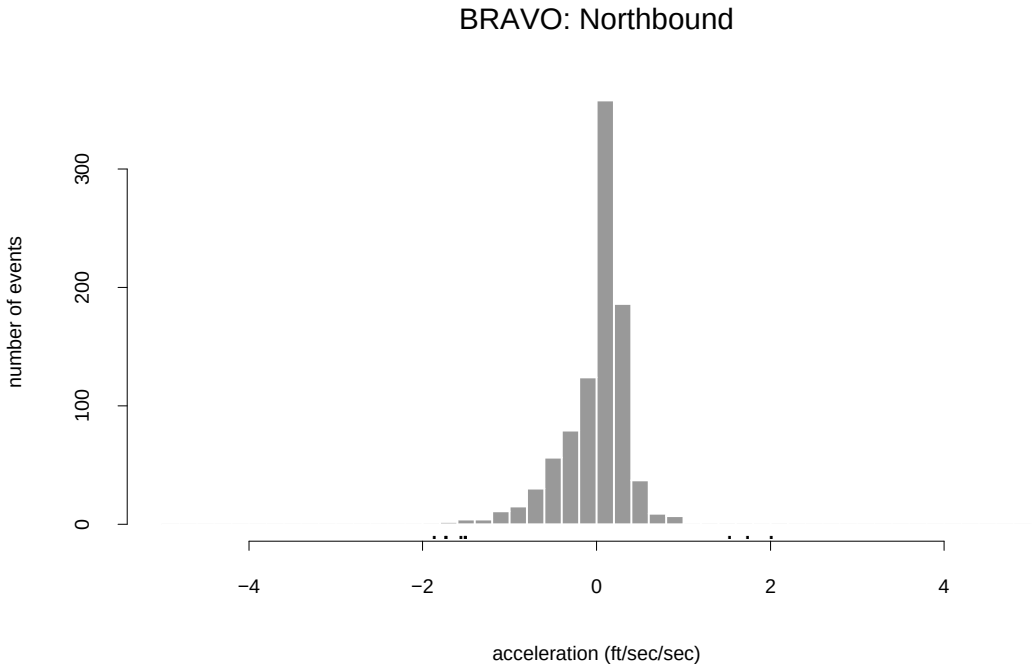
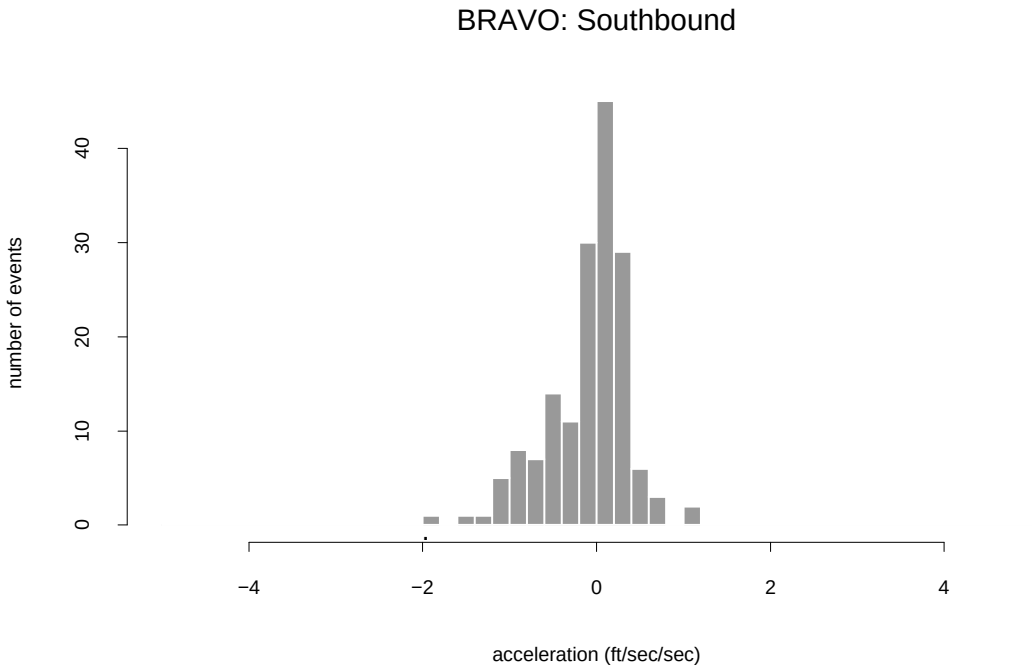


Figure 12: Acceleration Check After Cleaning of Data





3 Deviations From Taxiway Centerline

Since the raw data gave the average distances from the lasers to the outside tire edges of nose and main gears the "deviations from centerline" were then derived for the nose and the main gear at both lasers. These "deviations from centerline" represent the deviation of the airplane center at that lateral gear location from the taxiway center line. These derivations used the following formulas, where X represents the average measured distance from laser to outside tire edge and Y denotes the derived "deviation from centerline."

		Deviation Y of AC Center from Taxiway Centerline in Relation to Measured Distance X from Laser to Outer Gear Tire Edge	
Taxiway	Laser	At Nose Gear Location	At Main Gear Location
ALPHA	1	$Y = (X + 2.21) - 102.6$	$Y = (X + 20.665) - 102.6$
	2	$Y = (X + 2.21) - 102.5$	$Y = (X + 20.665) - 102.5$
BRAVO	1	$Y = (X + 2.21) - 406.5$	$Y = (X + 20.665) - 406.5$
	2	$Y = (X + 2.21) - 405.6$	$Y = (X + 20.665) - 405.6$

Here the values $102.5, 102.6$ and $406.5, 405.6$ are evident from Figure 1. The quantities $\Delta_N = 2.21$ and $\Delta_M = 20.665$ represent the lateral distances of the outer nose and main gear tire edge to the aircraft centerline. This calculation of centerline deviation results in negative (positive) deviations when the deviation is toward (away from) the laser. A zero deviation means that the aircraft centerline is directly above the taxiway centerline.

Here it should be pointed out that these distances Δ_N and Δ_M depend very much on the tire width which varies from 747-100 to 747-400. As seen from Figure 5, $\Delta_M = 18'1/2'' + 22'' + \delta_M/2 = 19.875' + \delta_M/2$, where δ_M is the tire width, which according to Kaz Konya can vary from $16''$ (747-100), $17''$ (747-200), $18''$ (747-300) to $19''$ (747-400). The value of $20.665'$ used above is based on an $19''$ tire. According to Kaz Konya about half of the 747s in service are 747-400s and may show up in even higher percentages at JFK and ANC. The possibility that a certain fraction of 747s has a smaller main gear tire width (while we assumed a $19''$ tire in the centerline deviation conversion) may cause a small bias away from the lasers.

4 Correlation and Time Plots

As with the ANC data a first look at the centerline deviation data consists of diagnostic correlation plots to see whether the data are consistent with expectations. These plots are shown as correlation scatter plots in Figures 13-16.

On each plot the main diagonal is shown as a dotted line. If the measurements correlate well they should follow this main diagonal closely. The plots all have the same abscissa

and ordinate scales to facilitate comparisons across plots. Also shown as solid line is the first principal component axis. This is the axis along which the data points show the most variation when projected onto it. Equivalently, it is the line for which the sum of the squared perpendicular distances of the points to that line is minimized.

The closeness of this clustering is captured to some extent by the indicated correlation coefficient r. For a close clustering around the main diagonal we would expect $r \approx 1$, with $r = 1$ indicating a perfect linear relationship. However, it should be kept in mind that a perfect linear relationship does not necessarily mean that this relationship is the main diagonal. For that reason one should also examine visually how well the plotted points cluster around the main diagonal and not some other line. For this reason we also plotted the principal component axis which was explained previously. One can take this axis as an indication of the line around which the points appear to cluster most tightly. For a good consistency measure one would expect that this axis more or less agrees with the main diagonal, at least within the range of the point cloud.

As with the ANC correlation plots one sees that the nose to main gear correlation at the same laser is very tight in all cases and not quite so tight when correlating the same gear for the two lasers. We do not seem to have any tilt behavior as was observed for the ANC data in the nose to main gear correlation plot at KILO Laser 1 Eastbound, see[10].

The plots correlating the same gear at different lasers all show wider scatter (somewhat wider for nose gear than for main gear). This may result from the distance between the lasers (150 ft) being longer than the distance between nose and main gear (76.54 ft along the centerline). The somewhat wider scatter in the nose gear correlation plot when compared to the main gear correlation plot also makes sense since the nose gear is where the steering changes occur and the main gear just follows in somewhat dampened fashion.

The deviation data were also plotted against the date and time of day. This plotting was done separately by nose and main gear, by laser, and by direction. The resulting plots are shown in Figures 17-24. Also shown in these plots are the number of positive and negative deviations.

It is quite evident by the indicated counts that positive deviations far outnumber the negative ones, except for the nose and main gear deviations Northbound at Laser 1 for BRAVO. A possible cause for bias in the deviations from the centerline could be a parallax effect. This would be due to the pilot sitting on one side of the aircraft centerline. However, with parallax bias alone one would expect opposite effects for opposite aircraft headings, i.e., positive biases in one direction and negative ones in the other, but roughly equal in absolute size. Since here the biases are all positive, regardless of aircraft heading, another bias effect must be at work here. This other bias is away from the lasers since positive deviations tend to be farther away from the laser than negative ones.

The same phenomenon of consistent positive bias in the centerline deviations was observed in the ANC data. As likely cause for such bias behavior was mentioned the taxiway centerlights. At ANC they were offset from the centerline by approximately 12 inches on the side away from the lasers. Here at JFK the offset is approximately 21 inches, on the side away from

