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

Summary

This report describes the analysis of 747 taxiway centerline deviation data that were collected from 9/24/2000 to 9/27/2001 at Anchorage International Airport. Deviations were measured for nose and main gear at two laser locations for each of two 75 ft straight taxiway segments with shoulder called KILO and ROMEO, respectively. The initial discrimination of 747 aircraft was based on the perceived lateral distance between nose gear and main gear. Unfortunately that may also have led to the inclusion of some MD-11s or DC-10s in this group.

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 10,000 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. The used extrapolation methodology [8] had previously been applied by Andrew Booker [1] to similar data collected from Schipol Airport at Amsterdam.

Altogether 10133 747-events were recorded of which only 9796 were eventually used for the final analysis. This 3.3% reduction in the data resulted from data cleaning using various filters, involving timing of laser hits for nose and main gear, velocity and acceleration, and lateral nose to main gear distance. For this cleaning it was very important to have accompanying information (timing, velocity, acceleration, and lateral nose to main gear distances) available, since screening based on just excessive extremeness of deviations would be very suspect for the following two reasons. First, one would have difficulty setting appropriate thresholds for excessive extremeness as opposed to natural extreme behavior of deviations and second, it is this extreme behavior we wish to characterize. The resulting 3.3% reduction in data volume hardly matters for the final analysis.

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

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

gears at the two lasers against each other. These plots showed great consistency between deviations for the same events and they also showed up some minor peculiarities that were attributed to residual effects of turning maneuvers. This led to choosing the main gear deviations at Laser 2 for each taxiway as the focus for further analysis since those deviations would be least affected by such effects and since the object of the study was to characterize the deviation behavior for straight taxiway segments. Also, 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" the sample size this way would be quite misleading. Furthermore, the methodology in [8] 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. We were able to isolate 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 .27 ft. This should be seen as an average offset behavior over the 9796 observed events. The other bias, estimated as 0.48 ft, was attributed to the centerlights whose centers are located about 12 inches away from the center of the centerline stripe on the far side from the lasers. This larger bias resulted in a consistent deviation away from the laser. Sometimes the parallax bias would reenforce it and lead to a total bias of .75 ft and sometimes it would diminish it to a total bias of .21 ft.

The reason for this bias behavior attributed to the centerlights is that the latter protrude. This can lead to bumping of the nose gear wheels which the pilot wishes to avoid. If the pilot positions the aircraft exactly on the centerline there will be bumping at each centerlight. Thus the preferred strategy may be to straddle the centerlights. This does not allow for much margin of deviation since the diameter of the centerlights is 10 inches while the inside distance between nose gear wheels is 19 inches. As soon as this margin is exceeded the bumping will start and sometimes the corrective maneuver may be wrong. This leads to a back and forth weaving motion that greatly adds to the natural deviation behavior that would exist if there were no centerlights. The point is that these offset centerlights not only create some amount of bias but they also add variability to the process. This issue may well deserve further study and steps for resolving this issue may be called for.

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 9796 deviations. The range of this set is $[-8.225, 8.863]$ 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 required 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 [8]. 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 answers. The analyses for separate positive and negative deviations were carried out to examine consistency. The behavior for positive deviations was more along the lines expected for extreme value analysis while the negative deviations showed some atypical behavior. Presumably this is caused by the centerlights, but we have no further insights to support this. For this reason we feel that the extrapolation of the absolute deviations represents a form of compromise and it is a practical position to take.

The extrapolation deviation thresholds for various risk levels were then back-corrected for the previously subtracted biases. This 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

	two-sided exceedance risk				
	$2 \cdot 10^{-7}$	$2 \cdot 10^{-6}$	$2 \cdot 10^{-5}$	$2 \cdot 10^{-4}$	$2 \cdot 10^{-3}$
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

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 same 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.

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	$1.66\ 10^{-7}$	$1.23\ 10^{-8}$	$4.13\ 10^{-11}$
	95% Upper Bound on Risk	$3.09\ 10^{-7}$	$2.50\ 10^{-8}$	$9.81\ 10^{-11}$
from A380	Estimated Risk	$1.69\ 10^{-6}$	$9.99\ 10^{-8}$	$2.19\ 10^{-10}$
	95% Upper Bound on Risk	$2.91\ 10^{-6}$	$1.90\ 10^{-7}$	$4.99\ 10^{-10}$
from NLA	Estimated Risk	$1.95\ 10^{-5}$	$8.85\ 10^{-7}$	$1.21\ 10^{-9}$
	95% Upper Bound on Risk	$3.05\ 10^{-5}$	$1.56\ 10^{-6}$	$2.63\ 10^{-9}$

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.

Since the results reported here are based on data collected for 747s at ANC 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 short answer to this is: Not until one has conducted similar studies under conditions that vary the above factors. Only then can one say what changes could occur and whether any are

practically significant. 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 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 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 ANC with a different collection of $\approx 10,000$ 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.

The data set analyzed here was collected by the FAA and is sometimes referred to as the Phase I data as opposed to another data set called the Phase II data. The latter was collected during subsequent time spans of 7/2/2001-7/2/2002 for ROMEO and 9/27/2001-7/2/2002 for KILO through the University of Alaska. This phase II data set captured 14318 events of a 747 passing the lasers. It is planned to analyze this data set as well and compare findings. If the data are reasonably comparable an analysis of the pooled data would also be feasible.

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. I thank Pete Sparacino of the FAA for the beautiful cover photo of the 747 taxiing at Anchorage International Airport. 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 from the same airport, but covering a different time period (7/2/2001-4/4/2002). This report caused us to look at the data quality more critically.

1 Data Description

At Anchorage International Airport (ANC) data on taxiway deviations from the taxiway centerline were collected from 9/24/2000 to 9/27/2001 at two laser locations for each of two straight 75-foot taxiway segments with shoulder called KILO (9/24/2000-9/27/2001) and ROMEO (9/24/2000-7/2/2001), respectively. The layout of the lasers is shown in Figure 1 which is based on Figure 2 of [10]. The lasers are capable of measuring the distance to anything hit by the laser beam. They were set up to intercept passing aircraft at the landing gear wheel level. The measurement error for this process was given as 1-2 inches.

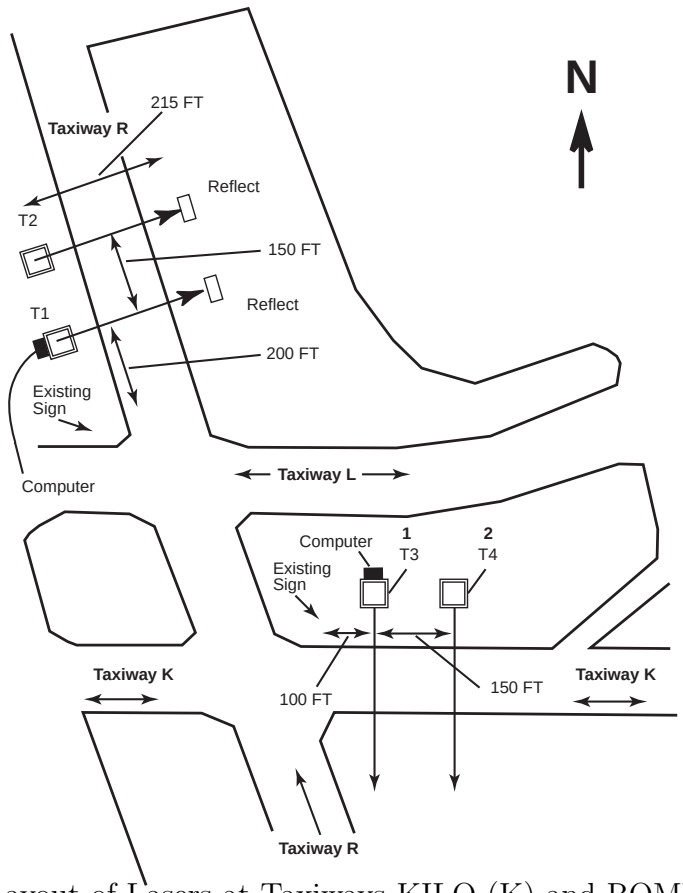


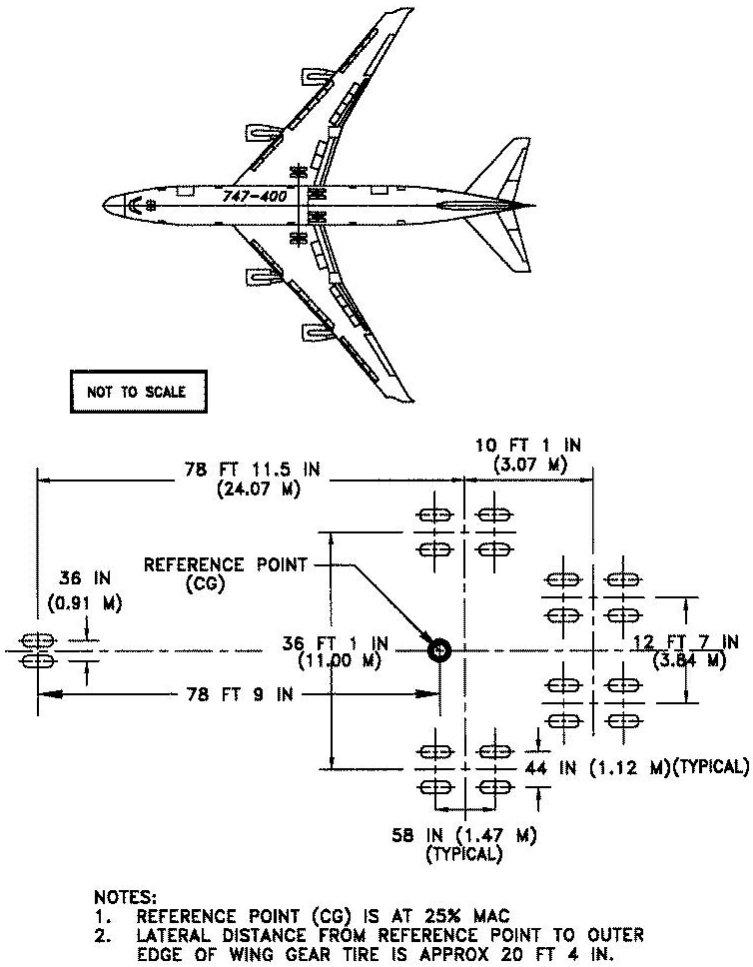
Figure 1: Layout of Lasers at Taxiways KILO (K) and ROMEO (R) at ANC

The data for 747 deviations were separated out using a discrimination criterion based on the lateral nose to main gear distance. The analysis of this 747 data set is the subject of this report. The main objective is to characterize the extreme deviation behavior of 747 aircraft as they taxi on straight 75-foot taxiway segments.

According to [10, pages 4 & 8] a passing aircraft was judged to be a 747 when the lateral difference between measured distances to nose and main gear exceeded 17 ft. This is presumably motivated by the fact that for a 747-400 the lateral distance between outer tire edges

of nose and main gear is roughly 18.375 ft, see Figure 2. It turns out that the MD-11 and DC-10 also have a lateral separation between nose and main gear in excess of 17 ft and thus they may be lumped in with the 747 events.

Figure 2: 747 Aircraft Dimensions



For each distance measurement there also is a time recording. These time recordings were used to decide whether a laser hit (measurement) corresponded to a real event, namely an aircraft passing the two lasers. The time measurements also make it possible to calculate aircraft velocities at each laser and between lasers. The velocity between lasers can be based on the 150 ft distance between the lasers while the velocity at the lasers requires the longitudinal distance between nose and main gear, i.e., one will need to have made a prior decision on aircraft type.

The taxiway deviations are accompanied by the direction in which the 747s were moving when their nose and main gears crossed the two laser beams. Deviations from centerline were calculated at both laser locations for nose and main gears. These deviations represent the deviation of the airplane centerline (also the aircraft longitudinal axis) at that gear location from the center of the taxiway centerline stripe. They were calculated from the raw measurement data for the lateral distances to the outer nose and main gears, from the aircraft dimensions, and from the known distances of the lasers to the taxiway centerlines.

For taxiway KILO Laser 1 was triggered first when the aircraft was moving eastbound while Laser 2 was triggered first while it was moving westbound. For taxiway ROMEO Laser 1 was triggered first in the northbound direction while Laser 2 was triggered first in the southbound direction.

Although it is stated in [10, page 4] that the taxiway width is 95 ft (halfwidth 47.5 ft) the actual width according to [6] is 75 ft. The extra 10 ft on either side just defined the zone for measuring aircraft, on or off the taxiway.

For taxiway ROMEO 6540 events (passings of 747 aircraft) were recorded while for taxiway KILO this number was 3593 for a total of 10133 events on both taxiways combined. Each event gave rise to nose gear and main centerline deviations at Laser 1 and Laser 2 and is represented by two consecutive records in the provided Excel spreadsheets for KILO and ROMEO.

2 Data Cleaning

In the following we examine the data integrity from various angles with the result that some event data are repaired and some are simply omitted. After all such checks there were 73 events from KILO and 264 events from ROMEO that were omitted from further analysis for good cause. This represents 337 cases or 3.3% of the orignial data cases. The 3.3% reduction in data volume hardly matters for the purpose of the analysis.

2.1 Direction Screen

There were 7 cases among the KILO events for which the heading was recorded as northbound and 2 cases where the heading was recorded as southbound. Apparently [6] this abnormality was caused by a computer program update that migrated from ROMEO to KILO and in the migration not all the hard coded directions were initially changed. For these 9 cases we converted northbound and southbound to eastbound and westbound, respectively.

2.2 Timing Screen

In our first analysis the data were taken more or less at face value. In particular not much attention was paid to the timing data except to look at the velocities between lasers. At that point we noted a small portion of cases for which the velocities were negative.

We did not verify the time screening rules laid out in [10, pages 4 & 7]. It was assumed that these rules had been applied. This first analysis resulted in several extreme abnormalities when comparing nose and main gear deviations from the taxiway centerline at the two lasers for the same aircraft. We will revisit these later. A preliminary report by Dan Cohen-Nir and Richard Marchi [2] (based on similar if not the same data) showed up a fair number of problem cases based on various filtering criteria. Many of these cases exhibited timing irregularities. Based on this we decided to apply a timing screen as well.

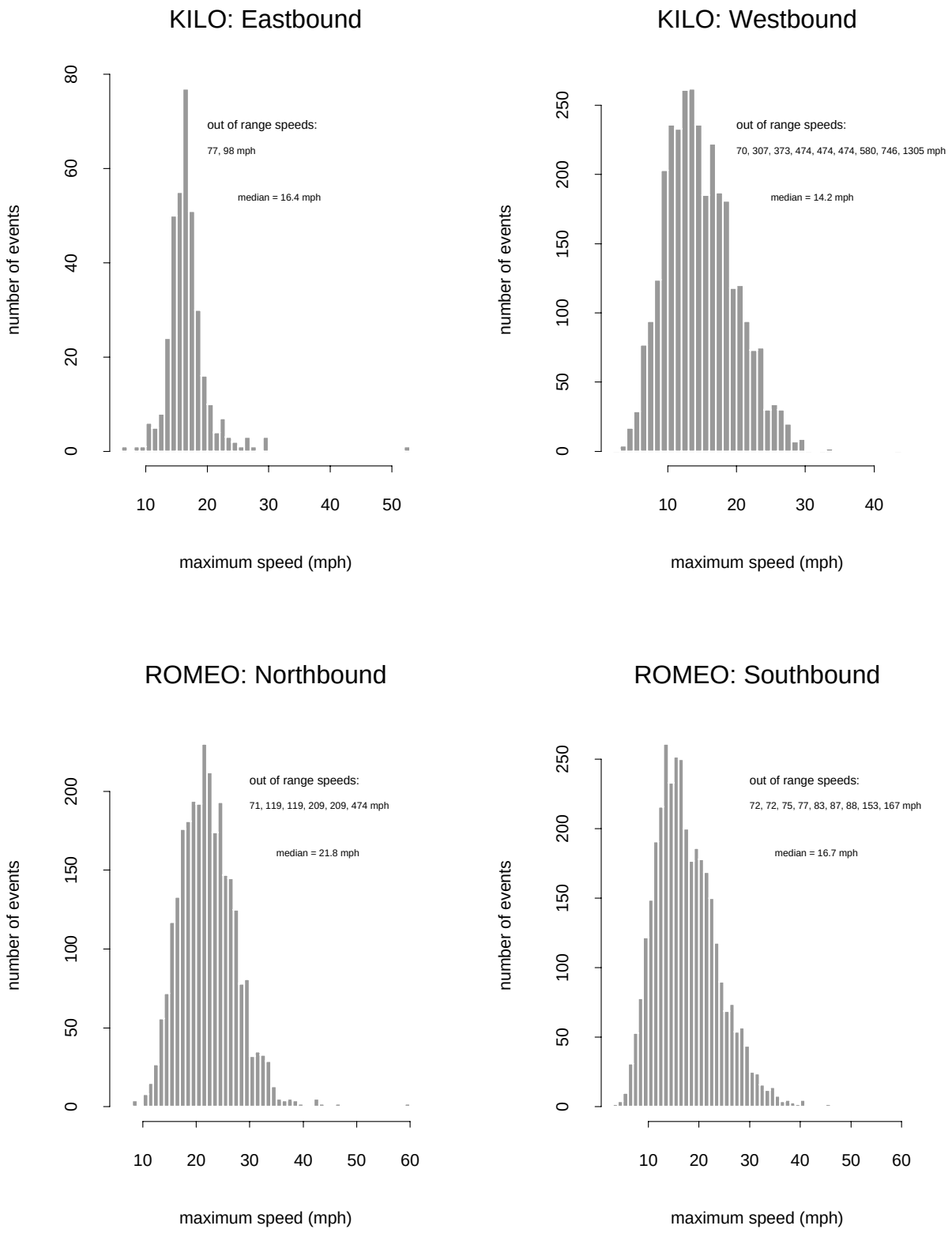
The time screening rules laid out in [10, pages 4 & 7] are not entirely consistent. There are differences between what is in the text on page 4 of [10] and what is shown in the flow chart on page 7 of [10]. We have been told [6] that the rules in the text apply when there is a contradiction. According to these rules not more than 30 seconds can elapse between nose and main gear at each laser. Also, not more than 60 seconds can elapse between the second hits on the first and second laser. The inconsistency arises in the case of this 60 second rule, where the flow chart indicates that not more than 60 seconds should elapse between the second hit on the first laser and the first hit on the second laser. In addition to these timing rules we also screened the timing data for being in increasing order. In particular, the nose gear time at the second laser should not come before the main gear time at the first laser, since the distance between nose and main gear of about 76.54 ft and the distance between lasers of 150 ft would make this a stretch. Based on this timing screen 55 cases from KILO and 243 cases from ROMEO were disqualified. In particular, the screen for increasing time order ruled out the previous 204 cases with negative velocities.

2.3 Velocity Screen

Since the timings between the various laser hits were recorded it was possible to convert these into velocities. Thus we can associate four basic velocities with each event, the velocities at the first and second laser (based on the times of nose and main gear passing the first and second laser and on the longitudinal distance between nose and main gear, i.e., 76.54 ft) and the nose and main gear velocities between lasers (based on the times of the nose and main gears crossing each laser and the distance between the lasers, i.e., 150 ft). We screened out any event with any of the velocities ≥ 100 ft/sec, which corresponds to 68.2 mph. Figure 3 shows the histograms of the maximum velocities (of the available four) for each event and identifies the screened out maximum speeds. After the timing screen there were additionally 11 events for KILO and 15 events for ROMEO that were omitted based on this velocity screen.

In each plot of Figure 3 the median speed is indicated. The speeds on taxiway ROMEO northbound appear to be at least 5 mph higher when compared to the speeds at KILO or in the opposite direction at ROMEO. This may be explainable by the fact that the northbound direction at ROMEO is outbound and with no apparent intersection traffic.

Figure 3: Maximum Speeds (at Each Laser & from Laser to Laser) cleaned using timing rules



2.4 Acceleration Screen

Cohen-Nir and Marchi [2] used the timings to check consistency of calculated nose to main gear distances with those of a 747. These calculations are based on assuming a more or less constant velocity as the aircraft passes both lasers. Our reverse approach, assuming a given 747 landing gear geometry, allowed us to screen for abnormal velocities and accelerations. This may run into a slight problem for the occasionally misclassified MD-11 or DC-10 which have respective nose to main gear distances of 78.08 ft and 69.71 ft. This is different from the corresponding distance 76.54 ft for a 747, but should not affect the velocity or acceleration screen dramatically.

Based on the two velocities at each laser and the average time between laser measurements we also calculated an "average" acceleration. Based on this any event was screened out for which the absolute acceleration/deceleration exceeded 7 ft/sec^2. Figures 4-5 show the acceleration histograms separated by direction and by taxiway.

Also indicated are all accelerations outside the range $\pm$5ft/sec^2. However, only those events with acceleration outside the range $\pm$7ft/sec^2 were screened out. Based on the bulk of the accelerations one could have been more drastic in screening out the stragglers shown as dots below the histograms. As it is we respectively removed 8 and 7 additional events (after the timing screen) from the KILO and ROMEO data based on this acceleration screen.

2.5 Lateral Nose to Main Gear Separation Screen

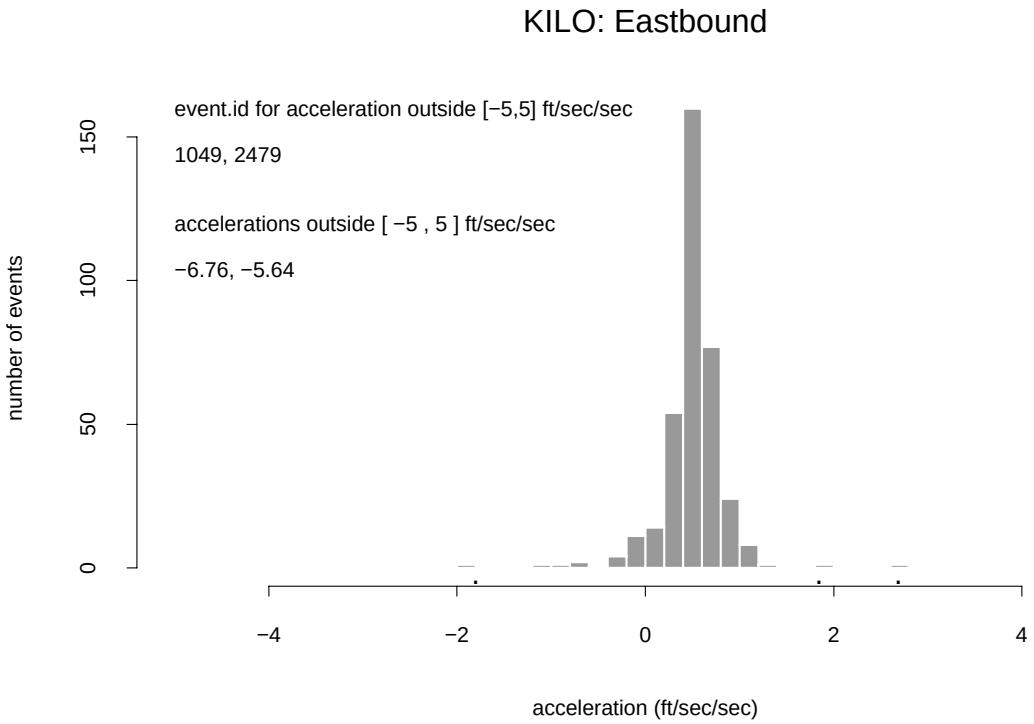
Previously it was noted that the lateral distance between nose and main gear was used to single out aircraft of 747 (MD-11 & DC-10) type. This was done by classifying an event as a 747 (MD-11 & DC-10) for which this difference between lateral distances to nose and main gears exceeded 17 ft, based on a nominal difference of about 18.5 ft (18.375 ft). The 1.5 ft allowance accounts for some measurement error (given as 1-2 inches in [10]) and for other sources of variation due to not moving parallel to the taxiway centerline. As will be seen later the 1.5 ft allowance was chosen very well. However, no upper limit was set for these differences between lateral distances to nose and main gears.