Figure 13: Correlation Plots for ALPHA Northbound

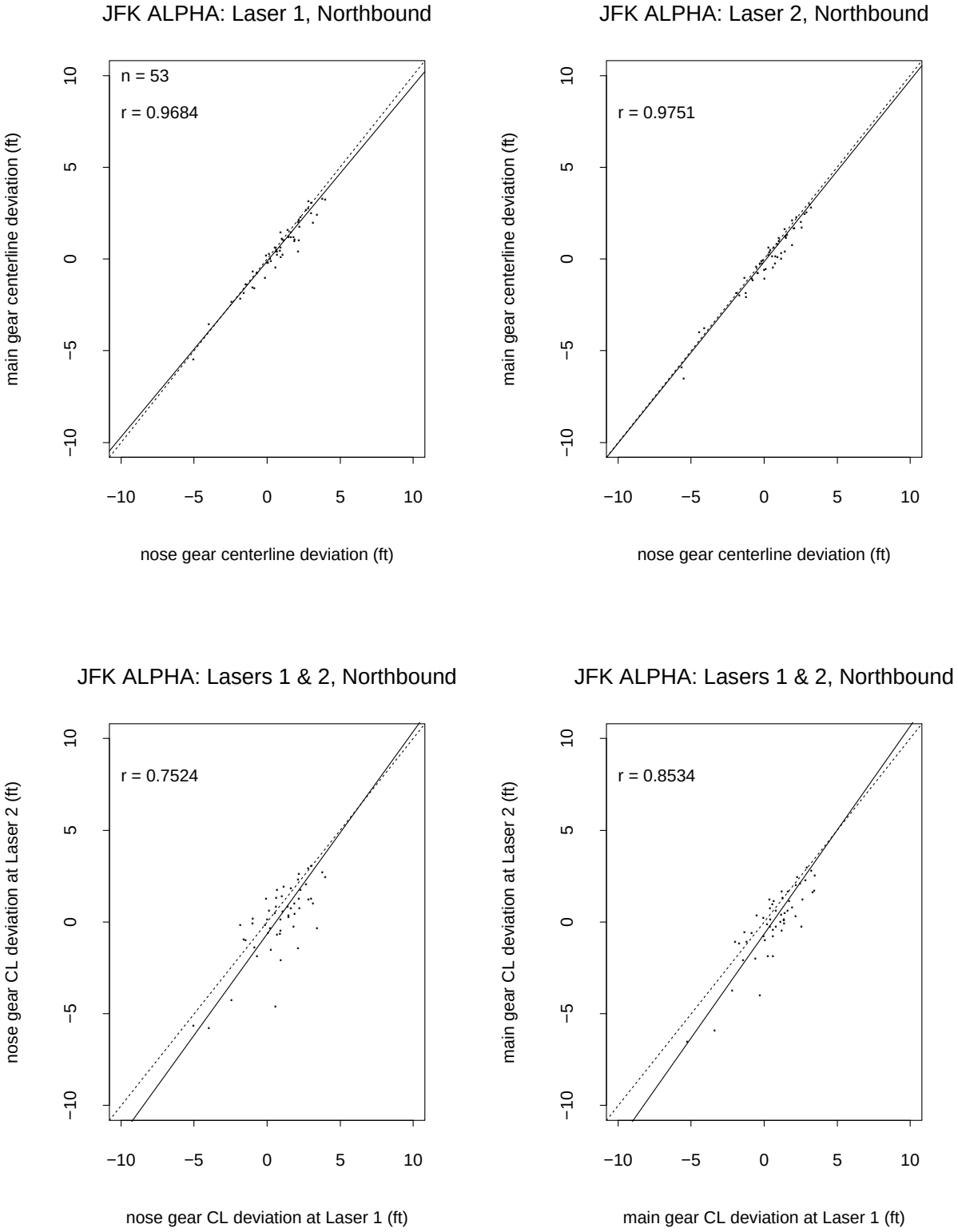


Figure 14: Correlation Plots for ALPHA Southbound

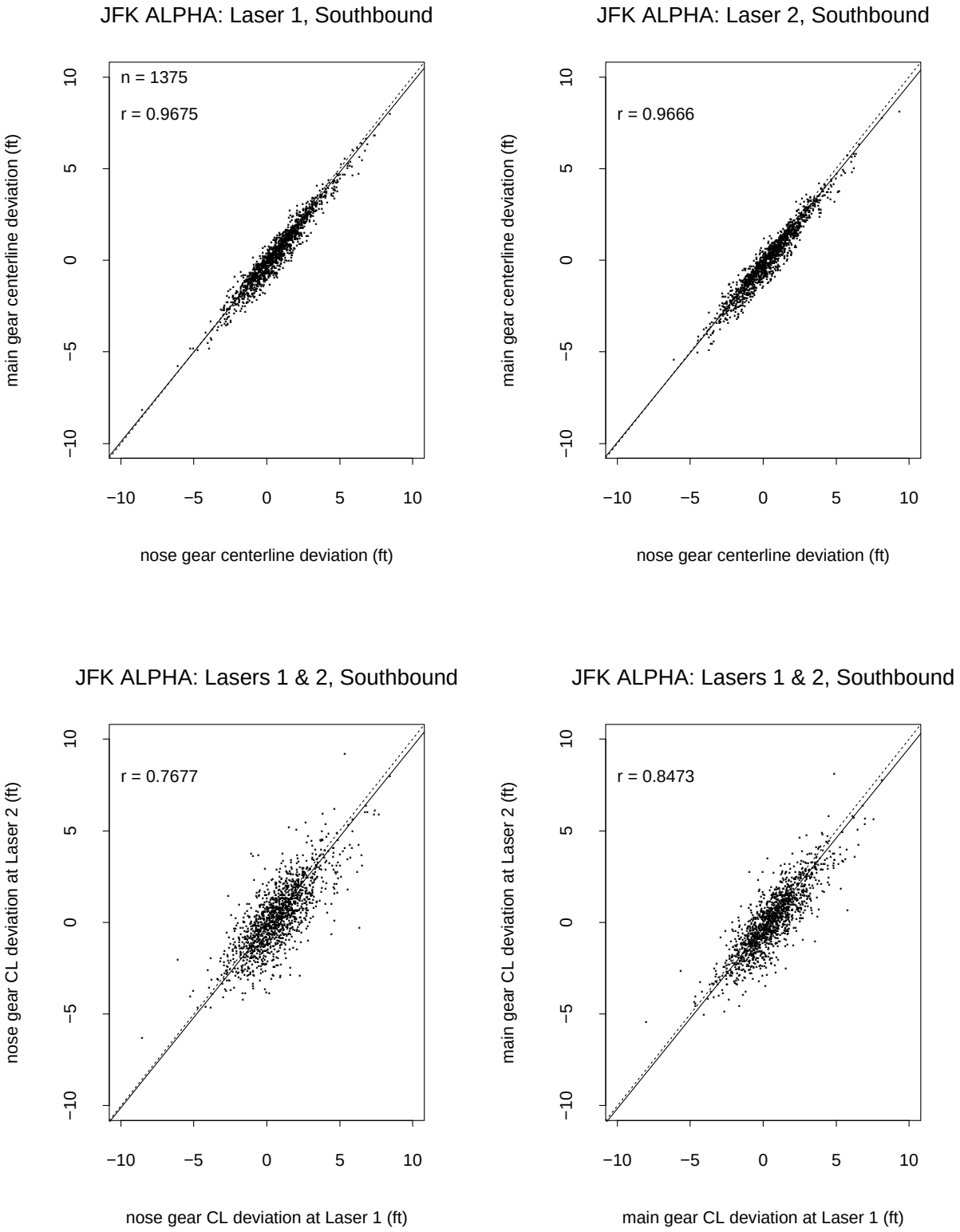


Figure 15: Correlation Plots for BRAVO Northbound

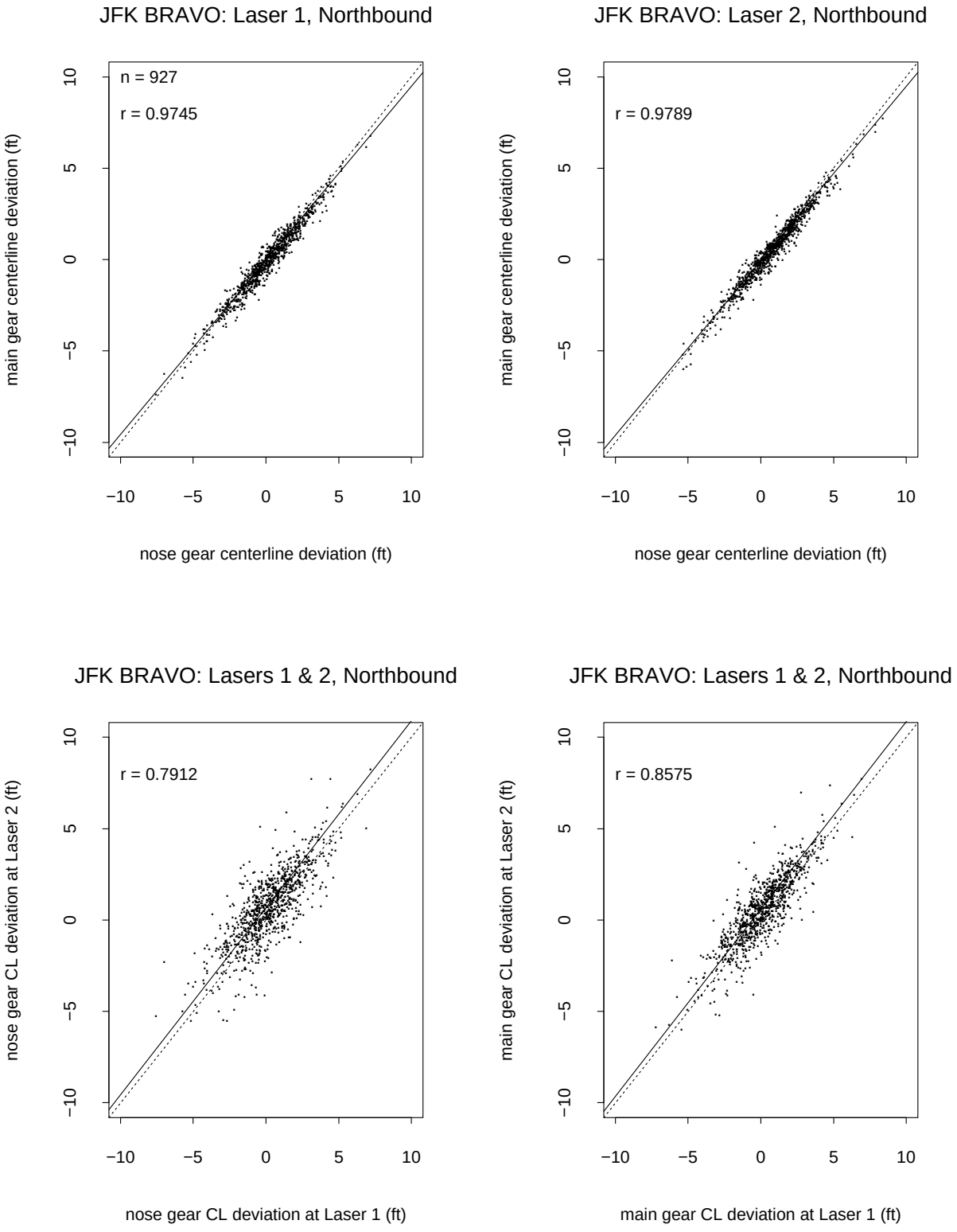


Figure 16: Correlation Plots for BRAVO Southbound

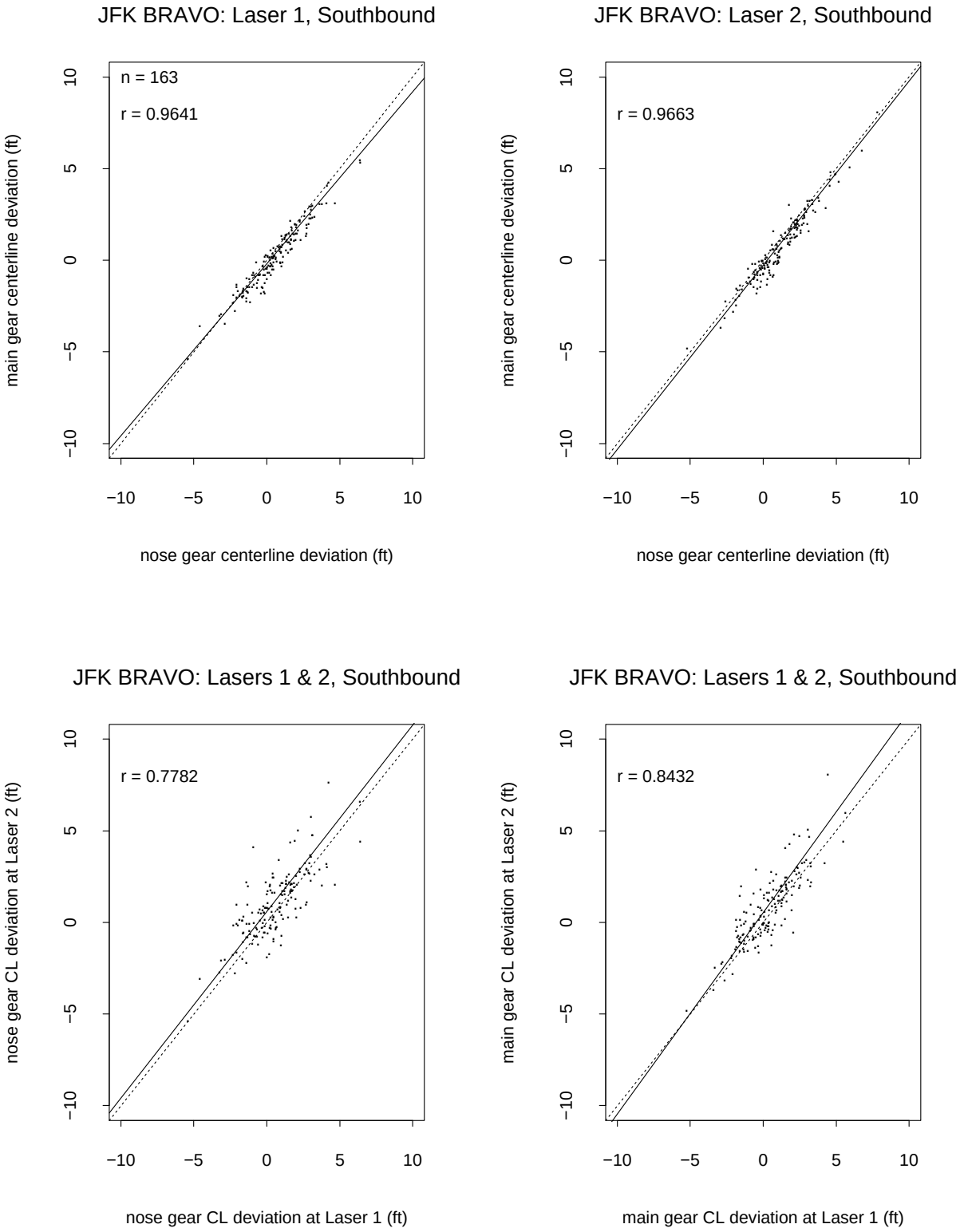


Figure 17: Time Plots for Nose & Main Gears at ALPHA Northbound, Laser 1

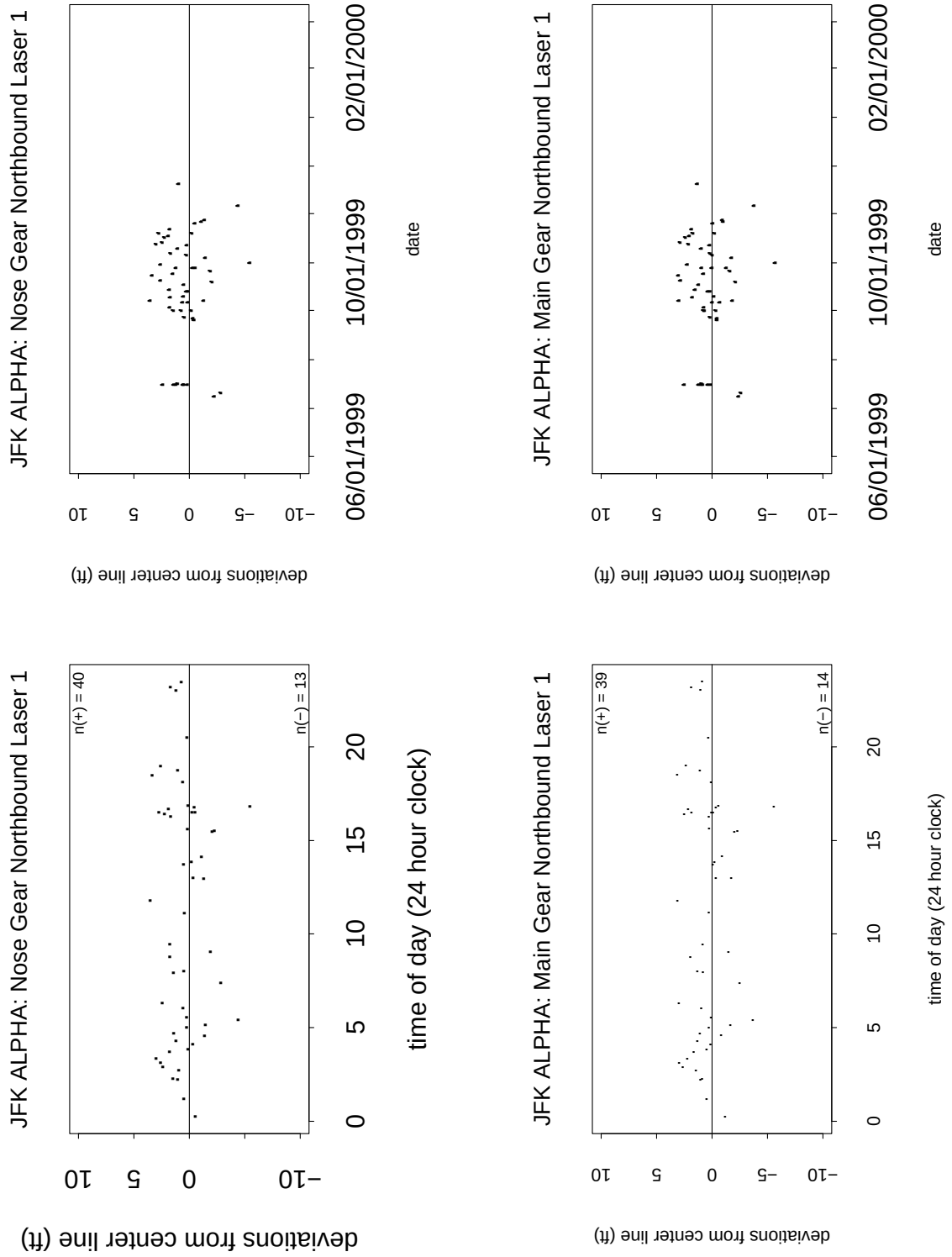


Figure 18: Time Plots for Nose & Main Gears at ALPHA Northbound, Laser 2

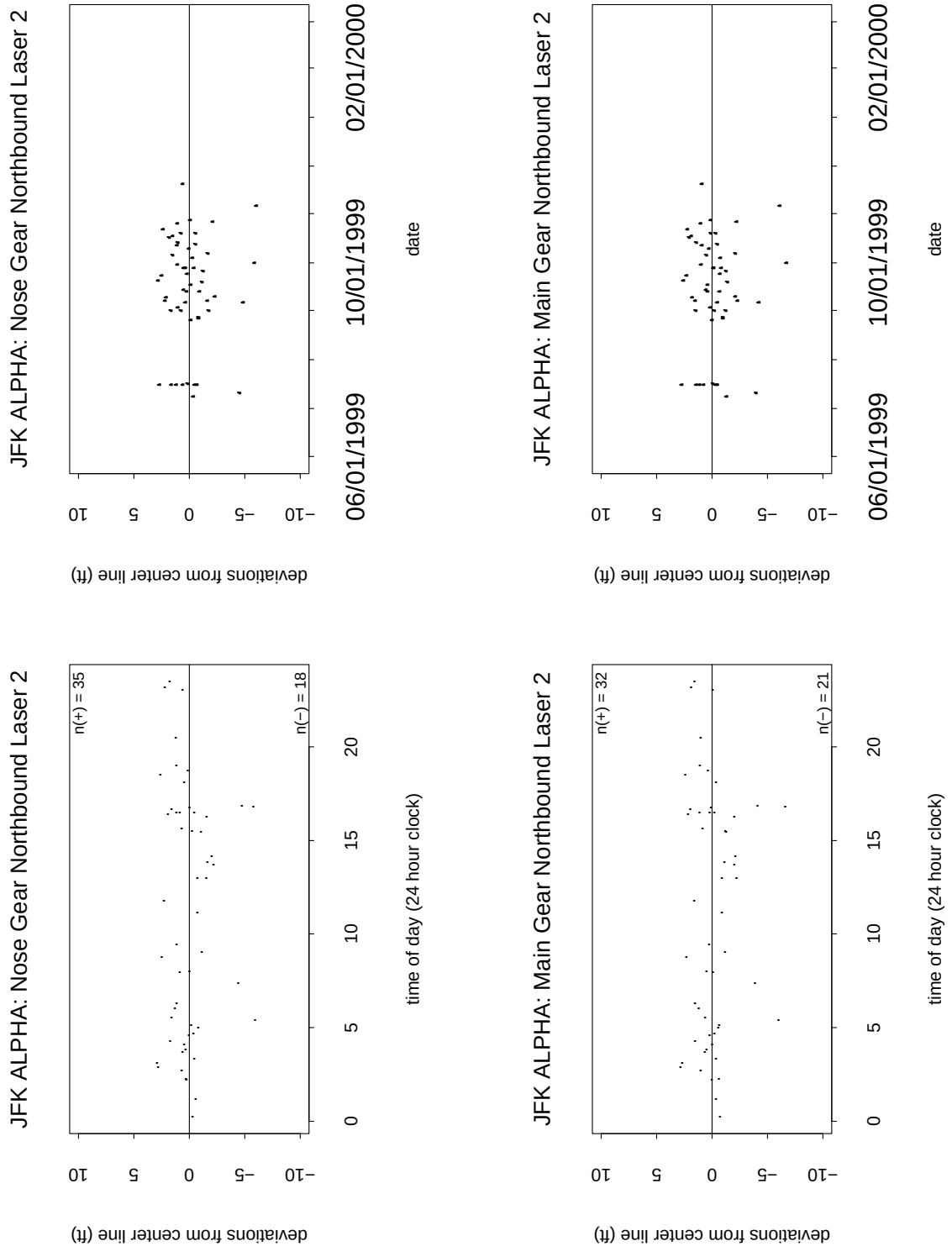


Figure 19: Time Plots for Nose & Main Gears at ALPHA Southbound, Laser 1

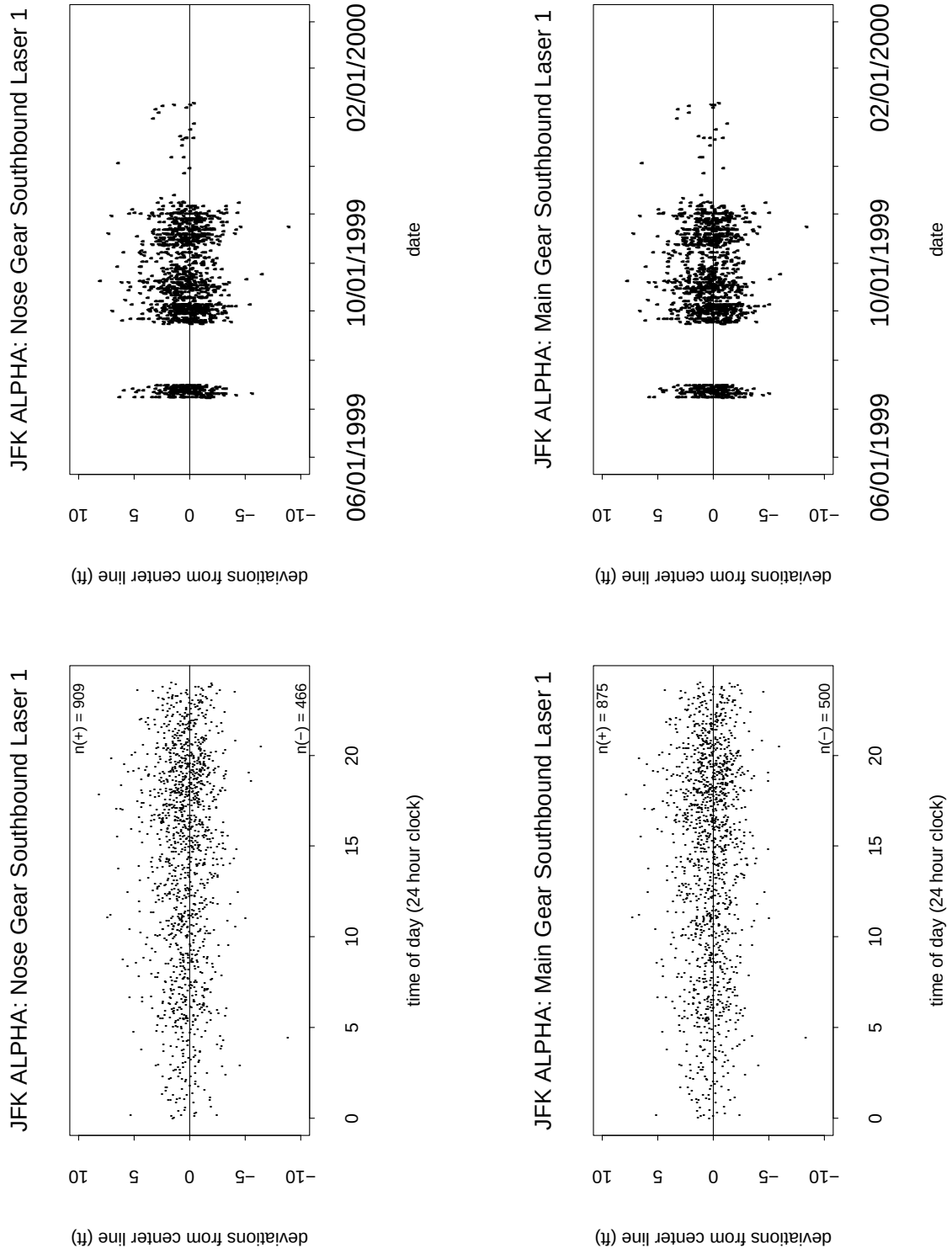


Figure 20: Time Plots for Nose & Main Gears at ALPHA Southbound, Laser 2

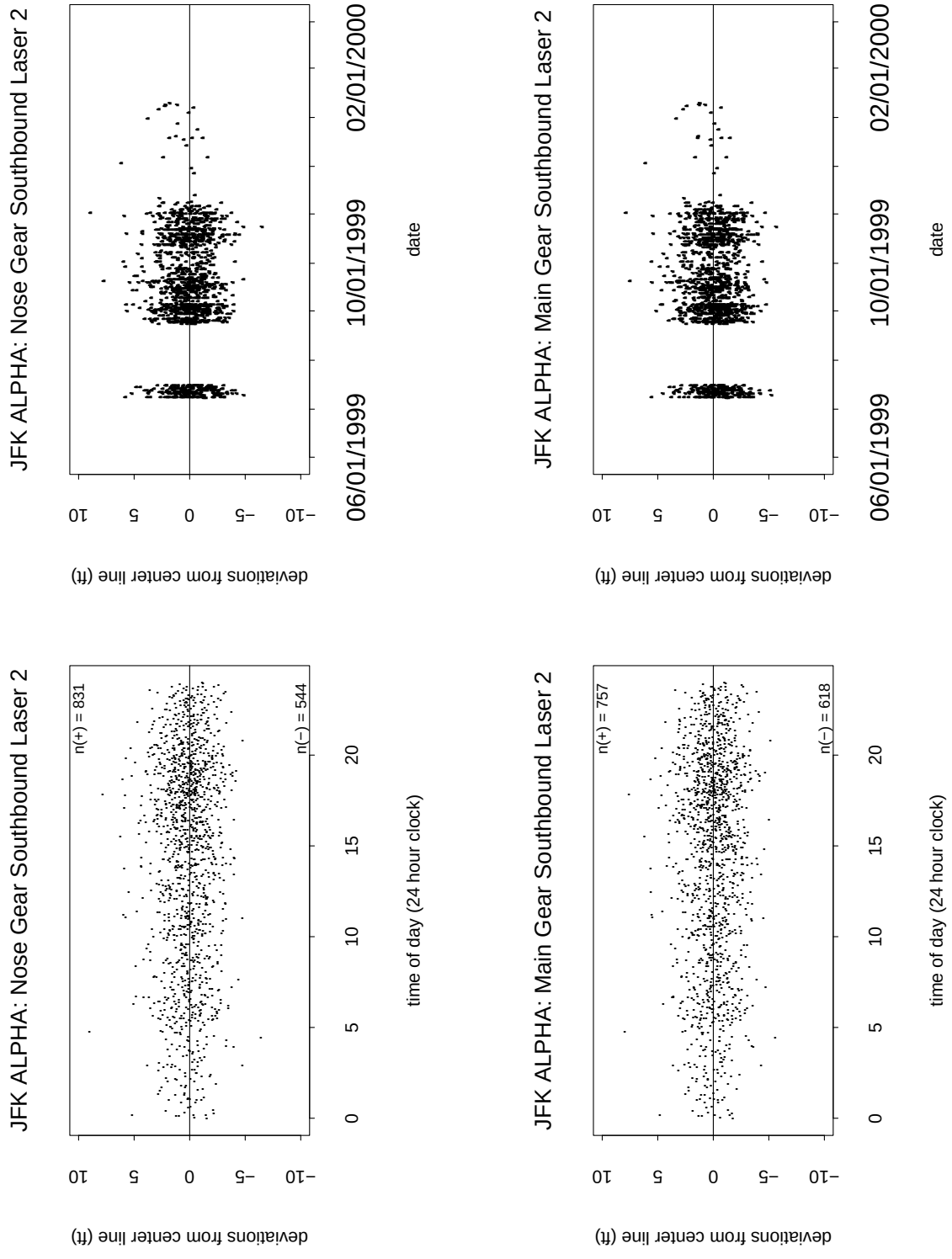


Figure 21: Time Plots for Nose & Main Gears at BRAVO Northbound, Laser 1

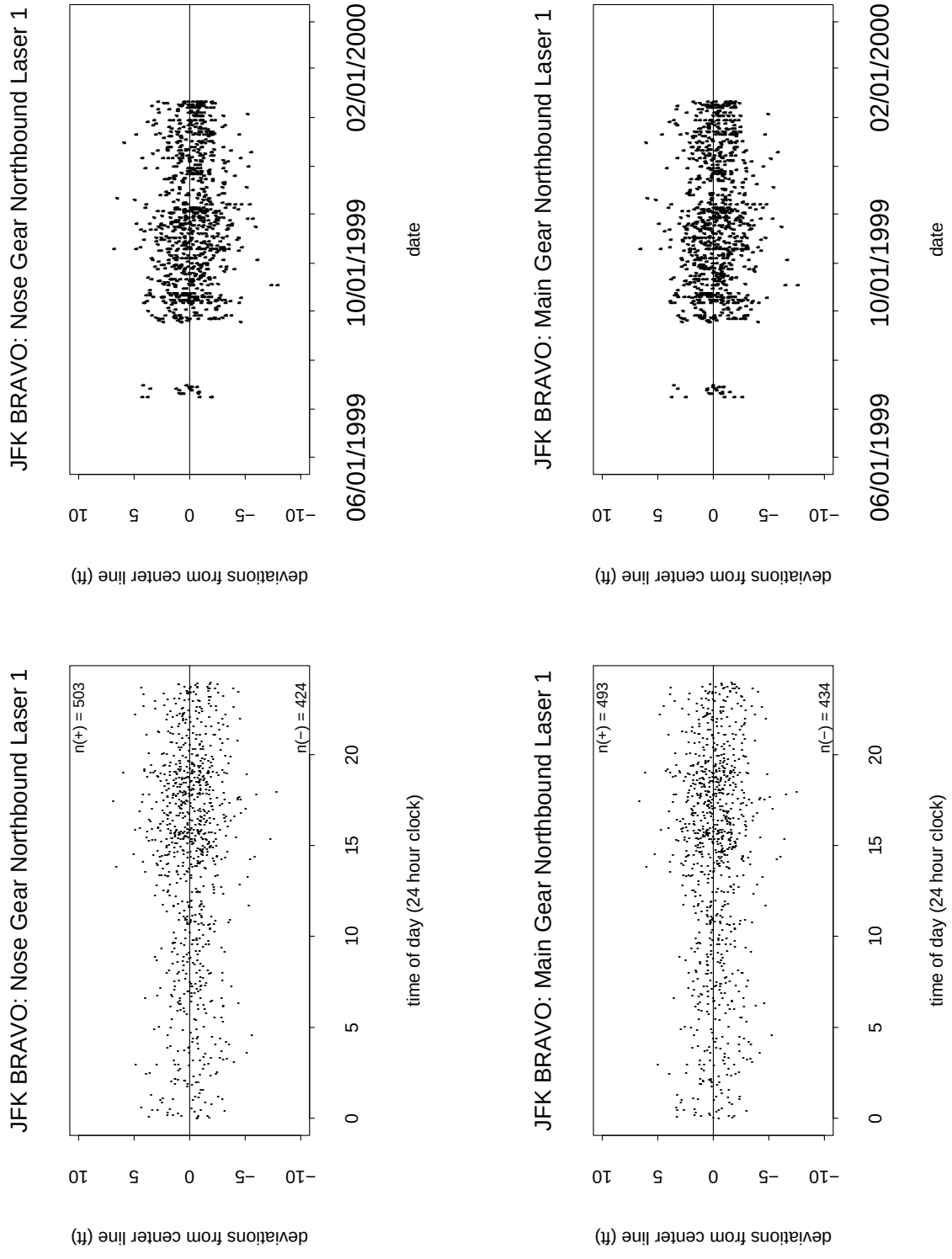


Figure 22: Time Plots for Nose & Main Gears at BRAVO Northbound, Laser 2

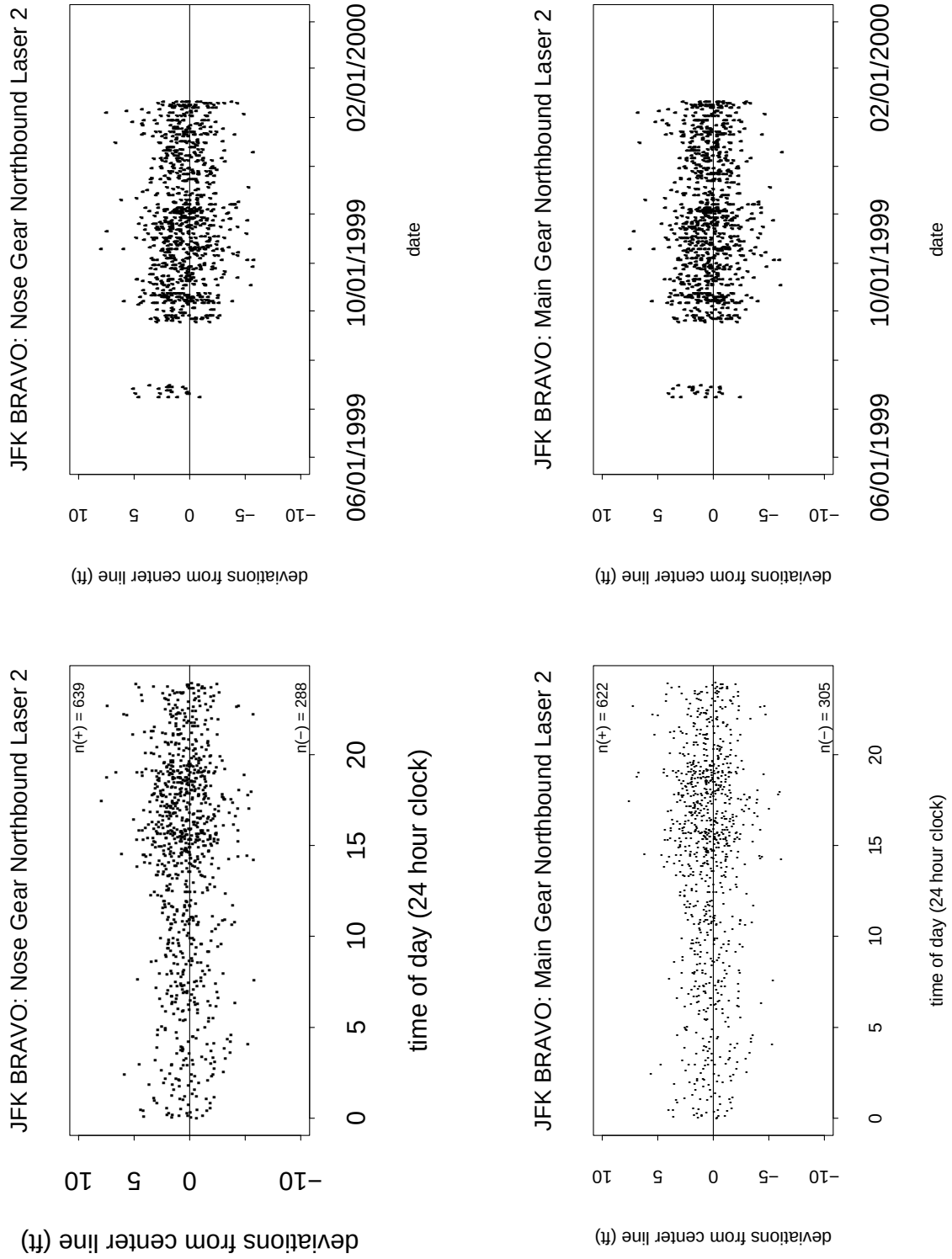


Figure 23: Time Plots for Nose & Main Gears at BRAVO Southbound, Laser 1

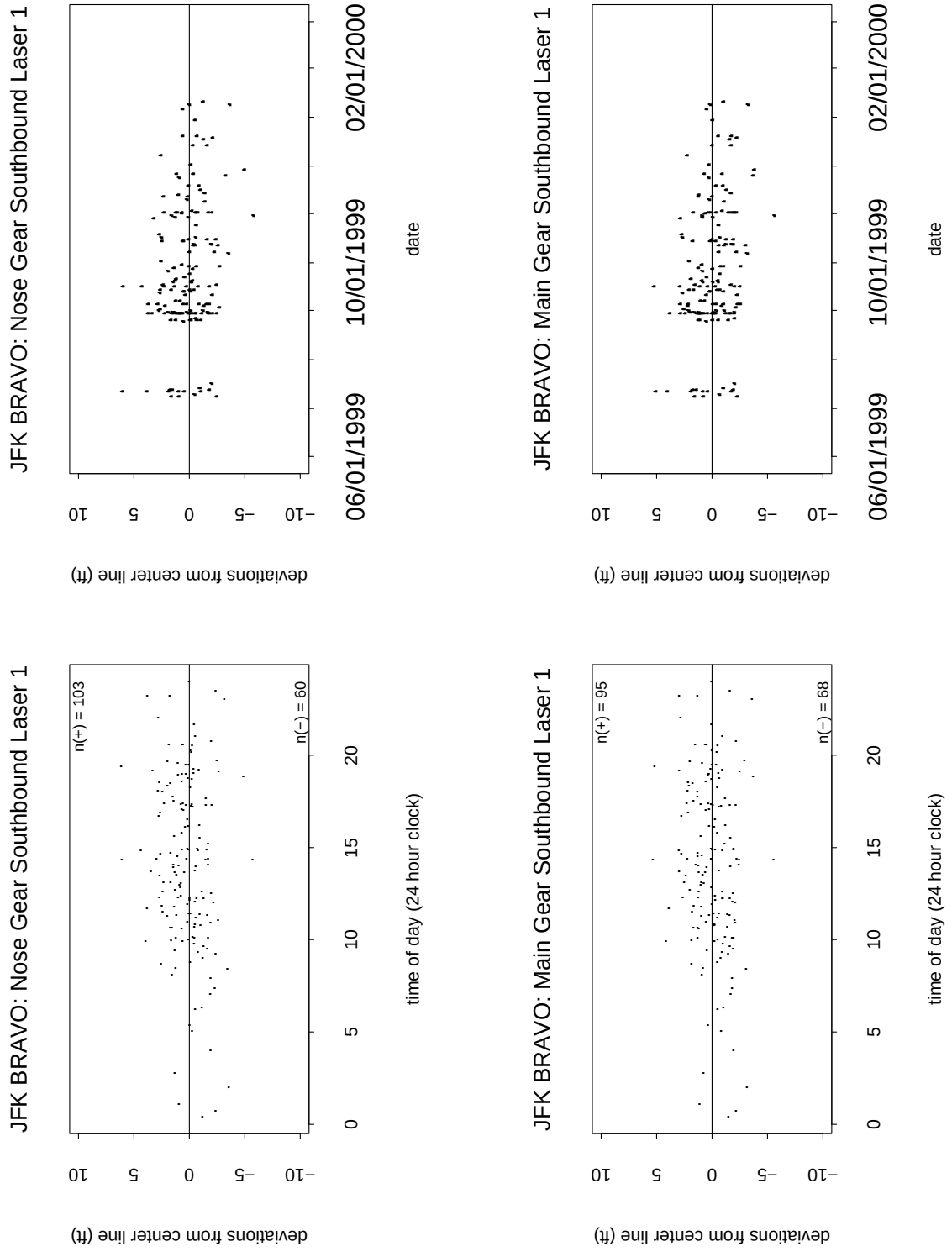
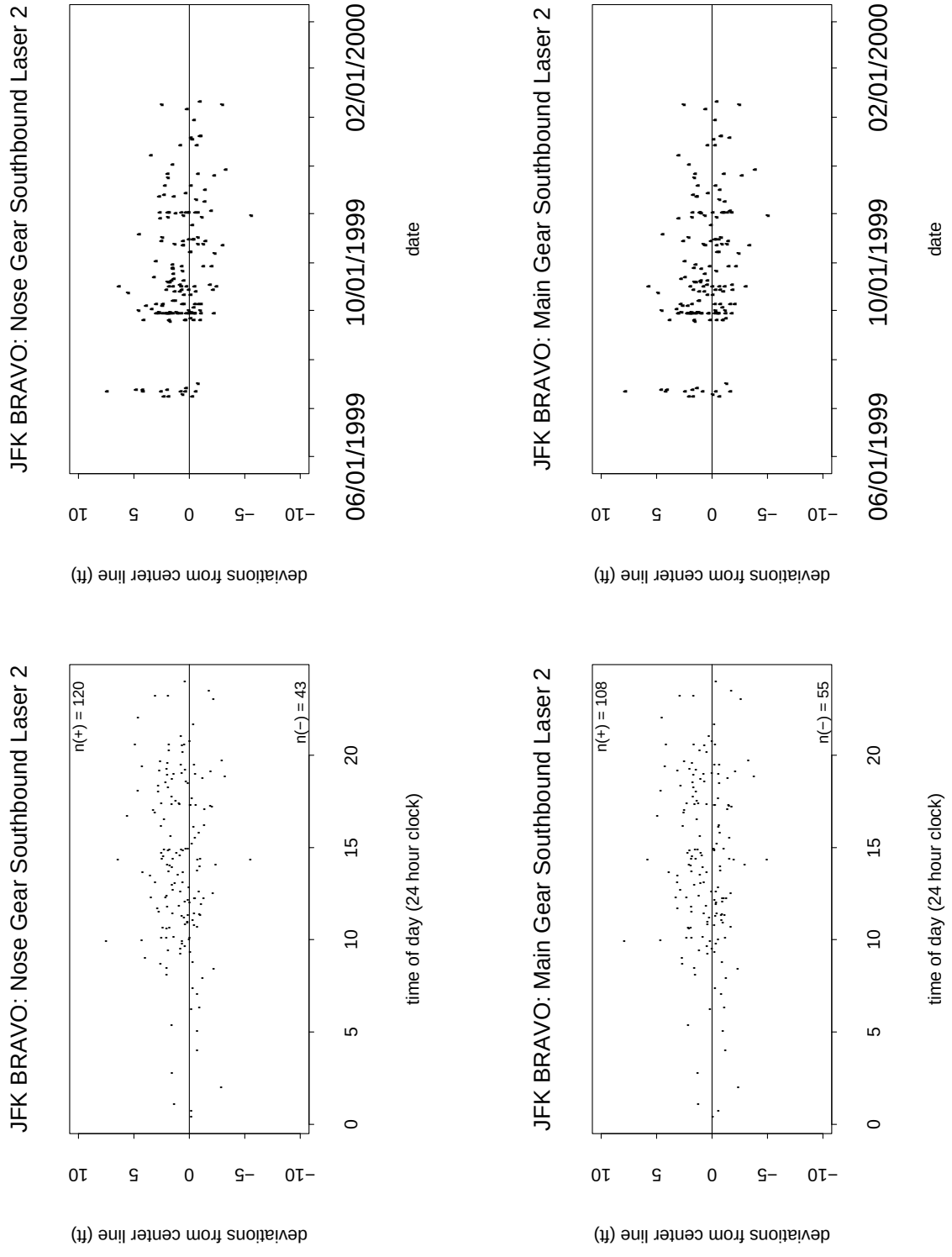


Figure 24: Time Plots for Nose & Main Gears at BRAVO Southbound, Laser 2



the lasers for ALPHA while for BRAVO the offset was in the direction toward the lasers, see Figure 1.

The three schematic views in Figures 25-27, provided by Ryan King, show the dimensional relationships between laser, centerline, centerlights, and the nose gear as appropriate for taxiway ALPHA. For BRAVO one would need to think of the laser as being on the other side of the centerline.

The taxiway lights are relatively flush with the pavement surface, less than $1/2''$ or so above the surface. They have a diameter of about $10''$ and are inset into the pavement. However, they do cause a "bump" effect to a tire, which the pilot may wish to avoid. It is reasonable to speculate that the pilots aim to taxi between the centerlights and the centerline stripe. The pilot could straddle the lights so that they pass between the nose gear tires. However, that would allow for only a very limited range of variation, much less than that which is experienced. It is more likely that the steering behavior with respect to the lights is a mixture, some proportion of pilots straddling them with the nose gear as in Figure 25, some staying on the near side from the laser as in Figure 26 (preferred choice for [7]) and some on the far side from the laser as in Figure 27. This mixture may not be symmetrically centered on the centerlights because of the centerline stripe which is supposed to serve as main guide. Since the lights are on different sides of the centerline for ALPHA and BRAVO we may see different bias effects in case there is another source of bias that points consistently to one side of the taxiway in relation to the laser position. This possibility was no present for the ANC data because there the centerlights were on the side away from the lasers for both taxiways. Another difference is the offset of the centerlights from the centerline, 21" at JFK and 12" at ANC.

5 Data Selection and Adjustments for Risk Analysis

For each event we have four deviations from the centerline, namely nose and main gear at each of two lasers. These readings are highly correlated, as demonstrated in the previous section. Since our risk extrapolation methodology is predicated on a sample of independent measurements we have to make a choice as to which of these readings we should choose for the risk analysis.

The risk analysis is to address the deviation risk for a straight taxiway segment. Therefore it was found most reasonable to take those readings that were least affected by possible prior or impending turning actions. Also, the main gear and the wings are most exposed to any extreme deviation risk, either in running off the taxiway or in colliding with fixed structures or with other aircraft on adjacent taxiways. In view of this it was decided to focus on the main gear deviations at Laser 1 for both ALPHA and BRAVO since Laser 1 is farther away from any turning opportunity than Laser 2. However, we point out that such turning effects were not observed for this data set.

We will treat these deviations separately by heading and taxiway to correct for possible parallax and other biases and in order to check for differences by heading and across taxiways.

Figure 25: Nose Gear Straddling Centerline Lights

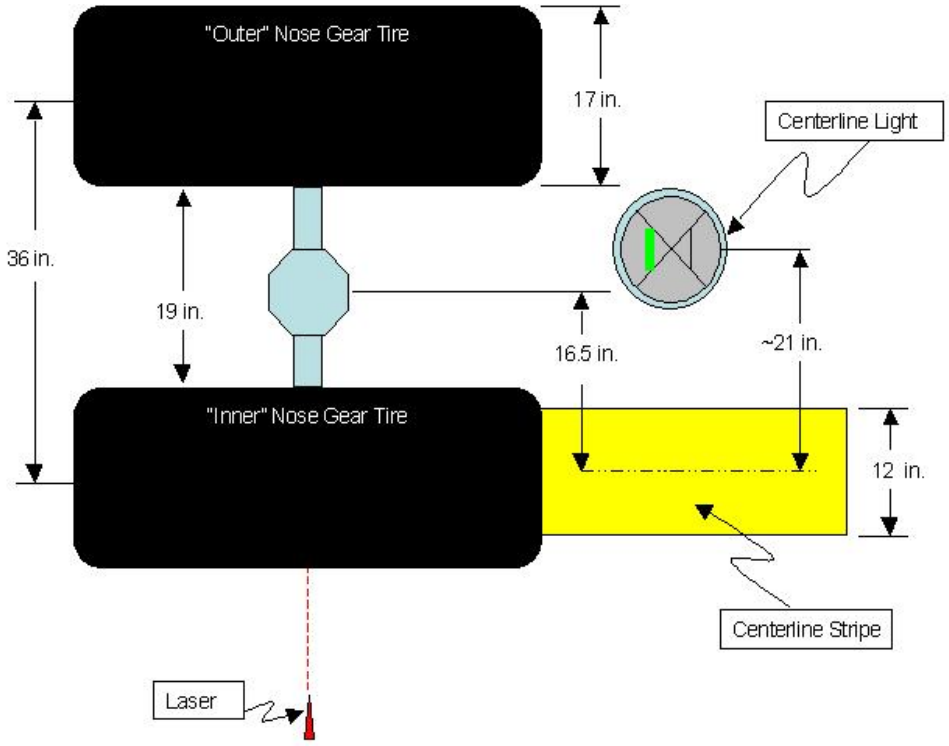


Figure 26: Nose Gear Not Straddling Centerline Lights, Close to Laser

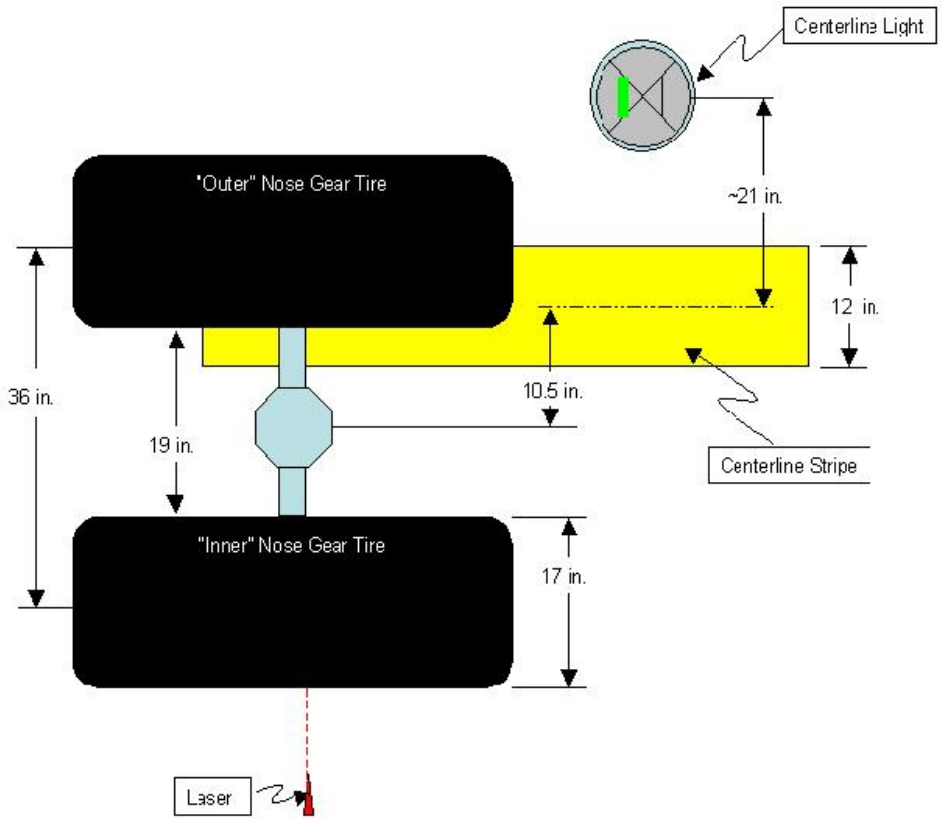
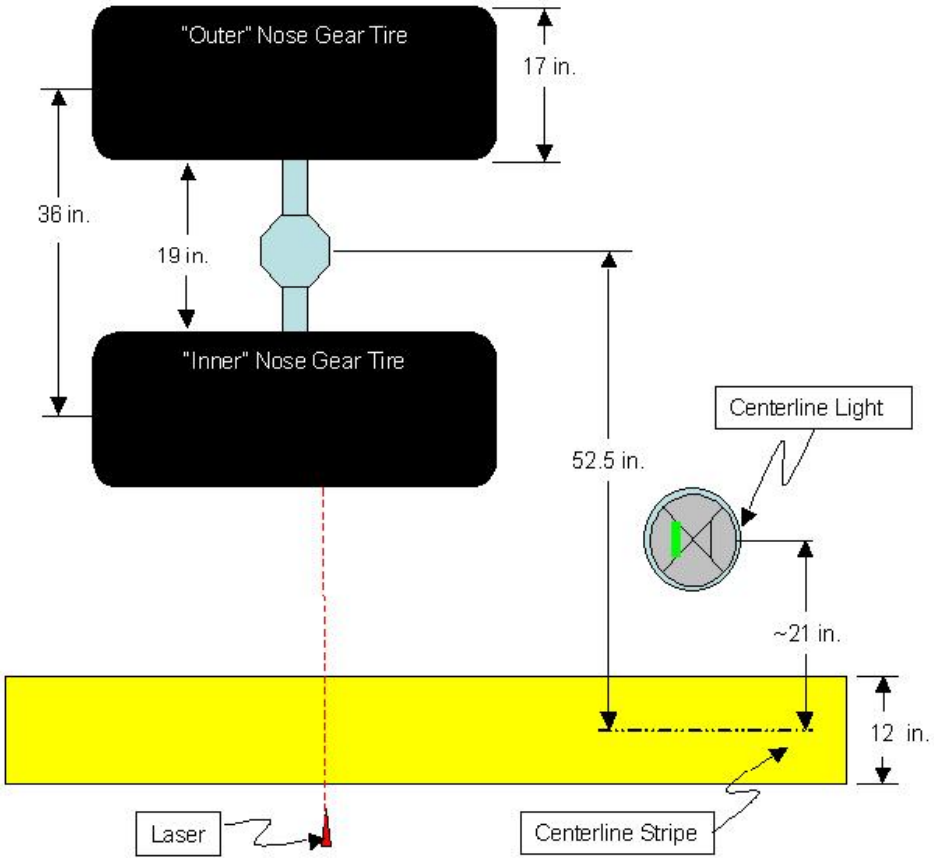


Figure 27: Nose Gear Not Straddling Centerline Lights, Far From Laser



It is hoped that the bias corrections will then lead to combined data that distribute more or less symmetrically around zero. This would then enable us to base the risk extrapolation on the data in both tails by using the absolute deviations, hence doubling the sample size that comes to bear on this issue.

As a first step we present in Figure 28 the four histograms of the main gear centerline deviations by heading at ALPHA and BRAVO. Shown by dots below the histograms are those extremes that exceed ±5 ft and in the upper left and right plot corners their number is indicated. Also indicated is the sample size n in each case and the mean (solid vertical line) and median (dashed vertical line) for each histogram, although one can hardly distinguish them. It is apparent that all means/medians are positive but a consistent pattern, as it was evident for the ANC data, does not seem to emerge.

If only a parallax bias were involved one would see opposite sign biases when going in opposite directions. Since biases are positive in all four cases other forms of biases seem indicated in the direction away from the lasers. Note that positive deviations are farther away from the laser than negative deviations. To sort out these biases and to correct for them we introduced the following measurement model for the four types of measurements:

$$A_{Nb} = \mu_P + \mu_O + e\ , \quad A_{Sb} = -\mu_P + \mu_O + e \quad B_{Nb} = -\mu_P + \mu_O + e\ , \quad B_{Sb} = \mu_P + \mu_O + e\ .$$

Here A_{Nb}, for example, stands for a measurement at ALPHA northbound (Nb), μ_P denotes the parallax bias, which changes sign with the heading, and μ_O is some other bias which does not change signs with the heading. Here the choice of signs on the μ_P contribution to the bias is motivated by the consideration that on ALPHA northbound the pilot views the centerline/centerlights arrangement the same way as on BRAVO southbound. Finally, e denotes a generic random centerline deviation term, that is without bias and is assumed to have a distribution that is symmetric around zero or at least to have a mean of zero. One can estimate μ_P and μ_O separately for each taxiway as follows from the averages ($\bar{A}_{Nb}, \bar{A}_{Sb}, \bar{B}_{Nb}$, and $\bar{B}_{Sb}$) of these four groups of data:

$$\text{for ALPHA:} \quad \widehat{\mu}_{P,A} = \frac{1}{2}\left(\bar{A}_{Nb} - \bar{A}_{Sb}\right) = .04\text{ft} \quad \text{and} \quad \widehat{\mu}_{O,A} = \frac{1}{2}\left(\bar{A}_{Nb} + \bar{A}_{Sb}\right) = .66\text{ft}$$

and

$$\text{for BRAVO:} \quad \widehat{\mu}_{P,B} = \frac{1}{2}\left(\bar{B}_{Sb} - \bar{B}_{Nb}\right) = .10\text{ft} \quad \text{and} \quad \widehat{\mu}_{O,B} = \frac{1}{2}\left(\bar{B}_{Nb} + \bar{B}_{Sb}\right) = .24\text{ft}\ .$$

In contrast, at ANC the two parallax bias estimates were .24ft and .29ft, repectively. This difference in parallax bias may well be the result of the different offsets of the centerlights from the centerline. The estimates for the other bias at ANC were .58ft and .42ft, respectively. The difference between these two values was smaller than the one observed for JFK. At this point we can only speculate that the larger difference between .66ft and .24ft results from a combination of various factors: i) the different offsets of the centerlights, ii) the centerlights are on opposite sides of the centerline, and iii) the taxiways are parallel to each other. We

note that on ALPHA the traffic was predominantly to the south while on BRAVO it was mainly to the north. Thus the centerlights would mostly appear to the left of the centerline from the pilot's point of view. It would seem that the position of the centerlights relative to the centerline may therefore be ruled out as a factor. This seems to agree with the findings at ANC where the centerlights were on the same side of the centerline for either taxiway (ROMEO and KILO) and while the traffic on KILO was predominant in one direction it was heavy in either direction on ROMEO. At this point we don't have a clear explanation for the different bias behaviors at ANC and JFK.

Figure 29 shows the deviation data after the biases were subtracted out. Note that the new averages are essentially zero. After subtracting out these biases from each of the four groups of data we also examined the symmetry around zero for each group using QQ-plots of the negative lower half of the each sample versus the upper half, as explained in detail in [10]. This was done for each bias corrected data set and the results are shown in Figure 30.

Given the natural sampling variation in the extremes it would be unlikely for the extremes in either tail of a sample to be close negatives of each other. For the not so extreme observations in each sample half the symmetry appears quite reasonable, although for ALPHA southbound the pattern appears to drift away from the diagonal.

One needs to be careful before reading too much meaning into such observed features. To put things in perspective and as we did in [10], we again generated samples of corresponding sample sizes from a perfectly symmetric distribution, namely the standard normal distribution, and performed similar diagnostic symmetry plots. The results are shown in Figure 31. The q in the upper left of each plot is half the sample size and indicates the number of plotted points. From these plots it should be clear that our observed deviation features (even the noted drifting away) could equally well have arisen due to sampling variation and not necessarily from any true asymmetry in the sampled distribution.

Next we examined whether the heading had any effect on the bias corrected deviations. Thus we compare the deviations from one direction with those in the opposite direction. Since the sample sizes in the two directions are not the same we employ a QQ-plot, details of which were explained in [10].

Such a comparison was done for each pair of opposing headings at ALPHA and BRAVO and is shown in the top two plots of Figure 32. The point patterns seems reasonably close to the main diagonal so that it makes sense to pool the adjusted deviation data for the two headings at each taxiway. The two bottom plots in Figure 32 show the symmetry check for these combined deviations for each taxiway. Symmetry seems to be very reasonable for the full range at BRAVO and at ALPHA for absolute deviations below 3 ft. Above that one sees higher positive than absolute negative deviations. At this point it is difficult to say whether such deviations from the main diagonal are statistically significant.

The top left plot in Figure 33 compares the deviation data (combined by heading) from BRAVO with the corresponding data from ALPHA using a QQ-plot. The point pattern follows the main diagonal reasonably well, suggesting that the deviation data from both taxiways may be combined into one overall sample of size $n = 1090 + 1428 = 2518$ for

Figure 28: Main Gear Centerline Deviations for Laser 1 at ALPHA & BRAVO

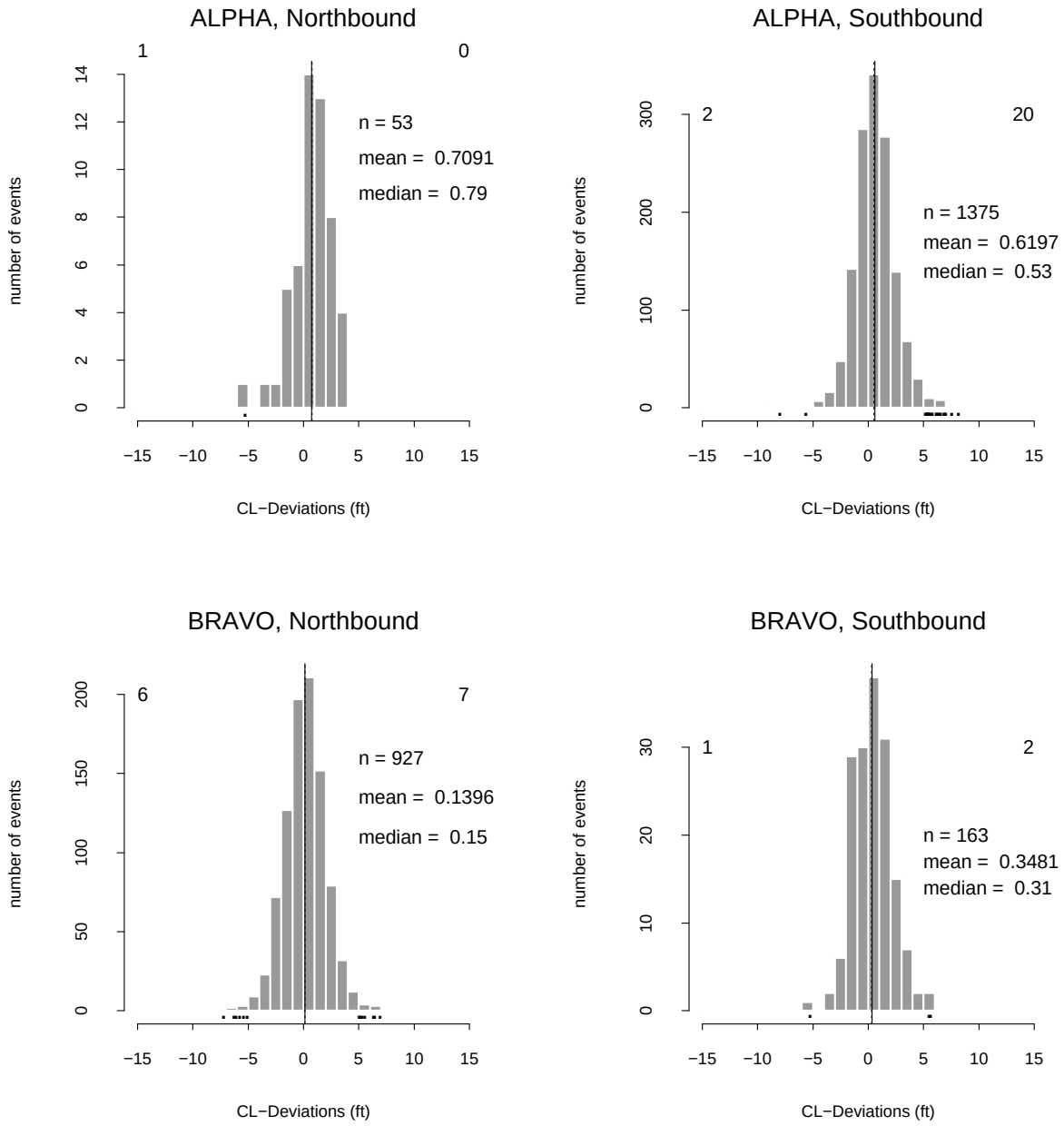


Figure 29: Adjusted Main Gear Centerline Deviations for Laser 1 at ALPHA & BRAVO

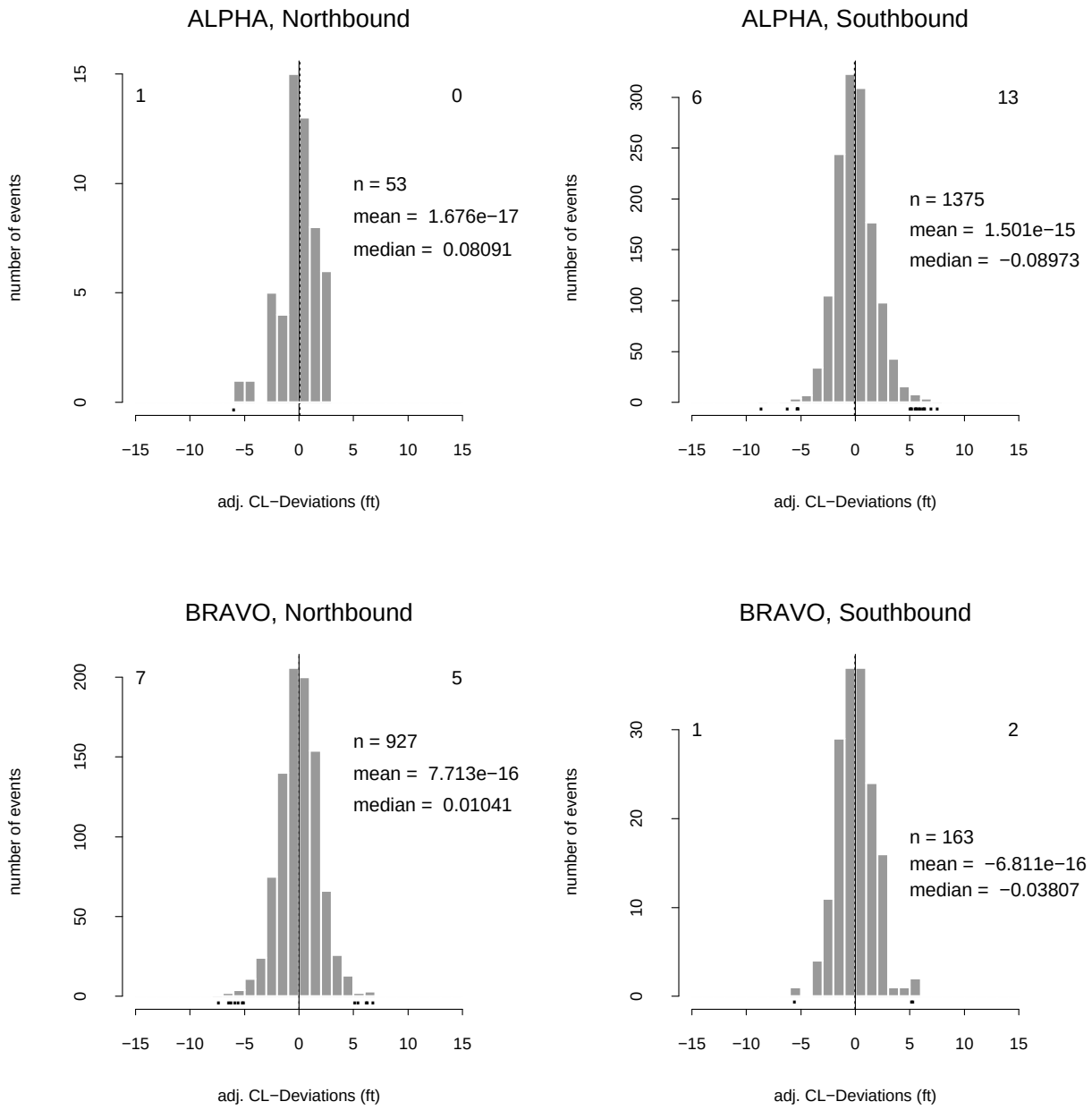
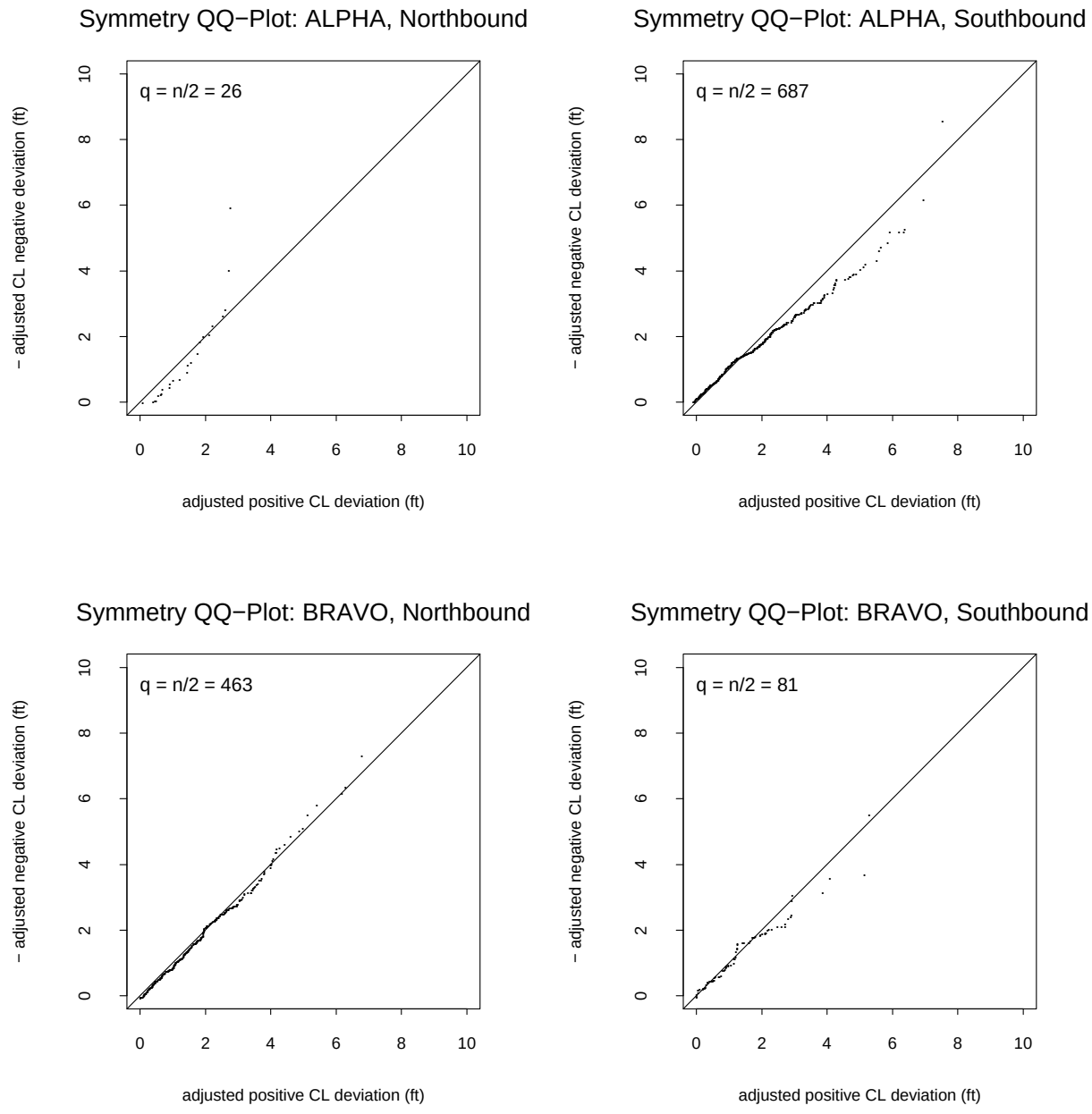


Figure 30: Testing Symmetry of Main Gear Centerline Deviations
for Laser 1 at ALPHA & BRAVO by Heading



purposes of risk extrapolation. The top right plot of Figure 33 examines the symmetry issue for this combined data set and mild deviations from the diagonal appear only at the high end. The Anderson-Darling two-sample test [8] applied to the absolute negative deviations as one sample and the positive deviations as the second sample gave a p-value of 0.211, which is not significant. Thus one can reasonably accept the symmetry hypothesis.

The histogram in the lower left of Figure 33 represents the combined sample of $n = 2518$ adjusted main gear deviations from ALPHA and BRAVO at laser 1. Deviations beyond ± 5 ft are indicated by dots below the histogram and their number is noted in the upper left and right corners, respectively.

The QQ-plot in the lower right of Figure 33 examines whether the combined sample of $n = 2518$ deviations follows a normal distribution. If this were the case then the ordered deviation values, $x_1 \leq \ldots \leq x_n$, when plotted against corresponding normal p_i-quantiles ($p_i = (i - .5)/n, i = 1, \ldots, n$) should show a roughly linear pattern.

Such linear behavior is indeed the case within the ± 1.5 sigma portion of the normal distribution. However, beyond that (corresponding to deviations beyond $2.5 - 3$ ft) the actual observations tend to fall further outward from the centrally fitted straight line. Hence the tails of the combined sample are heavier (reach out more) than indicated by a normal model.

The Anderson-Darling test for normality applied to the combined, bias adjusted set of taxiway deviations yields a p-value of $3.2\ 10^{-8}$. Thus the normality assumption can be clearly rejected. Since the QQ-plot for normality indicates that the deviations from normality manifest mainly in the tails, suggesting larger taxiway deviations than indicated by the normal distribution, it would result in overly optimistic results when using the normal distribution for tail extrapolation purposes.

6 Risk Extrapolation

6.1 Overview

We now focus on the combined set of adjusted main gear centerline deviations from Laser 1 at both ALPHA and BRAVO, as portrayed in the lower left histogram of Figure 33. The range of these $n = 2518$ adjusted deviations is $[-8.63, 7.53]$ ft. Noteworthy here is the more than 1 foot difference in the absolute values of the range endpoints. It underscores the fickleness in the behavior of extremes.

The primary concern is to assess the risk of exceeding thresholds beyond this observed range. Conversely one can ask for the deviation thresholds that are exceeded with various specified small risks, say $10^{-5}, 10^{-6}, 10^{-7}$. Methodology for this was developed in [9] and other methods are discussed in the recent reference [3]. In essence all of these methods deal with the k most extreme observations in either tail of a given random sample, $X_1, \ldots, X_n$. A major question is how large to choose k. Too large a k will introduce influences from the center of the data that have little to do with the extreme behavior of such deviations and thus could lead to bias. Too small a k will leave us open to the typically strong fluctuations

Figure 31: Simulated Symmetry Test Plots from Standard Normal Samples for Laser 1 at ALPHA & BRAVO by Heading

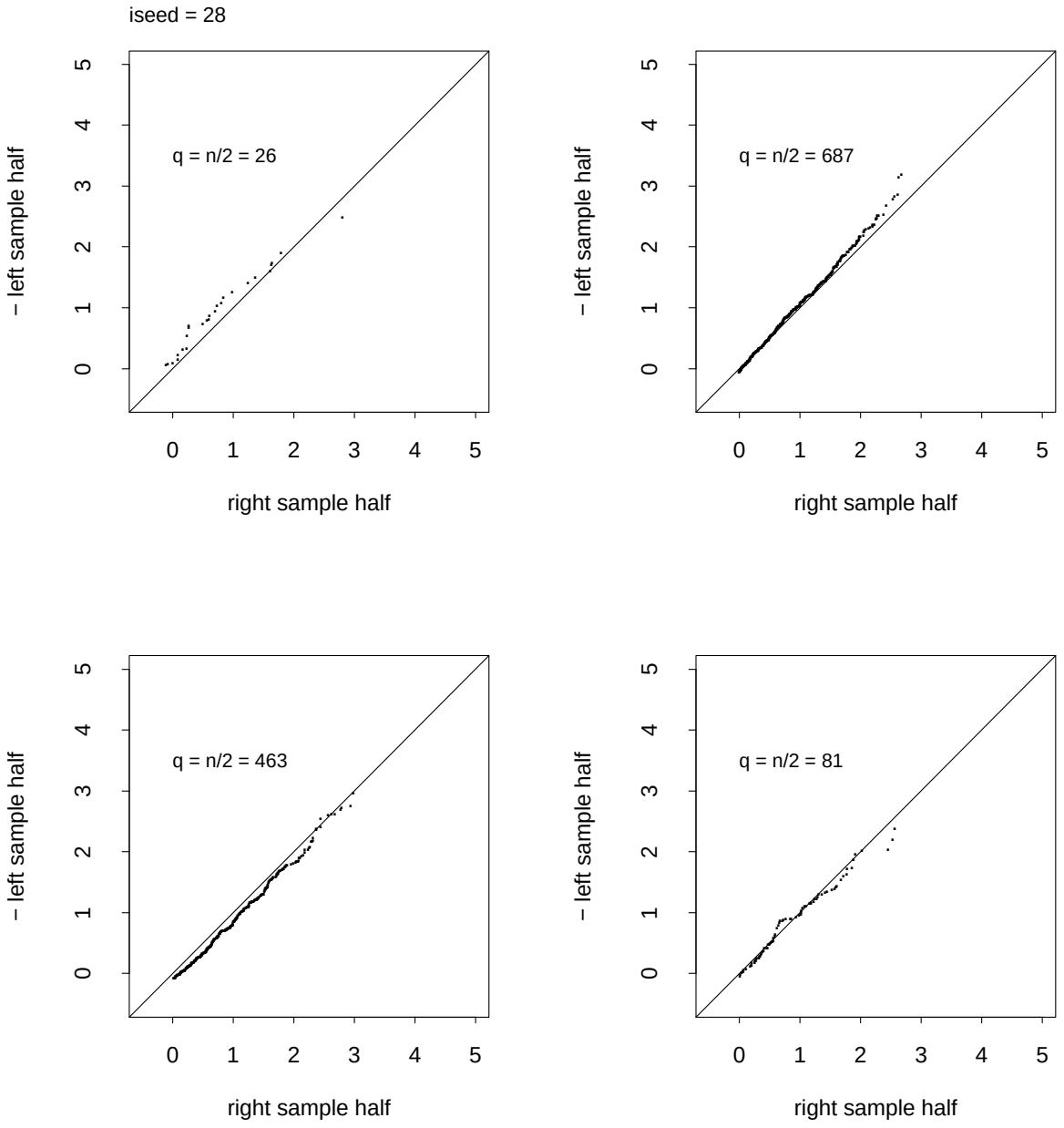


Figure 32: Testing Equality by Direction and Symmetry by Taxiway
of Main Gear Centerline Deviations for Laser 1 at ALPHA & BRAVO

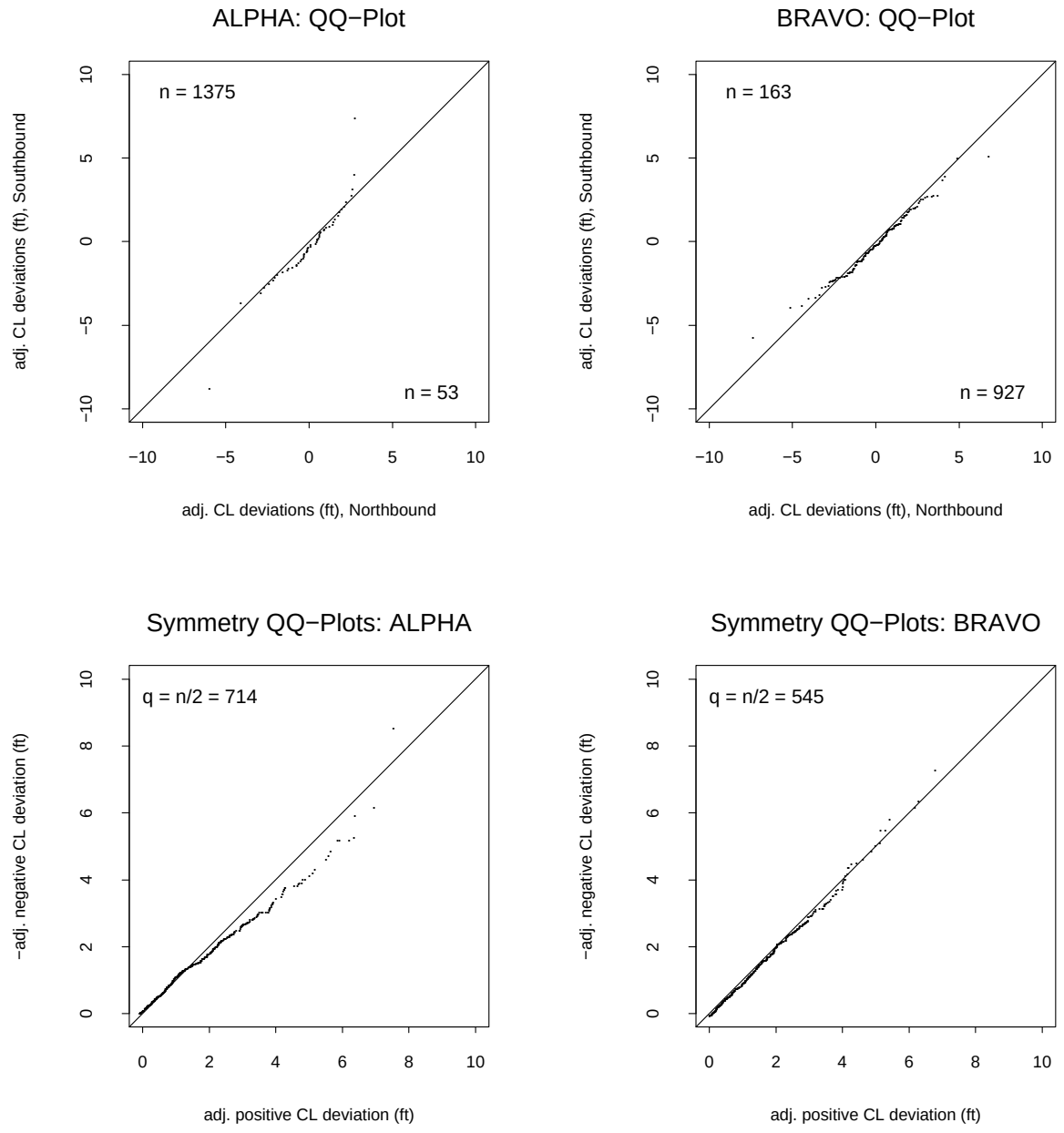
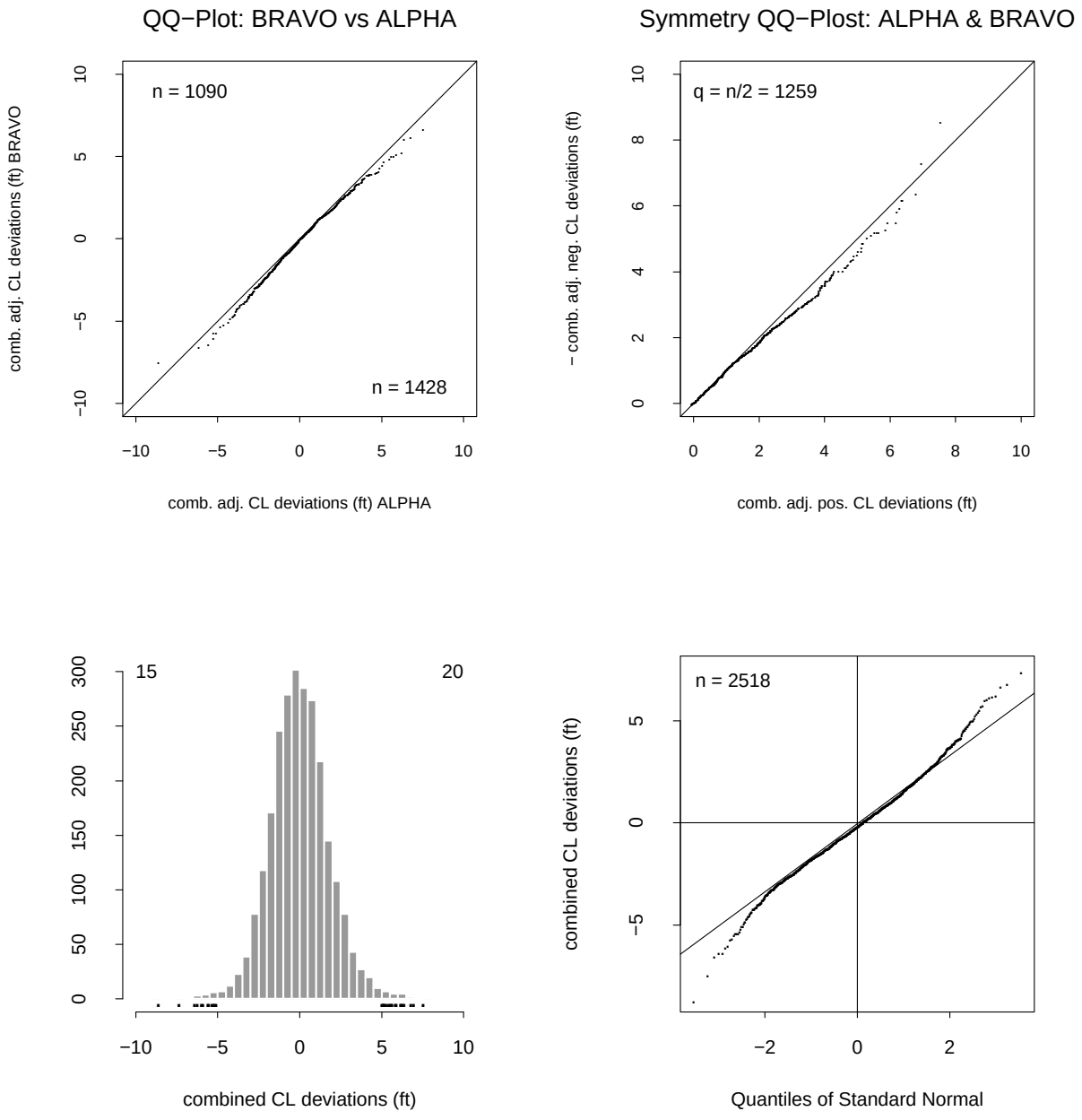


Figure 33: QQ-Plot of ALPHA vs BRAVO, Symmetry Plot for ALPHA & BRAVO Combined
Histogram and Normal QQ-Plot for ALPHA & BRAVO Combined
Main Gear Adjusted Centerline Deviations for Laser 1 at ALPHA & BRAVO



in the tail of the data and will thus result in too much uncertainty concerning our risk extrapolations. Although [9] proposed a method for choosing k we will not use it here. Instead we take a close look at the tail behavior of the data and make a judgment call on the proper choice of k.