If a 747 passes a laser following a straight line trajectory but possibly at some angle $\pm\alpha$ to the taxiway centerline then the measured difference (without measurement error) would always be 18.5 ft (18.375 ft) or larger. In fact, under this assumption one can either back out the angle α (either sign would do) which would produce the measured lateral aspect difference d between nose and main gear or one can calculate the aspect difference d based on the angle α. The relationships are as follows

$$\alpha = \arccos\left(18.5/d\right) \quad \text{or} \quad d = \frac{18.5}{\cos(\alpha)} \; .$$

For example, setting an upper limit of 24 ft and allowing for the measurement error size of 2/12 ft one could take $d = 24 - 2/12 = 23.83$ ft and arrive at $\alpha = 39.1^\circ$ as the direction angle

Figure 4: Acceleration Check (KILO)



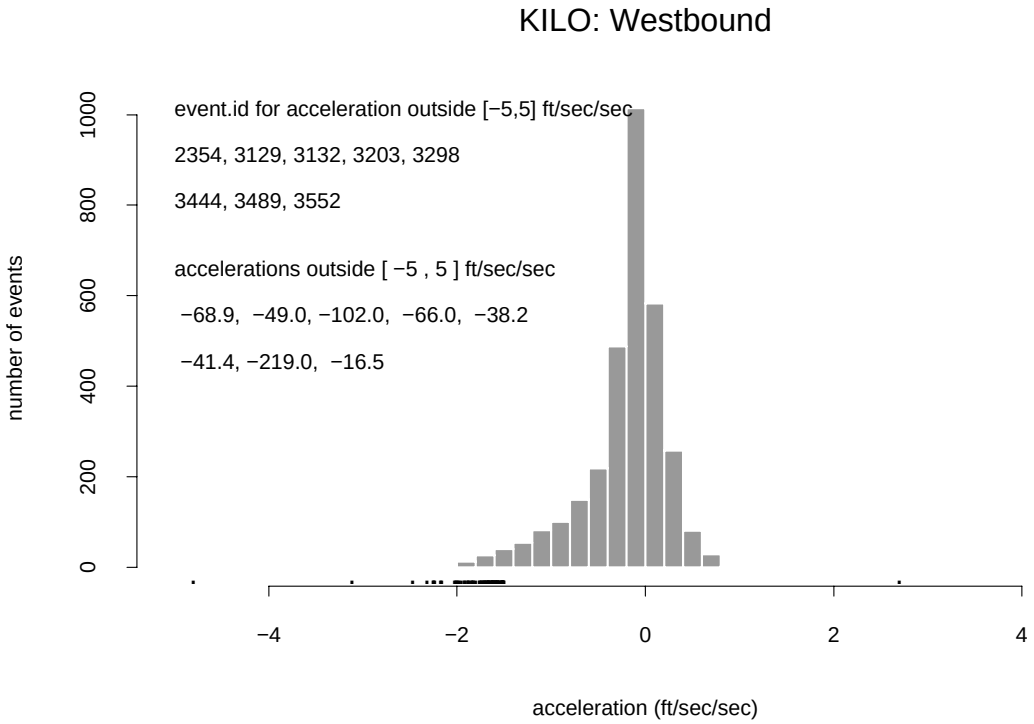
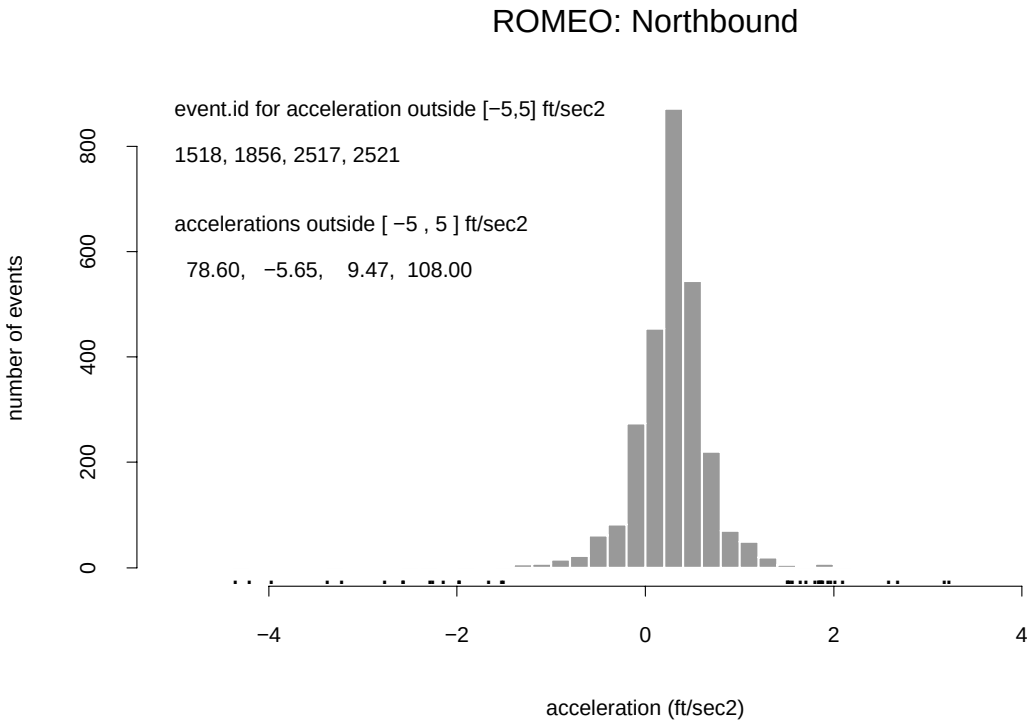
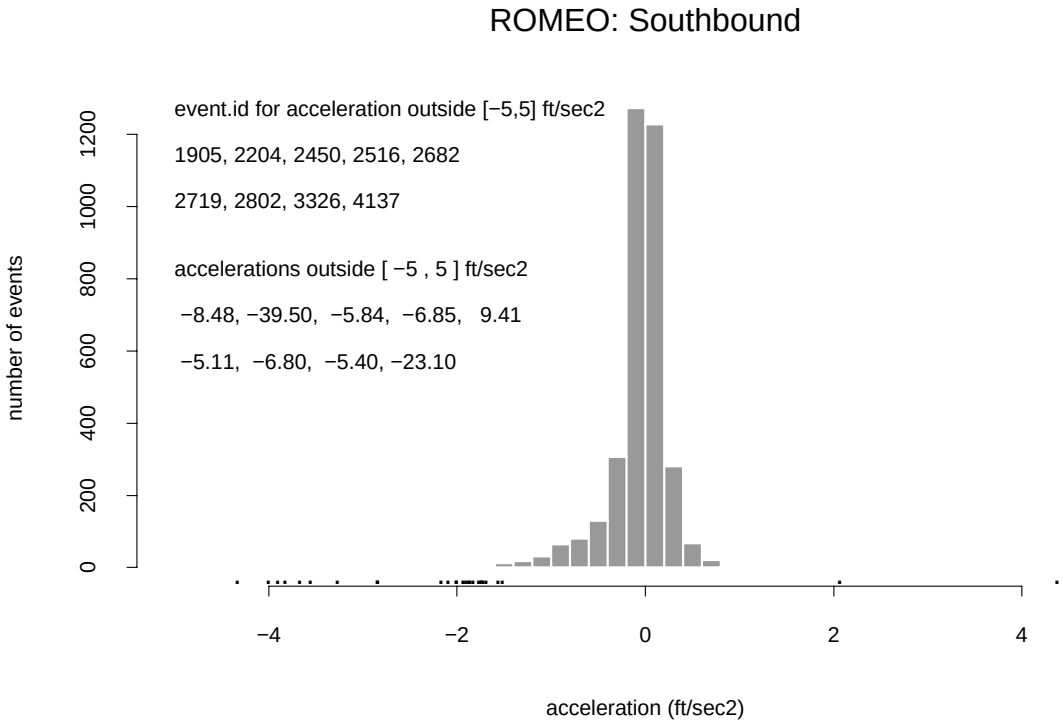


Figure 5: Acceleration Check (ROMEO)





by which a straight moving 747 would have to move to produce such an aspect difference between nose and main gear. As easy as such straight line analyses are they mostly do not apply, at least not for extreme angles. Typically the aircraft will move in an arcing or meandering fashion with low amplitudes except when doing real turns. Some such turning maneuvers may well be carrying their effect into the straight sections following such turns. On intuitive grounds it would seem that an arc bending away from the laser would reduce the lateral nose to mean gear distance from the nominal value of 18.5 ft while an arc bending toward the laser would increase it. It would be of some interest to understand this better on kinematic grounds.

Figure 6 shows scatter plots of these measured lateral distances between nose and main gear at the first and second laser, separated by direction and taxiway. In each scatter plot there are two rectangles. The solid rectangle (actually a square) contains all those distance pairs for which each distance is within 18.5 ± 1.5 ft while the dashed rectangle (square) contains all distance pairs for which each distance is between 16 ft and 24 ft. This latter rectangle contains all events that were not screened out based on this lateral distance criterion. The reason for dropping the lower screening bound from 17 ft to 16 ft was to admit one additional case which is visible below the solid line rectangle in the upper right plot. That event had a lateral distance above 17 ft when passing the first laser (and thus was admitted as an event) but apparently showed a lateral distance below 17 ft at the second laser. It appears that this lateral distance at the second laser was not used for screening. Another such case is indicated by its coordinates $(18.83, -1.01)$ in the upper right plot of Figure 6. This obvious nonsense case was screened out. The choice of 24 ft as upper limit as opposed to using no upper limit at all is explained below.

Figure 7 shows the same scatter plots after the data have been cleaned, using all the cleaning criteria (timing, velocity, acceleration, and lateral nose to main gear distance). Now the scale is magnified because the outliers are gone. With respect to these plots one can make some interesting observations. The point cloud in the upper right plot is mostly contained within the solid rectangle, i.e., almost all those lateral nose to main gear distances are within 18.5 ± 1.5 ft. Similar but not quite so strong patterns exist for the bottom two plots. There one sees some spillage of points outside the solid line rectangle. A much stronger spillage can be seen in the top left scatter plot. This indicates that at the first laser eastbound we tend to show a fair number of lateral distances in excess of 20 ft while these same events show lateral distances within 18.5 ± 1.5 ft at the second laser.

Our interpretation of these observations is as follows. The events captured in the upper right plot correspond to a section of KILO which is not much affected by turns into and from other taxiways. This appears to be especially true for laser 2 which is passed as the first laser westbound and in most cases it also appears to be true westbound for laser 1, the second laser. There does not appear to be much preparation effect for possible turns onto taxiway Romeo. Of course, this would depend strongly on how much of the traffic in this westbound direction would actually turn onto ROMEO. In contrast, the eastbound traffic on KILO appears to be affected by turns from ROMEO at laser 1 and these effects have died

Figure 6: Lateral Distances Between Nose Gear and Main Gear prior to cleaning of data

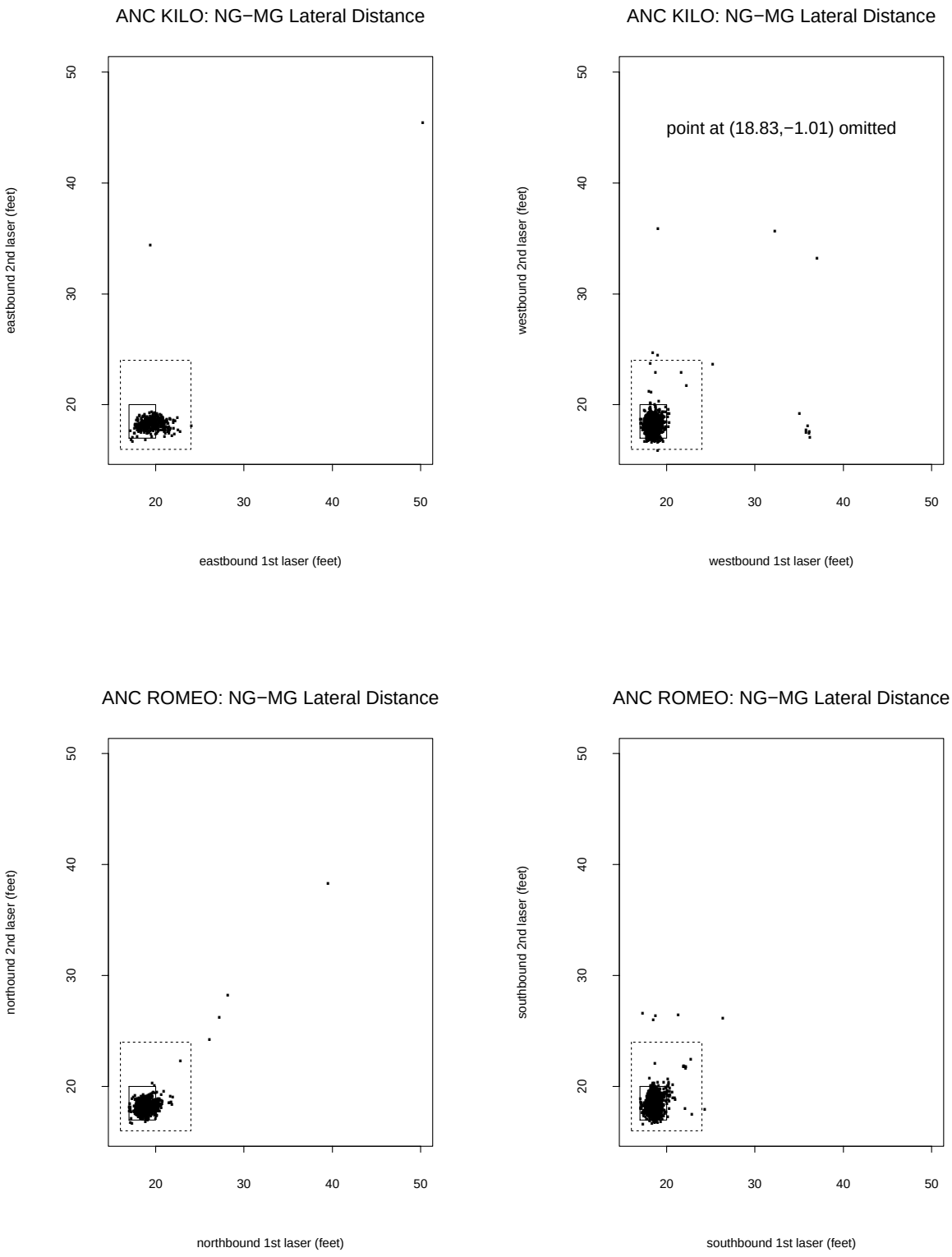
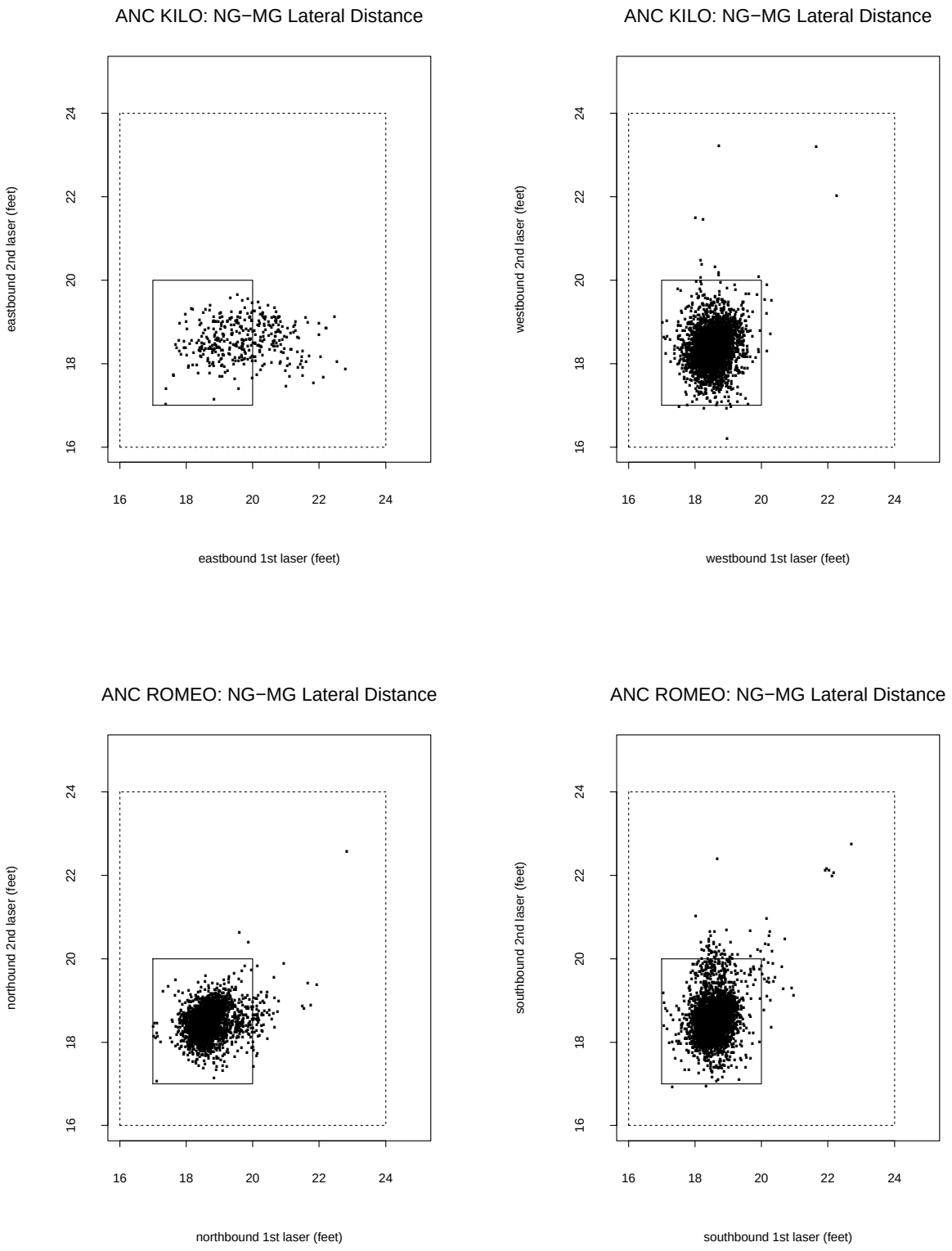


Figure 7: Lateral Distances Between Nose Gear and Main Gear after cleaning of data



down by the time this traffic reaches laser 2. This effect will show up again later in one of our correlation plots.

It was this natural wider scatter eastbound at laser 1 of KILO that motivated the choice of 24 ft as upper limit for our screen. Setting this limit so high we allow for these turning effects which will have died down by the time the second laser is reached. This allows more data to be used from that second laser.

Similar, but weaker, interpretations would apply to the bottom two plots of Figure 7. In each case the mild spillage could be linked to turns to and from KILO.

After all the screening was done a total of 9796 events remained as valid cases on which the remainder of the analysis will focus. This represents roughly 96.7% of the original data.

Similar considerations lead to choosing the deviations at laser 2 for ROMEO, both northbound and southbound, since at that laser the lateral nose to main gear separation is best contained within the 18.5 ± 1.5 ft. This is an indication that the aircraft are moving in a mostly straight fashion and are less influenced by turns.

The reason for focussing on the main gear deviations is that again a choice had to be made since nose and main gear deviations are highly correlated. The main gear deviations were judged to be most relevant for the purpose of the study, namely assessing the risk of leaving the taxiway and the risk of collision, either with a fixed structure or with another aircaft on neighboring taxiways.

3 Correlation and Time Plots

Originally the following diagnostic correlation and time plots were our only attempt at checking the data integrity. Clearly nose and main gear deviations for the same event and for the same laser should correlate very well (correlation coefficient $r \approx 1$), especially since the main gear follows the nose gear in close succession (76.54 ft) across the laser beam. This expectation is based on the fact that when the aircraft moves in a straight line it should show the same centerline deviations at nose and main gear location when it passes the same laser. Both lasers at KILO and ROMEO were on straight taxiway sections and thus there is some expectation that the aircraft move in a more or less straight line fashion.

Similarly, one should see good correlation between the nose gear deviations of the same event at the two lasers and again for the two main gear deviations since the lasers are only 150 ft apart. These plots are shown as correlation scatter plots in Figures 8-11 on pages 15-18. The data in these plots were cleaned using the previously discussed screening criteria. Prior to this cleaning the same plots showed rogue outlier points very far away from the main diagonal. These have been captured in Table 1 on page 19, with the last column indicating the removal reason for that particular case and the other columns giving other specifics in each case. The codes for the removal reasons are a follows: “a” for acceleration, “d” for nose to main gear lateral distance, “s” for speed, and “t” for timing.

About half of the outlier patterns are similar in that only one gear at one laser fell out of

line. The rest of the cases present a mixed bag. Of the 14 events 10 fell within the ten day time window 09/17/2001-09/26/2001. Also note that some of these outliers show a high taxi speed. For example, 53.57 ft/sec amounts to 36.5 mph and 73.89 ft/sec is 50.4 mph.

The correlation plots all have the same abscissa and ordinate scales to facilitate comparisons across plots. On each plot the main diagonal is shown as a dotted line. 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.

Since the plotted pairs of measurements (x, y) are supposed to be close to each other (i.e., $x \approx y$) we expect the points to cluster closely around this main diagonal, which describes all points (x, y) for which $x = y$ holds. 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 would expect that it more or less agrees with the main diagonal, at least within the range of the point cloud.