All of the extreme value extrapolation methods rely on a basic limiting assumption, namely that the most extreme data points (maximum or minimum) behave in a particular way for large sample sizes. For the maximum $M_n = \max(X_1, \ldots, X_n)$ it can be stated as follows:

> **Extreme Value Limiting Assumption:**
> There are deterministic sequences of normalization constants $\{a_n > 0\}$ and $\{b_n\}$ so that for all z
>
> $$P\left\{\left(M_n - b_n\right)/a_n \leq z\right\} \longrightarrow G(z) \quad \text{as } n \to \infty$$
>
> for some cumulative distribution function G.

If this convergence holds then $G(z)$ can only be of the following form

$$G(z) = \exp\left\{-\left[1 + c\left(\frac{z-\mu}{\sigma}\right)\right]^{-1/c}\right\}$$

for all z with $1 + c(z-\mu)/\sigma > 0$. The parameters μ and $\sigma > 0$ act as location and scale parameters and the parameter c is called extreme value index or shape parameter. These three parameters allow for great flexibility in modeling the distributional behavior of M_n in large samples and thus also the right tail behavior of the sampled distribution function F since

$$P(M_n \leq x) = F^n(x) \; .$$

To explain the connection between the distributional behavior of M_n and that of the k largest sample values we digress briefly. Originally the above distributional limit result was used as follows. View the sample $X_1, \ldots, X_n$ as a collection of k subsamples of size m where $k \times m = n$. If the original n cannot be cleanly subdivided into k subsamples of size m one may have to discard some observations. If m is large enough one may regard the k maxima from the k subsamples as having roughly the above limiting distribution. Thus one could model the sample of k maxima by that distribution and estimate its three parameters by the method of maximum likelihood. If k is small one deals with the issues of greater variability/uncertainty accompanying small samples. If m is not large enough then the approximating distribution G may not be a good approximation yet and thus there will be some bias which persists no matter how large k is for fixed m. Given that n is fixed and we have to work within the constraint $n = k \times m$ we are stuck between the trade-off of bias and variability.

This view and approach through subsamples has a disconcerting aspect. Namely, how should the allocation of the original sample of size n into k subsamples of size m be done? Furthermore, different allocations would lead to different analysis results. Given that one wants to

treat the k maxima from these k subsamples as a random sample (independent) one would have to insist on a random subdivision of $X_1, \ldots, X_n$ into k subsamples. This would render the final analysis results subject to extraneous randomness, i.e., different analysts would arrive at different answers using the same method but not the same random number generator and/or random seed.

Typically the k maxima so obtained will come fairly close to the k largest observations $Y_1 \geq Y_2 \geq \ldots \geq Y_k$ in the original sample of size n. It is likely but not guaranteed that a random subdivision of $X_1, \ldots, X_n$ into k subsamples will allocate one of those $Y_1 \geq Y_2 \geq \ldots \geq Y_k$ to each of these k subsamples. When that does not happen one uses a lesser extreme in the analysis and one does not make most efficient use of the extreme data in the original sample. This concludes our digression.

In determining the proper tail depth k we employ several approaches. In the first two we compute two types of estimates of the extreme value index c for various values of k and examine when, in terms of k, the variation of these estimates around some level transitions into a drifting off behavior. A third diagnostic examines whether an expected consequence of the assumed extreme value limiting assumption does show up in the data to a reasonable tail depth.

We note here that the above domain condition $1+c(z-\mu)/\sigma > 0$ for the limiting distribution $G(z)$ implies that its upper range is unlimited when $c \geq 0$ and that it is limited by $z < \mu - \sigma/c$ when $c < 0$. In the latter case one therefore models a hard upper limit for the behavior of $(M_n - b_n)/a_n$, and thus presumably for M_n since one uses the approximation for some reasonably large n for fixed a_n and b_n. However, a_n and b_n could be quite large so that for practical purposes this upper bound could still be quite far off and appear as infinite. In any case, we thought it worthwhile to point out this special status of $c < 0$, in particular with respect to the current application of 747 deviations from the taxiway centerline because of the somewhat limiting effect of a taxiway edge when it is perceived by the pilot.

After having fixed on the proper number k of extremes we then ran the EXTRAP program on the respective number of extremes in the adjusted deviation sample. The resulting extrapolated deviation thresholds were then tabulated for various specified risk levels.

6.2 Examination of Tail Depth *k* Using the Hill Estimator

In [9] the modified Hill moment estimator, based on the k largest sample values, was used to estimate the extreme value index c. Another estimate, based on the exceedances of sample values over a given threshold u, is discussed in [3], which also cites a web site for freely available software.

We looked at both methods of estimation to see when the estimates for c start to deteriorate. This may then give us an indication of how large a k to choose or how high a threshold u to employ. Determining this point of deterioration is somewhat subtle. Typically the extreme value index estimates will fluctuate strongly when based on small values of k. As k gets larger these estimates settle down to a more stable value until they start to drift off to possibly

other levels. Sometimes this first settling down range of k is short and may be difficult to distinguish from the wild fluctuation and the drifting off behavior.

A preferred way of viewing these estimates as a function of k is to plot k on a logarithmic scale. Such plots for the modified Hill estimator are shown in Figures 34-36 on pages 51-53, where we took either the k extremes in the right (positive) tail of the data or the k extremes in the left (negative) tail of the data or the k extremes in the right tail of the absolute data. The latter allows us to double the amount of data characterizing the tail behavior of the taxiway deviations. The use of this strategy seemed indicated by the apparent symmetry examined earlier. The reason for also looking at the individual tails was to see whether there is some consistency with what one finds with the absolute deviation data. For the individual tail data we subtracted the median from the data prior to using the modified Hill estimator (not that it changed much) but for the absolute data we did not. For a different perspective we replotted these estimated extreme value indices also against a straight scale for k (not logarithmic) in Figures 37-39 on pages 54-56.

In examining the proper tail depth k for Figure 34 (positive deviations, right tail data) one notes for very small $k \leq 18$ strong fluctuations of the estimated extreme value index around no particular horizontal level. After that, for $18 \leq k \leq 30$, one sees fluctuations around some horizontal level near $c = -.32$ after which there is sharp drop followed by a steep rise toward a new horizontal level at roughly $c = -.06$ for $50 \leq k \leq 96$. This is followed by another drop and rise to a new level near $c = .02$ for $140 \leq k \leq 390$, and so on.

The difficulty in selecting the appropriate k is in deciding when fluctuations oscillate sufficiently long around "some level" (to be able to call it a level) and further when they start drifting away from that level to possibly another level or for good. Another difficulty is in deciding what that "some level" should be, given that we need to select it within possibly strong fluctuations. Our view is that the first and second horizontal level of $c = -.32$ and $c = -.06$ appear to be reasonably central choices for the first two levels. Ordinarily one would choose the first level $c = -.32$, but since it is strongly negative it would imply a hard upper bound on the deviations, as will be confirmed later. Furthermore, that level is maintained only for a relatively short range of $18 \leq k \leq 30$. Therefore we also entertain the second level of $c = -.06$. This is still negative, implying again a hard upper bound but probably much further away for practical purposes.

Based on these level choices for the estimated extreme value index one finds that $k = 30$ and $k = 96$ are roughly the largest values producing these level values (actually we find $c = -.3046$ for $k = 30$ and $c = -.06216$ for $k = 96$). Recall that we want to pick k large enough to downplay the uncertainty of the fluctuations, but not too large to be affected unduly by the possible bias from the middle of the data. The choices of $k = 30$ and $k = 96$ can be examined from a different perspective in Figure 37 without the logarithmic scale of k.

The fact that we have several apparent levels could be influenced by the centerlight offset from the centerline and it could also result from the taxiway edge acting as an inhibitor beyond some deviation level. The latter may not be experienced often in the current data

set and it may well be reflected by the first level choice of $c = -.32$. Because of the sparcity of evidence we thought it prudent to also consider the next level.

The pattern in Figure 35 (negative deviations, left tail data) seems similar to that in Figure 34. Here the strong fluctuations around no particular level appear to last up to about $k = 19$. For $19 \leq k \leq 31$ the extreme value index estimates appear to fluctuate around a level of $c = -.16$ after which there is a strong rise to another level of about $c = 0$ maintained for $36 \leq k \leq 77$. Again, for reasons explained previously, we chose to consider both levels as viable possibilities, resulting in choices of $k = 31$ and $k = 77$ with respective actual extreme value index estimates of $c = -.1485$ and $c = .005792$. For a different perspective see also Figure 38 without the logarithmic scale of k.

The pattern in Figure 36 (absolute deviations, right tail data) appears to show a first fluctuation level near $-.18$ for $27 \leq k \leq 52$. This is followed by a strong dip and rise to a level near $-.04$ for $85 \leq k \leq 200$. Again we chose to consider both levels as viable possibilities, resulting in choices of $k = 52$ and $k = 200$ with respective actual extreme value index estimates of $c = -.1766$ and $c = -.02774$. For a different perspective see also Figure 39 without the logarithmic scale of k.

6.3 Examination of Tail Depth k Using the Excess Over Threshold Estimator

Next we examine the behavior of the estimates for the extreme value index when these estimates are based on deviation exceedances over the threshold u, as explained in [3]. We used the software available on the website referenced in [3]. These estimates were plotted against the respective thresholds u in Figures 40-42 on pages 57-59 for the positive and negative deviations, and for the right tail of the absolute deviations over the threshold u.

At the top of each plot one can also see the number of sample values exceeding the respective thresholds. These numbers allow us to view these estimates also in terms of k, our previously employed number of extremes used in the modified Hill estimator.

For the negative deviations we changed their sign and took again a positive threshold u. Also shown around each index estimate is an approximate 95% confidence interval which gives some measure of the uncertainty in each estimate, explaining the stronger fluctuations for high u or for small k. Note that the fluctuations in these estimates don't appear as strong as for the modified Hill estimator. However, that is just a function of the larger ordinate scale which was chosen to accommodate the confidence intervals. To illustrate this we also plotted the estimates without the confidence bounds with a much reduced ordinate scale in Figures 43-45 on pages 60-62. The fluctuation behavior is then seen to be quite similar to that seen for the Hill estimator.

In Figures 40 and 43 we see levels of stability near $c = -.42$ for $u \in [4.3, 5.2]$ and around $c = -.15$ for $u \in [3, 4]$. The corresponding number of exceedances, as indicated at the top edge of the plots, are $18 \leq k \leq 34$ and $51 \leq k \leq 119$. These ranges are not too different from what was seen with the Hill estimates. However, the levels around which these estimates vary are somewhat different from those obtained from the Hill estimates. This should not

be surprising since a different estimation method was employed and given the amount of uncertainty expressed by the confidence bounds.

In Figures 41 and 44 we see levels of stability near $c = -.12$ for $u \in [4.2, 4.8]$ and around $c = -.02$ for $u \in [3.2, 4]$. The corresponding number of exceedances, as indicated at the top edge of the plots, are $17 \leq k \leq 29$ and $37 \leq k \leq 76$. Again this is very similar to what was seen with the Hill estimates for c.

In Figures 42 and 45 we see levels of stability near $c = -.16$ for $u \in [4.7, 5.2]$ and around $c = -.08$ for $u \in [3, 4]$. The corresponding number of exceedances, as indicated at the top edge of the plots, are $28 \leq k \leq 48$ and $88 \leq k \leq 191$. Again this is very similar to what was seen with the Hill estimates for c.

6.4 Diagnostic for Extreme Value Limiting Assumption

While the previous examination used extreme value index estimates based on the amount of exceedance over the threshold u, [3] also suggests using a plot of the mean values of these excess amounts as a function of u. The motivation for this lies in the fact that the limiting assumption behind extreme value theory implies that these mean values should roughly follow a straight line for large values of u. [3] also gives confidence intervals for these means but we felt that the assumptions for their validity are unlikely to be met. Instead we modified the idea to using the medians of the excesses in place of their means. For medians we have nonparametric confidence bounds which are derived under minimal assumptions, that should be valid here. These medians as a function of u are also expected to follow a roughly linear pattern for large u provided the limiting assumption behind extreme value theory holds. These plots are shown in Figures 46-48 and one sees that the confidence funnels can be penetrated linearly to a fair amount of depth. This provides some affirmation that the extreme value limiting assumption is not unreasonable.

Figure 34: Estimation of Extreme Value Index for Various Tail Depths k Using Positive Deviations

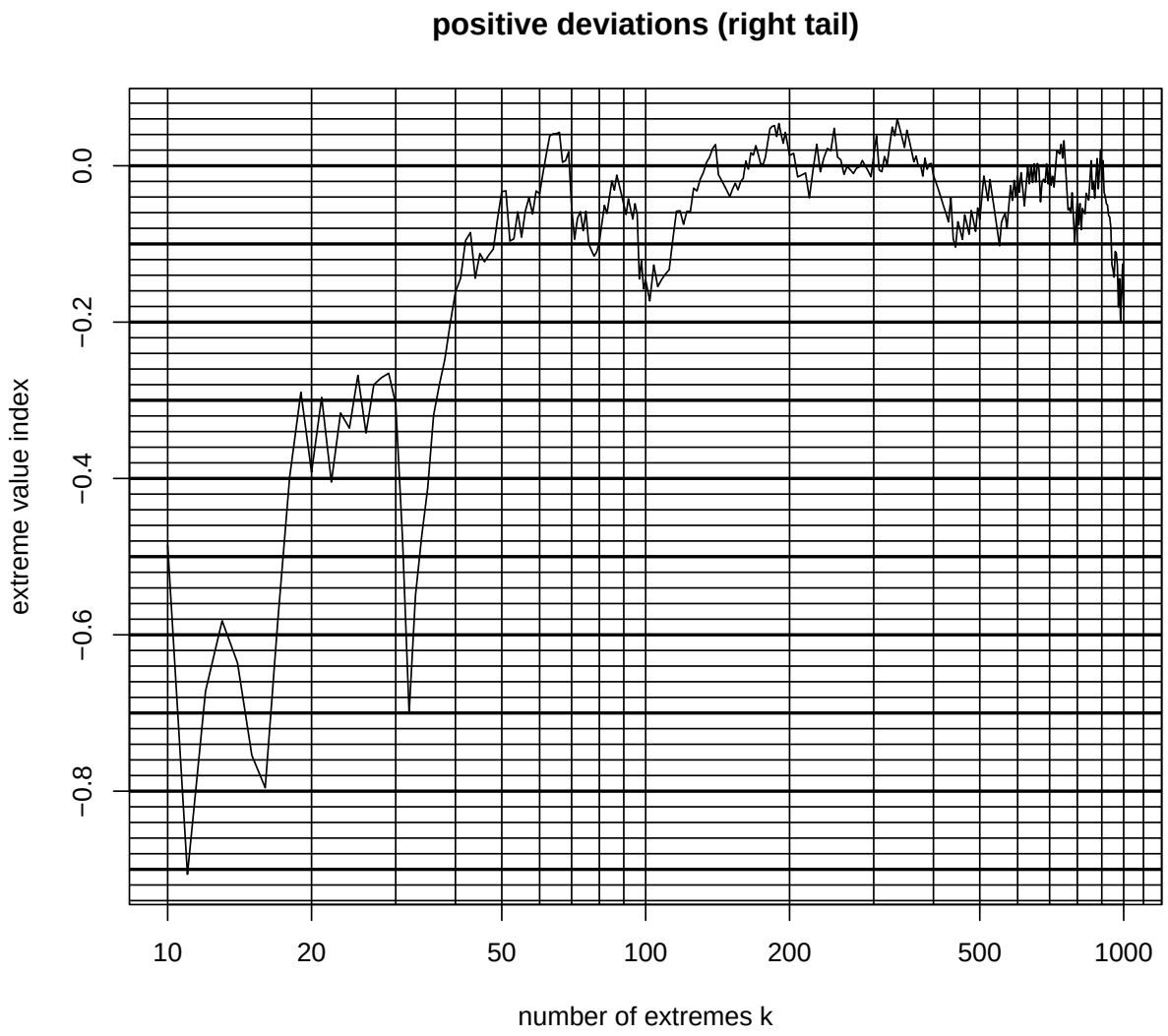


Figure 35: Estimation of Extreme Value Index for Various Tail Depths k Using Negative Deviations

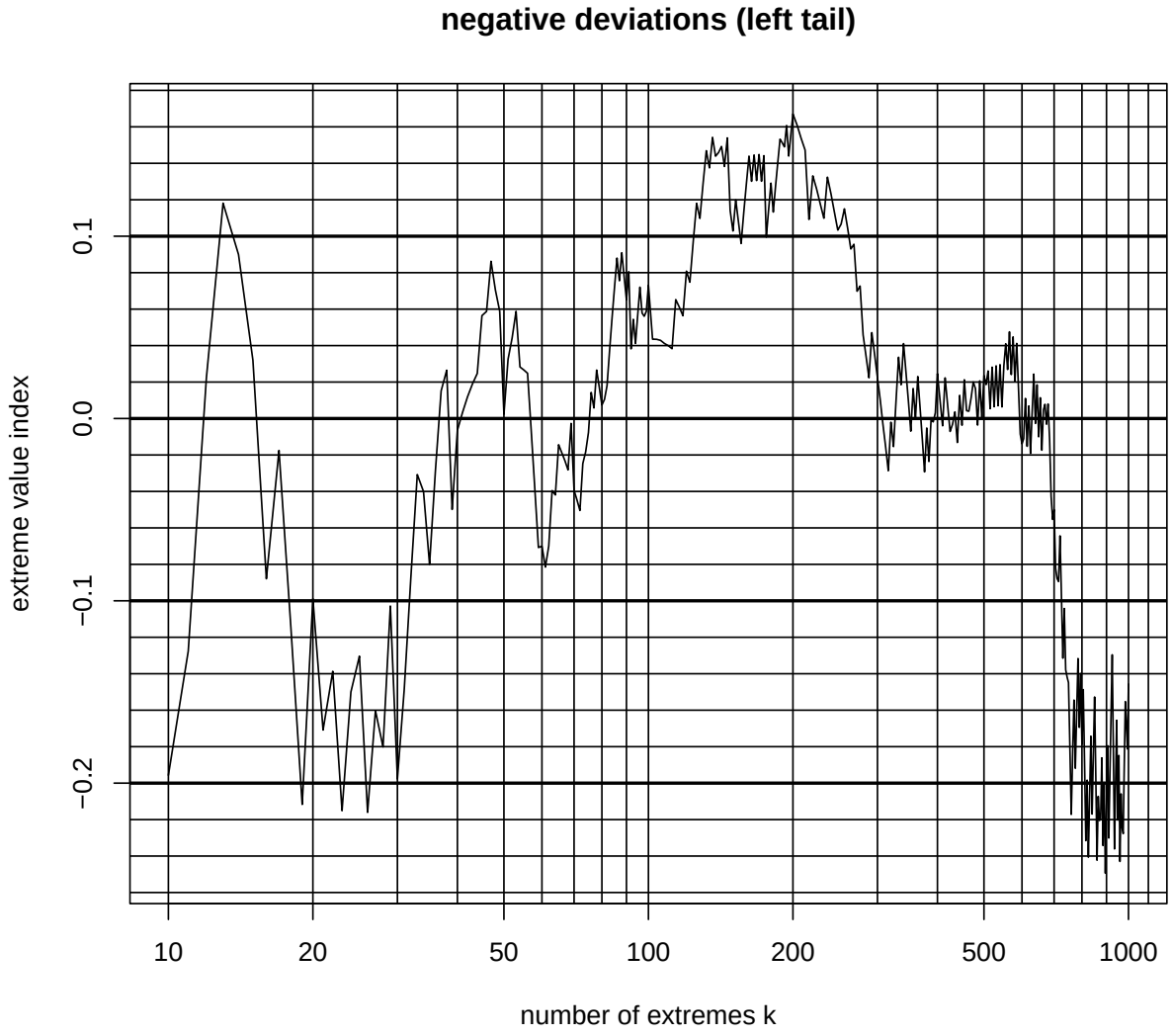


Figure 36: Estimation of Extreme Value Index for Various Tail Depths k Using Absolute Deviations

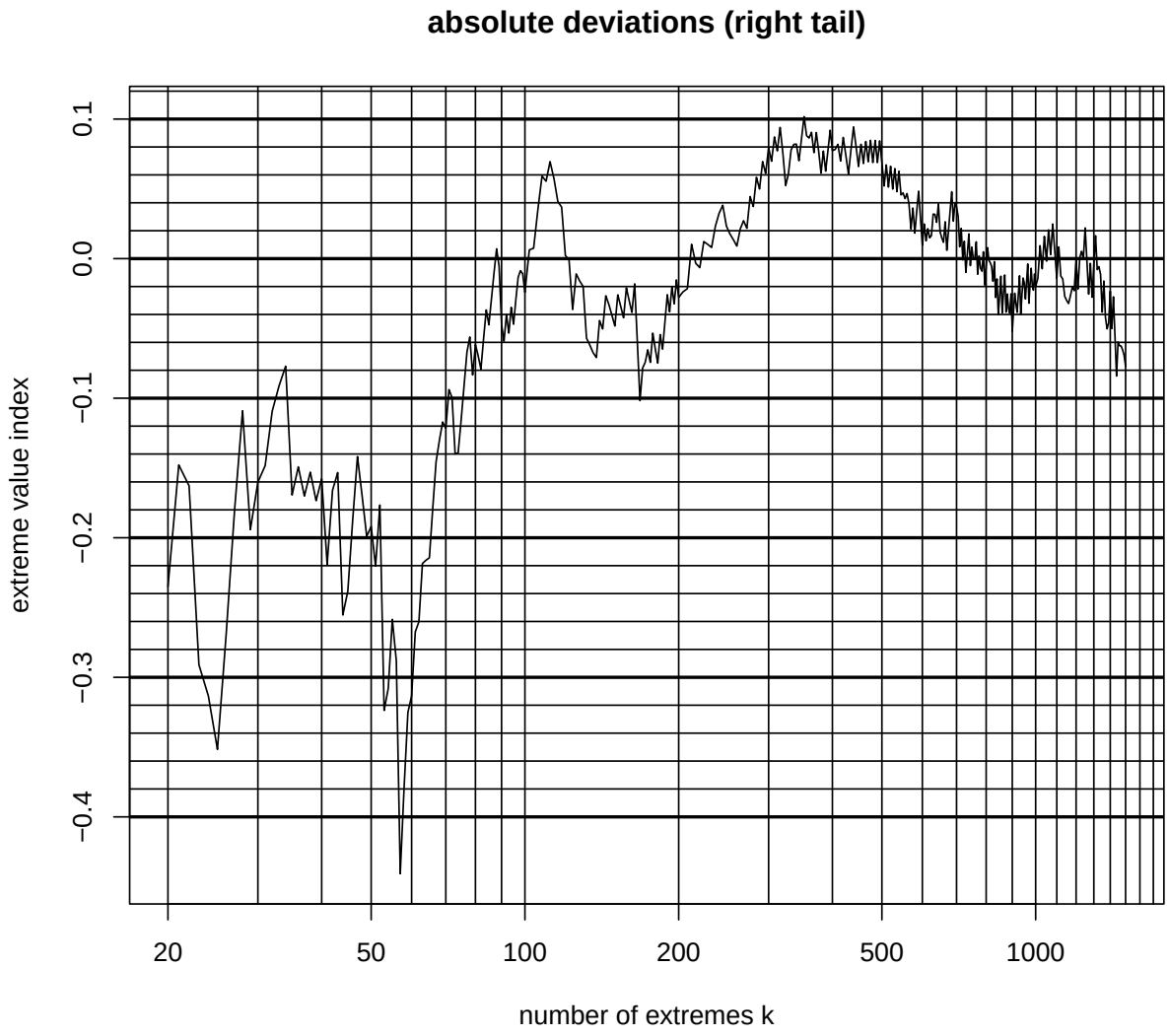


Figure 37: Estimation of Extreme Value Index for Various Tail Depths k Using Positive Deviations (Detail)

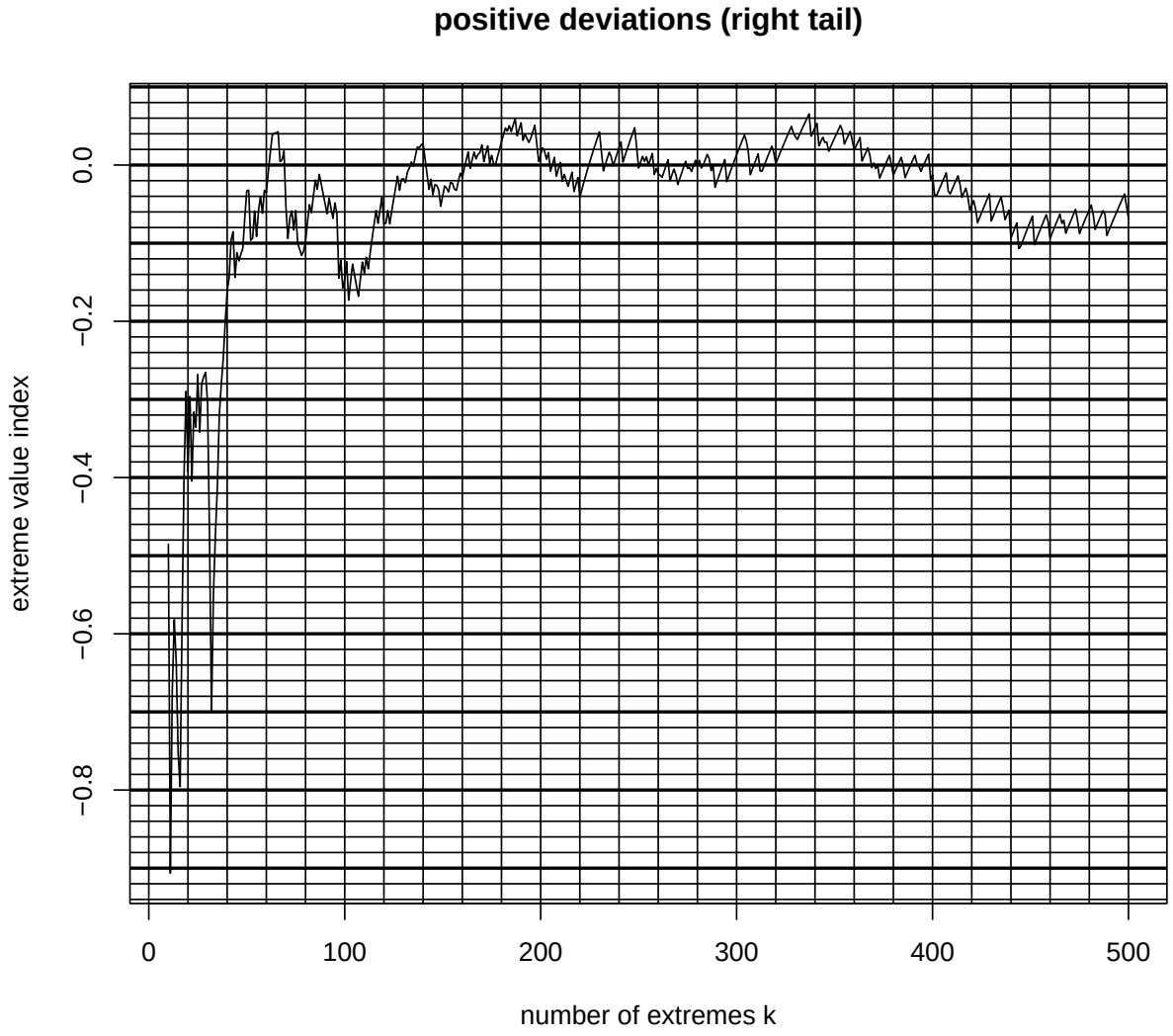


Figure 38: Estimation of Extreme Value Index for Various Tail Depths k Using Negative Deviations (Detail)

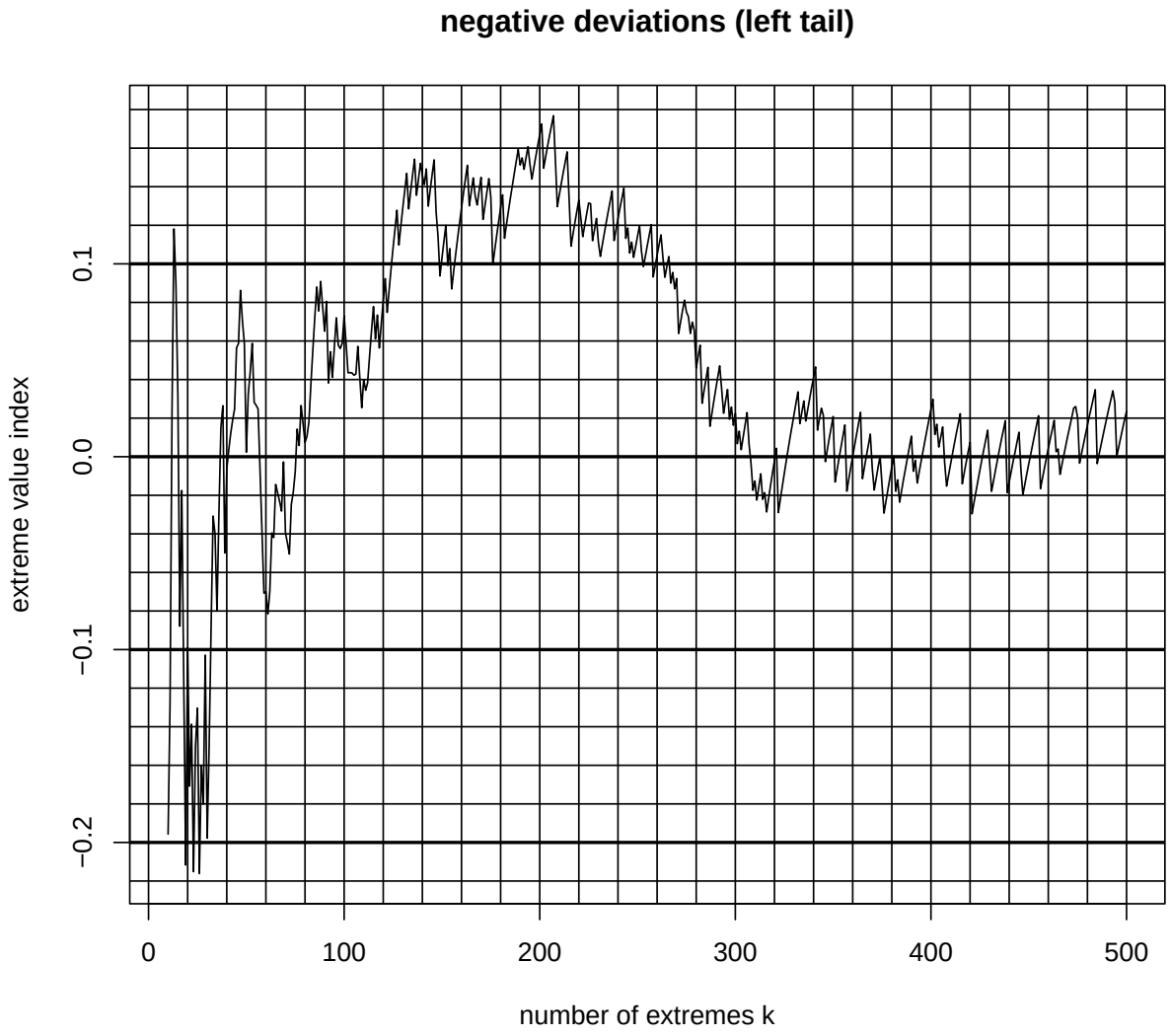


Figure 39: Estimation of Extreme Value Index for Various Tail Depths k Using Absolute Deviations (Detail)

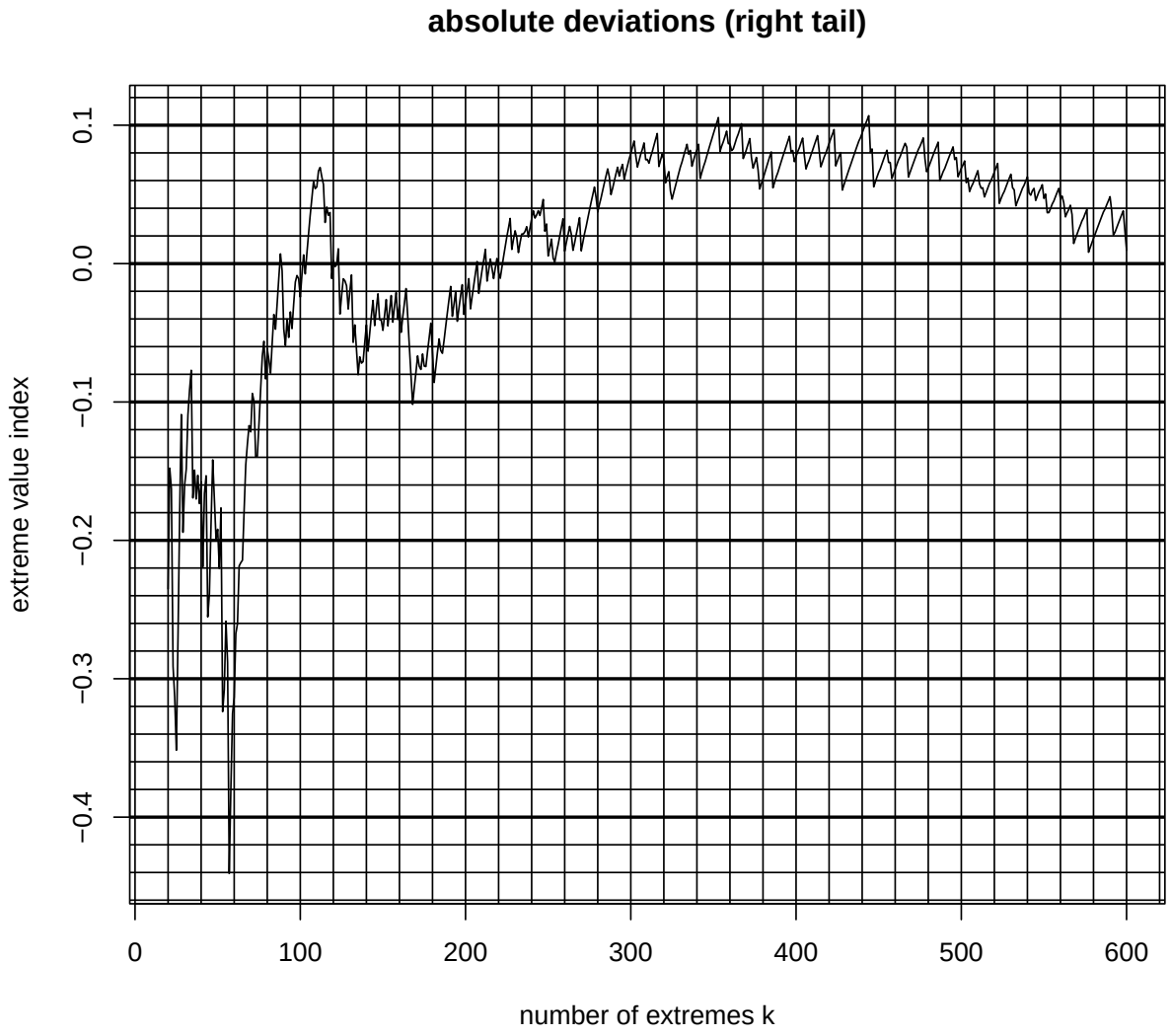


Figure 40: Estimation of Extreme Value Index for Various Thresholds
Using Positive Deviations Over Threshold u

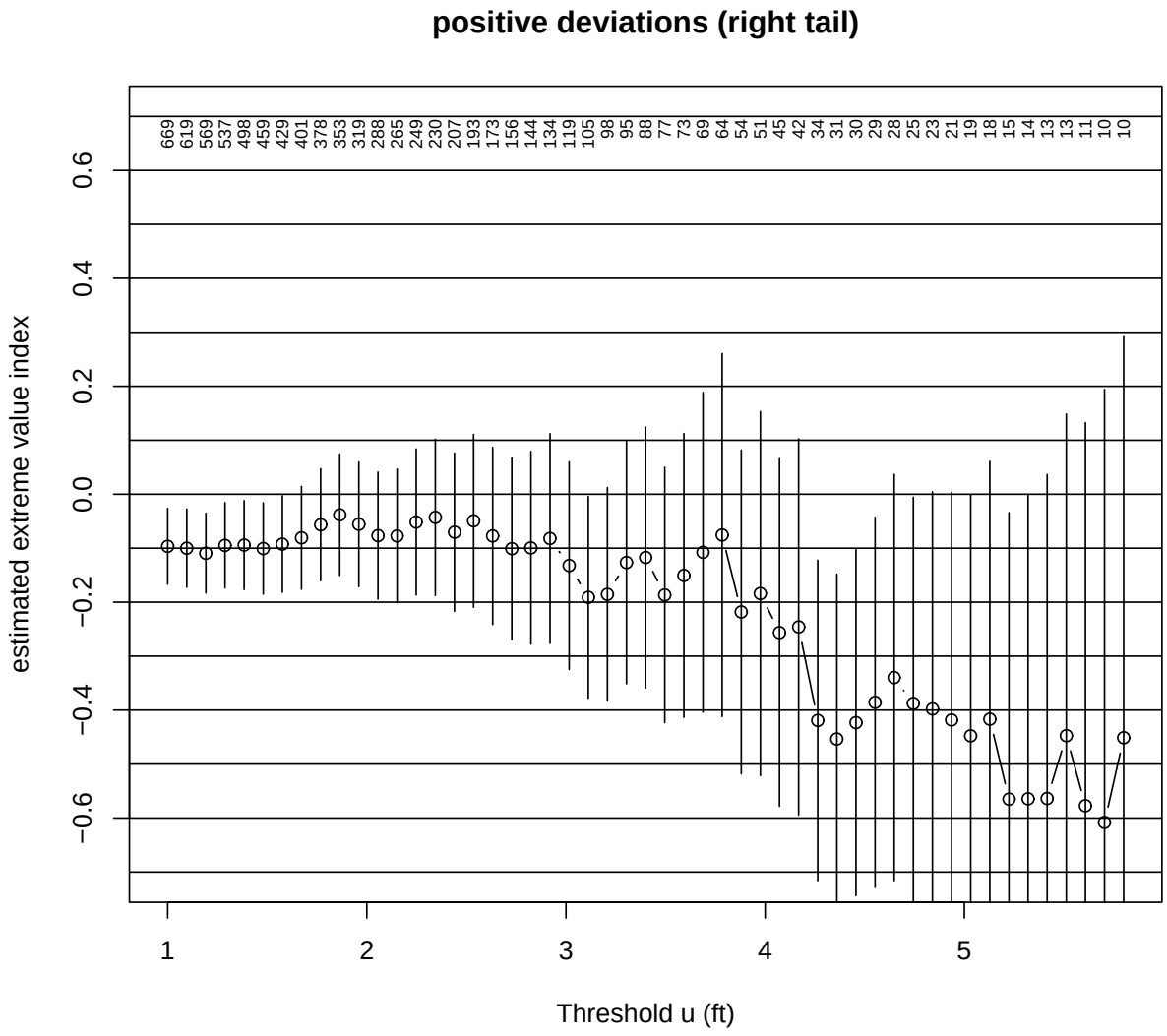


Figure 41: Estimation of Extreme Value Index for Various Thresholds
Using Negative Deviations Over Threshold u

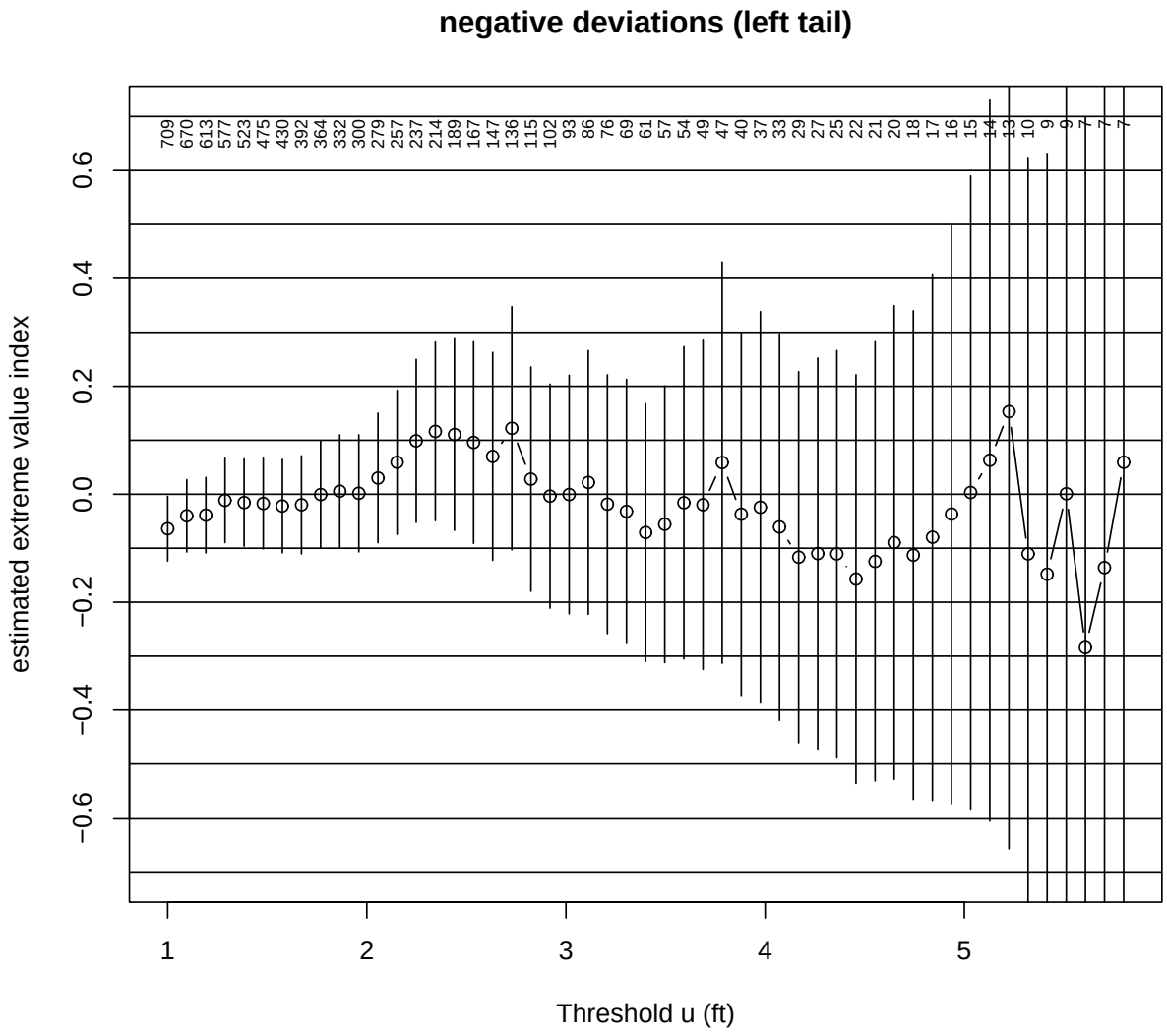


Figure 42: Estimation of Extreme Value Index for Various Thresholds
Using Absolute Deviations Over Threshold u

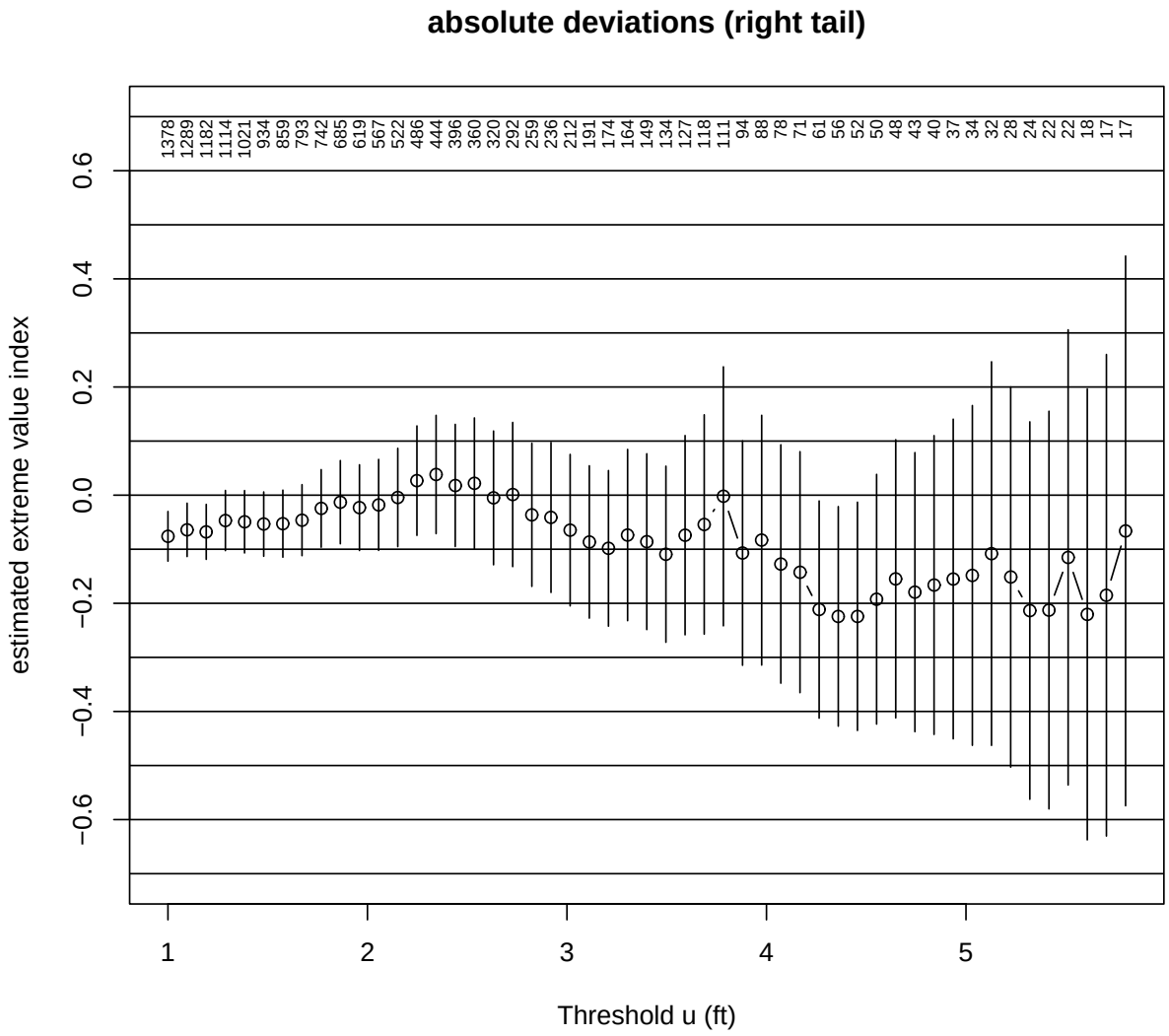


Figure 43: Estimation of Extreme Value Index for Various Thresholds
Using Positive Deviations Over Threshold u

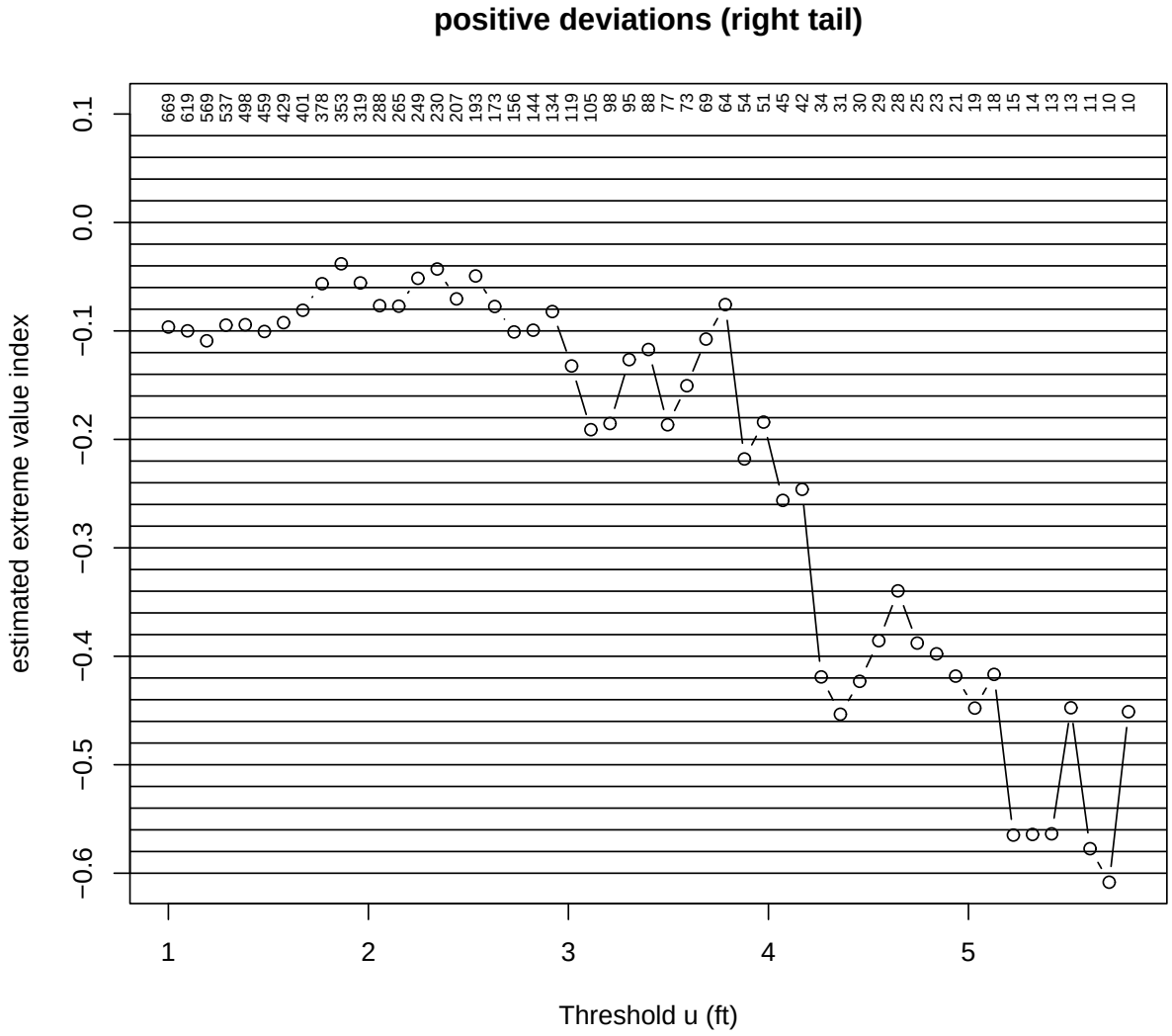


Figure 44: Estimation of Extreme Value Index for Various Thresholds
Using Negative Deviations Over Threshold u

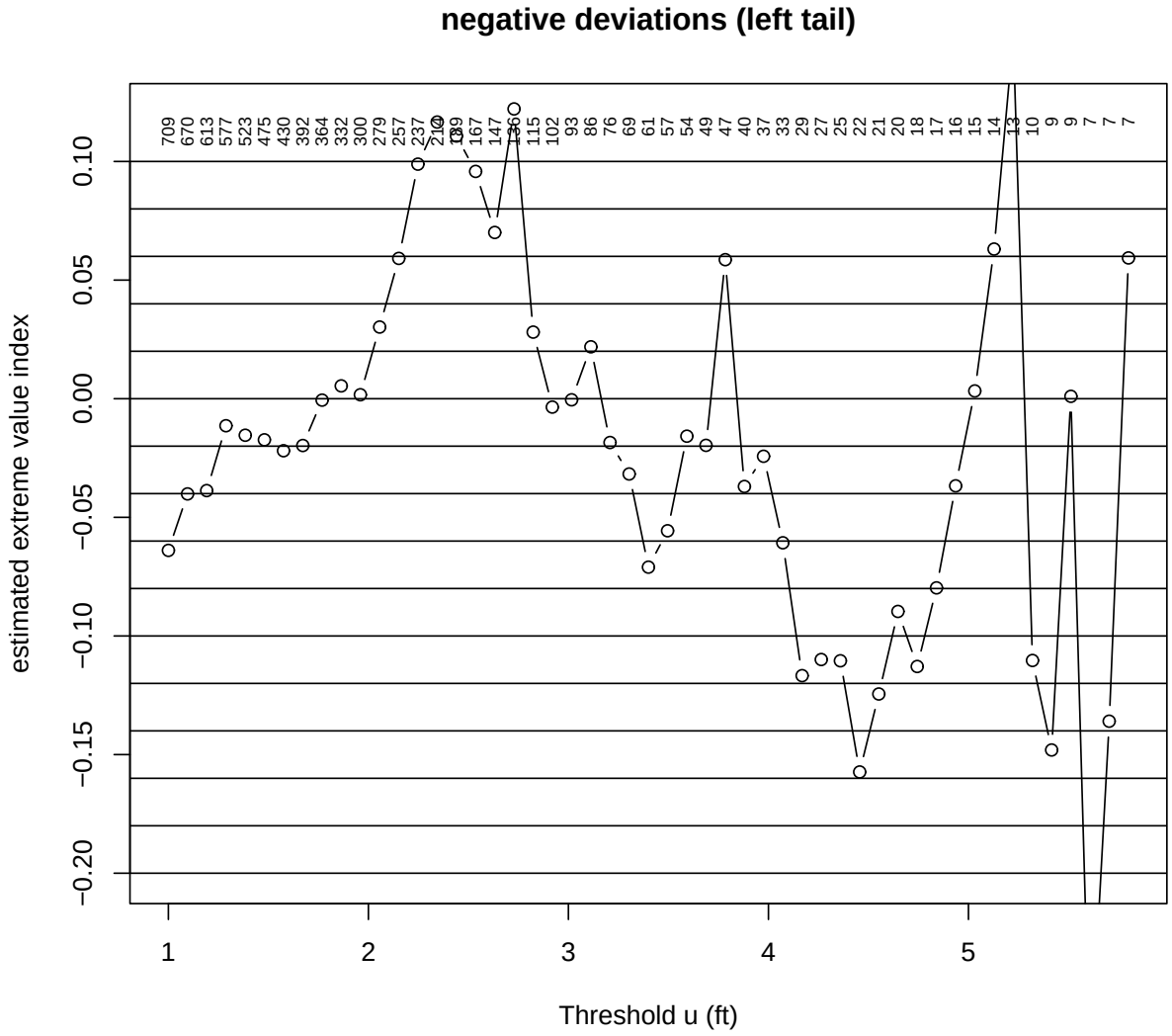


Figure 45: Estimation of Extreme Value Index for Various Thresholds
Using Absolute Deviations Over Threshold u

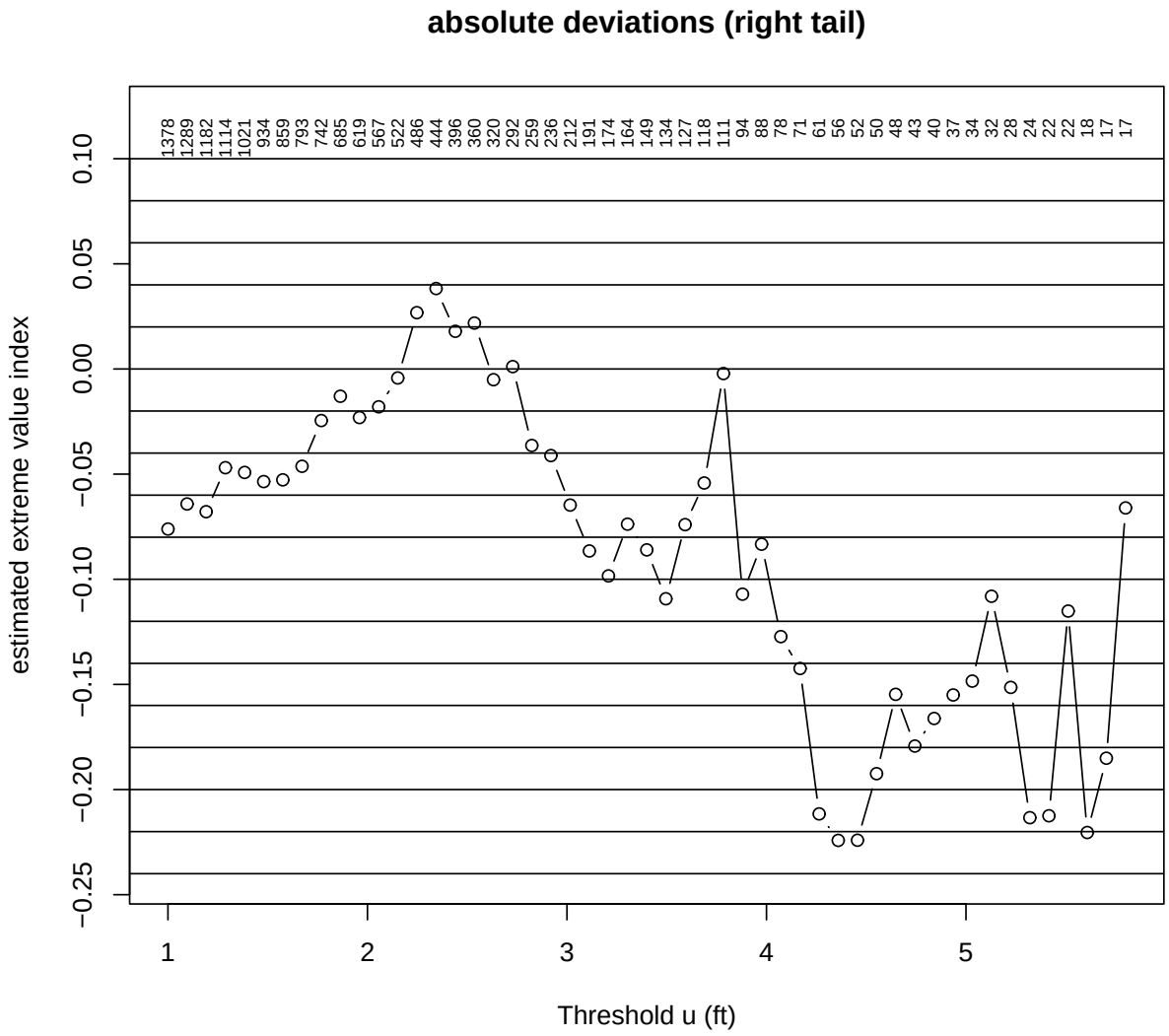


Figure 46: Estimated Median Excess over Threshold for Various Thresholds Using Positive Deviations Over Threshold u

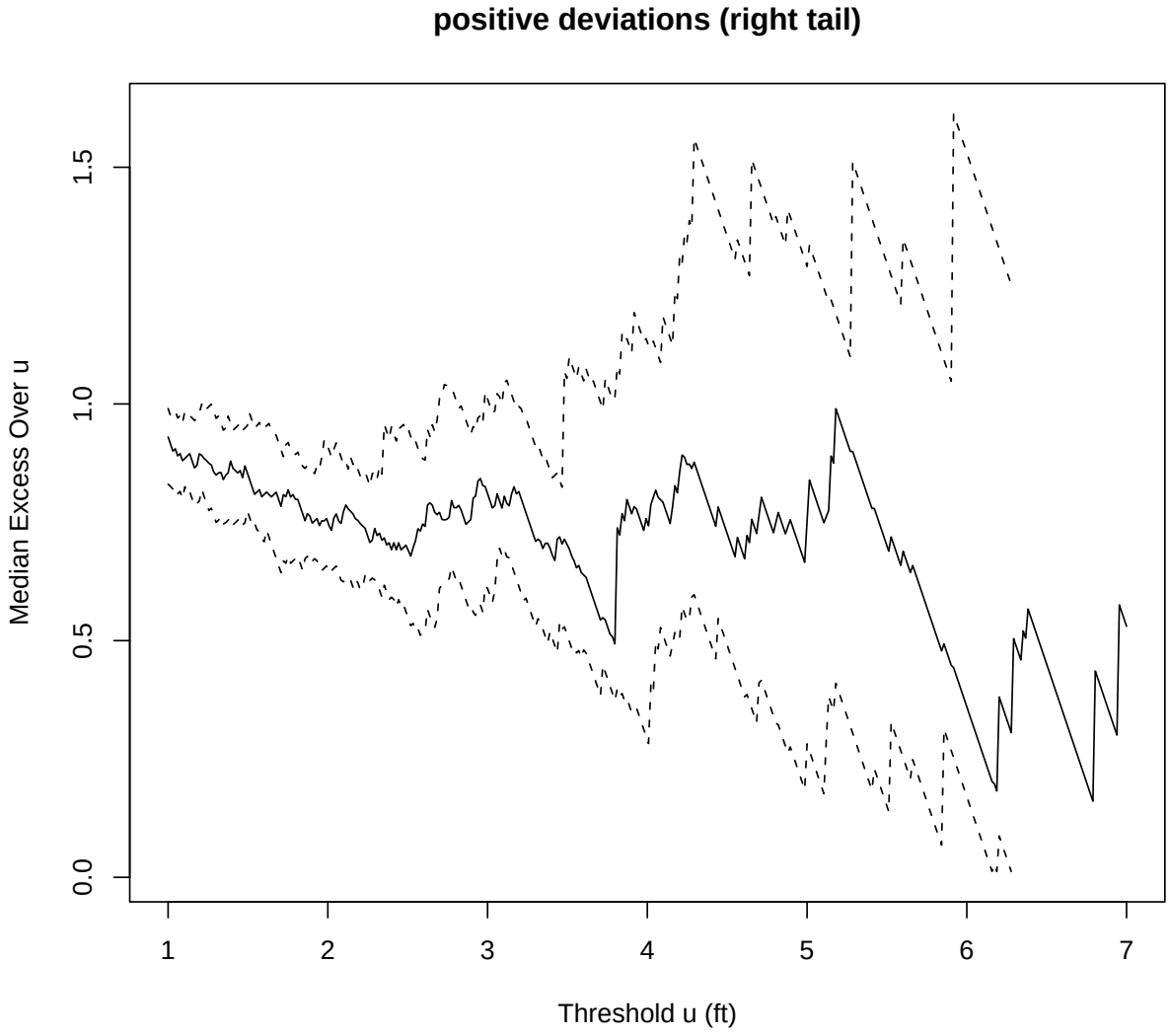


Figure 47: Estimated Median Excess over Threshold for Various Thresholds Using (−) Negative Deviations Over Threshold u