The nose to main gear correlation is very tight ($r \approx 1$) in all cases when we compare nose and main gear deviations at the same laser. The correlation between nose and main gear in the top left plot of Figure 8 is a bit looser ($r \approx .91$) than in the other same laser plots ($r > .95$). Also note that in this case the linear relationship appears somewhat offset/tilted from the main diagonal. This is emphasized analytically through the principal component axis. This could be caused by the fact that Laser 1 for KILO comes shortly after a merging fork (eastbound) of two taxiways, one going straight and the other merging from the northwest. Thus there may be some effect from taking the turn, whenever it was taken. It would show up predominantly eastbound at Laser 1 and the correlation plot seems to confirm this. This view agrees with a similar interpretation made earlier in the context of the lateral nose to main gear distance at that laser for this particular direction. This offset/tilt effect from laser 1 eastbound also shows up in the bottom two plots of Figure 8. This effect appears stonger here since the correlation is no longer so high when comparing deviations at different lasers.

The plots correlating the same gear at different lasers all show similar but wider scatter than were observed for different gears at the same laser. This may simply be due to the larger distance between lasers (150 ft) when compared to the distance between nose and main gear (76.54 ft along the centerline).

Furthermore, the scatters of the nose gear deviations at the two lasers are wider than those for the main gear. This is indicated by consistently lower correlations for nose gears than for main gears. The somewhat wider scatter in the correlation plots for the nose gear when compared to those of the main gear appears plausible since steering changes occur at the nose gear and the corresponding changes in the main gear follow in a somewhat dampened

fashion.

In fact, it can be shown [9] that if an aircaft travels at constant velocity and its lateral nose gear deviation $Y_{NG}(t)$ from the taxiway centerline as a funtion of time t (sec) follows a sinusoidal path with amplitude A (ft) then the main gear lateral deviation $Y_{MG}(t)$ will also follow a sinusoidal path (after quickly settling into a steady state condition) with a somewhat reduced amplitude A' (ft) and a phase shift φ, namely

$$Y_{NG}(t) = \Delta + A\sin(2\pi t/T + \theta) \qquad \text{and} \qquad Y_{MG}(t) = \Delta + A'\sin(2\pi t/T + \theta + \varphi)$$

with

$$A' = A \Big/ \sqrt{(2\pi L\,/TV\,)^2 + 1} < A \qquad \text{and} \qquad \varphi = -\arctan\left(2\pi L\,/TV\,\right)\ , \tag{1}$$

where T (seconds) represents the period of oscillation, V (ft/sec) the velocity of the aircraft, and L (ft) the distance between the centroids of nose gear and main gear. The phase shift θ controls the nose gear deviation at time $t = 0$, namely $Y_{NG}(t) = \Delta + A\sin(\theta)$. Although the lasers would not necessarily measure the distance to the wheels at the aircaft path amplitudes this reduction in amplitudes would still mean that measurements at the main gear position would generally lead to shorter distances than would be measured for the nose gear, if indeed the nose gear follows a sinusoidal path somewhat parallel to the taxiway centerline. According to [7], long period sinusoidal type paths are possible but superimposed on that would be many small steering corrections.

In some of these correlation plots, in particular in the two top plots of Figure 11, one may still perceive a string of outliers parallel to the main diagonal. We did not remove those. They were certainly far less extreme than those documented in Table 1.

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 12-19 on pages 20-27. For each set of deviations we also note the count of positive (≥ 0) and negative (< 0) deviations.

Looking at these plots there are no evident deviation level changes due to time of day or time of the year. One can see definite date gaps in deviation recordings but that was presumably quite deliberate, possibly because of needed maintenance actions. As far as time of day is concerned, one can see a naturally varying intensity of cases depending on the time of day. The wider spread in the deviations during high intensity time slots is a natural outcome of sampling variation. The more one samples the farther out one will experience extremes.

It is quite evident by the indicated counts that positive deviations far outnumber the negative ones. 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 equal in absolute size. Since 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.

Figure 8: Correlation Plots for KILO Eastbound

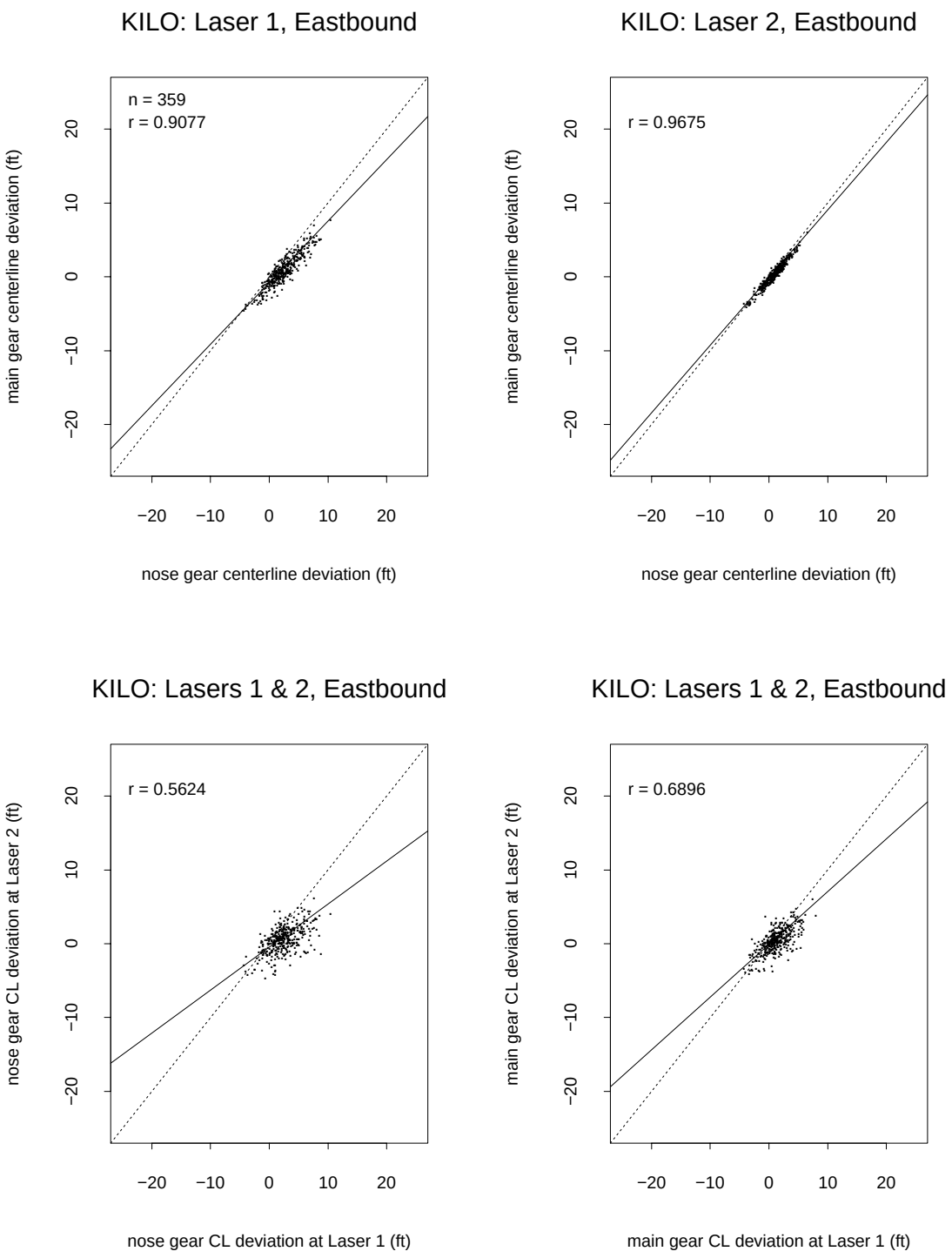


Figure 9: Correlation Plots for KILO Westbound

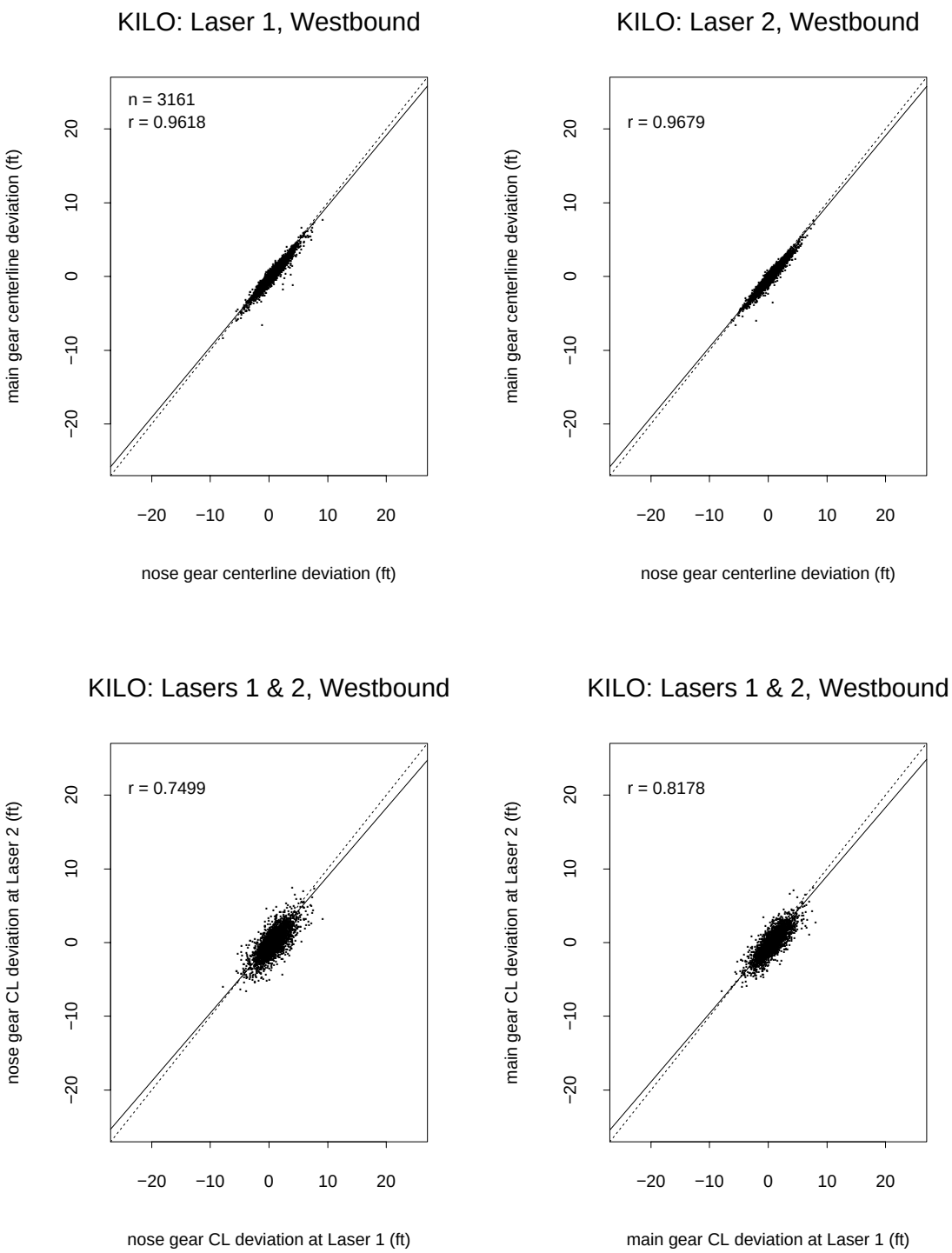


Figure 10: Correlation Plots for ROMEO Northbound

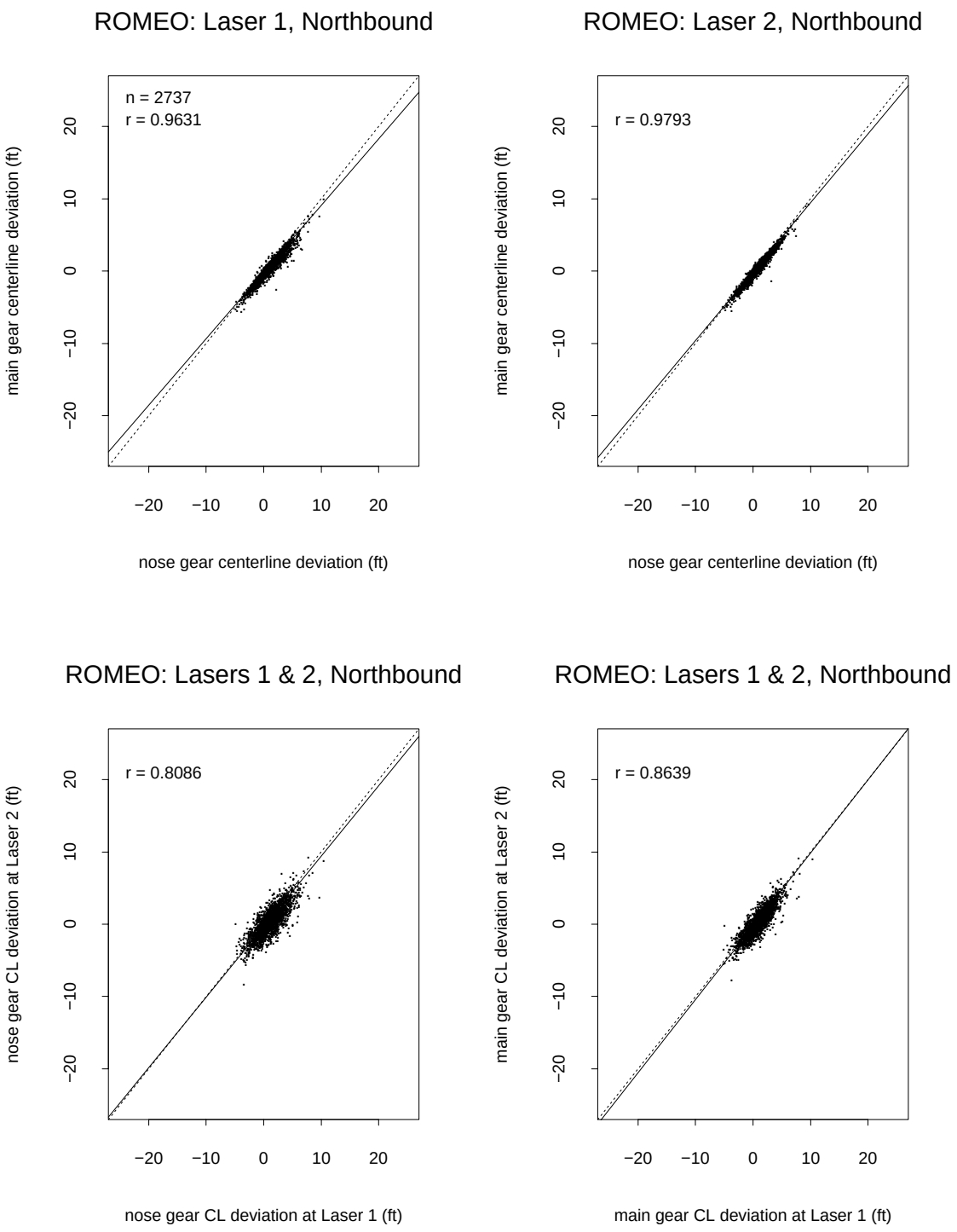


Figure 11: Correlation Plots for ROMEO Southbound

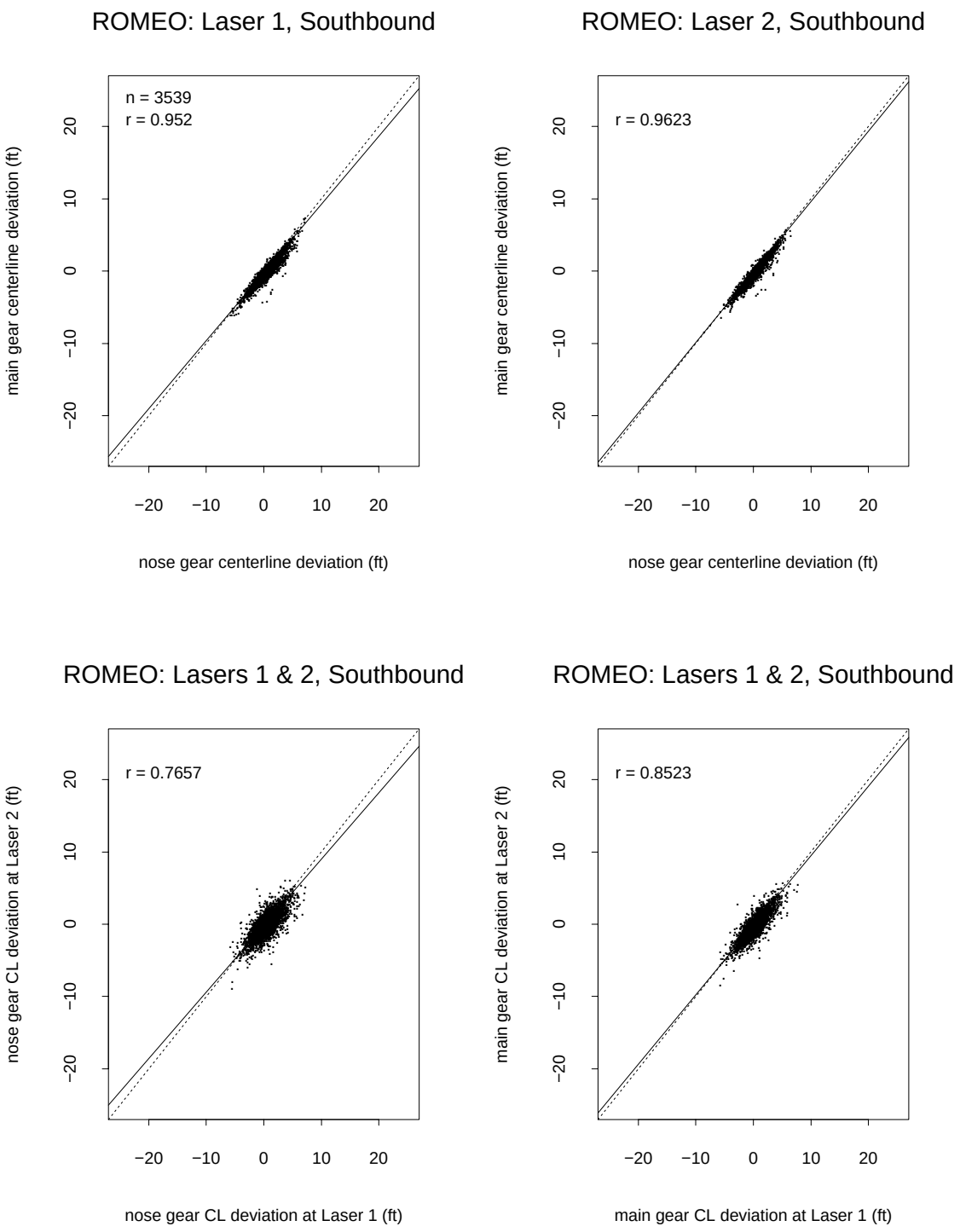


Table 1: Outlier Cases

Laser	delta (ft) nose gear	delta (ft) main gear	heading	speed ft/sec	event time	event date	event id	clean code†
ROMEO								
1	9.32	-11.69	Nb	48.70	4.509167	06/07/2001	4146	t
2	9.16	-11.09	Nb	48.70	4.510000	06/07/2001	4146	t
KILO								
1	-4.37	-5.28	Eb	27.88	22.000278	04/27/2001	1176	d, t
2	9.57	-6.76	Eb	27.88	22.001667	04/27/2001	1176	d, t
1	16.18	16.79	Eb	30.67	13.896667	06/23/2001	2308	a, t
2	-0.87	0.08	Eb	30.67	13.898056	06/23/2001	2308	a, t
1	7.67	-24.07	Eb	51.37	12.712222	08/03/2001	3109	d, t
2	13.90	-13.39	Eb	51.37	12.713056	08/03/2001	3109	d, t
2	18.26	0.53	Wb	17.73	11.675833	09/15/2001	3129	a, d, s
1	-1.50	-0.49	Wb	17.73	11.678333	09/15/2001	3129	a, d, s
2	16.34	-1.30	Wb	19.79	12.412222	09/15/2001	3132	a, d, s
1	-2.30	-1.66	Wb	19.79	12.414444	09/15/2001	3132	a, d, s
2	17.13	-0.53	Wb	15.69	6.597500	09/17/2001	3203	a, d, s
1	-2.07	-1.57	Wb	15.69	6.600000	09/17/2001	3203	a, d, s
2	12.39	-6.17	Wb	53.57	10.643611	09/17/2001	3212	d, t
1	8.25	-6.87	Wb	53.57	10.644444	09/17/2001	3212	d, t
2	1.01	-12.74	Wb	9.93	10.798889	09/18/2001	3254	d
1	1.28	-16.28	Wb	9.93	10.803056	09/18/2001	3254	d
2	-1.56	-2.09	Wb	4.36	16.209722	09/18/2001	3262	d
1	-2.73	-20.50	Wb	4.36	16.219444	09/18/2001	3262	d
2	19.35	2.05	Wb	12.94	14.560556	09/19/2001	3298	a, d, s
1	1.92	2.32	Wb	12.94	14.563611	09/19/2001	3298	a, d, s
2	19.48	2.94	Wb	12.41	6.357500	09/23/2001	3444	a, d, s
1	3.08	1.95	Wb	12.41	6.360833	09/23/2001	3444	a, d, s
2	15.94	-1.35	Wb	73.89	11.780278	09/24/2001	3489	a, d, s
1	-2.72	-2.11	Wb	73.89	11.780833	09/24/2001	3489	a, d, s
2	18.44	0.99	Wb	7.78	9.703056	09/26/2001	3552	a, d, s
1	1.63	1.61	Wb	7.78	9.708333	09/26/2001	3552	a, d, s

† clean code: a = acceleration, d = nose to main gear lateral distance, s = speed, t = timing.