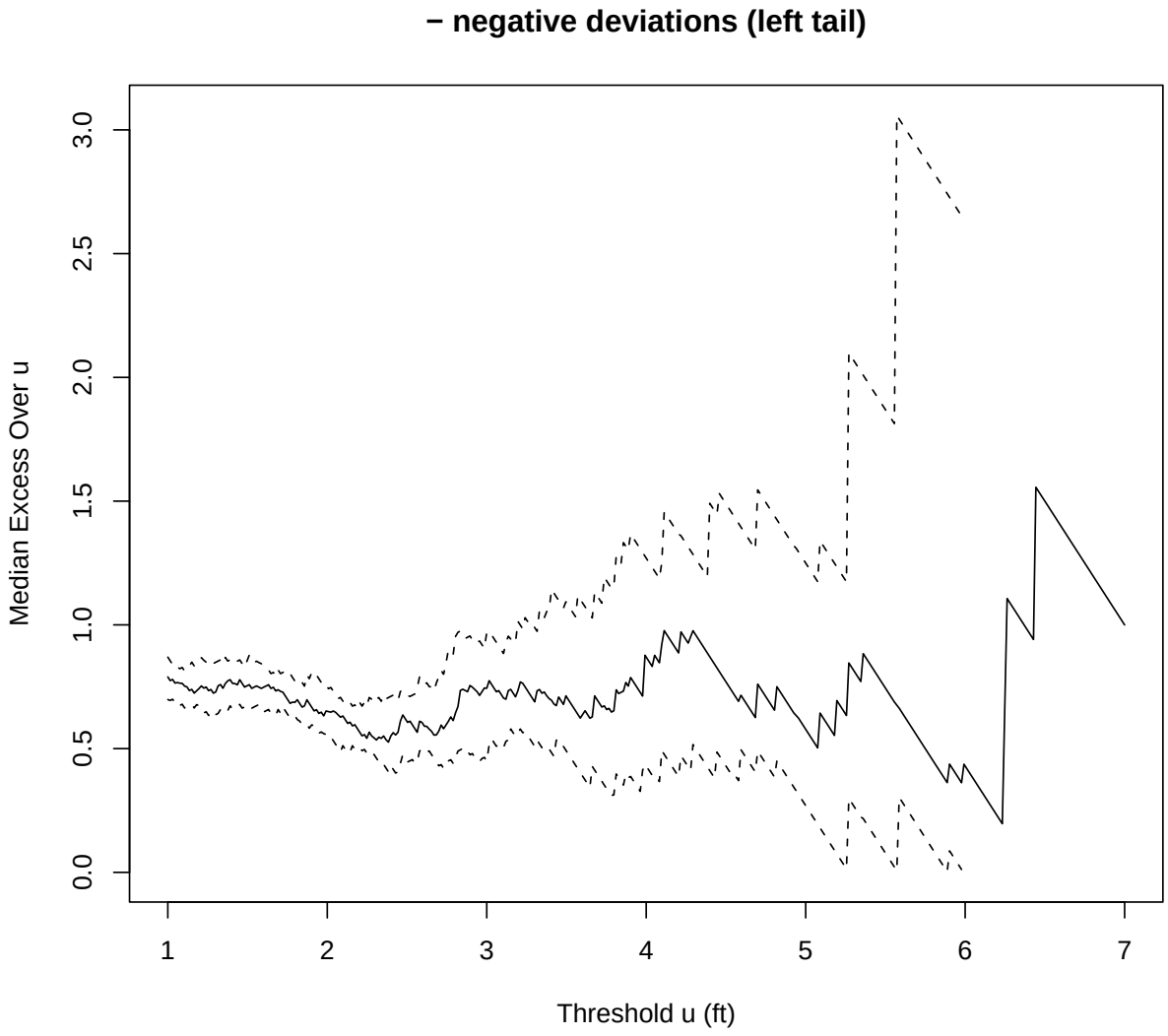
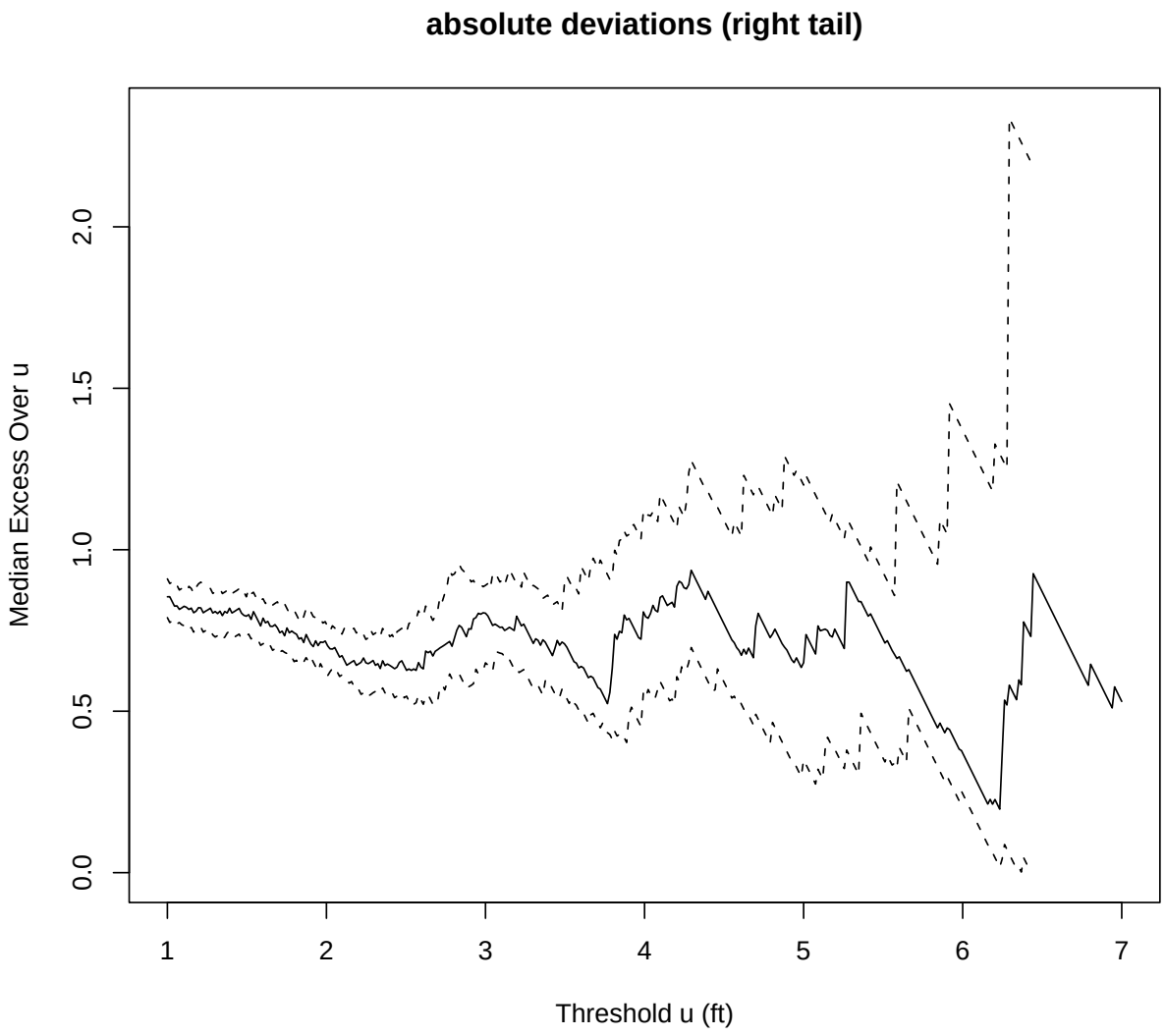


Figure 48: Estimated Median Excess over Threshold for Various Thresholds Using Absolute Deviations Over Threshold u



6.5 Results Using EXTRAP

Based on the above examinations we decided to use the EXTRAP program (described in [9]) to extrapolate the positive tail of the deviation data using $k = 30$ and $k = 96$ extremes, the negative tail of the deviation data using $k = 31$ and $k = 77$ extremes, and the right tail of the absolute deviations using $k = 52$ and $k = 200$ extremes. The dual choices of k in each case should give us some appreciation of the sensitivity of the extrapolations, especially for extremely small exceedance risks.

The resulting estimates and 95% upper bounds are shown in Figures 49-54 on pages 68-73 (explained below) and the risk thresholds for various exceedance risk levels are tabulated in Tables 1 and 2 on page 74, respectively. The reason for doubling the exceedance risks in the case of absolute deviations is to make the deviation thresholds comparable, since for $x \geq 0$

$$P(|X| > x) = 2\ P(X > x) = 2\ P(X < -x)\ ,$$

where X denotes the unbiased random deviation from the centerline, which is identified with zero.

Note that the abscissas of the plots in Figures 49-54 have two scales, a distorted probability scale (indicating the probability of exceeding a certain 747 deviation level from the taxiway centerline as indicated on the ordinate) and a linear $h(p)$-scale, labeled "transformed p(i)," which relates to the exceedance probabilities p through the following transformation formula

$$h(p) = \frac{(-n \ln(1-p))^{-c} - 1}{c}$$

with ln denoting the natural logarithm, $n = 2518$ the total sample size, and c the estimated extreme value coefficient or index as given in the plots.

Two types of points are plotted, crossed circles and dotted diamonds. The crossed circles plot Y_i against $h(p_i)$, where Y_i is the i^{th}-largest sample deviation value and p_i is chosen with the requirement $P(Y_i \geq x_{1-p_i}) = .5$, where x_{1-p} is the $(1-p)$-quantile of the continuous distribution function F underlying the original sample $X_1, \ldots, X_n$. One can meet this requirement through the binomial distribution without knowing F or the functional form of x_{1-p}. That makes this approach nonparametric. If the limiting extreme value assumption holds then one expects the x_{1-p} to be a roughly linear function of $h(p)$ for small values of p. Thus one would expect the point pattern given by the $(h(p_i), Y_i)$ to scatter around such a linear pattern, at least for the high extremes.

The line through this point pattern of crossed circles was fitted by the method of weighted least squares, weighted to account for the greater variability of Y_i for lower values of i and for the correlation among the monotone Y_i ($Y_1 \geq Y_2 \geq Y_3 \geq \ldots$). Because of the requirement $P(Y_i \geq x_{1-p_i}) = .5$ one can view the Y_i as reasonable estimates of x_{1-p_i}, with equal chances of falling below or above x_{1-p_i}, namely .5. Since the point pattern is expected to be linear, one can use the fitted line for extrapolation purposes for any small p which would extend way beyond the observed data extremes. However, one should treat such extrapolations with

caution. Although the fitted line may be well anchored in the data its slope variability has growing impact the further out one dares to go. The only reassurance for extrapolating at all is the perceived linearity in the plotted points.

For the diamonds the situation is similar. One plots Y_i against $h(\tilde{p}_i)$, where $\tilde{p}_i$ is chosen with the requirement $P(Y_i \geq x_{1-\tilde{p}_i}) = .95$. Thus one can view Y_i as a 95% upper confidence bound for $x_{1-\tilde{p}_i}$ for each i. Since the points $(h(\tilde{p}_i), Y_i)$ for small p follow the expected linear function $x_{1-p} = \alpha + \beta h(p)$ in a consistent pattern that falls above that line with probability .95 and below it with probability .05, it is suggested to fit a line to this point pattern as well. This was done again by the method of weighted least squares. This line can then be used in extrapolated fashion beyond the data, i.e., for any p on the abscissa one reads off the ordinate value from this line and interprets it as a 95% upper confidence bound for x_{1-p}. Here it should be pointed out that this confidence bound assumes that the appropriate number k of extremes used and the value c employed in the abscissa scale transformation $h(p)$ are known. The fact that they are only estimated is not accounted for in the confidence cushion that the bound provides. This confidence cushion is illustrated graphically by the growing vertical gap between the fitted extrapolation lines.

The plotted linear relationships for the deviation quantile estimates and 95% upper bounds relate to $h(p)$ as follows: $x_p = \delta h(p) + \lambda$. The corresponding fitted slopes and intercepts (δ, λ) are indicated in the plots so that threshold estimates or 95% upper bounds can be obtained for other values of the exceedance risk p.

They can also be used for the inverse problem, namely finding the estimated exceedance risk or upper confidence bound for that risk for a specified exceedance threshold. This inversion can be implemented via the following formula

$$p = 1 - \exp\left(-\frac{1}{n}\left[1 + c\left(\frac{y-\lambda}{\delta}\right)\right]^{-1/c}\right) , \tag{1}$$

where y is the specified threshold, p is the risk estimate or upper bound for the risk of exceeding the threshold y, and λ, δ, n and c are as given in the extrapolation plots for the estimation or upper bound extrapolation lines. Note that in the case of absolute deviations p represents the risk $p = P(|X| \geq y) = P(X \geq y) + P(X \leq -y) = 2 \cdot P(X \geq y)$.

The extrapolations appear to be different in the two tails and the gap widens as the risk of exceedance becomes smaller. This is most clearly seen in Tables 1 and 2 which summarize the results from the six extrapolation plots.

Figures 49 and 53 illustrate our earlier comment about the hard boundary when c is negative and sufficiently different from zero. Clearly the tick marks on the horizontal probability scale tend to bunch up as the exceedance risk gets smaller. Because of the bunching the exceedance risk labels are not shown when they would overlap but they can be inferred from the other extrapolation plots where the bunching is not so strong. Of all the extrapolation plots only the one in Figure 52 shows widening gaps between the risk level tick marks. This results from estimate for c being positive in this case.

It appears that of the two choices for k in each extrapolation the larger value is more appropriate and also a conservative choice. As an example consider the positive right tail where for the choice of $k = 30$ the estimated threshold of 8.32 for an exceedance risk of 10^{-5} only moves up to 8.57 and 8.69 by reducing that exceedance risk by a factor of .10 and .01. This kind of behavior does not look like a credible view of the situation at hand. Thus one should opt for the larger value of k. Similar considerations can be made for the negative tail and the right tail of the absolute deviations. The results obtained from the absolute deviations are chosen for the ultimate tail extrapolation since they are based on twice the amount of data (extremes from both tails).

Figure 49: Extrapolation for Positive Deviations

Using 30 Most Extreme Positive Deviations

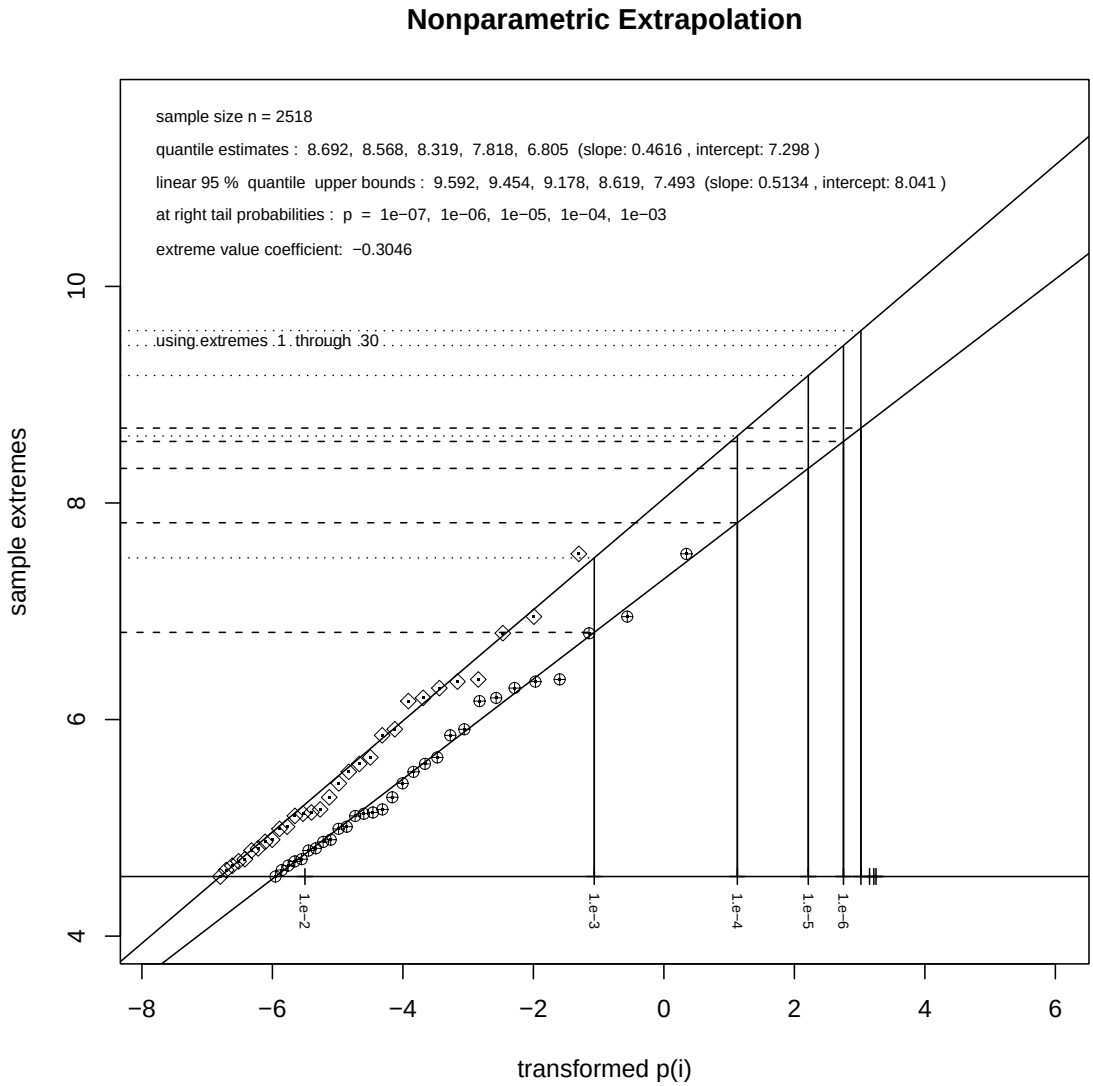


Figure 50: Extrapolation for Positive Deviations
Using 96 Most Extreme Positive Deviations

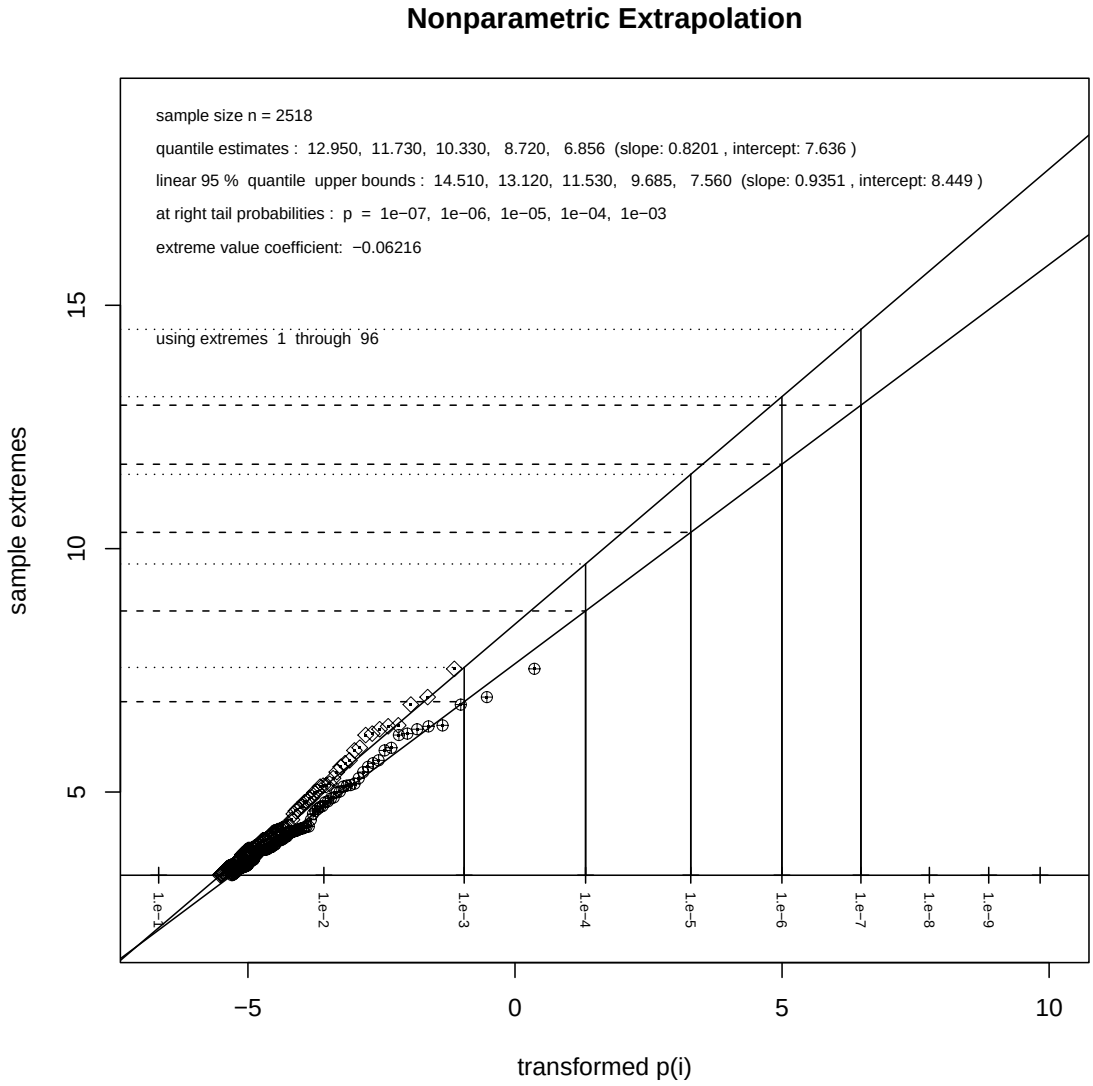


Figure 51: Extrapolation for Negative Deviations
Using 31 Most Extreme Negative Deviations (sign reversed)

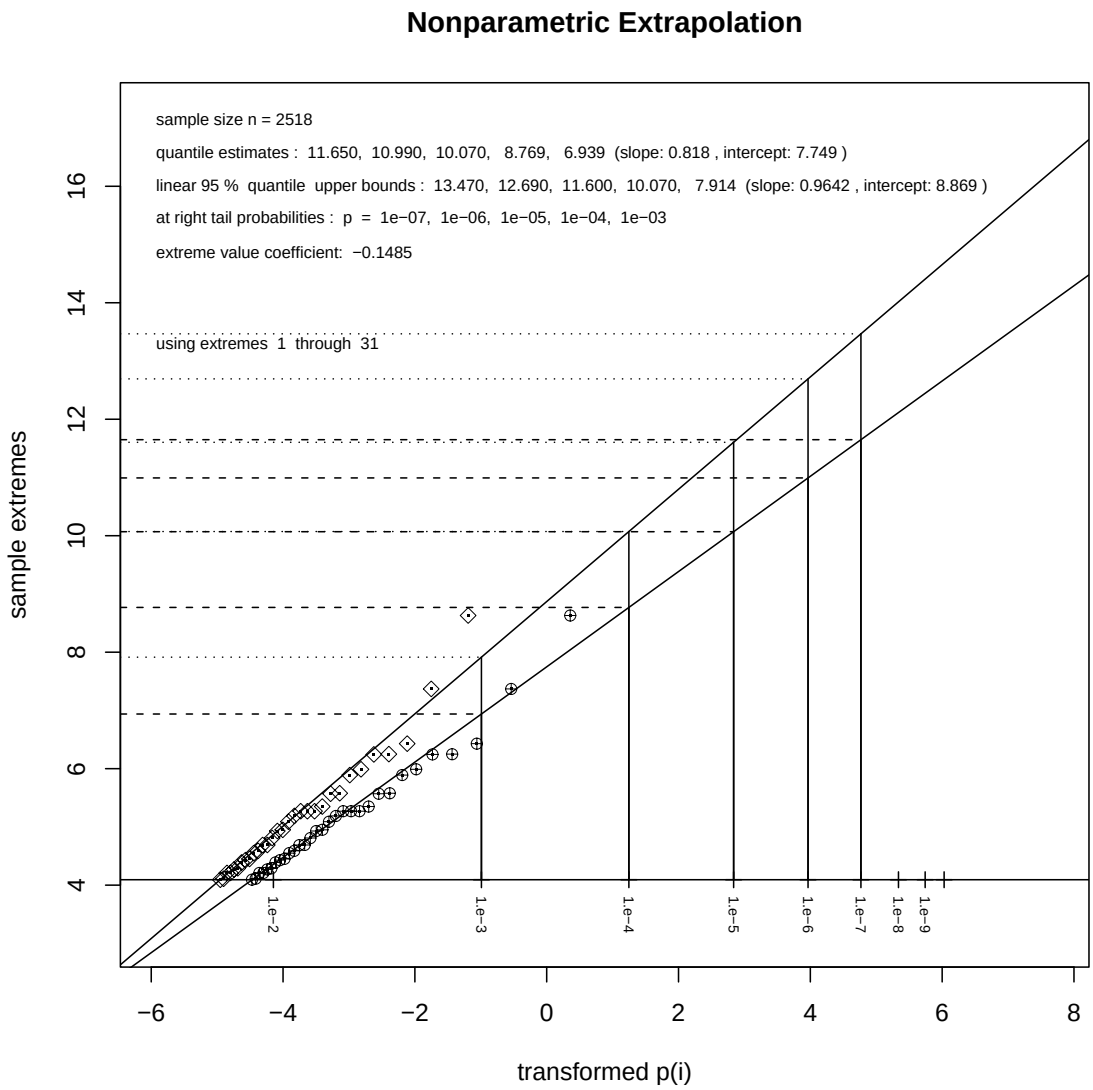


Figure 52: Extrapolation for Negative Deviations
Using 77 Most Extreme Negative Deviations (sign reversed)

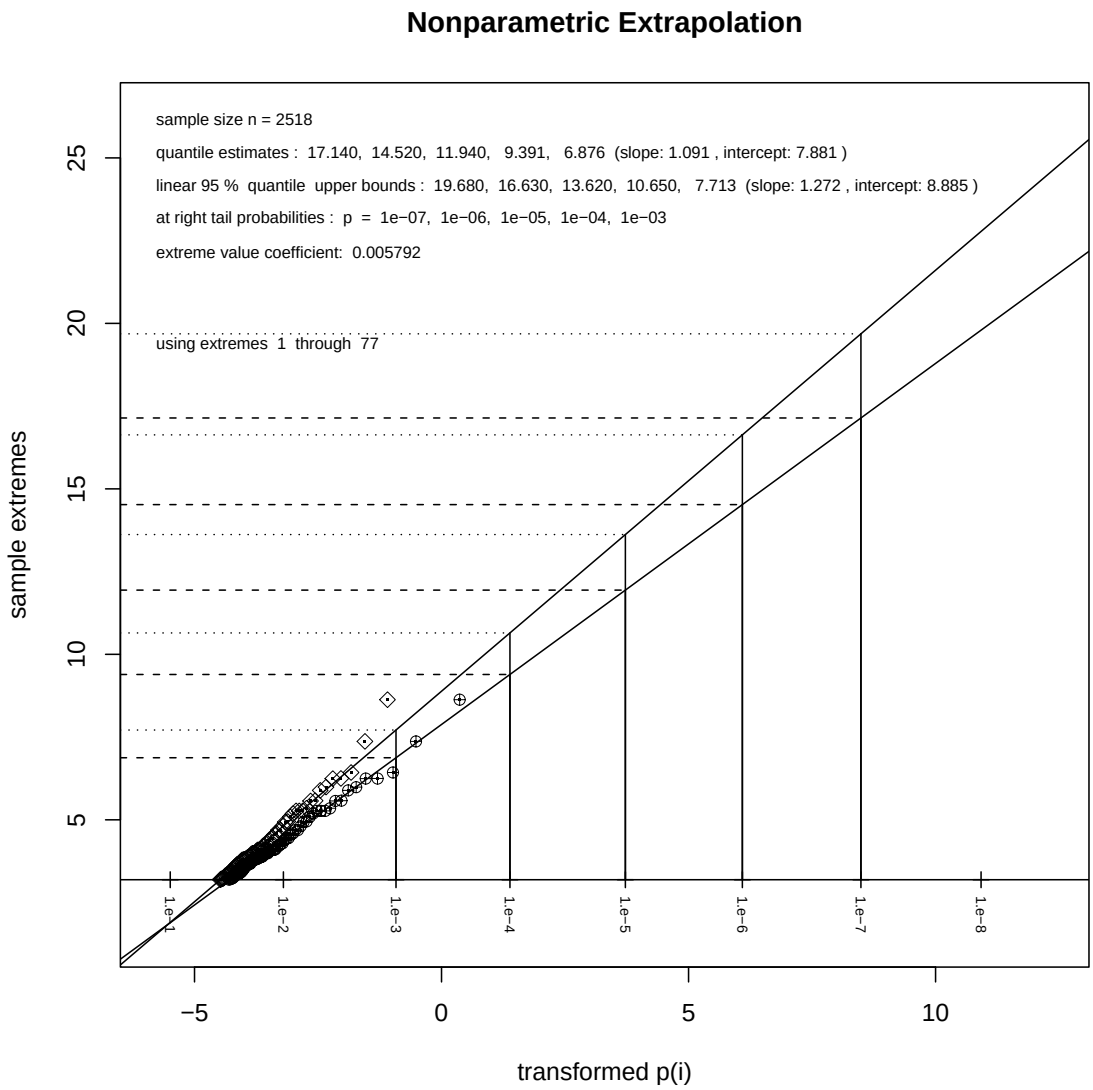


Figure 53: Extrapolation for Absolute Deviations
Using 52 Most Extreme Absolute Deviations

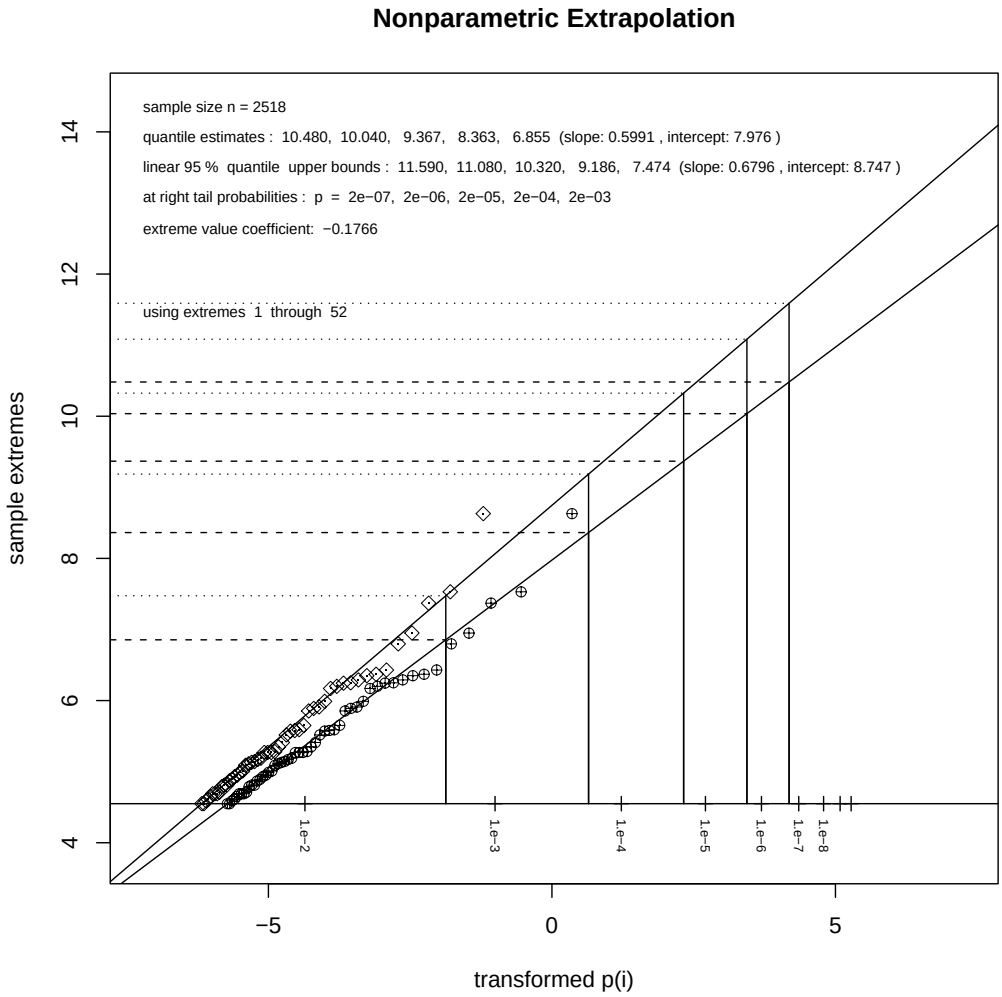


Figure 54: Extrapolation for Absolute Deviations
Using 200 Most Extreme Absolute Deviations

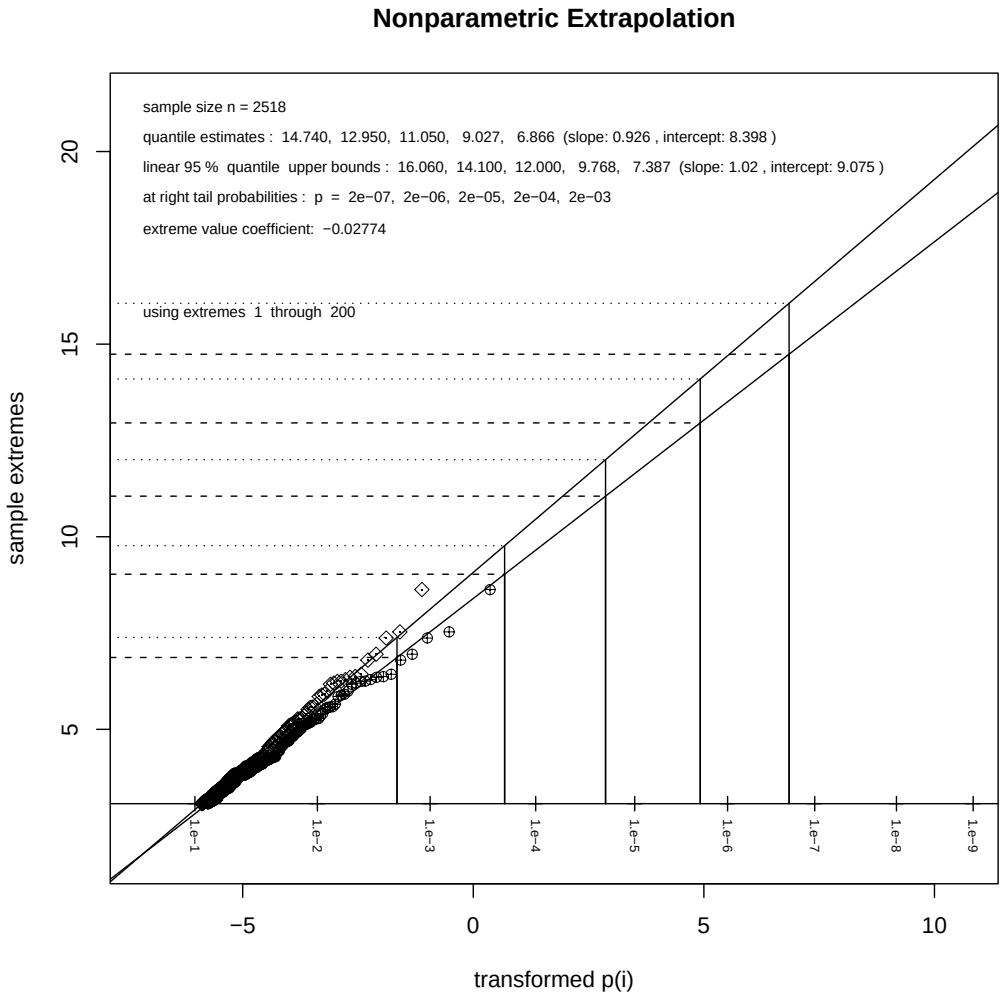


Table 1: Estimated Thresholds by Exceedance Risk
for Adjusted 747 Centerline Deviations from the Taxiway Centerline

	exceedance risk				
	10^{-7}	10^{-6}	10^{-5}	10^{-4}	10^{-3}
positive right tail $k = 30$	8.69 ft	8.57 ft	8.32 ft	7.82 ft	6.81 ft
positive right tail $k = 96$	12.95 ft	11.73 ft	10.33 ft	8.72 ft	6.86 ft
negative left tail $k = 31$	11.65 ft	10.99 ft	10.07 ft	8.77 ft	6.94 ft
negative left tail $k = 77$	17.14 ft	14.52 ft	11.94 ft	9.39 ft	6.88 ft
	$2 \cdot 10^{-7}$	$2 \cdot 10^{-6}$	$2 \cdot 10^{-5}$	$2 \cdot 10^{-4}$	$2 \cdot 10^{-3}$
absolute deviations $k = 52$	10.48 ft	10.04 ft	9.37 ft	8.36 ft	6.85 ft
absolute deviations $k = 200$	14.74 ft	12.95 ft	11.05 ft	9.03 ft	6.87 ft

Table 2: Threshold 95% Upper Bounds by Exceedance Risk
for Adjusted 747 Centerline Deviations from the Taxiway Centerline

	exceedance risk				
	10^{-7}	10^{-6}	10^{-5}	10^{-4}	10^{-3}
positive right tail $k = 30$	9.59 ft	9.45 ft	9.18 ft	8.62 ft	7.49 ft
positive right tail $k = 96$	14.51 ft	13.12 ft	11.53 ft	9.68 ft	7.56 ft
negative left tail $k = 31$	13.47 ft	12.69 ft	11.60 ft	10.07 ft	7.91 ft
negative left tail $k = 77$	19.68 ft	16.63 ft	13.62 ft	10.65 ft	7.71 ft
	$2 \cdot 10^{-7}$	$2 \cdot 10^{-6}$	$2 \cdot 10^{-5}$	$2 \cdot 10^{-4}$	$2 \cdot 10^{-3}$
absolute deviations $k = 52$	11.59 ft	11.08 ft	10.32 ft	9.19 ft	7.47 ft
absolute deviations $k = 200$	16.06 ft	14.10 ft	12.00 ft	9.77 ft	7.39 ft

7 Back-Adjustment for Bias

Previously we carefully identified and adjusted for possible biases, namely parallax bias and the other bias. The former is attributed to the pilot's cockpit position relative to the 747 centerline while the latter could possibly be explained by the various pilots dealing in different ways with the offset of the centerlights from the centerline. This centerlight offset from the centerline is away from the laser for ALPHA and towards the laser for BRAVO. Parallax bias changes sign depending on the heading while the other bias is mainly independent of the heading of the aircraft, although there is a potential that the parallax issue and the offset issue interact within the pilots.