Figure 12: Time Plots for Nose & Main Gear at KILO Eastbound, Laser 1

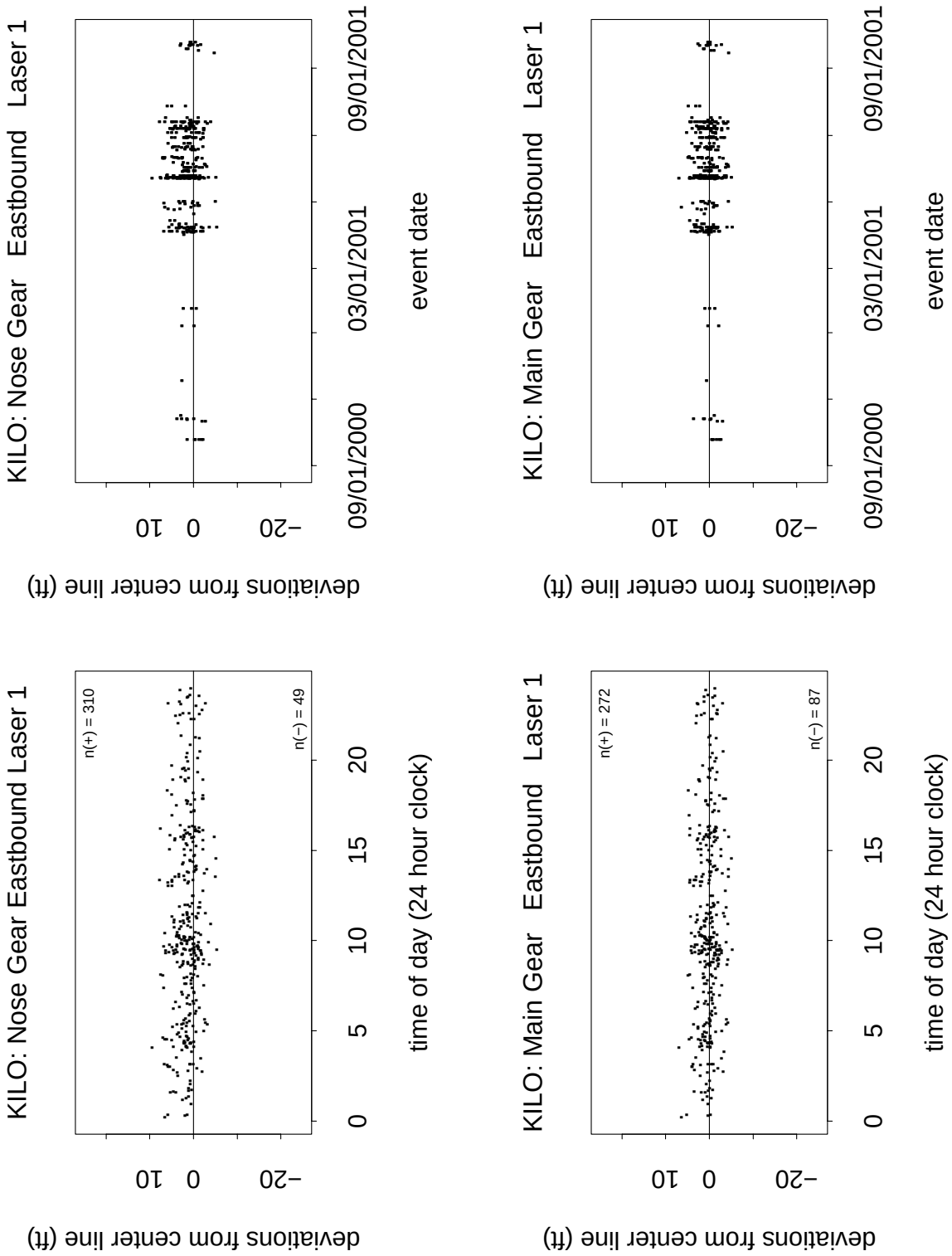


Figure 13: Time Plots for Nose & Main Gear at KILO Eastbound, Laser 2

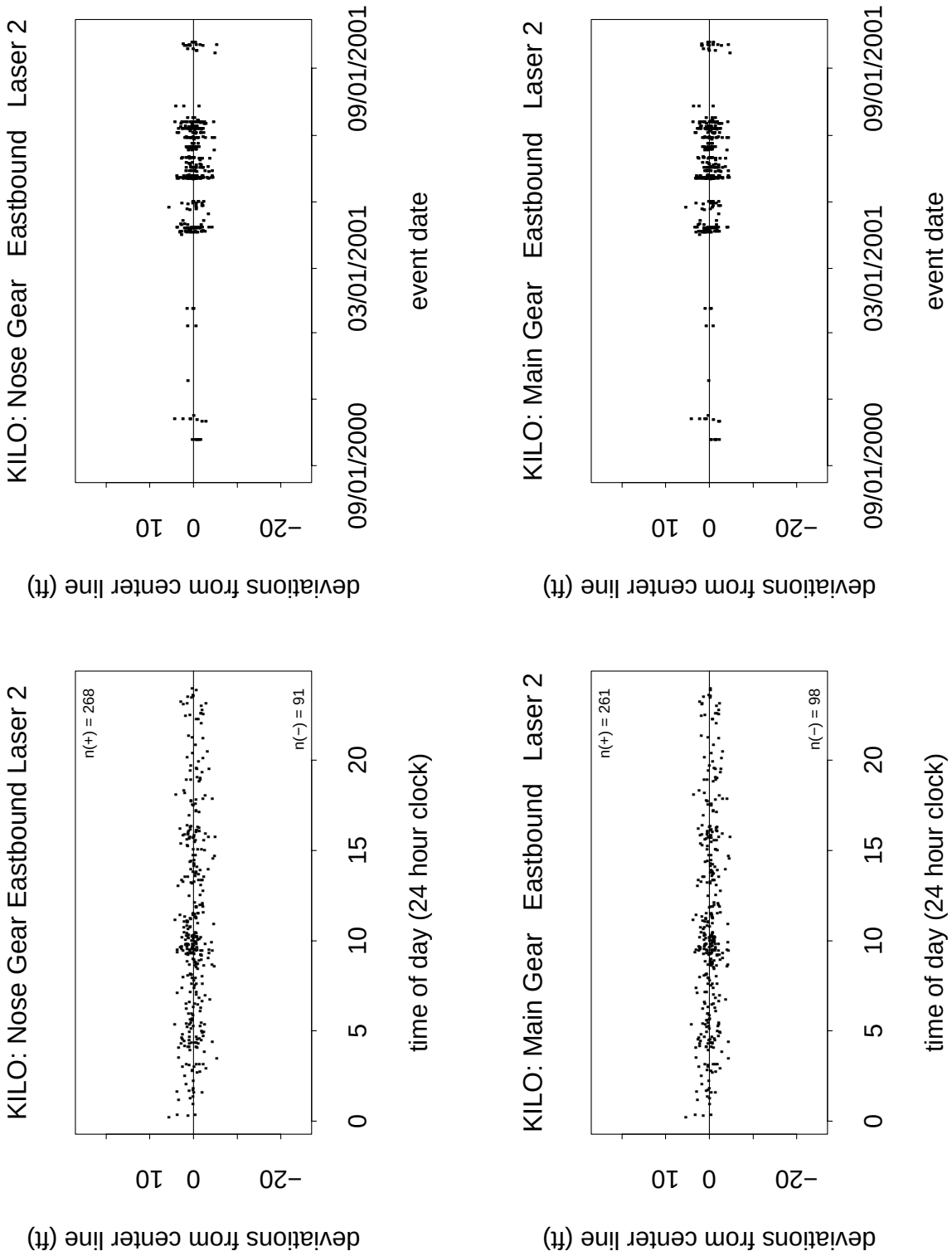


Figure 14: Time Plots for Nose & Main Gear at KILO Westbound, Laser 1

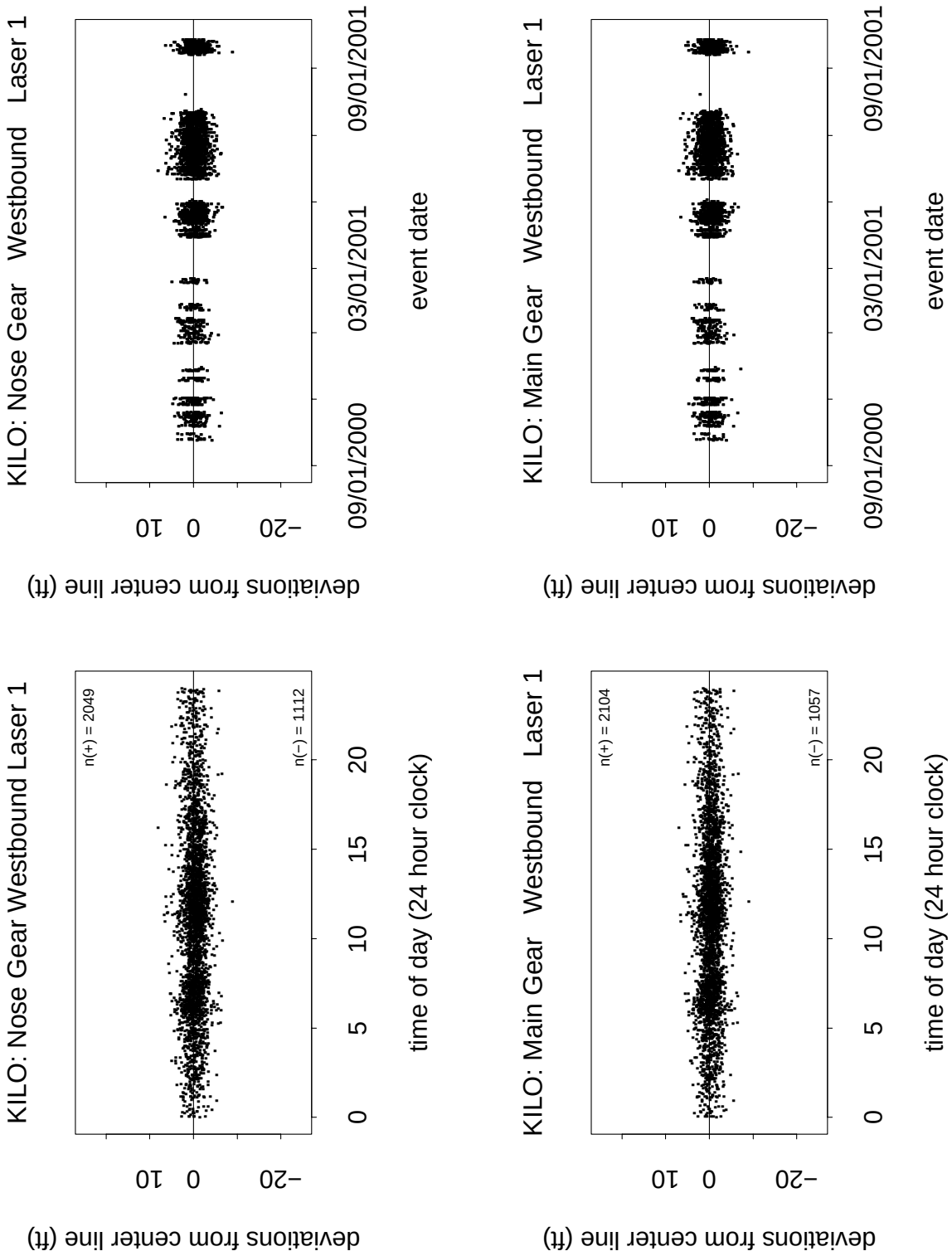


Figure 15: Time Plots for Nose & Main Gear at KILO Westbound, Laser 2

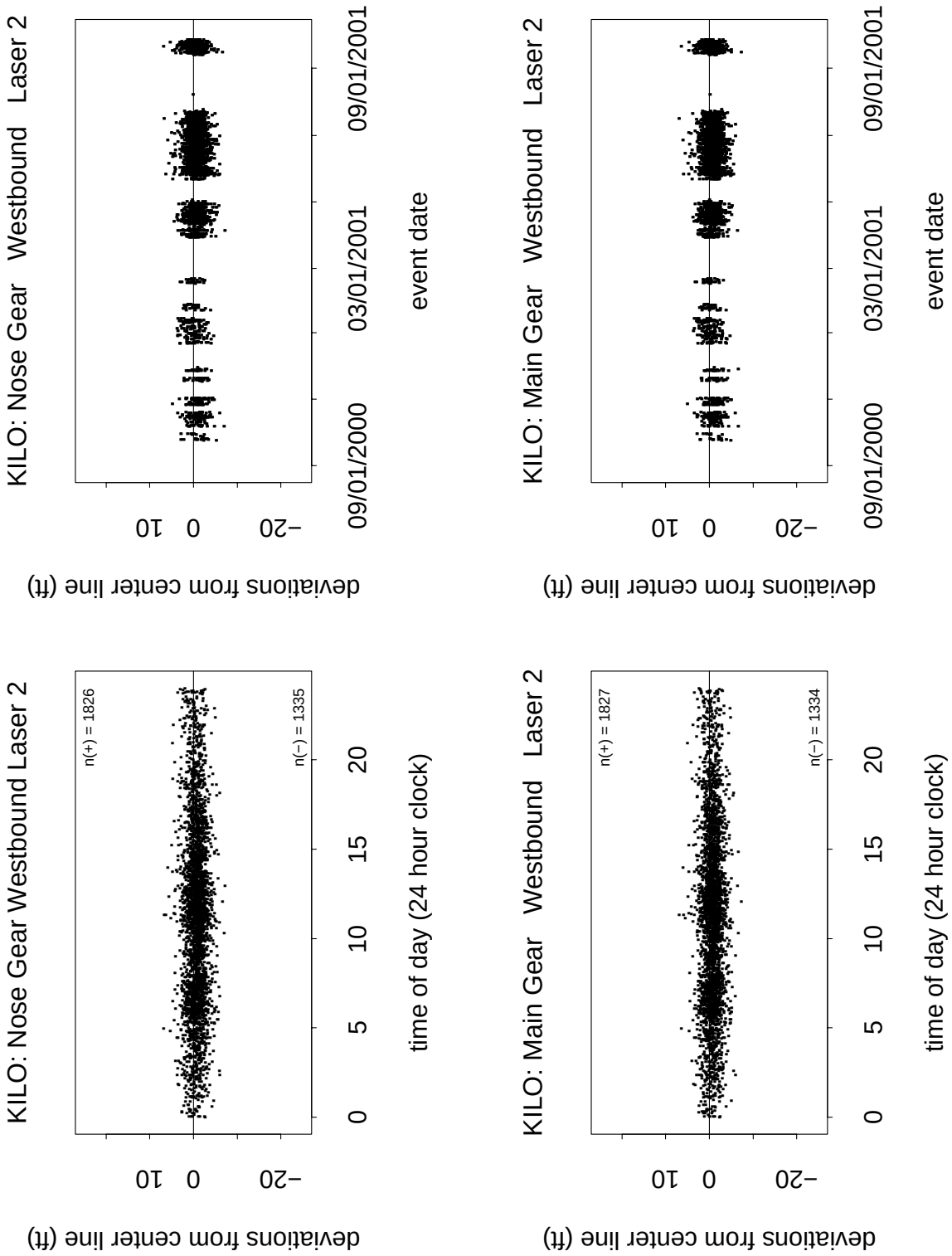


Figure 16: Time Plots for Nose & Main Gear at ROMEO Northbound, Laser 1

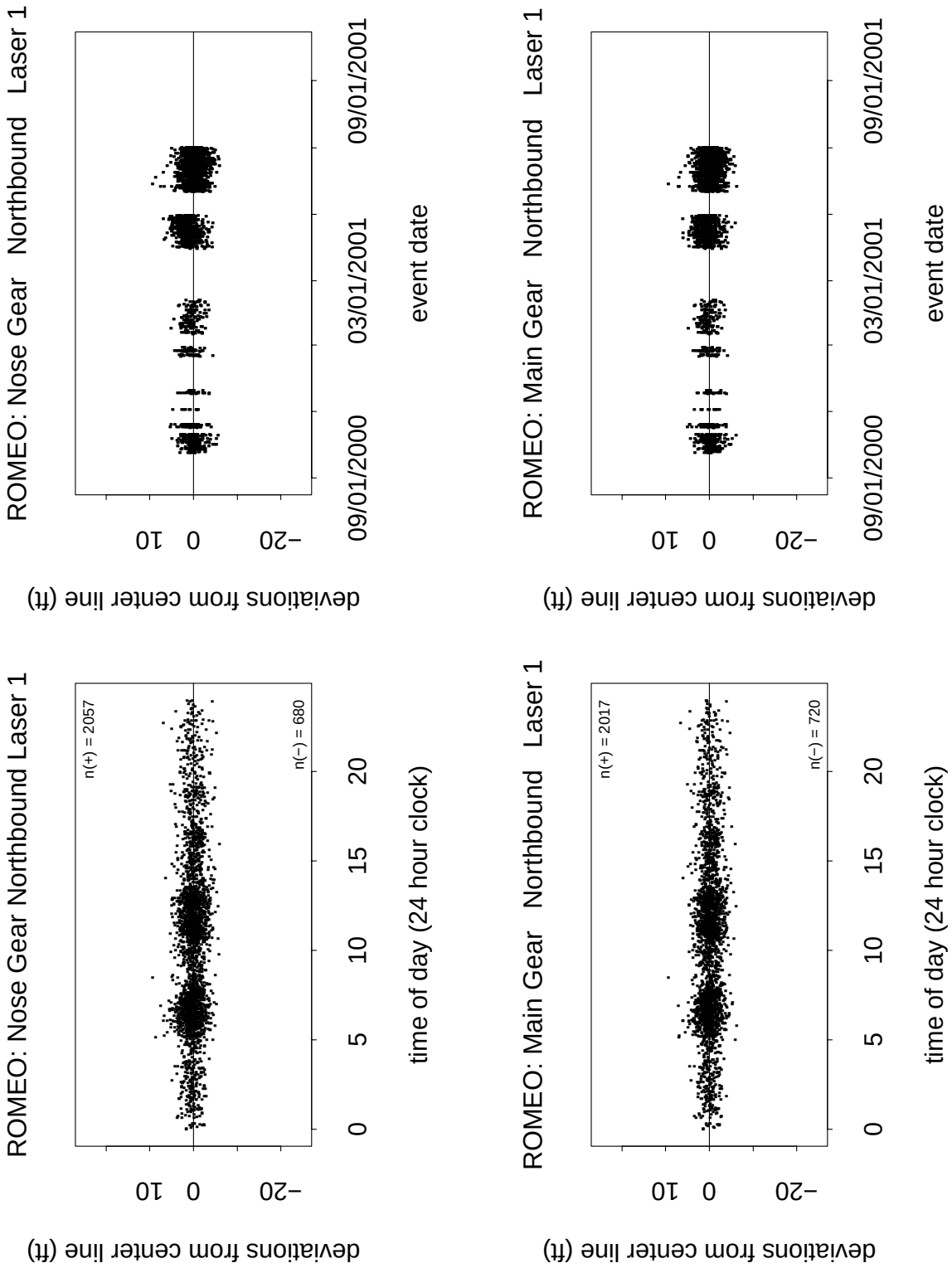


Figure 17: Time Plots for Nose & Main Gear at ROMEO Northbound, Laser 2

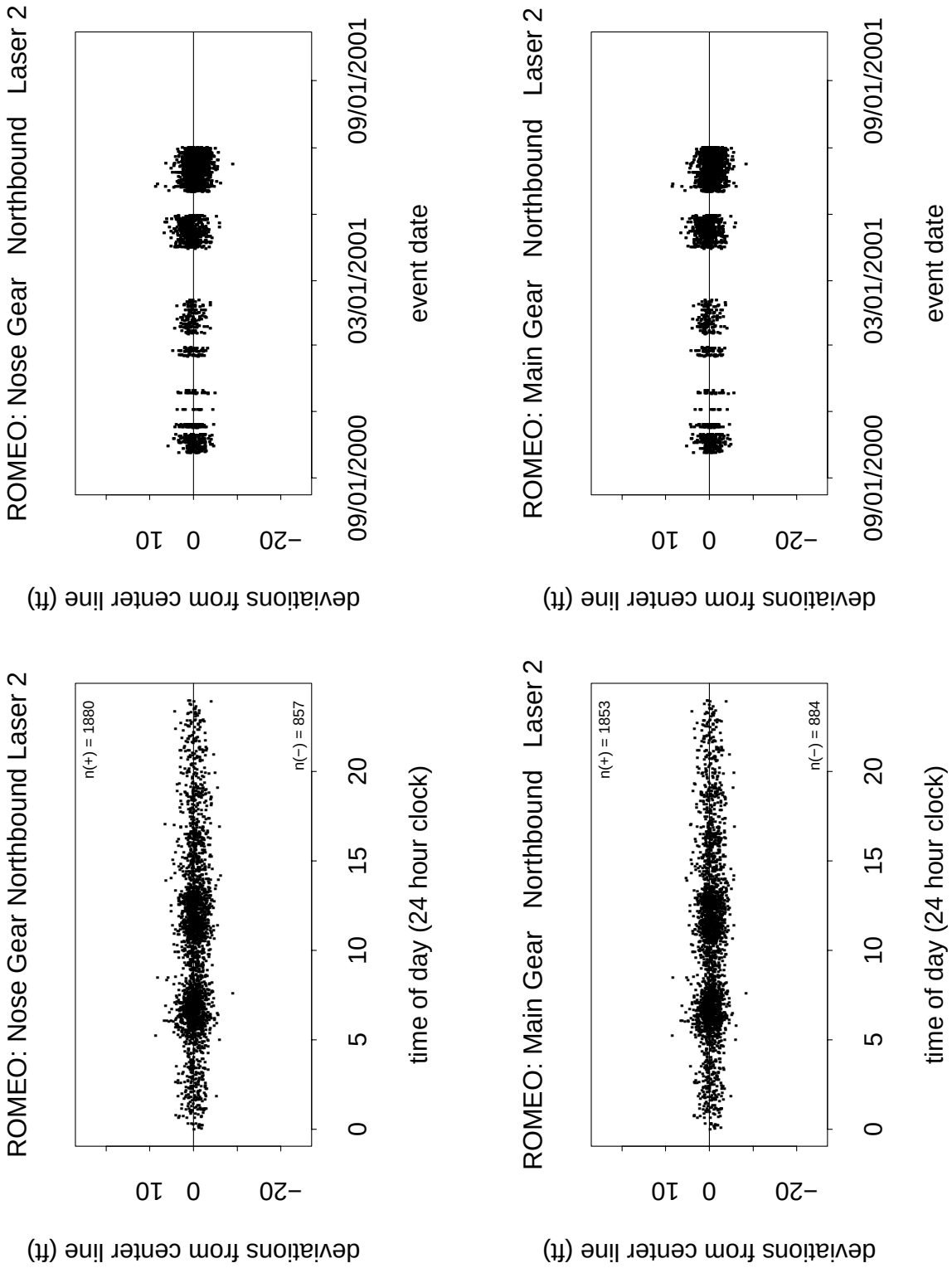


Figure 18: Time Plots for Nose & Main Gear at ROMEO Southbound, Laser 1

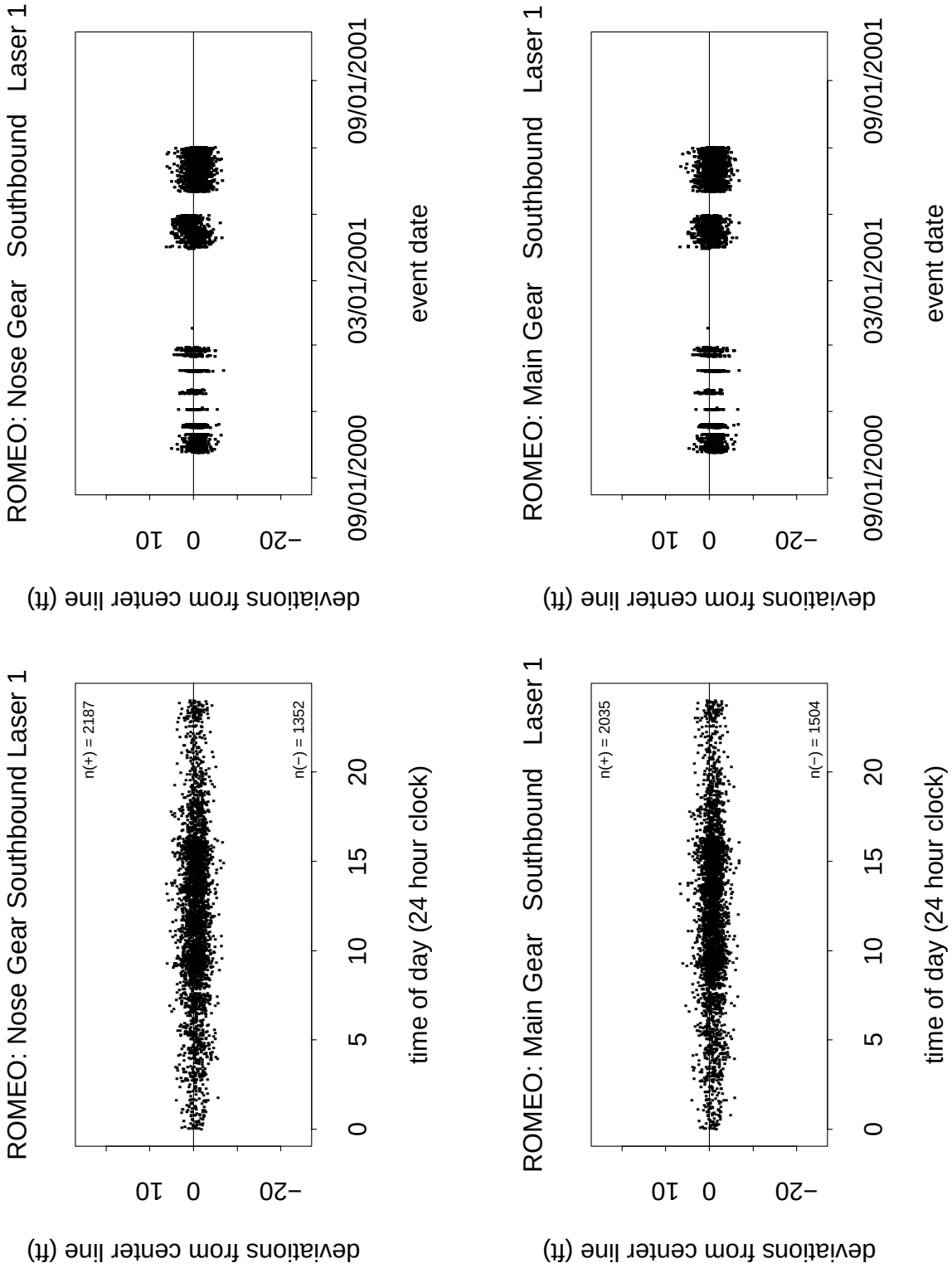
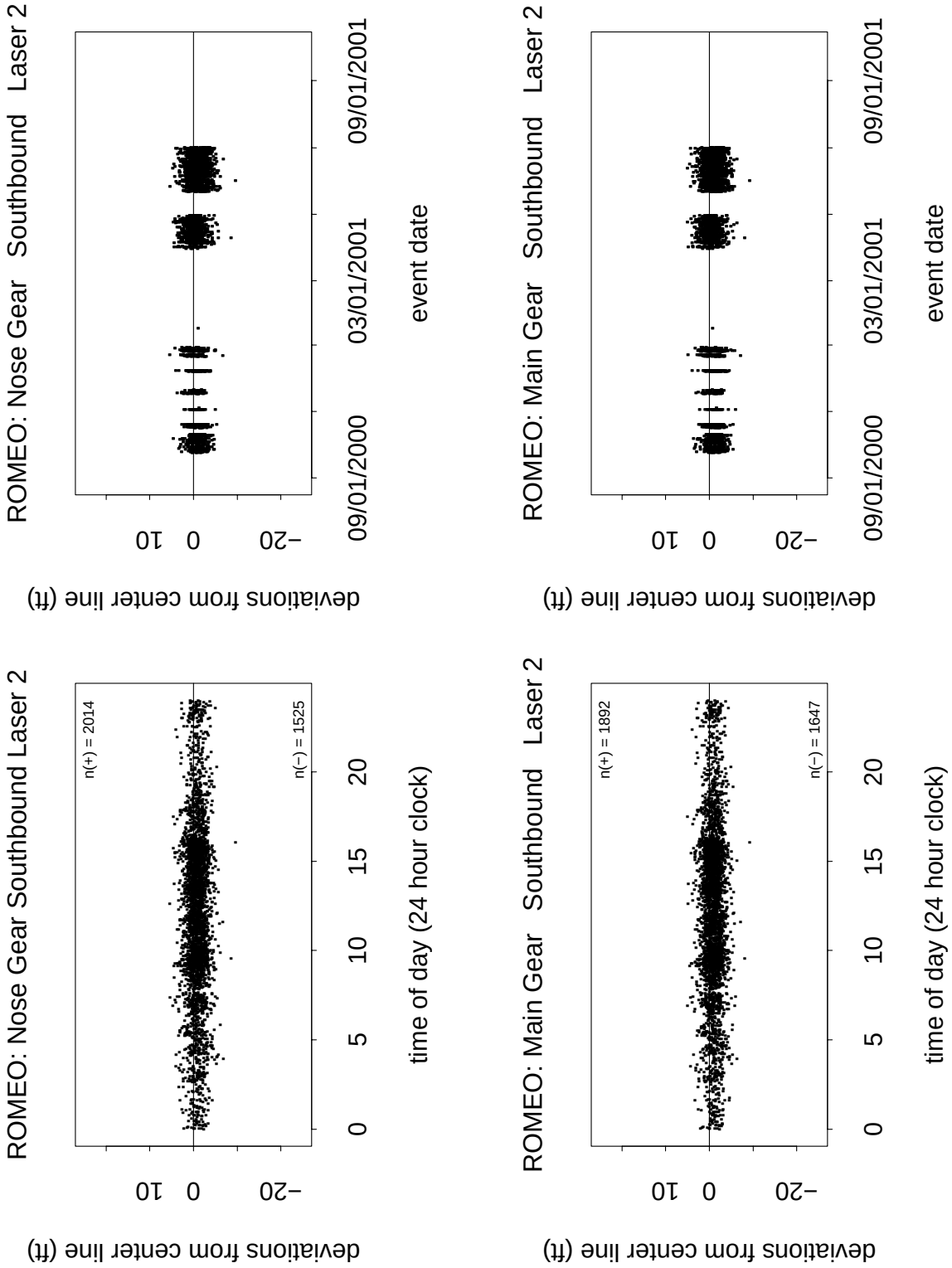


Figure 19: Time Plots for Nose & Main Gear at ROMEO Southbound, Laser 2



Initially this additional bias was interpreted as an avoidance bias. The lasers were covered by a roughly 6ft high protective structure and even though this is not high enough to be a threat it may still act on the subconscious of pilots. According to [7] such subconscious behavior should not be ruled out. Another more plausible explanation may come from the fact that parallel to the painted taxiway centerline there also is a line of centerlights to guide the aircraft in the dark. At both KILO and ROMEO these centerlights are on the side of the centerline away from the lasers.

According to [6], who also provided the three schematic views in Figures 20-22 (pages 29-31) of the dimensional relationships between laser, centerline, centerlights, and the nose gear, 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 20, some staying on the near side from the laser as in Figure 21 (preferred choice for [7]) and some on the far side from the laser as in Figure 22. 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 the side away from the laser on both KILO and ROMEO one may see a positive bias somewhere between the centerline and the centerlights. The size of this bias will be assessed later.

The three views of the nose gear in relation to the taxiway centerlights and the centerline illustrate three borderline positions of the nose gear wheels relative to the centerlights.

In Figure 20 the nose gear straddles the centerlights barely without bumping on the lights while trying to be as close to the centerline as possible. This results in an offset of $12'' + 5'' - 9.5'' = 7.5''$ of the aircraft centerline from the taxiway centerline in the direction away from the laser.

The distance between the center of the centerlights and the center of the centerline stripe is shown as $12''$. According to Ryan King and in response to a follow-up request by him measurements were taken at just one location at ROMEO and KILO and resulted in $10''$ and $12''$ for this distance, respectively. Supplemental previous measurements by Pete Sparacino and Jim White resulted in $10.8''$. Jim White recalled that the centerline stripe was off from a straight line pattern by $\pm 2''$ over the length of the data collection site. We chose to stay with a $12''$, mainly to have a definite dimension for illustration purposes.

In Figure 21 the nose gear does not straddle the centerlights but comes as close as possible to the taxiway centerline without bumping the lights. This leads to a $12'' - 5'' - 17'' - 9.5'' = -19.5''$ offset of the aircraft centerline from the taxiway centerline. The negative sign indicates that the offset is toward the laser.

Figure 22 shows the nose gear not straddling the centerlights but being on the side away from

Figure 20: Nose Gear Straddling Centerline Lights

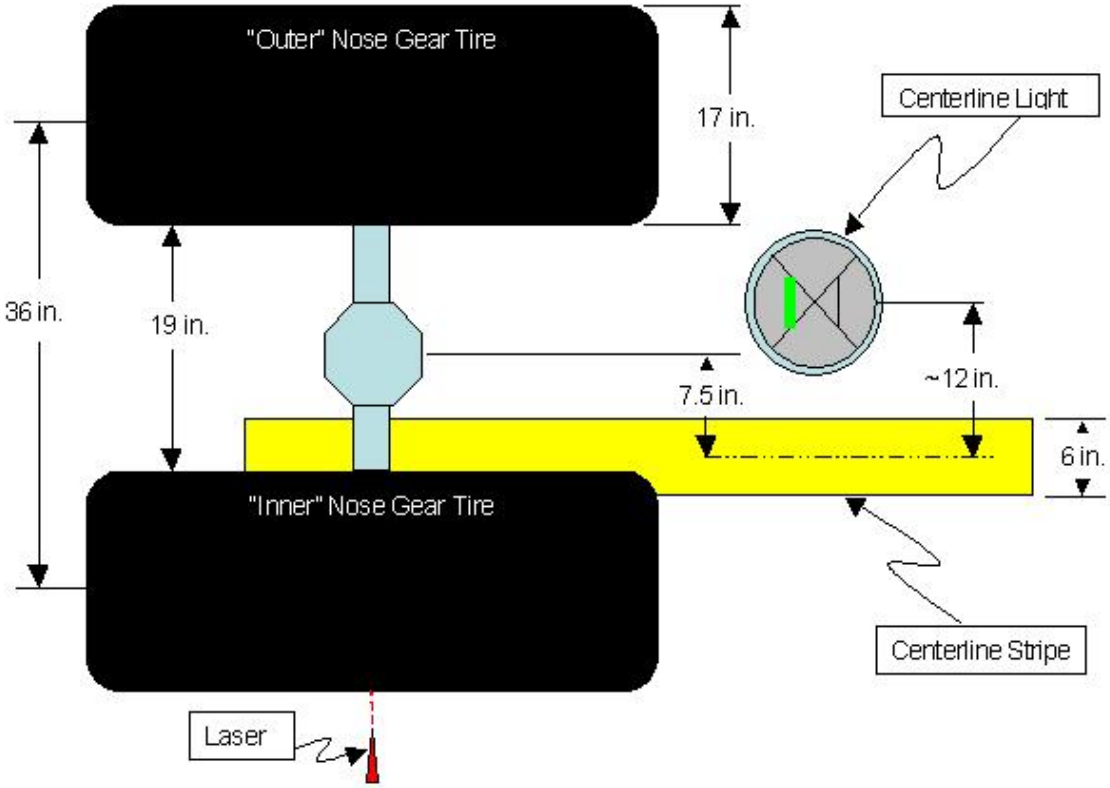
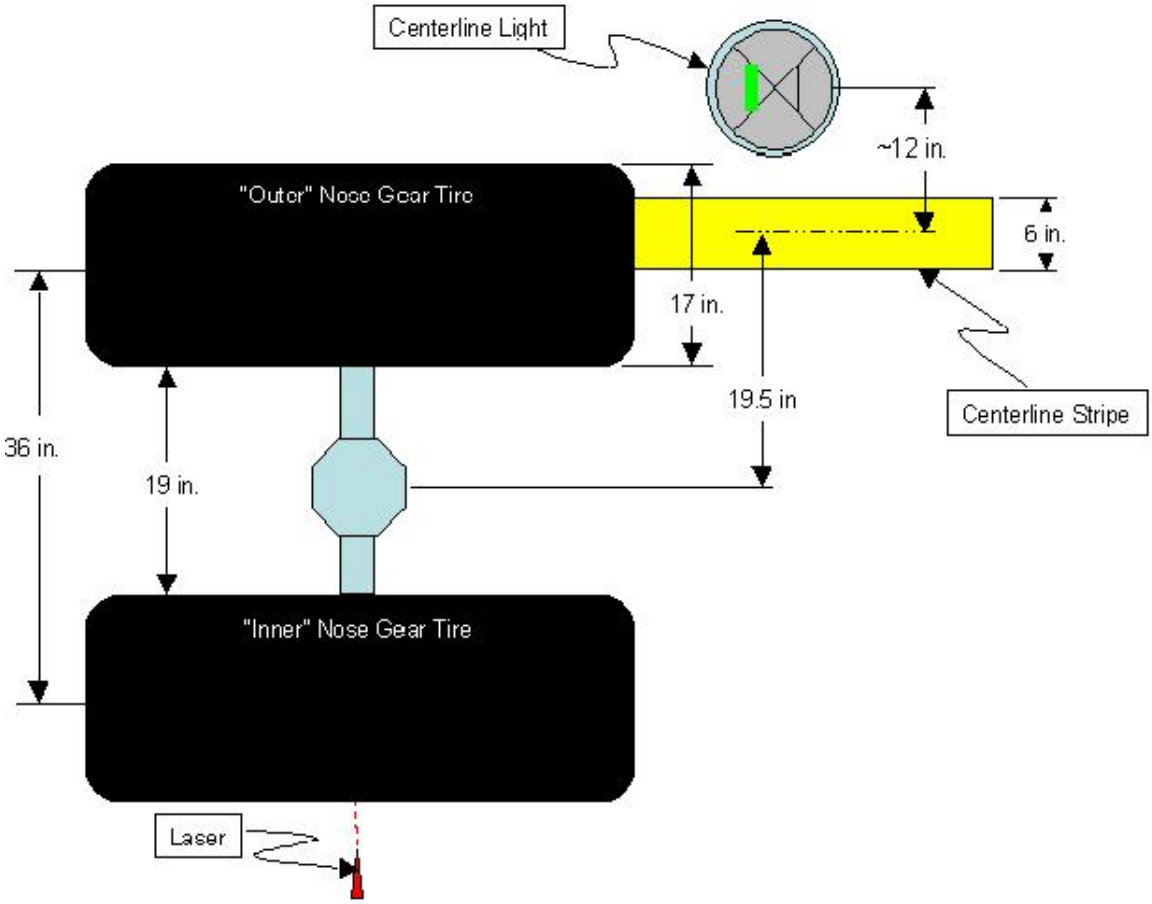
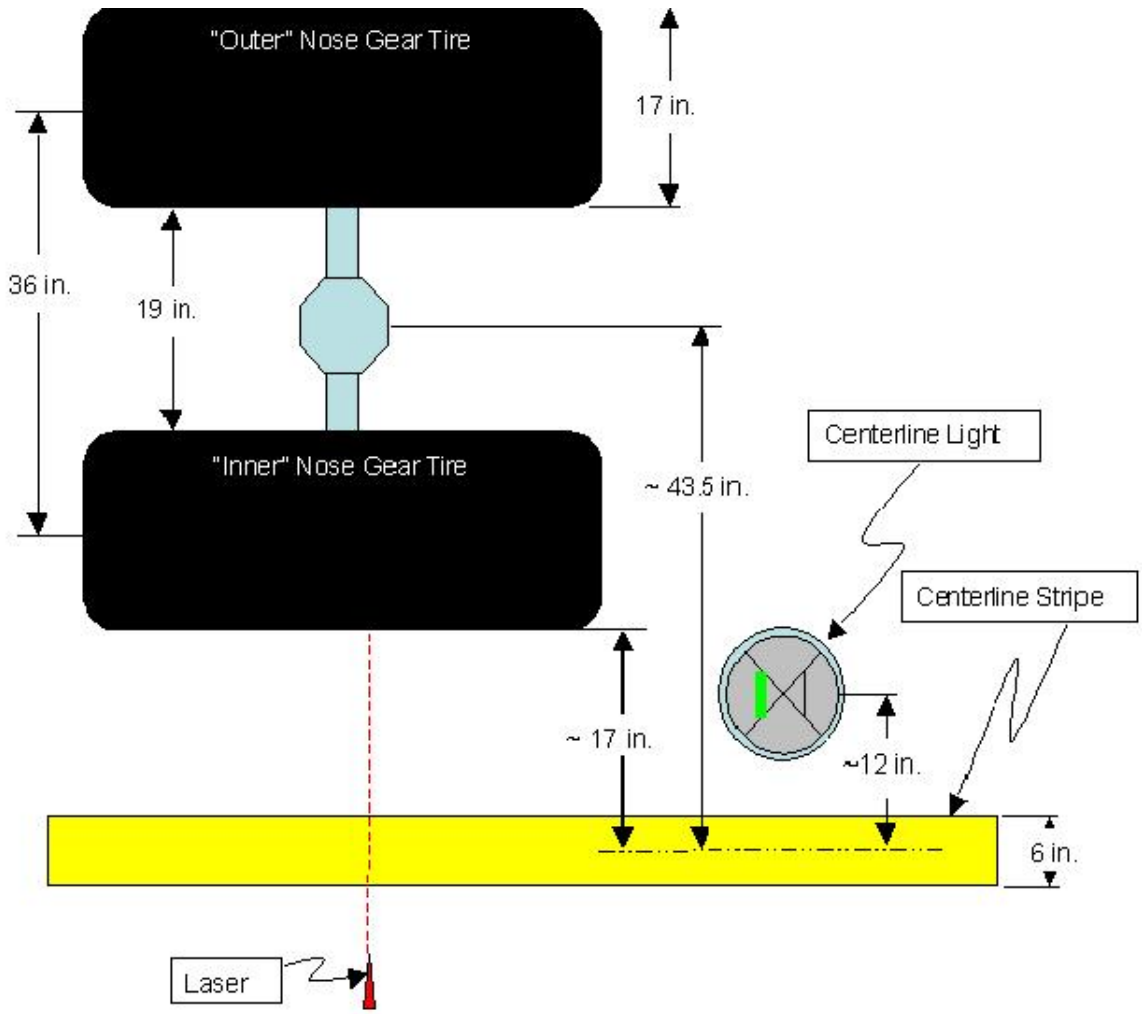


Figure 21: Nosegear Not Straddling Centerline Lights, Close to Laser



the centerline and coming as close as possible to the taxiway centerline without bumping the lights. It would result in an offset of $12'' + 5'' + 17'' + 9.5'' = 43.5''$, which is 3.625 ft. These three cases may explain why there is a bias away from the laser. Given the large difference in the offsets of $-19.5''$ toward the laser and $43.5''$ away from the laser one may wonder whether the pilots follow a meandering path just to avoid the constant bumping. This meandering would definitely add to the variability of the deviations and thus to their extremes.

Figure 22: Nosegear Not Straddling Centerline Lights, Far From Laser



It would therefore be interesting to know what the deviation pattern would be on taxiways where the centerlights are positioned on the centerline or where there are no centerlights. We expect the deviation pattern to be much narrower. Other than ease of repainting a worn centerline we see no reason to avoid positioning the centerlights on the centerline. If our expectation of reduced variation in centerline deviations is correct, it would be worthwhile to

compare the savings from ease of painting with the cost of wider taxiway deviations resulting in wider taxiways to minimize collision or taxiway runoff risks.

Figure 23: JFK May 2002: Following an Air China 747 on Taxiway

Figure 23 illustrates the avoidance bias through a single example. The photo is not from ANC but was taken at John F. Kennedy International Airport when Ryan King and Jerry Robinson were following the aircraft. Here the 747 is positioned to the right of the taxiway centerline with an in-pavement centerlight shown in the foreground on the left side of the centerline. Of course, a single instance does not represent a consistent bias, but it is interesting that the one random event captured here illustrates the point.

4 Data Selection and Adjustments for Risk Analysis

For each event we have four deviations from the centerline, 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 2 for both KILO and ROMEO.

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 across taxiways. It is hoped that the bias corrections will then lead to combined data that distributes 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.

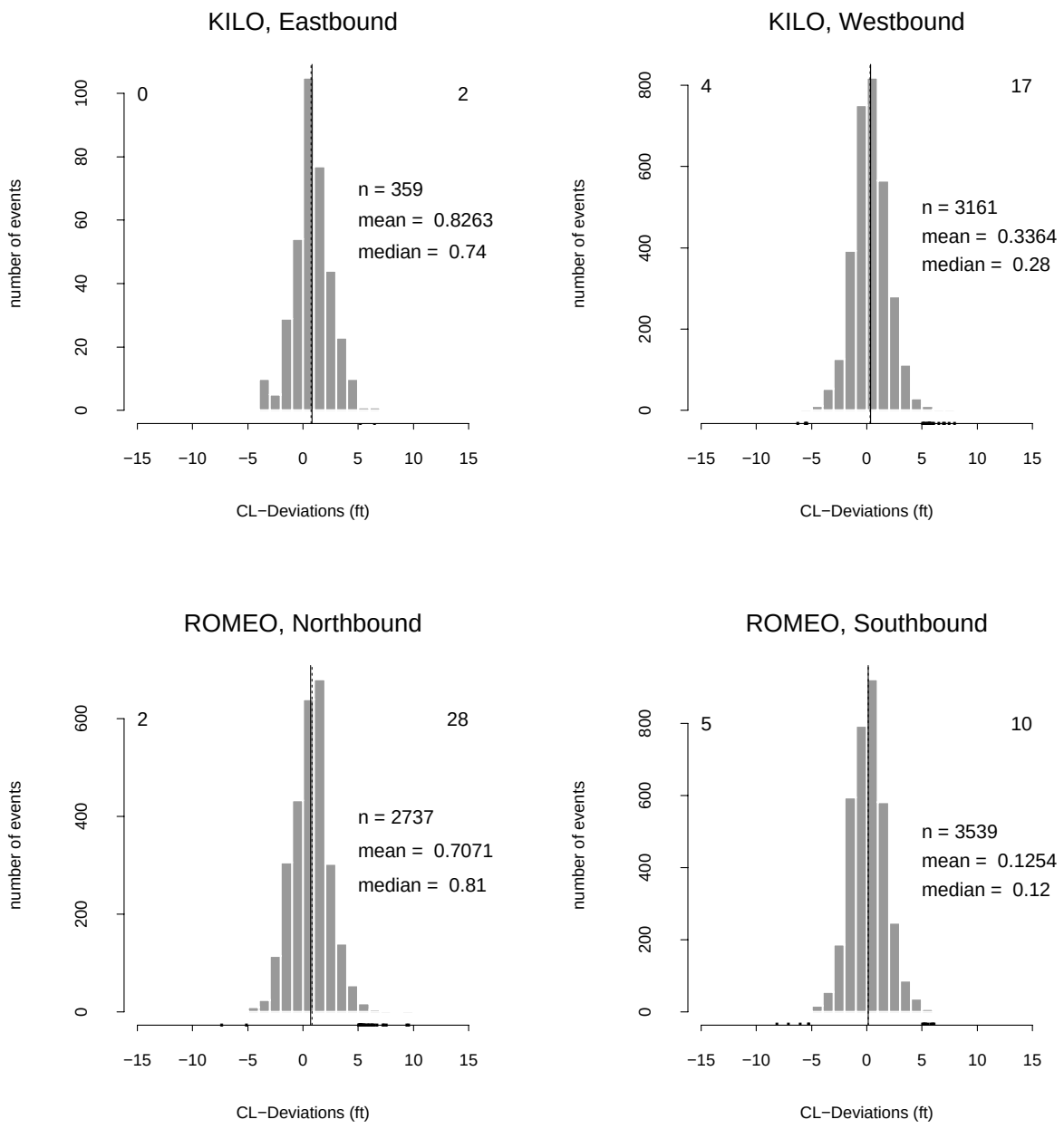
As a first step we present in Figure 24 the four histograms of the main gear centerline deviations by heading at KILO and ROMEO. Shown by dots below the histograms are those extremes that exceed ± 5 ft and in the upper left and right plot corners their number is indicated. Also indicated is the sample size n in each case and the mean (solid vertical line) and median (dashed vertical line) for each histogram. It is apparent that all means/medians are positive, more so for the eastbound and northbound headings when the laser is to the left. For the opposite headings the positive bias is smaller for both taxiways.