The bias correction was intended to make the deviation data look as symmetrical around zero as possible so that ultimately we could work with absolute adjusted deviations and thus have twice the sample size for the extremes. The presented deviation exceedance values for given risk levels, as given in Tables 1 and 2, are based on the bias adjusted deviation data. These values are meaningful only if we pretend that there is no bias in the deviations from the centerline under normal operation. As it is, the biases exist and we need to account for them when assessing operational exceedance thresholds corresponding to given risk levels.

Depending on the heading of the 747 and the taxiway one would have to add different corrections to the derived exceedance values. These corrections are respectively of the form $\hat{\mu}_O + \hat{\mu}_P$ or $\hat{\mu}_O - \hat{\mu}_P$ depending on the heading.

For setting taxiway width standards one would not want to make different bias adjustments based on the heading of the aircraft, since presumably one would use that taxiway in either direction. Also, one would not want to make the assumption that possible obstacles or adjacent taxiways are only on a particular side of the taxiway. Finally, one would not want to assume that the offset of the centerlights from the centerline is always on one particular side of the centerline. Having said all this, we should therefore add the sum of absolute biases to the threshold values given in the Tables 1 and 2, i.e., add $|\hat{\mu}_P| + |\hat{\mu}_O|$ to these values.

Previously we identified two sets of bias corrections, one set for each taxiway. These are slightly different from each other for the parallax bias and somewhat different for the other bias. The difference in parallax bias is presumably just due to sampling variation. It would make sense to make a parallax bias correction that is independent of the taxiway. One could simply average the two values for $\hat{\mu}_{P,A}$ and $\hat{\mu}_{P,B}$, but since different sample sizes were involved at ALPHA and BRAVO, namely $n = 1428$ for ALPHA and $m = 1090$ for BRAVO, it would make more sense to use a weighted average, using as weights $n/(m+n)$ and $m/(m+n)$. Hence we obtain the following taxiway-independent parallax bias correction

$$\hat{\mu_P} = \frac{n}{m+n}\hat{\mu}_{P,A} + \frac{m}{m+n}\hat{\mu}_{P,B} = .07 \text{ ft} .$$

Since the other biases are so different from each other we chose to be conservative by taking the larger of the two bias values, namely .66 ft, instead of averaging .66 ft and .24 ft in some weighted manner. Adding these two bias components .07 ft and .66 ft yields .73 ft which comes close to the corresponding value of .75 ft that resulted for the ANC data. We thus

decided to take this slightly larger value of .75 ft as a conservative adjustment and added it to any threshold given in the Tables 1 and 2.

On one side of the taxiway and for one direction of travel this correction will actually yield the correct threshold for the one-sided exceedance risk, while on the other side the correction will add an unnecessary amount and thus the exceedance risk will be smaller.

This is illustrated in Figure 55 where the top density shows the distribution of centerline deviations without bias. The dashed heavy vertical lines represent the 10^{-4} estimated risk threshold on either side of that density. The solid heavy vertical lines represent the corresponding back-adjusted thresholds, i.e., they are offset by .75 ft away from zero.

The middle density represents the distribution of centerline deviations with maximal bias of $.68 + .07 = .75$ ft. (For the sake of illustration we have split up .75 ft into .07 ft and .68 ft, as opposed to originally .66 ft.) Here the other bias and the parallax bias compound. Note that the risk of exceeding the back-adjusted threshold on the right is still 10^{-4}, since both the adjusted threshold and the density have undergone the same shift. However, on the left it is only $1.84 \cdot 10^{-5}$ because the center of the distribution is $.75 + .75 = 1.5$ ft further to the right of the back-adjusted threshold on the left.

The bottom density represents the distribution of centerline deviations with the smaller bias of $.68 - .07 = .61$ ft. Here the other bias and the parallax bias cancel to some extent. Note that the risk of exceeding the back-adjusted threshold on the right now is $9.23 \cdot 10^{-5}$, while on the left it is only $2.16 \cdot 10^{-5}$ because the center of the distribution is $.75 + .61 = 1.36$ ft to the right of the back-adjusted threshold on the left. These risk estimates were derived using the previously explained process of converting threshold values to exceedance risks using equation 1 on page 67.

For example, when dealing with a risk of 10^{-6} of the 747 centerline at the main gear location exceeding a deviation threshold to a particular side of the taxiway, we find from Table 1 the estimated threshold 12.95 ft. This was based on the analysis of the absolute values of the bias corrected deviations. To this estimated threshold we now have to add .75 ft and obtain $12.95 + .75 = 13.70$ ft as the desired threshold with the back-adjusted bias correction. The interpretation is as follows. We fix a particular location along a straight taxiway segment (ALPHA or BRAVO) and a particular direction perpendicular to the taxiway centerline. Then 13.70 ft is our estimate for the threshold distance from the taxiway centerline, that is conservativly expected to be exceeded in that chosen direction by the 747 centerline (at the landing gear location) about once in 10^6 passes. The conservative nature is the same as was discussed in the context of Figure 55, it depends on the heading of the aircraft and on the side chosen for the deviation. For other risk levels these back-adjusted estimated thresholds and their confidence bounds are given in Table 3. For comparison we also repeat the corresponding table as it was obtained from the analysis of the ANC data. The results appear to be in reasonable agreement. For higher risk levels the threshold estimates derived from the JFK data are further out than those derived from the ANC data. For lower risk levels the situation is reversed. For a two-sided risk level of $2 \cdot 10^{-6}$ the threshold estimates are almost identical while the confidence bounds are less than .5 ft apart.

Figure 55: Exceedance Probabilities for Various Bias Scenarios

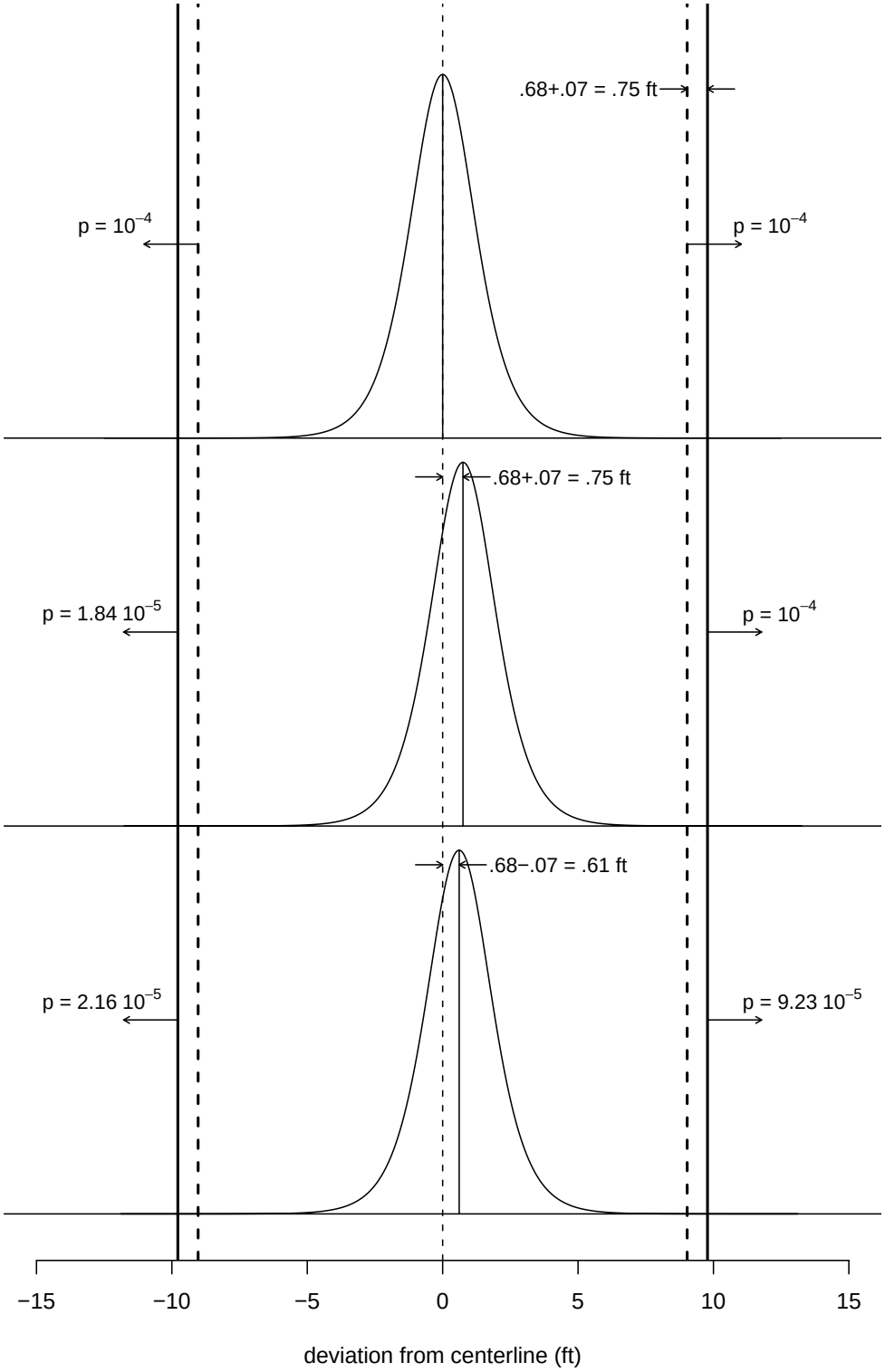


Table 3: Back-Adjusted Thresholds by Exceedance Risk
for 747 Centerline Deviations from Taxiway Centerline

From JFK Data	two-sided exceedance risk				
$n = 2518$	$2 \cdot 10^{-7}$	$2 \cdot 10^{-6}$	$2 \cdot 10^{-5}$	$2 \cdot 10^{-4}$	$2 \cdot 10^{-3}$
estimate	15.49 ft	13.70 ft	11.80 ft	9.78 ft	7.62 ft
95% upper bounds	16.81 ft	14.85 ft	12.75 ft	10.52 ft	8.14 ft

Table 4: Back-Adjusted Thresholds by Exceedance Risk
for 747 Centerline Deviations from Taxiway Centerline

From ANC Data	two-sided exceedance risk				
$n = 9796$	$2 \cdot 10^{-7}$	$2 \cdot 10^{-6}$	$2 \cdot 10^{-5}$	$2 \cdot 10^{-4}$	$2 \cdot 10^{-3}$
estimate	16.59 ft	13.79 ft	11.23 ft	8.88 ft	6.74 ft
95% upper bounds	17.41 ft	14.44 ft	11.73 ft	9.24 ft	6.97 ft

Above we made a statement concerning the exceedance of 13.70 ft in one particular direction, namely that it was expected to happen about once in 10^6 passes. This somewhat vague statement concerning the number of occurrences of extreme and rare events will benefit from some further explanation. This should also clarify the kind of precision one can expect in statements about such rare events. If the extreme event has chance $p = 10^{-6}$ of occurring and if we focus on a sequence of $n = 1/p = 10^6$ events we can ask how likely it is that we see $k = 0, 1, 2, 3, 4, 5, 6, \ldots$ extreme events among them. The respective probabilities for these counts are .3679, .3679, .1839, .0613, .0153, .0031, .0005, with the remaining counts having chance $< .000084$. Although one would expect 1 such rare event in n such trial events the probabilities themselves make it clear that there could be quite some variability in such a count, with most likely and equally probable counts of 0 and 1 and half that chance of seeing 2 extreme events.

The calculation of the above probabilities for extreme event counts is relatively insensitive to how small p is, as long as it is small, say $p < .01$. All that is additionally required is the link between p and n through $n = 1/p$ so that $np = 1$.

8 Results in Relation to Taxiway Width Standards

We now discuss the obtained extrapolation results for centerline deviations in relation to the current standards for taxiway widths within the FAA and ICAO, see [4] and [5]. Both documents also give a further design requirement called the Taxiway Edge Safety Margin (TESM), that is used to determine taxiway widths for all Design Groups/Code Letters. "TESM is the minimum acceptable distance between the outside of the airplane main wheels and the pavement edge," see [4], with similar text in [5]. These standards and the TESM are shown in Table 5. The standards are illustrated in Figure 56.

Table 5: Standards for Taxiway Width

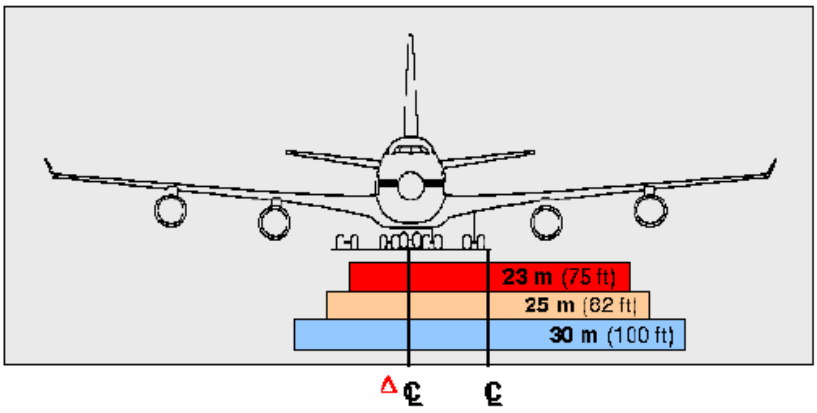
	FAA design group V	FAA design group VI	ICAO Code E	ICAO Code F
Width standard	75 ft	100 ft	23 m	25 m
TESM	15 ft	20 ft	4.5 m	4.5 m

The previously derived back-adjusted threshold of 13.70 ft corresponding conservatively to a 10^{-6} one-sided risk of exceedance translates into a $13.70 + 20.665 = 34.37$ ft deviation of the outside edge of the 747 landing gear relative to the taxiway centerline. Here we assume that the outside edge of the main landing gear is 20.665 ft from the aircraft centerline, which corresponds to a 747-400 with 19 inch wide tires. Older models of the 747 have 16, 17 or 19 inch wide tires. Doubling this threshold we see that 68.74 ft is well below the current standard width of 75 ft. However, the threshold exceeds the TESM clearance standard. This assessment is based on the estimated threshold for the one-sided target risk of 10^{-6}.

If we do the same calculation using the 95% upper bound on the threshold for the same one-sided target risk we arrive at $2 \times (14.10 + .75 + 20.665) = 71.04$ ft for the unexposed width of the taxiway, still well within the current standard of 75 ft. Since here we deal with the one-sided risk of 10^{-6}, the chance of exceeding the threshold on either side would conservatively be double that, i.e., $2 \cdot 10^{-6}$. The conservative aspect derives from the fact that while on one side the risk may be 10^{-6} under the right bias constellation on the other side it will be somewhat less because here the bias constellation will work to reduce the risk of exceedance.

If we start with a given taxiway width of 75 ft and subtract 20.665 ft and .75 ft from 75/2 ft we arrive at a threshold of $37.5 - 20.665 - .75 = 16.085$ ft for the bias corrected centerline deviations. Using equation 1 from page 67 we get the estimated two-sided risk of exceeding this threshold as

$$p = 1 - \exp\left(-\frac{1}{2518}\left[1 - .02774\left(\frac{16.08 - 8.398}{.926}\right)\right]^{1/.02774}\right) = 3.17 \cdot 10^{-8}\,,$$



22-Foot (6.7 m) Main Gear Displacement From TWY Centerline

Figure 56: Taxiway Centerline Deviation in Relation to Standards

where the relevant numbers were taken from Figure 54 for the estimated case. Similarly one proceeds for the 95% confidence bounds and for the other taxiway widths of 82 ft and 100 ft. The results are summarized in the top third of Table 6. The risks derived for a 747 using the A380 or NLA outer-to-outer main gear dimension of 47.03 ft or 52.49 ft are also shown in Table 6. In that case it is assumed that the 747 with such A380 or NLA main gear dimensions will show aircraft centerline to taxiway centerline deviation behavior similar to the unaltered 747. This assumption may be reasonable when the aircraft moves at a fixed deviation parallel to the centerline. The situation is somewhat different if the deviation path is sinusoidal around a line parallel to the taxiway centerline. The reason for this is that the A380 gear geometry not only has a wider outer-to-outer main gear dimension of 47.03 ft but it also has a greater separation between nose and main gear centroids, namely $L_2 = 93.34$ ft, as compared to $L_1 = 78$ ft for the 747. This issue was discussed at some greater length in [10] (pp. 75-77) and it is not repeated here.

Table 6: Two-Sided Exceedance Risks for 747

Risk of Outer Main Gear Tire Edge Exceeding the Taxiway Edge
for Different Taxiway Width Standards, Using 747, A380 & NLA Main Gear Dimensions.
Does Not Compensate for the Different Nose to Main Gear Distance for the A380 or NLA.

Main Gear Dimensions	Taxiway Width	75 ft	82 ft	100 ft
from 747	Estimated Risk	$3.17\ 10^{-8}$	$1.62\ 10^{-10}$	$1.17\ 10^{-18}$
	95% Upper Bound on Risk	$1.94\ 10^{-7}$	$2.13\ 10^{-9}$	$5.73\ 10^{-16}$
from A380	Estimated Risk	$1.41\ 10^{-6}$	$1.26\ 10^{-8}$	$1.34\ 10^{-15}$
	95% Upper Bound on Risk	$5.24\ 10^{-6}$	$8.74\ 10^{-8}$	$1.42\ 10^{-13}$
from NLA	Estimated Risk	$3.79\ 10^{-5}$	$5.25\ 10^{-7}$	$3.88\ 10^{-13}$
	95% Upper Bound on Risk	$9.50\ 10^{-5}$	$2.21\ 10^{-6}$	$1.39\ 10^{-11}$

These exceedance risks are pointwise, i.e., they refer to the exceedance at one particular prespecified point along a straight taxiway. Such pointwise assessments may be relevant when dealing with the risk of collision with some stationary structure. However, in that case the risk would need to be recalculated to take into account the distance of the structure from the taxiway centerline and the wingspan of the 747. This would be very situation specific. On the other hand one could subtract half the wingspan and .75 ft from the distance of structure to the taxiway centerline to define the danger threshold distance from the taxiway centerline that should not be exceeded by the aircraft centerline. For that threshold one

would then employ equation 1 from page 67 to obtain the two-sided exceedance risk value p which would then be divided by 2 to get the one-sided value, since presumably the structure is to just one side of the taxiway.

The pointwise exceedance risks given in Table 6 are not the same as the lengthwise risk of the outside main gear tire edge exceeding the taxiway width somewhere along the length of the taxiway. This risk is bound to be higher than the pointwise risk. This lengthwise risk is currently out of reach since data at JFK (as at ANC) were collected at just two points along the taxiway. The measurement design used at Frankfurt airport does address this point to some extent and data from this collection effort may be useful in answering this broader and more relevant question of lengthwise risk.

9 Combined ANC/JFK Data

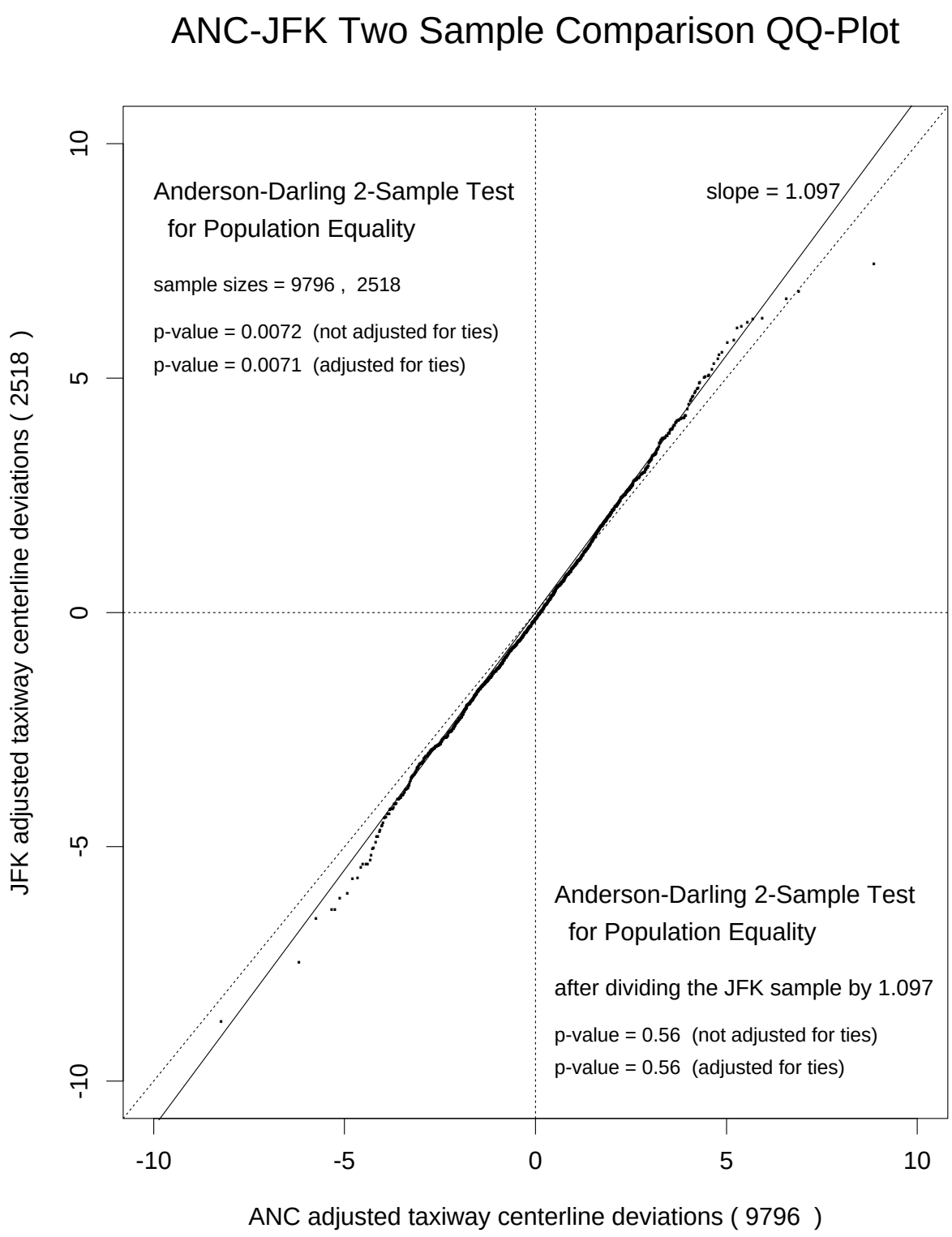
Here we examine to what extent the deviation experience at JFK agrees with that at ANC. As a first step we construct a QQ-plot, see Figure 57. If the two samples of adjusted deviations come from the same population one would expect the point pattern to hug the main diagonal reasonably well. That main diagonal is shown as a dashed line and it appears that the point pattern tilts away from it with a slightly higher slope. That slope, estimated by least squares and indicated by the solid line, is 1.097. Thus it appears that the adjusted deviations from JFK are roughly 10% more spread out than the adjusted deviations from ANC. Why that is the case one can only speculate. One major factor could be the greater separation between centerline and centerlights at JFK.

The fact that the point pattern follows a straight line reasonably well is confirmation that the distributions of the adjusted deviations from ANC and JFK are very much the same in character, differing in location and scale only. As it turns out, the only difference appears to be a scale factor. Any location difference were preempted due to the bias correction leading to the adjusted deviations. This sameness in distribution character is reassuring since we deal with similar phenomena, 747s taxiing on straight taxiway segments at two different airports.

Also shown in Figure 57 are the results of two formal tests. The test employed is the Anderson-Darling k-sample test (here $k = 2$), see [8]. In its first application we compare the adjusted deviation data from ANC with those of JFK, the tested hypothesis being that both samples come from the same population. The results are shown in the upper left part of the plot in Figure 57. The p-values are around .007. This means that when assuming the hypothesis to be true the chance of seeing an Anderson-Darling discrepancy metric as large or larger than was seen with the current two samples is .007. This does not support the hypothesis. In fact one would say that the result is significant at the .01 level since $.007 < .01$.

In the second application of the Anderson-Darling two-sample test we first divided the JFK adjusted deviations by 1.097, the apparent factor by which the JFK adjusted deviations appear to be more dispersed than the ANC adjusted deviations. This division by 1.097

Figure 57: Comparison of ANC and JFK Adjusted Deviation Data



should align these two samples better. The results of the Anderson-Darling two-sample test applied to the ANC adjusted deviations and the scaled JFK adjusted deviations are shown in the lower right part of the plot in Figure 57. Now the p-values are .56, which no longer speaks against the hypothesis of both samples, the ANC sample and the scaled JFK sample, arising from the same population.

We point out that this application of the Anderson-Darling k-sample test does not adjust for the fact that both samples were adjusted in the first place (biases removed) and that the JFK sample was scaled. This would induce an Anderson-Darling discrepancy metric that is more favorable to the hypothesis. We feel that this effect is mild since all estimates (biases and scale factor) are based on fairly large samples.

Thus it appears that we can combine the adjusted JFK data with the adjusted ANC data provided we first scale the JFK data by dividing them by 1.097. However, when applying risk extrapolation results from the combined data to the JFK situation one should undo the scaling by multiplying any derived risk thresholds for the combined data by 1.097. For the ANC situation we don't have to do this multiplication. This form of combination benefits both JFK and ANC since we work with a larger sample size of $n = 9796 + 2518 = 12314$ in the combined sample. For the remainder of this section we will be working with the combined data as described above.

As in Section 6.2 we again compute the Hill estimator for the right tail of the absolute combined data for various tail depths k. We bypass the comparison with the right and left tail of the combined data since we will be working with the absolute values anyway in order to enhance the sample size that to comes to bear on the extrapolation.

Figures 58 and 59 give the plots of the extreme value index estimates by the Hill estimation method on the logarithmic and straight k-scale, respectively. From these plots it appears that the estimates fluctuate around a level of about $c = .03$ which last seems to be attained around $k = 860$ before a steady decrease. Closer examination yields $k = 859$ with $c = .0307$ and that k was used in the extrapolation. The latter is represented by Figure 63.

The corresponding plots for extreme value index estimates based on exceedances over threshold as examined previously in Section 6.3 are shown in Figure 60 and 61 with and without confidence bounds, respectively. The above choice of $k = 859$ seems reasonably well supported by this alternate view. It appears to roughly correspond to a threshold choice of $u = 3$ ft which is supported by the diagnostic plot in Figure 62 as it was first introduced and discussed in Section 6.4.

The results from the extrapolation in Figure 63 are tabulated in Table 7 and juxtaposed with the corresponding results from the separate extrapolations from the ANC data and the JFK data. The agreement between the ANC/JFK extrapolations and those derived from the ANC data alone is quite close.