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 (with some consistency between taxiways) it seems to point to some other form of bias 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:

$$K_{Eb} = \mu_P + \mu_O + e \ , \quad K_{Wb} = -\mu_P + \mu_O + e \quad R_{Nb} = \mu_P + \mu_O + e \ , \quad R_{Sb} = -\mu_P + \mu_O + e \ .$$

Here K_{Eb}, for example, stands for a measurement at KILO eastbound, μ_P denotes the parallax bias, which changes sign with the heading, and μ_O is the other bias which does not change signs with the heading. 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 for both taxiways as follows from the averages ($\bar{K}_{Eb}, \bar{K}_{Wb}, \bar{R}_{Nb}$, and $\bar{R}_{Sb}$) of these four groups of data:

Figure 24: Main Gear Centerline Deviations for Laser 2 at KILO & ROMEO



$$\text{for KILO:} \quad \widehat{\mu}_{P,K} = \frac{1}{2}\left(\bar{K}_{Eb} - \bar{K}_{Wb}\right) = .24\text{ft} \quad \text{and} \quad \widehat{\mu}_{O,K} = \frac{1}{2}\left(\bar{K}_{Eb} + \bar{K}_{Wb}\right) = .58\text{ft}$$

and

$$\text{for ROMEO:} \quad \widehat{\mu}_{P,R} = \frac{1}{2}\left(\bar{R}_{Nb} - \bar{R}_{Sb}\right) = .29\text{ft} \quad \text{and} \quad \widehat{\mu}_{O,R} = \frac{1}{2}\left(\bar{R}_{Nb} + \bar{R}_{Sb}\right) = .42\text{ft}\ .$$

Figure 25 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 as follows. If a set of ordered numbers is exactly symmetric around zero, say $x_1 = -7.5, x_2 = -3.4, x_3 = -2, x_4 = 2, x_5 = 3.4, x_6 = 7.5$, then the top half matches the negative bottom half, i.e., $7.5 = -(-7.5), 3.4 = -(-3.4)\ldots$. Plotting the two ordered halves (the bottom half with its sign reversed) against each other gives us three points on the main diagonal $(2, 2), (3.4, 3.4)$, and $(7.5, 7.5)$.

This same idea is used to examine the symmetry of our *ordered* sample $x_1 \leq \ldots \leq x_n$. Namely, plot $-x_1, -x_2, \ldots, -x_q$ against $y_1 = x_n, y_2 = x_{n-1}, \ldots, y_k = x_{n-q+1}$, where $q = n/2$. This was done for each bias corrected data set and the results are shown in Figure 26. 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 ROMEO northbound a meandering pattern is notable. We also point out that for the more extreme sample data the right half seems to be consistently larger than the left half in absolute value because there the points fall below the diagonal (the exception being KILO eastbound).

One needs to be careful before reading too much meaning into such observed features. To put things in perspective we 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 27. 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 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. If the sample sizes were equal such a QQ-plot would plot one ordered sample against the other ordered sample, i.e., smallest against smallest, ..., largest against largest.

When sample sizes are unequal one plots observations of equal rank within each ordered sample against each other. This may involve some degree of interpolation. If $x_1 \leq \ldots \leq x_n$ and $y_1 \leq \ldots \leq y_m$ are the two ordered samples with $m \geq n$, one plots $y_j^\star$ against x_i for $i = 1, \ldots, n$. Here $y_j^\star$ is interpolated between y_j and y_{j+1}, where $(j - .5)/m \leq (i - .5)/n \leq (j + 1 - .5)/m$.

Figure 25: Adjusted Main Gear Centerline Deviations for Laser 2 at KILO & ROMEO

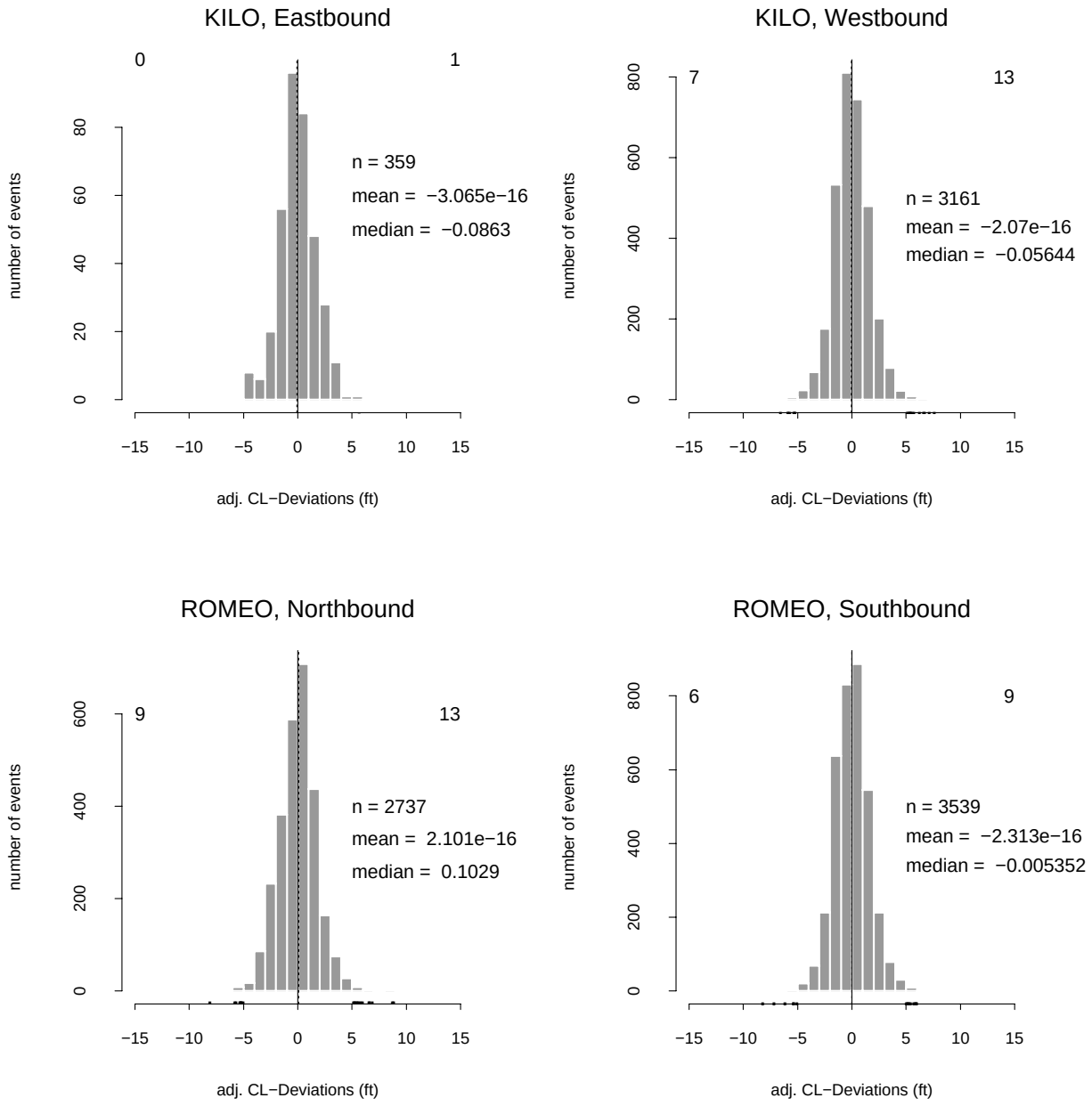


Figure 26: Testing Symmetry of Main Gear Centerline Deviations
for Laser 2 at KILO & ROMEO by Heading

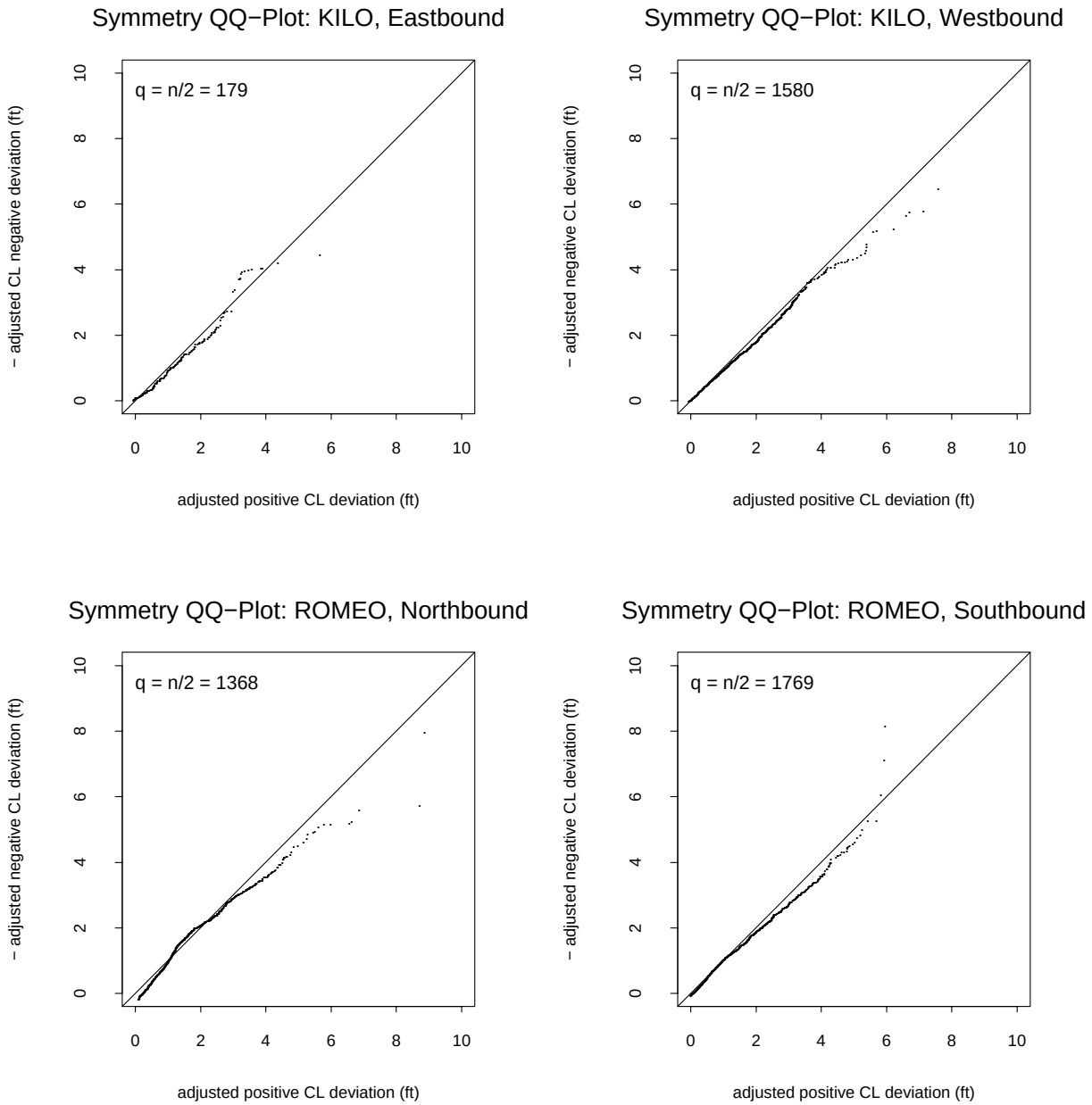
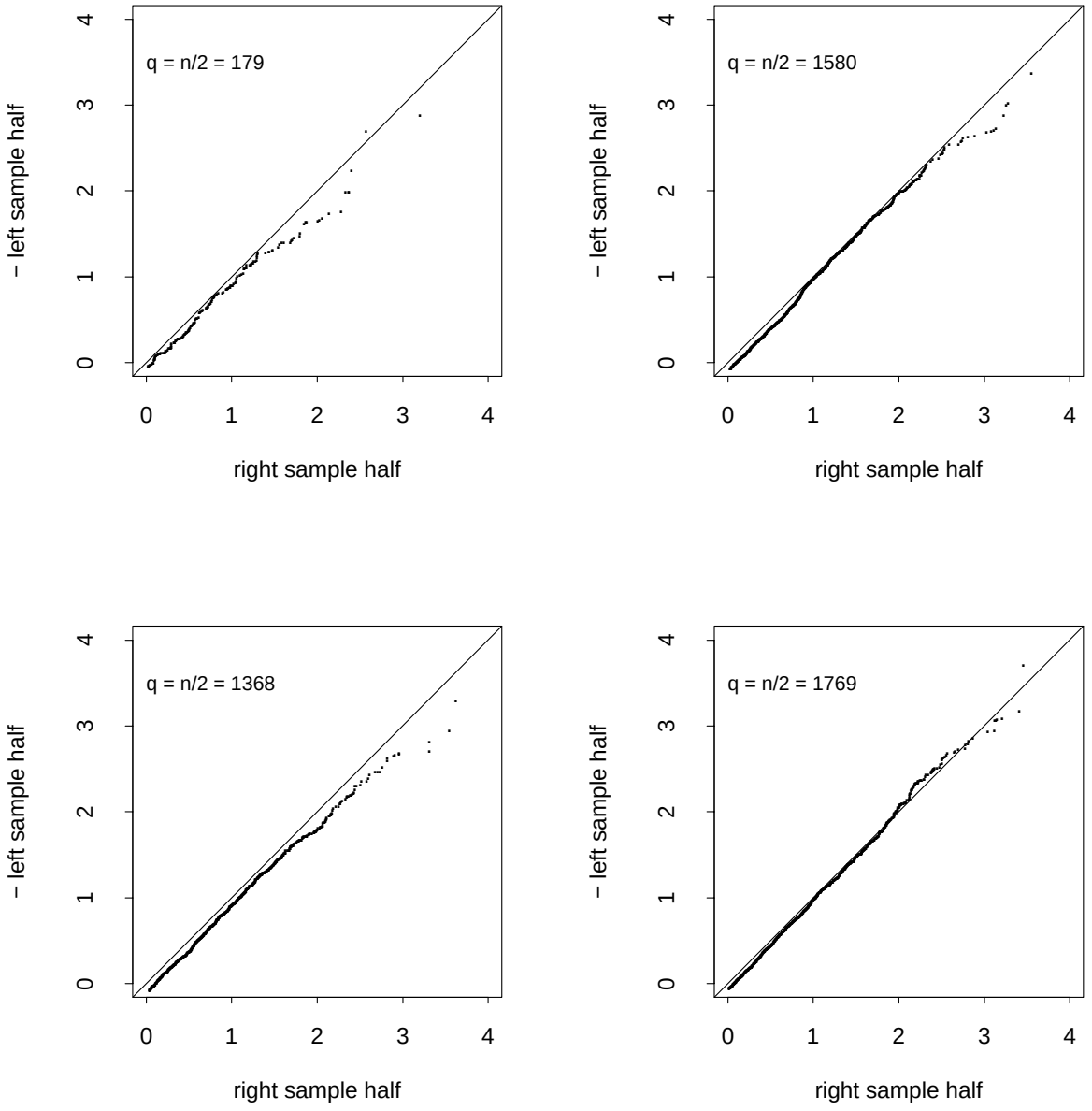


Figure 27: Simulated Symmetry Test Plots from Standard Normal Samples



Such a comparison was done for each pair of opposing headings at KILO and ROMEO and is shown in the top two plots of Figure 28. 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 28 show the symmetry check for these combined deviations for each taxiway. Symmetry seems to be very reasonable for absolute deviations below 3 ft at ROMEO and below 4 ft at KILO. Above that one sees again 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 29 compares the deviation data (combined by heading) from ROMEO with the corresponding data from KILO 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 $n = 3520 + 6276 = 9796$ for purposes of risk extrapolation. The top right plot of Figure 29 examines the symmetry issue for this combined data set and mild deviations from the diagonal appear only at the high end. Whether this has repercussions will have to be seen.

The histogram in the lower left of Figure 29 represents the combined sample of $n = 9796$ adjusted main gear deviations from KILO and ROMEO at laser 2. 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 29 examines whether the combined sample of $n = 9796$ 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 away from the centrally fitted straight line. Hence the tails of the combined sample are heavier (reach out more) than indicated by a normal model. This behavior in not just confined to the deviations beyond ± 5 ft, but starts earlier at about ± 3 ft. Using a normal model for exceedance risk assessments would lead to misleading and overly optimistic results.

Figure 28: Testing Equality by Direction and Symmetry by Taxiway of Main Gear Centerline Deviations for Laser 2 at KILO & ROMEO

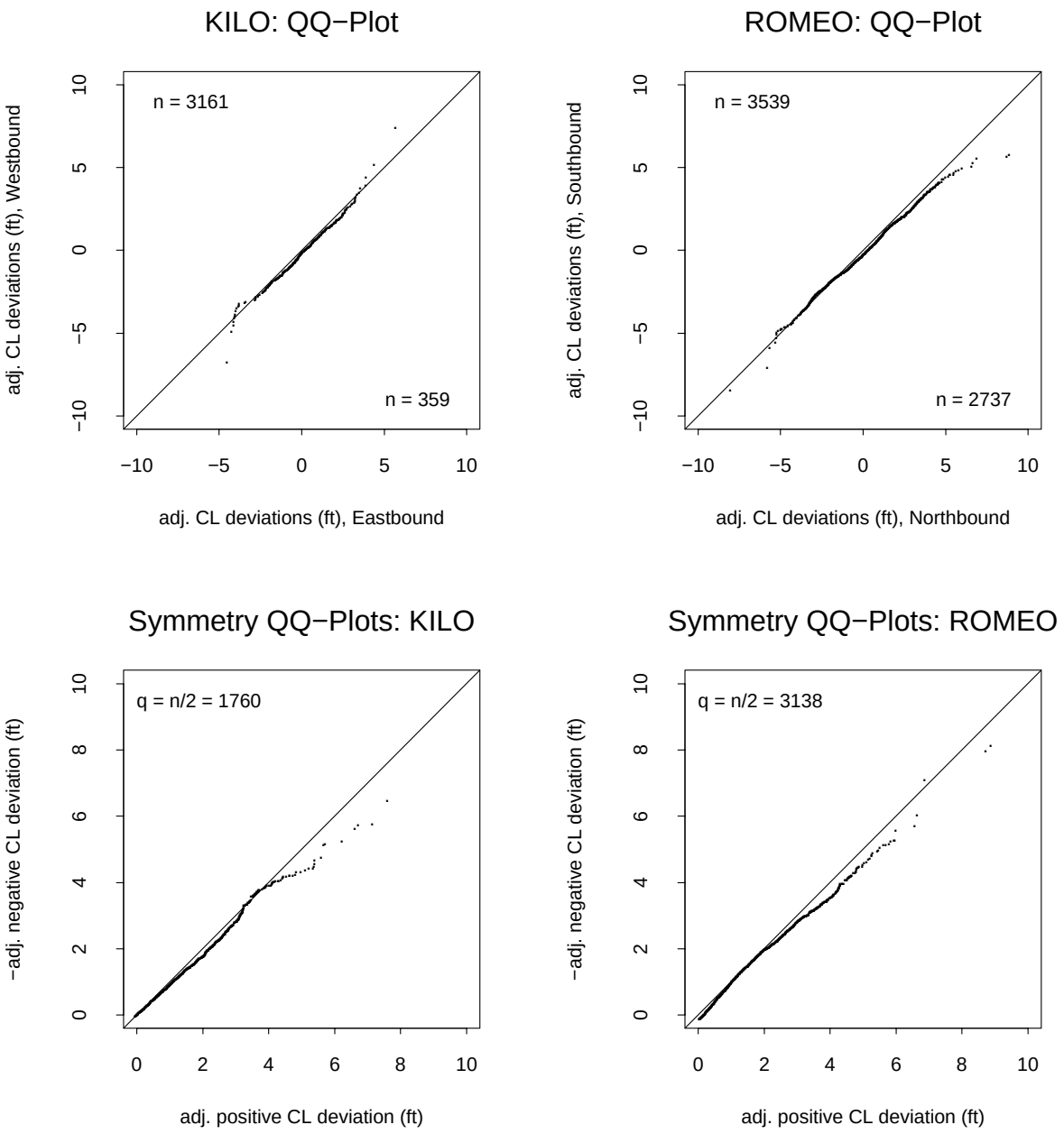
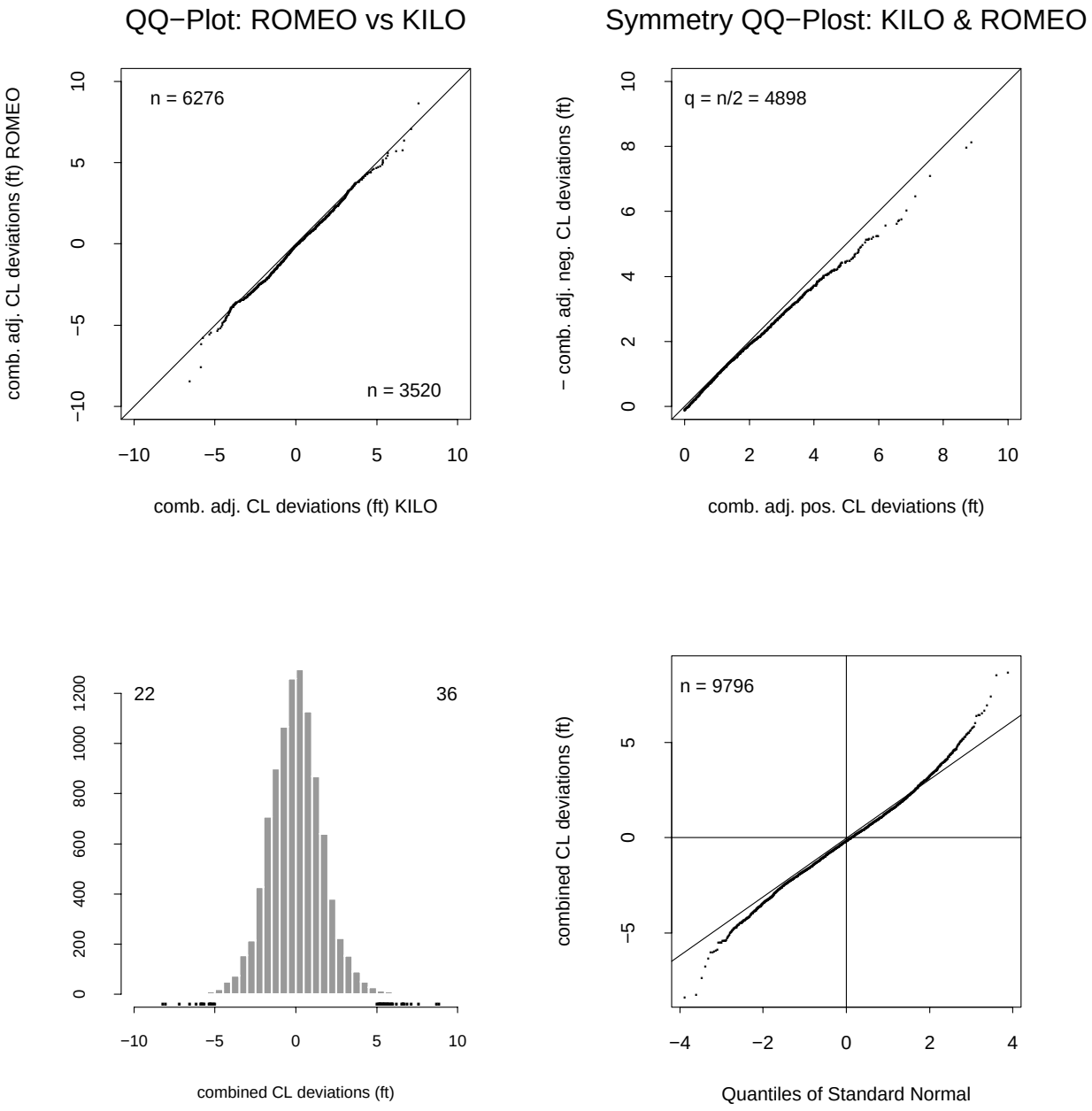


Figure 29: QQ-Plot of ROMEO vs KILO, Symmetry Plot for ROMEO & KILO Combined
Histogram and Normal QQ-Plot for ROMEO & KILO Combined

Main Gear Adjusted Centerline Deviations for Laser 2 at KILO & ROMEO



5 Risk Extrapolation

5.1 Overview

We will now focus on the combined set of adjusted main gear centerline deviations from Laser 2 at both KILO and ROMEO, as portrayed in the lower left histogram of Figure 29. All $n = 9796$ adjusted deviations were bounded within ± 9 ft. In fact, the range of these $n = 9796$ adjusted deviations is $[-8.225, 8.863]$ ft. The primary concern is to assess the risk of exceeding higher thresholds than 9 ft. Conversely one can ask what are the deviation thresholds associated with various small risks of threshold exceedance. Methodology for this was developed in [8] and other methods are discussed in the recent reference [3]. All of these methods deal with the k most extreme observations in a given random sample, $X_1, \ldots, X_n$, and a major question is how large to choose k. Too large a k will introduce influences from the center of the data that have nothing 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 in the tail of the data and will thus result in too much uncertainty concerning our risk extrapolations. Although [8] proposed a method for choosing k we will not use it here. Instead we will take a close look at the tail behavior of the data and make a judgment call as to 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

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

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 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.

After having fixed on the proper number k of extremes (in the case of negative extremes this came to four different values for k) we then ran the EXTRAP program on the respective number of extremes in the deviation sample.

5.2 Examination of Tail Depth k Using the Hill Estimator

In [8] 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. Sometimes this 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 30-32 on pages 45-47, 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 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 33-35 on pages 48-50.

In examining the proper tail depth k for Figure 30 (positive deviations, right tail data) one notes for small k the strong fluctuations of the estimated extreme value index around some horizontal level somewhere near .04. This is followed by a drop, starting shortly after $k = 385$ to a level of about $-.02$ with much tighter fluctuations. The difficulty in selecting the appropriate k is in deciding when fluctuations around "some level" start drifting away from that level for good. Another difficulty is in deciding what that "some level" should be, given that we need to select it within the strong fluctuations. Our view is that .04 is a reasonable central choice for that level. Other "round number" level choices would

be .06 and .03, which for some k-interval seem quite stable but these level values are not too centrally located within the fluctuation region. Based on a level choice of .04 for the estimated extreme value index one finds that $k = 385$ is roughly the largest value producing that level value before the estimates start drifting off. 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 choice of $k = 385$ becomes clearer in the finer detail of Figure 33 without the logarithmic scale of k.

The pattern in Figure 31 (negative deviations, left tail data) seems quite different from that in Figure 30. Here the strong fluctuations only appear to last up to $k = 55$. This is followed by much tighter fluctuations around some higher level near .14. Then there is a drop down to about .06, again with tight variation around that level. This is followed by a drop down to about 0 to .02, again with tight variation. Somewhere around $k = 325$ the pattern drifts down strongly and does not invite further consideration. Because of this rather inconclusive behavior we decided to do the extraplolation based $k = 98, 145, 270$, and 325 left tail deviation extremes. For a different perspective see also Figure 34 without the logarithmic scale of k.

The pattern in Figure 32 (absolute deviations, right tail data) seems to be a mix of the previous two situations. Initial strong fluctuations are followed by a meandering pattern of milder fluctuations. This pattern seems to settle down to a level around .04 from which it drifts down at around $k = 700$. Thus we chose $k = 700$ as the number of extremes to use for extrapolation.

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

Next we examined the behavior of the estimates for the extreme value index when these estimates were 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 36-38 on pages 51-53 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 39-41 on pages 54-56. The behavior then is seen to be quite similar to that seen for the Hill estimator.

Figure 30: Estimation of Extreme Value Index for Various Tail Depths k Using Positive Deviations

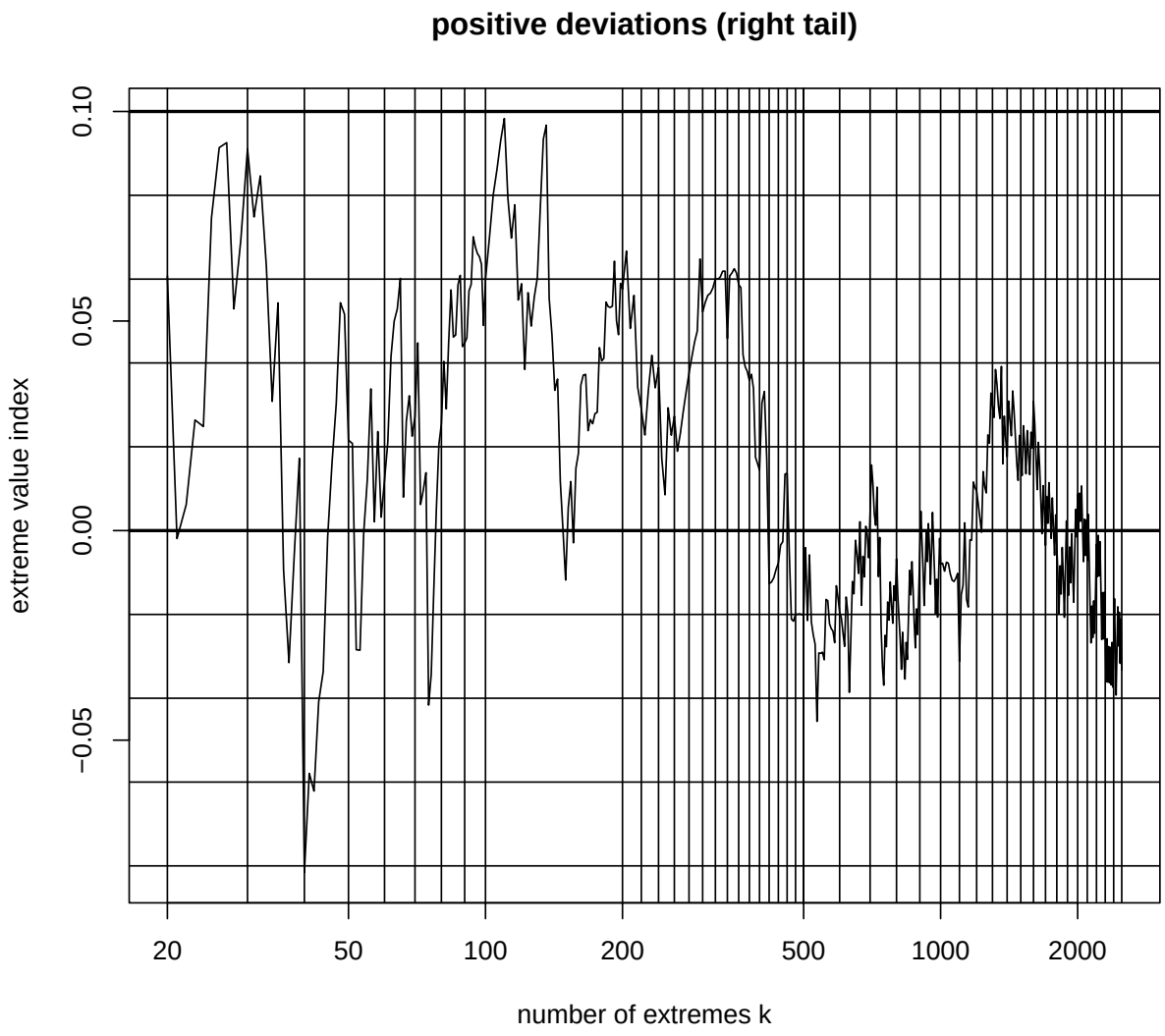


Figure 31: Estimation of Extreme Value Index for Various Tail Depths k Using Negative Deviations

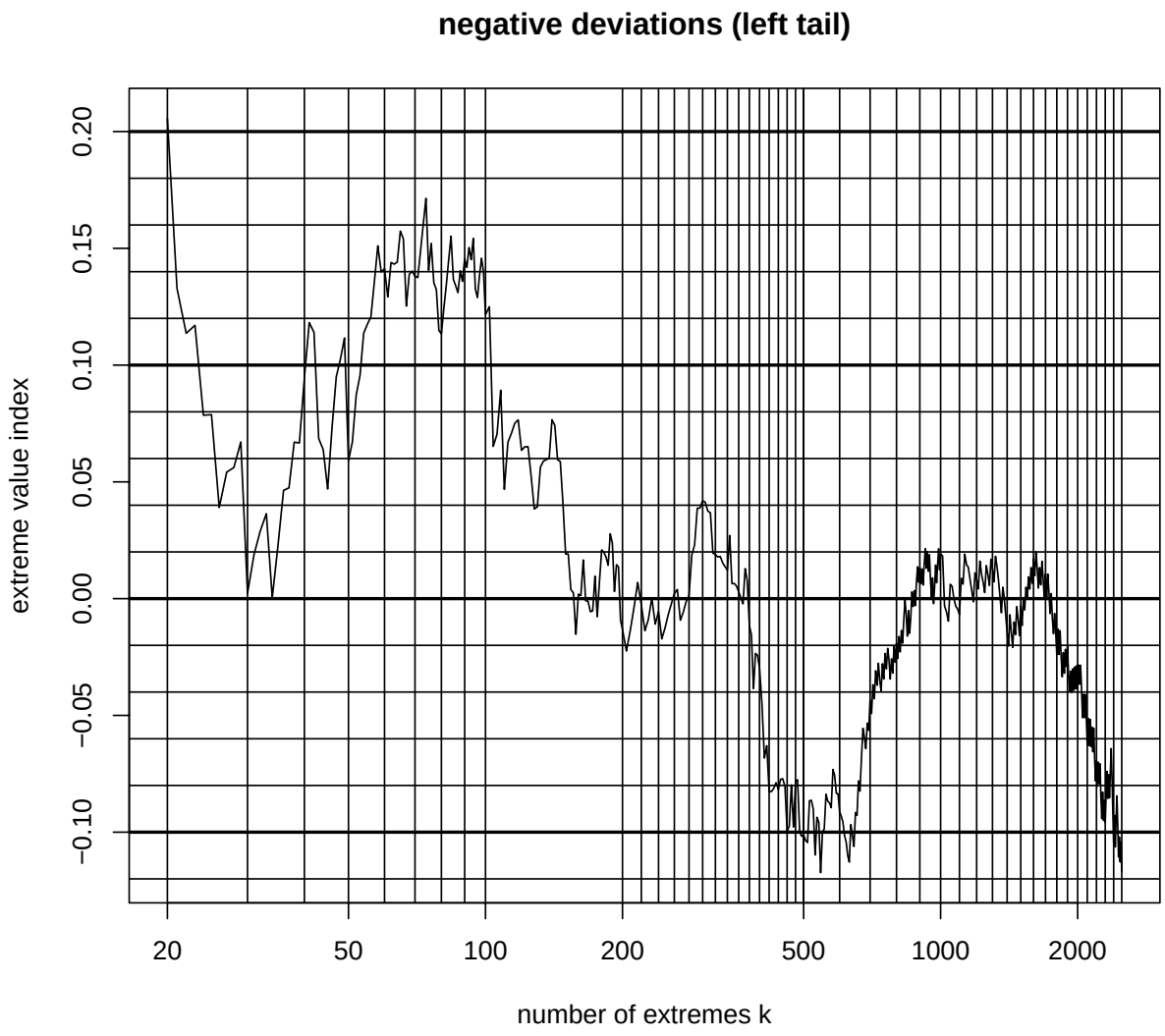


Figure 32: Estimation of Extreme Value Index for Various Tail Depths k Using Absolute Deviations

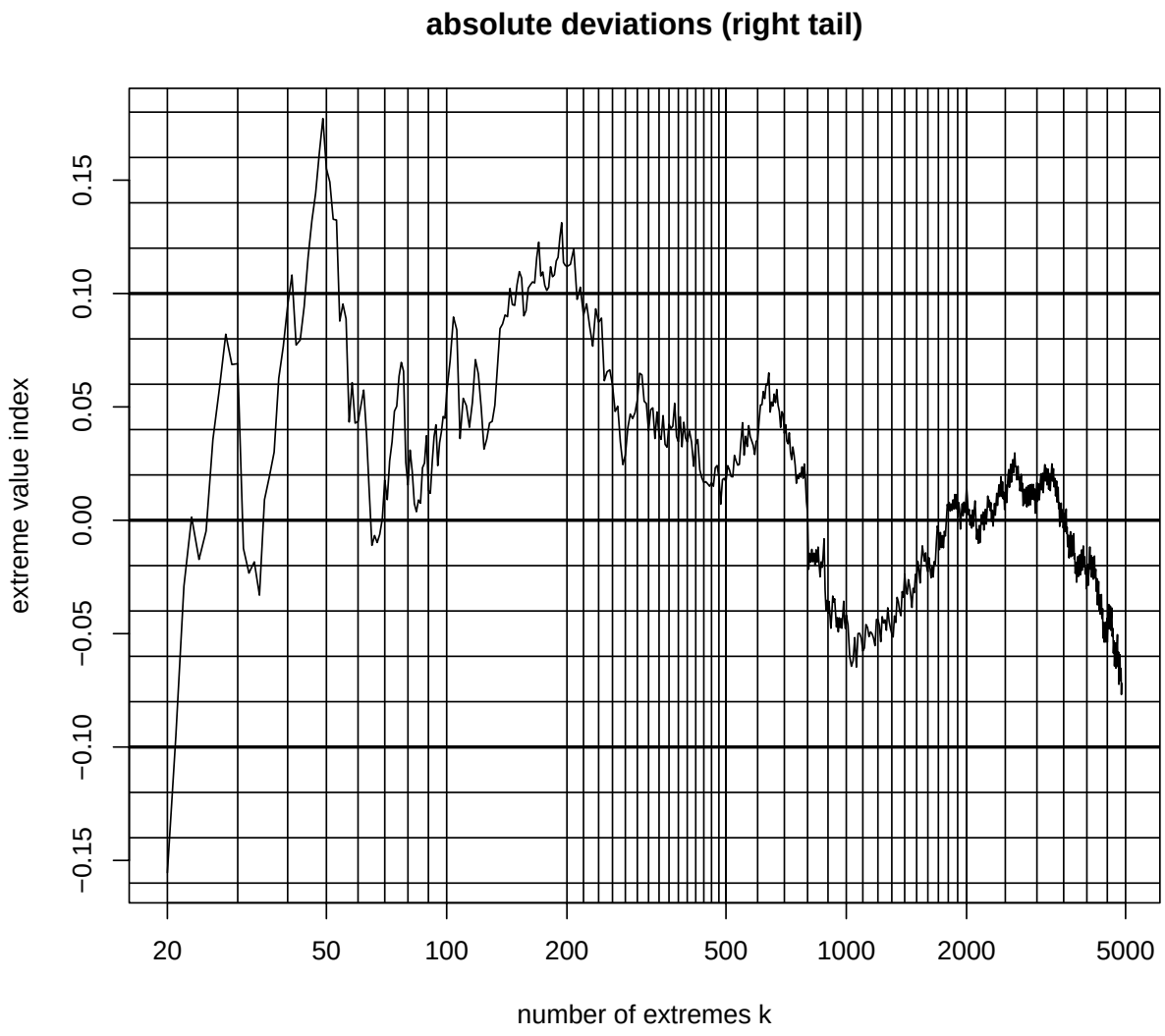


Figure 33: Estimation of Extreme Value Index for Various Tail Depths k
Using Positive Deviations (Detail)

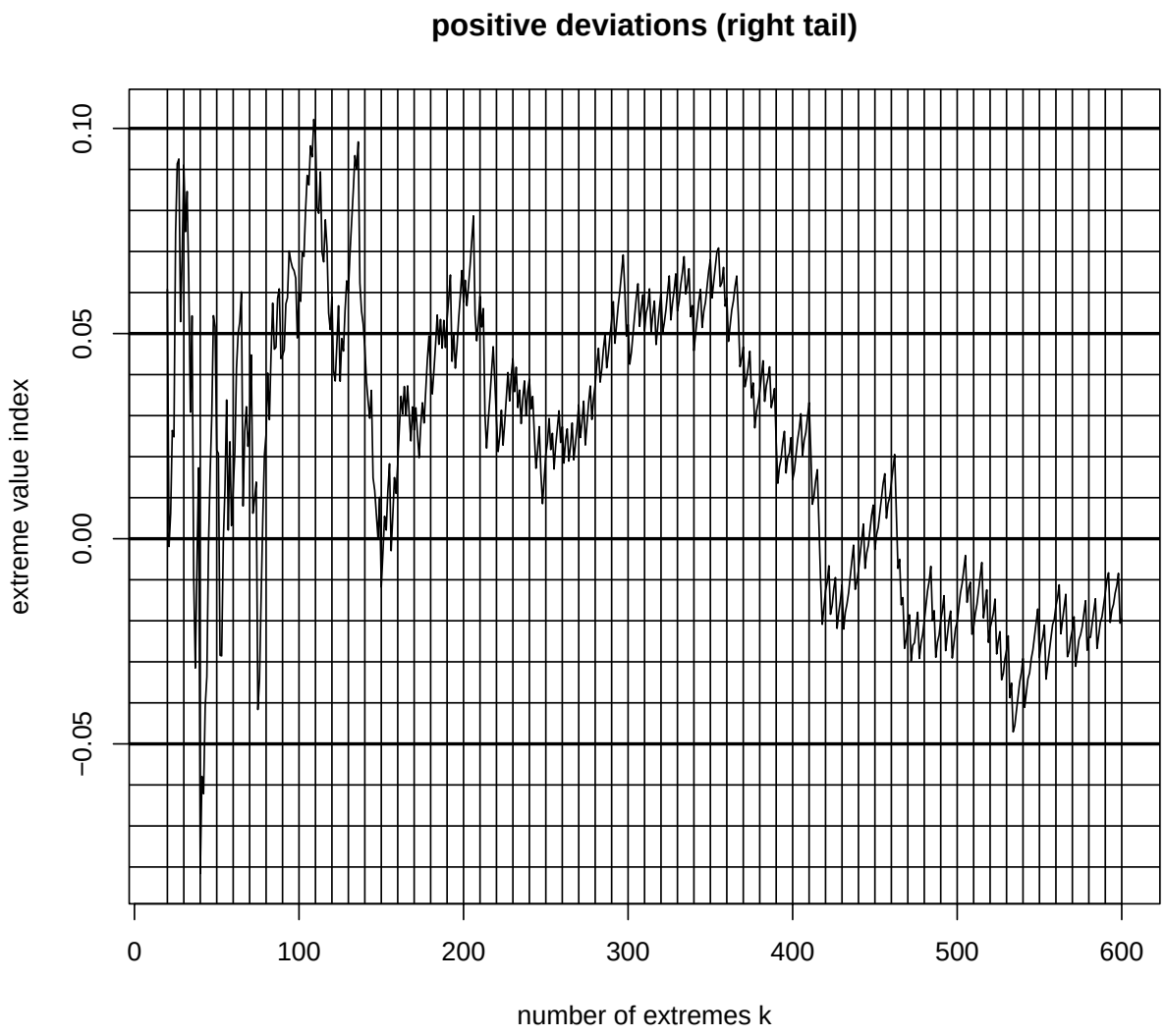


Figure 34: Estimation of Extreme Value Index for Various Tail Depths k Using Negative Deviations (Detail)