Figure 58: Estimation of Extreme Value Index for Various Tail Depths k Using Absolute Deviations from Combined Adjusted ANC/JFK Data

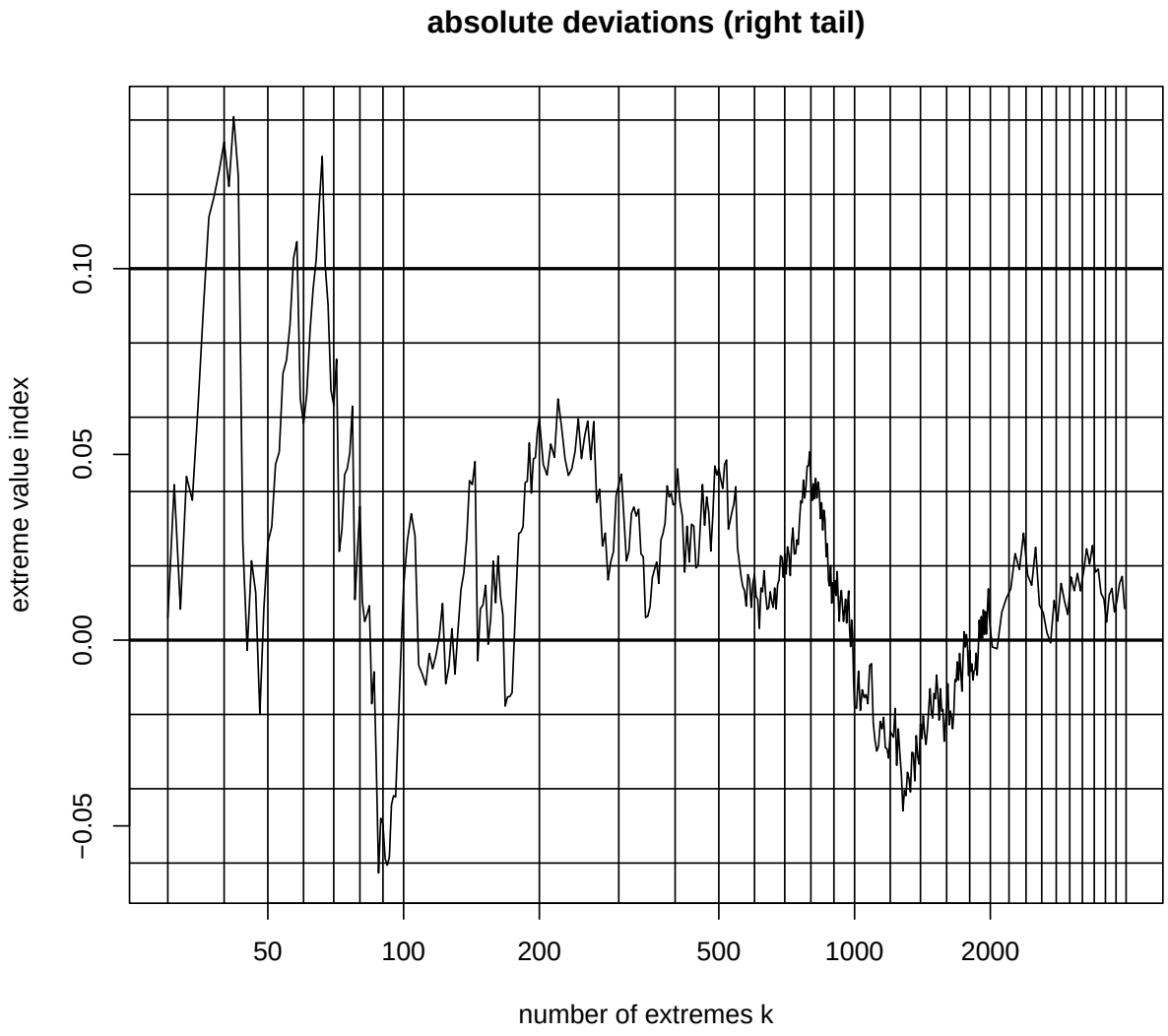


Figure 59: Estimation of Extreme Value Index for Various Tail Depths k
Using Absolute Deviations from Combined Adjusted ANC/JFK Data (Detail)

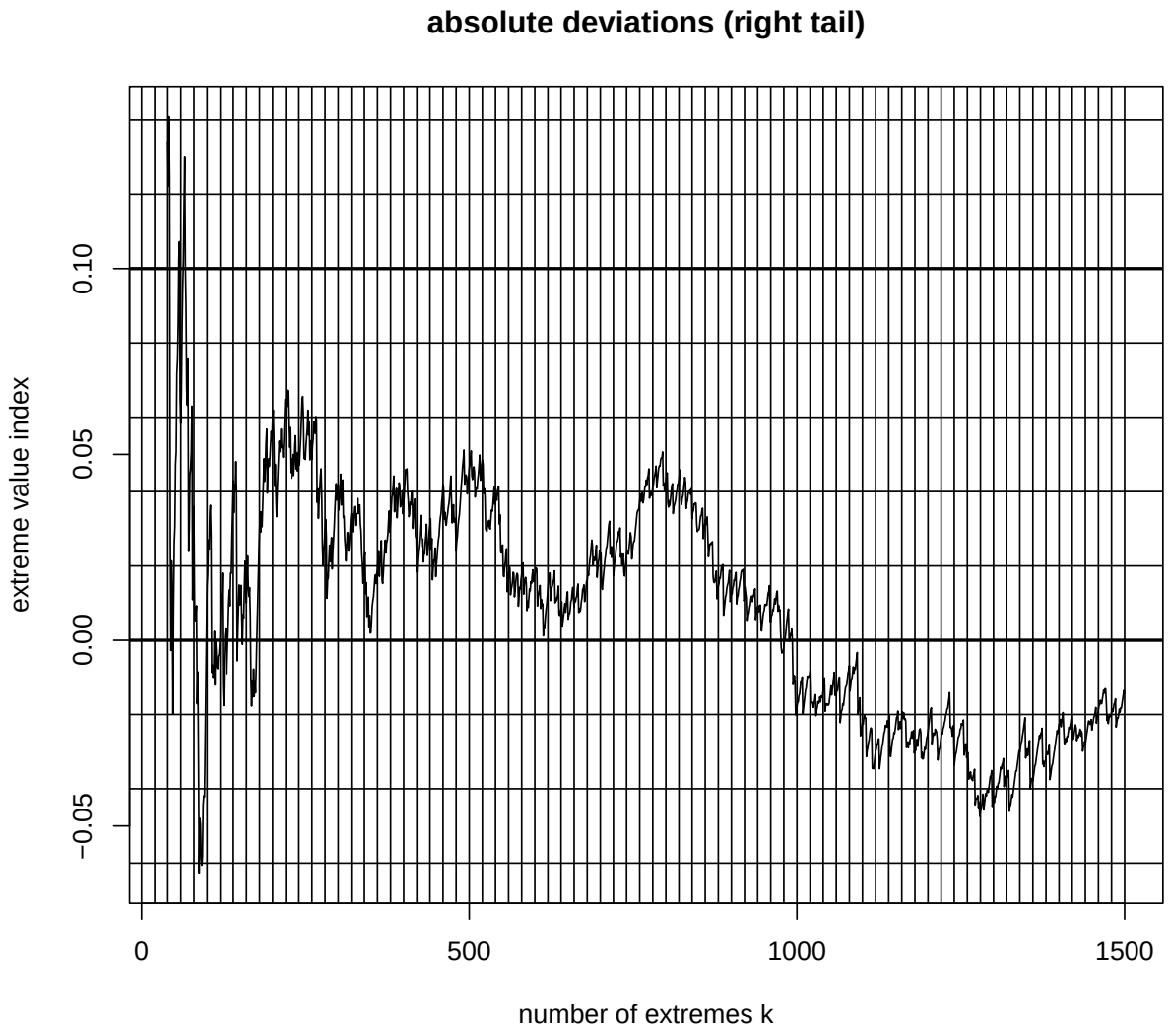


Figure 60: Estimation of Extreme Value Index for Various Thresholds Using Absolute Deviations Over Threshold u

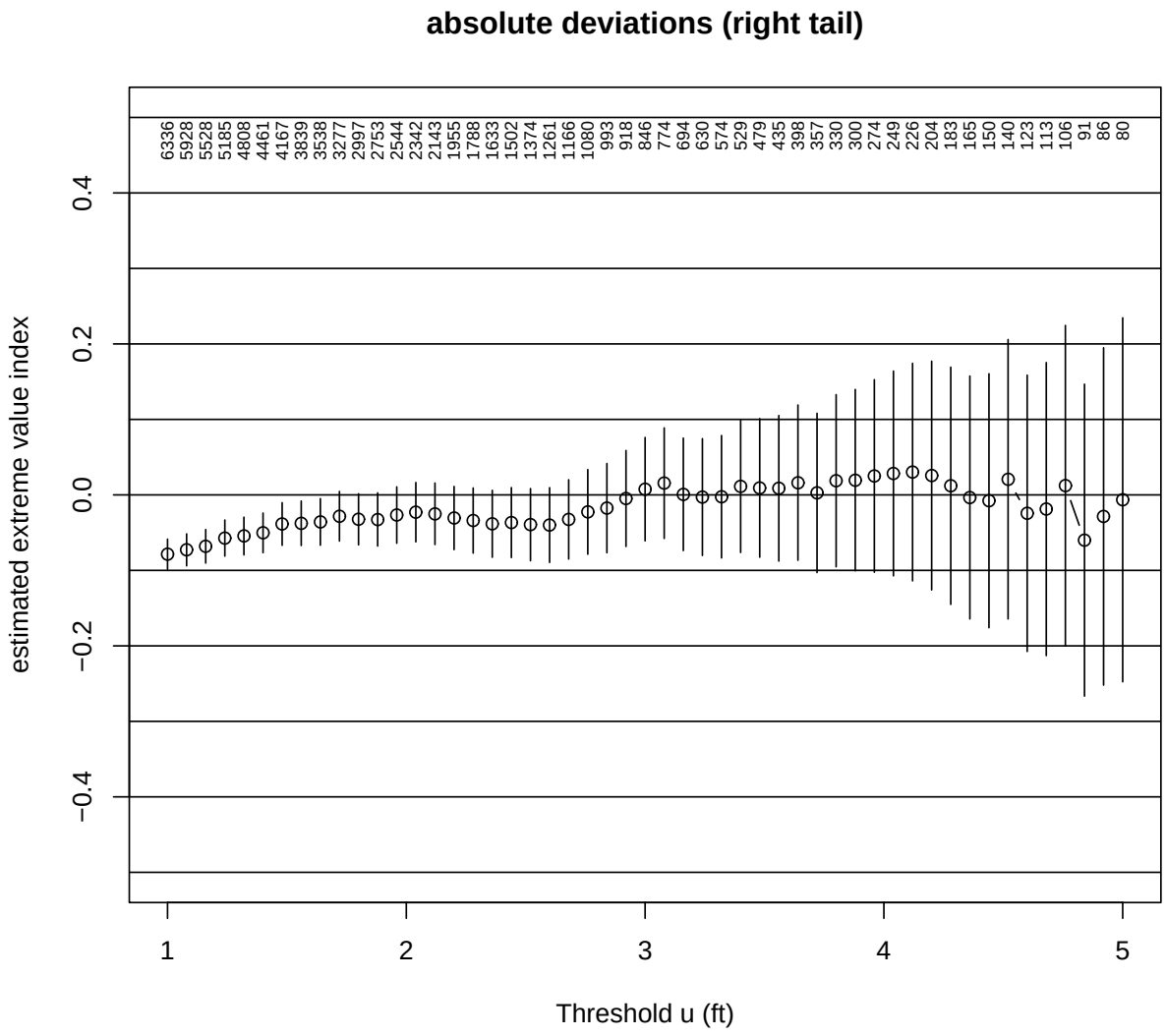


Figure 61: Estimation of Extreme Value Index for Various Thresholds
Using Absolute Deviations Over Threshold u

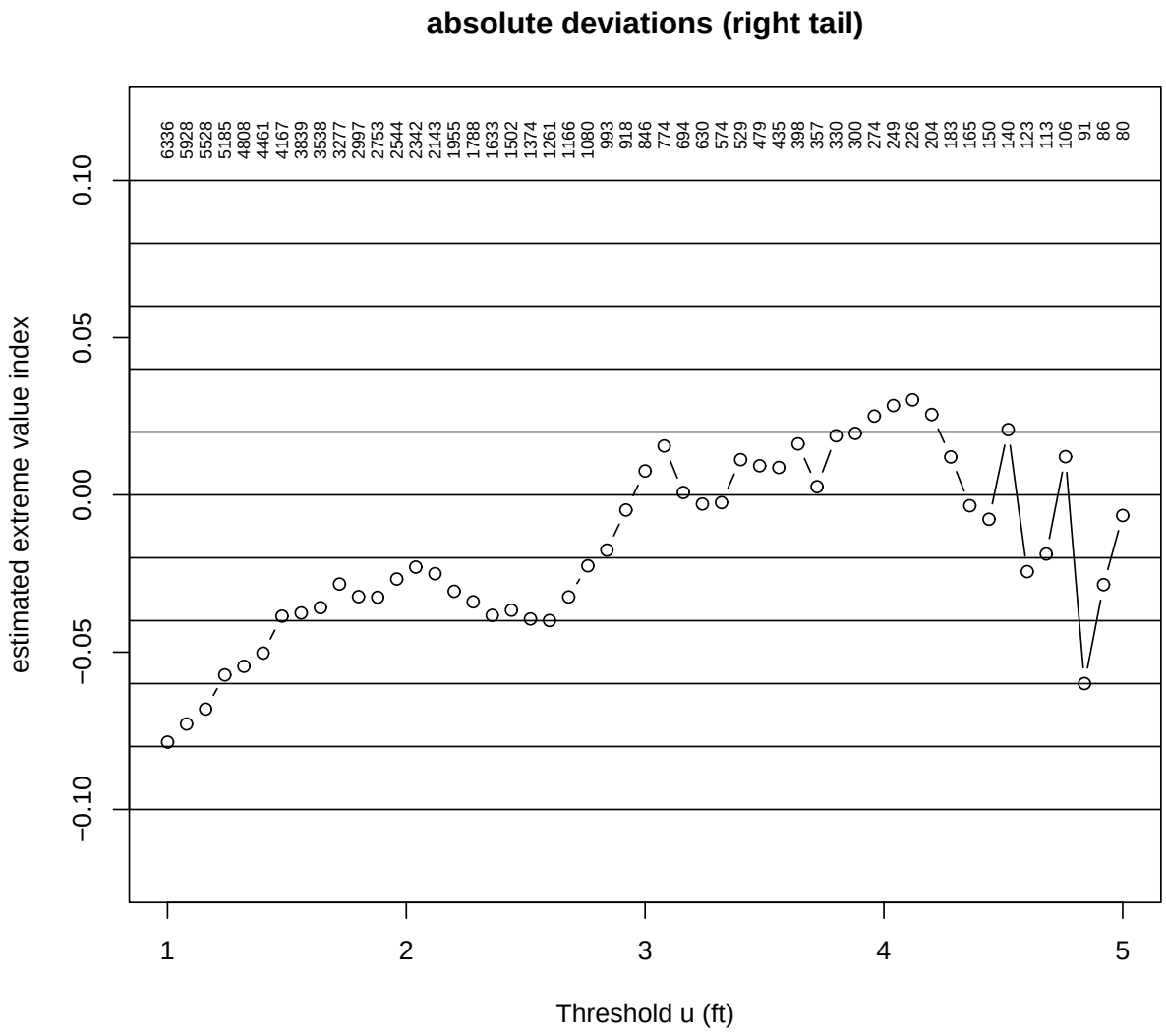


Figure 62: Estimated Median over Threshold for Various Thresholds Using Absolute Deviations Over Threshold u

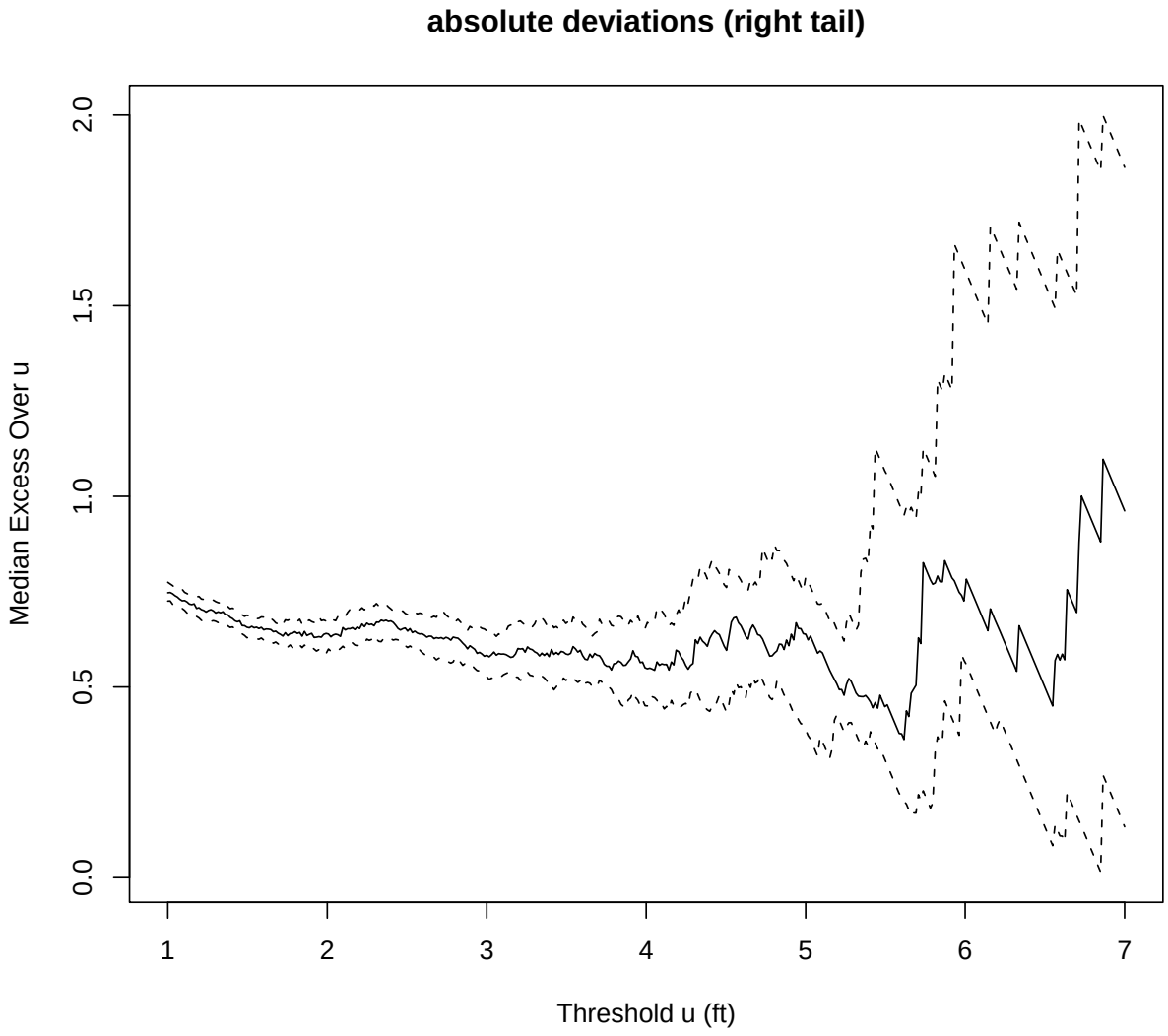


Figure 63: Extrapolation for Combined Adjusted ANC/JFK Data
Using 859 Most Extreme Absolute Deviations

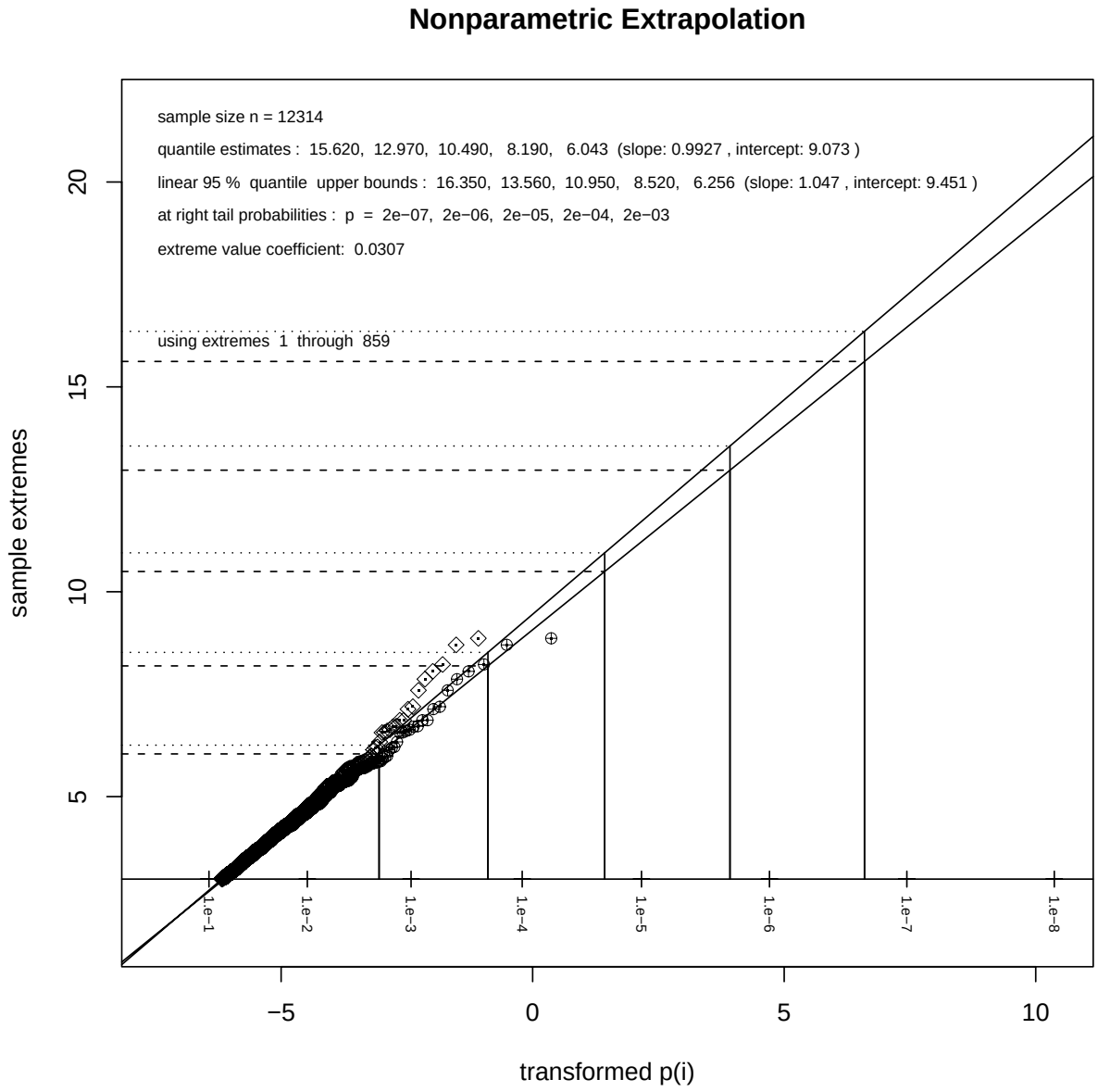


Table 7: ANC/JFK Estimated Thresholds by Exceedance Risk
for Adjusted 747 Centerline Deviations from Taxiway Centerline

based on	exceedance risk				
absolute deviations	$2 \cdot 10^{-7}$	$2 \cdot 10^{-6}$	$2 \cdot 10^{-5}$	$2 \cdot 10^{-4}$	$2 \cdot 10^{-3}$
	estimated thresholds				
JFK Data $k = 200$	14.74 ft	12.95 ft	11.05 ft	9.03 ft	6.87 ft
ANC Data $k = 700$	15.84 ft	13.03 ft	10.48 ft	8.13 ft	5.99 ft
ANC & JFK Data $k = 859$	15.62 ft	12.97 ft	10.49 ft	8.19 ft	6.04 ft
last row multiplied by 1.097	17.14 ft	14.22 ft	11.51 ft	8.98 ft	6.63 ft
	95% upper confidence bounds				
JFK Data $k = 200$	16.06 ft	14.10 ft	12.00 ft	9.77 ft	7.39 ft
ANC Data $k = 700$	16.66 ft	13.69 ft	10.98 ft	8.49 ft	6.22 ft
ANC & JFK Data $k = 859$	16.35 ft	13.56 ft	10.95 ft	8.52 ft	6.26 ft
last row multiplied by 1.097	17.94 ft	14.87 ft	12.01 ft	9.35 ft	6.86 ft

Also shown in Table 7 are thresholds derived from the ANC/JFK data after multiplication by 1.097 to render them applicable to the JFK context. It turns out that these scaled values agree reasonably well (within roughly half a foot) with the values obtained from the JFK data alone provided we limit the scope to two-sided risks from $2 \cdot 10^{-3}$ to $2 \cdot 10^{-5}$. For smaller risks the gap widens and one is faced with the question which is more appropriate. Recall that the JFK data alone yielded the extreme value index estimate $c = -0.02774$ which by its sign suggests a hard upper limit. On the other hand, the ANC/JFK data produced an estimate of $c = 0.0307$ which suggests that there is no upper limit. Also recall the good linear agreement between the ANC and the JFK data as portrayed in Figure 57. Based on this, the fact that the ANC/JFK analysis is based on more data, and the fact that the scaled values only extend by 1.3 ft beyond the ANC $2 \cdot 10^{-7}$ thresholds (estimated and 95% confidence bounds) we feel that the scaled values from the ANC/JFK analysis present a reasonable and conservative choice.

From Table 7 we can again derive the corresponding back-adjusted thresholds by adding .75 ft either to the ANC/JFK thresholds or to the rows which show the scaled thresholds, depending on whether the results should apply to the ANC or JFK context respectively.

Recall that we had settled on a common conservative bias adjustment of .75 ft for both ANC and JFK. The results are tabulated in Table 8.

Table 9 gives the corresponding risks for exceeding the taxiway edge for different taxiway widths and aircraft when based on the extrapolation results from the combined ANC/JFK data. That table breaks down those risks as they apply to the ANC and the JFK contexts. The reason for this differentiation is that JFK had wider scatter (by a factor of 1.097) than ANC.

Table 8: Back-Adjusted Thresholds by Exceedance Risk
for 747 Centerline Deviations from Taxiway Centerline

Based on Analysis of Combined ANC/JFK Data

$n = 12314$	two-sided exceedance risk				
	$2 \cdot 10^{-7}$	$2 \cdot 10^{-6}$	$2 \cdot 10^{-5}$	$2 \cdot 10^{-4}$	$2 \cdot 10^{-3}$
Applied to ANC Context	Thresholds				
estimate	16.37 ft	13.72 ft	11.24 ft	8.94 ft	6.79 ft
95% upper bounds	17.10 ft	14.31 ft	11.70 ft	9.27 ft	7.01 ft
Applied to JFK Context	Thresholds				
estimate	17.89 ft	14.97 ft	12.26 ft	9.73 ft	7.38 ft
95% upper bounds	18.69 ft	15.62 ft	12.76 ft	10.10 ft	7.61 ft

Table 9: Two-Sided Exceedance Risks for 747 Based on ANC/JFK Data

Risk of Outer Main Gear Tire Edge Exceeding the Taxiway Edge
for Different Taxiway Width Standards, Using 747, A380 & NLA Main Gear Dimensions.
Does Not Compensate for the Different Nose to Main Gear Distance for the A380 or NLA.

Main Gear Dimensions	Taxiway Width	75 ft	82 ft	100 ft
	as applied to ANC context			
from 747	Estimated Risk	$1.36\ 10^{-7}$	$8.46\ 10^{-9}$	$1.70\ 10^{-11}$
	95% Upper Bound on Risk	$2.48\ 10^{-7}$	$1.69\ 10^{-8}$	$4.07\ 10^{-11}$
from A380	Estimated Risk	$1.57\ 10^{-6}$	$7.97\ 10^{-8}$	$1.07\ 10^{-10}$
	95% Upper Bound on Risk	$2.64\ 10^{-6}$	$1.48\ 10^{-7}$	$2.44\ 10^{-10}$
from NLA	Estimated Risk	$1.98\ 10^{-5}$	$7.97\ 10^{-7}$	$6.94\ 10^{-10}$
	95% Upper Bound on Risk	$3.01\ 10^{-5}$	$1.37\ 10^{-6}$	$1.50\ 10^{-9}$
	as applied to JFK context			
from 747	Estimated Risk	$4.51\ 10^{-7}$	$3.25\ 10^{-8}$	$8.66\ 10^{-11}$
	95% Upper Bound on Risk	$7.90\ 10^{-7}$	$6.22\ 10^{-8}$	$1.98\ 10^{-10}$
from A380	Estimated Risk	$4.55\ 10^{-6}$	$2.72\ 10^{-7}$	$5.05\ 10^{-10}$
	95% Upper Bound on Risk	$7.33\ 10^{-6}$	$4.86\ 10^{-7}$	$1.10\ 10^{-9}$
from NLA	Estimated Risk	$4.91\ 10^{-5}$	$2.40\ 10^{-6}$	$3.00\ 10^{-9}$
	95% Upper Bound on Risk	$7.21\ 10^{-5}$	$3.95\ 10^{-6}$	$6.21\ 10^{-9}$

10 Do These Results Generalize?

The question arises what these analysis results mean beyond the taxiways (75 ft wide straight segment with shoulders), the aircraft type (747), and the location (JFK) for which the data were collected. Although we found that the deviation data from the two taxiways (ALPHA and BRAVO) showed comparable behavior once we corrected for any biases one cannot make any conclusions concerning similar behavior once the common factors (JFK, 747, 75 ft wide straight segment with shoulders and parallel taxiways) for both taxiways no longer apply. To draw any such conclusions one would need similar data collections from other airports, other types of taxiway segments (different widths and without shoulders), and for other types of aircraft, and show that the deviation behavior does not change in any appreciable way. One step in this direction will be the comparison with data and the analysis from the ANC study, although this is still for 747s and for 75 ft wide straight segments with shoulders, but at a different airport. One issue that could make a difference is the offset of centerlights from the taxiway centerline. At JFK the offset is about 18 to 21 inches while at ANC it was around 12 inches.

The taxiway width and presence/absence of shoulders could have a limiting effect on the deviations, i.e., the pilots would follow the centerline more closely or at least the very extreme deviations could be curtailed. That could change the extrapolation behavior significantly.

As for expecting similar deviation results for different aircraft there are several reasons why that may not be the case. Smaller aircraft have different steering response behavior and because of their smaller size there may be the opposite effect to self limiting. Until the appropriate data are collected and the relevant comparisons are made one should not make any judgments on this.

The form of data collection through stationary range finder lasers at ANC is sufficiently similar to that at JFK. Thus one might expect similar deviation behavior at ANC and JFK. However, the above mentioned offsets for the centerlight could make a difference. Furthermore, at ANC much of the 747 traffic consisted of freighters which could cause different behavior in piloting. To some extent the similarity is there, except for the scaling factor of 1.097. Making more such comparisons will give us a better idea about such systematic differences and how to bound their effects.

The issue of data collection method will come into play when comparing with data from other airports. The airplanes and pilots are international and deviations "should" be the same at all airports, all other factors being equal (weather patterns, centerlight offsets, shoulders, taxiway widths). However, each data collection method has its own idiosyncracies for identifying aircraft and each has its own measurement variations. Some misidentified aircraft or objects might slip through or some legitimate aircraft is screened out for some reason. If this screening out in any way relates to extreme deviation behavior it will cause different analysis results.

11 Cautionary Remarks on Extrapolation

The usual cautionary reminder is given with respect to the risk extrapolations given here. There may be additional extreme value behavior that has not yet had a chance to manifest itself and that may exhibit more extreme behavior than indicated by the observed linear pattern used for extrapolation. On the other hand there may be a natural feedback loop through the pilot's increased awareness of the approaching taxiway edge that would prohibit much larger extreme observations than already observed. As far as the confidence bounds are concerned, they do not take into account the uncertainty in the choice k of the number of extremes to be used and the estimation uncertainty of the extreme value index c that is used to make the extrapolation pattern look linear. These uncertainties arise from the inherent sampling variation in the data. This means that a different study at JFK with a different collection of $\approx 2,518$ deviations (different due to sampling variations) might have led to a somewhat different choice of k and a different estimate of the extreme value index. This might have led to different estimates and confidence bounds. To some extent this deficiency will be counteracted by the conservative nature of the nonparametric confidence bounds. They do not rely on the assumptions underlying extreme value analysis. It is hoped that future research will allow taking all these concerns into account and adjust for them.

References

[1] Booker, A. (1995), "Statistical Analysis of Aircraft Deviations from Taxiway Centerline," Boeing Information & Support Services.

[2] Cohen-Nir, D. and R. Marchi, "Preliminary Analysis of Taxiway Deviation Data and Estimates of Airplane Wingtip Collision Probability,"
Paper Number TRB-03-3302, Richard Marchi (corresponding author),
Senior Vice President, Technical and Environmental Affairs,
Airports Council International–North America,
1775 K Street, Suite 500 NW, Washington DC 20006, Email: rmarchi@aci-na.org

[3] Coles, S. (2001), *An Introduction to Statistical Modeling of Extreme Values*, Springer-Verlag, London.

[4] FAA Advisory Circular 150/5300-13, Airport Design (Table 4-1)

[5] ICAO ANNEX 14, Volume I, Aerodrome Design and Operations,
Chapter 3, Paragraph 3.8.3.

[6] King, Ryan, FAA, personal communication.

[7] Macdonald, J., Boeing 747 Chief Project Pilot.

[8] Scholz, F.W. and Stephens, M.A. (1987), "K-Sample Anderson-Darling Tests," *J. Amer. Statist. Assoc.* **82**, pp. 918-924.

[9] Scholz, F.W. (1995), "Nonparametric Tail Extrapolation," *ISSTECH-95-014*, ISS Technology, Boeing Information

[10] Scholz, F.W. (2003), "Statistical Extreme Value Analysis of ANC Taxiway Centerline Deviations for 747 Aircraft."

[11] Sparacino, P., July 2000, "Collect and Analyze Data on Centerline Deviations of Wide Body Aircraft on Airport Taxiways," DOT/FAA/AAR-410/xx July 2000 Draft, FAA William J. Hughes Technical Center, Atlantic City International Airport, NJ 08405