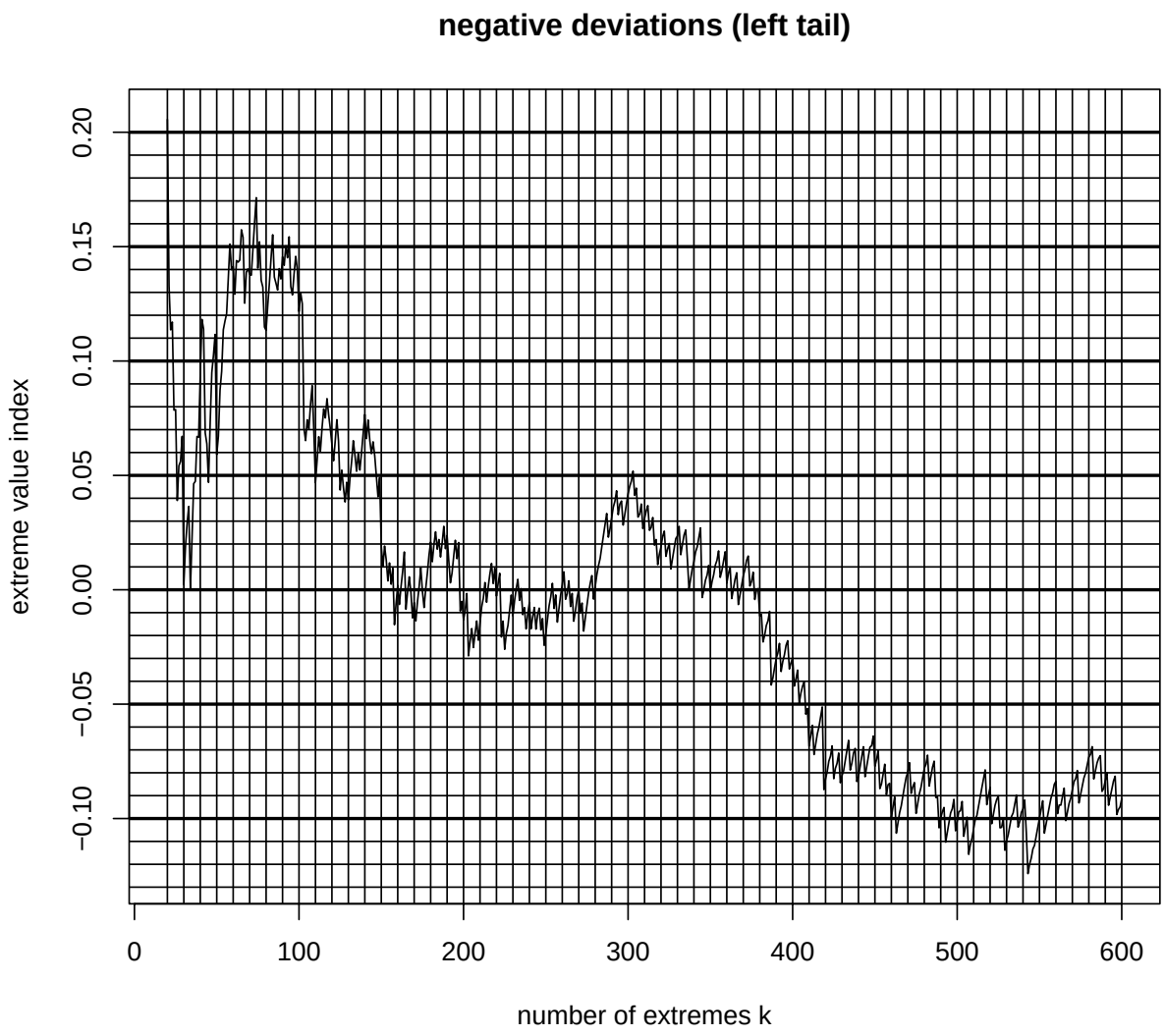


Figure 35: Estimation of Extreme Value Index for Various Tail Depths k Using Absolute Deviations (Detail)

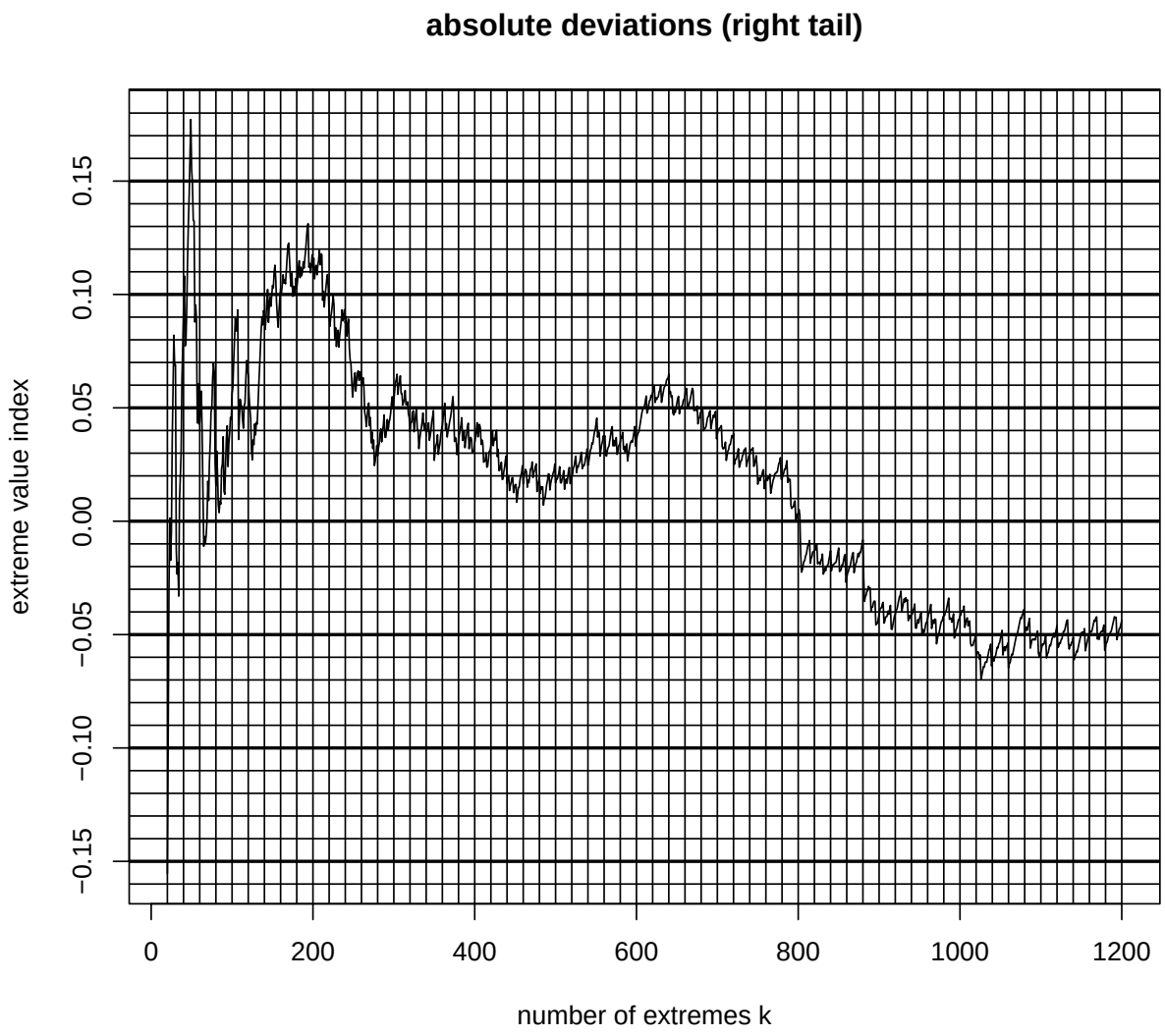


Figure 36: Estimation of Extreme Value Index for Various Thresholds Using Positive Deviations Over Threshold u

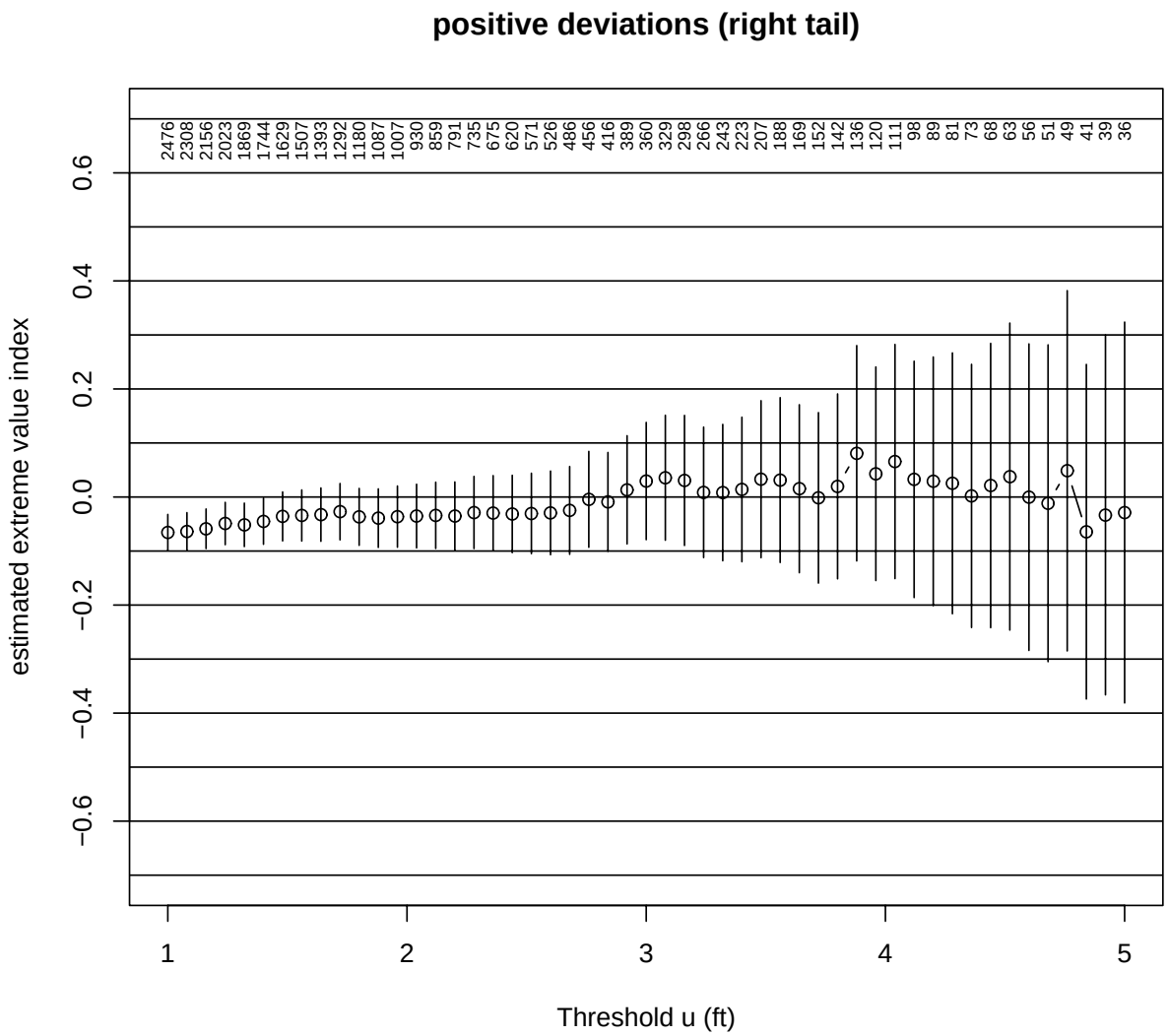


Figure 37: Estimation of Extreme Value Index for Various Thresholds Using Negative Deviations Over Threshold u

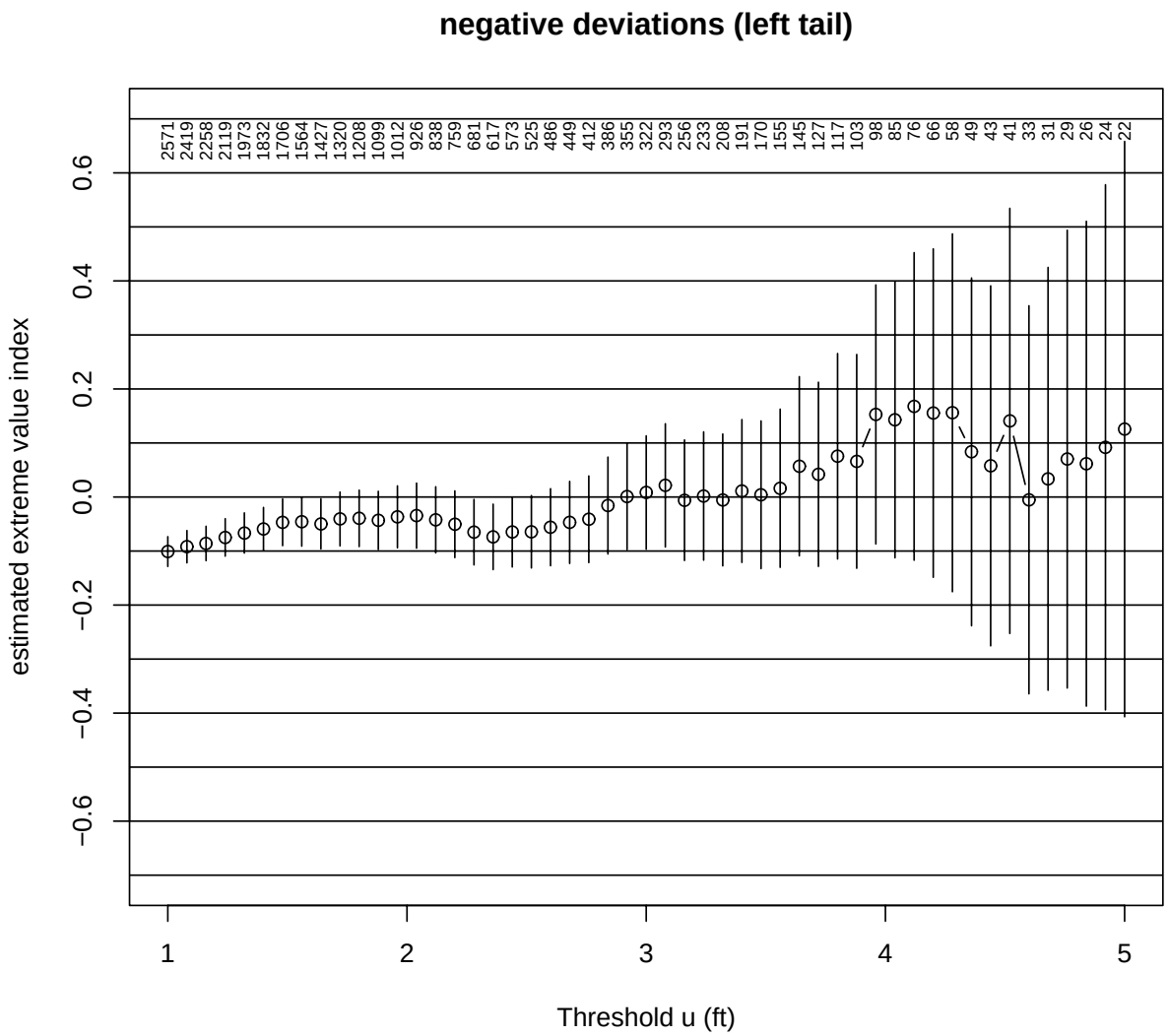


Figure 38: Estimation of Extreme Value Index for Various Thresholds Using Absolute Deviations Over Threshold u

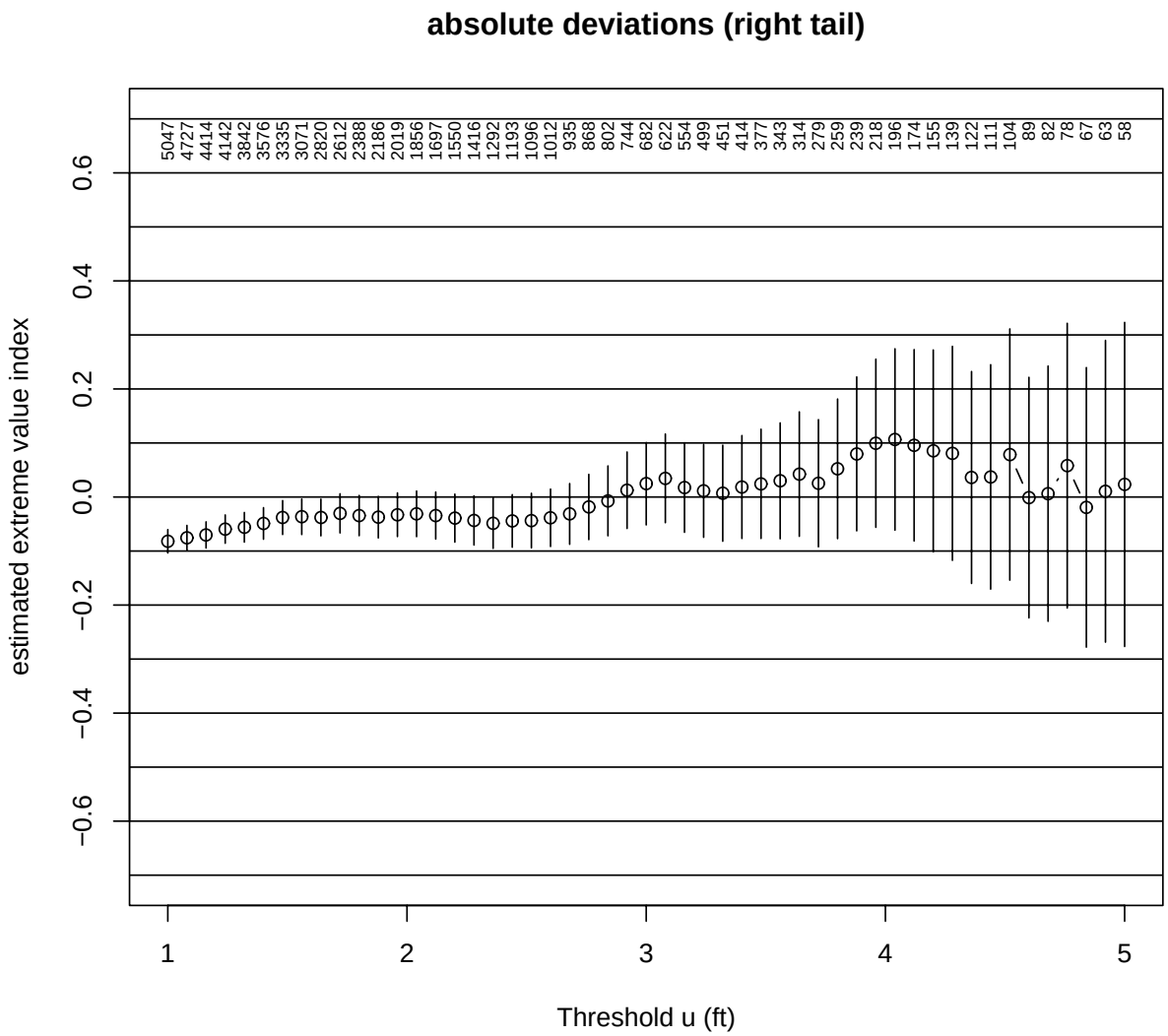


Figure 39: Estimation of Extreme Value Index for Various Thresholds Using Positive Deviations Over Threshold u

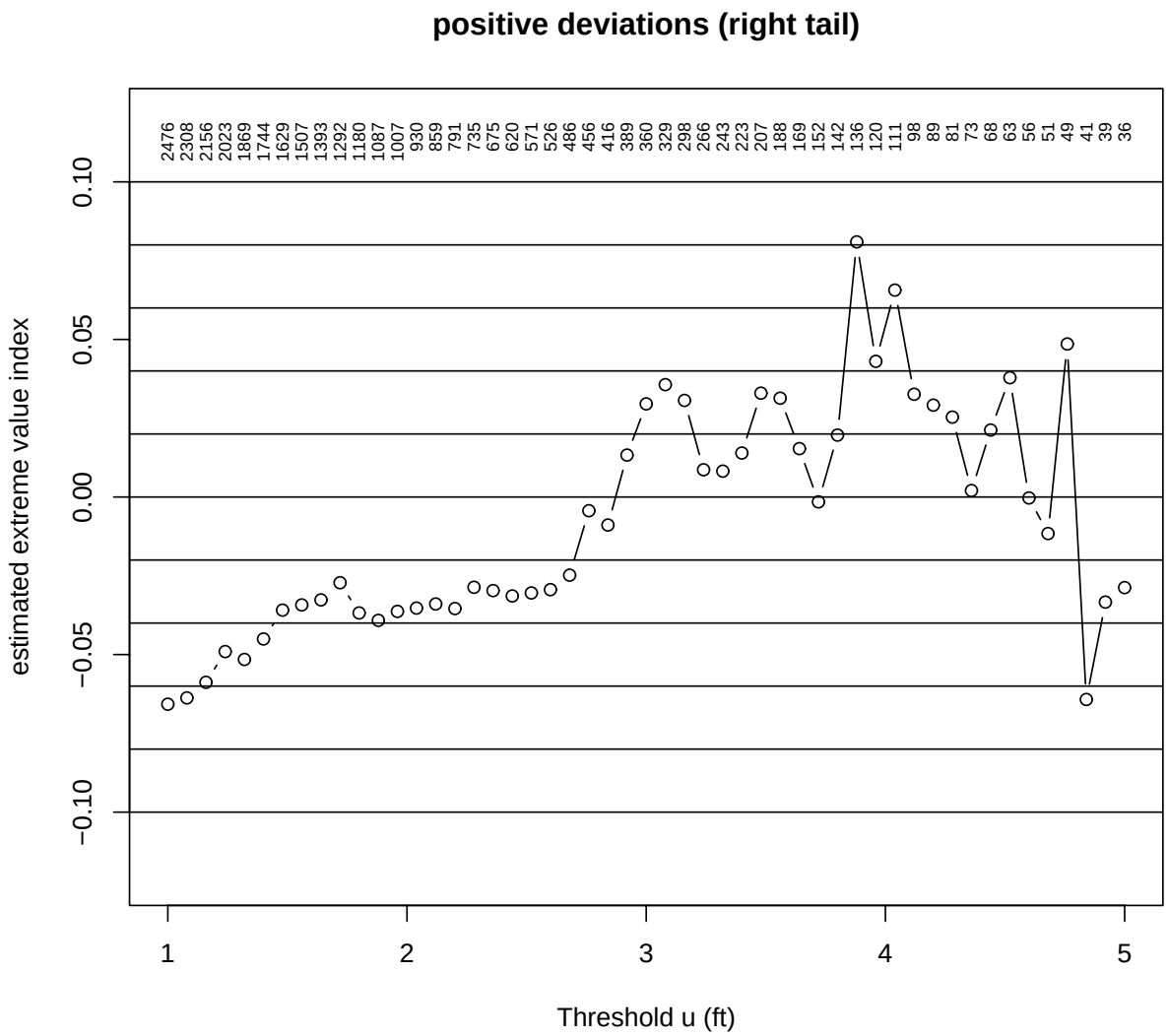


Figure 40: Estimation of Extreme Value Index for Various Thresholds Using Negative Deviations Over Threshold u

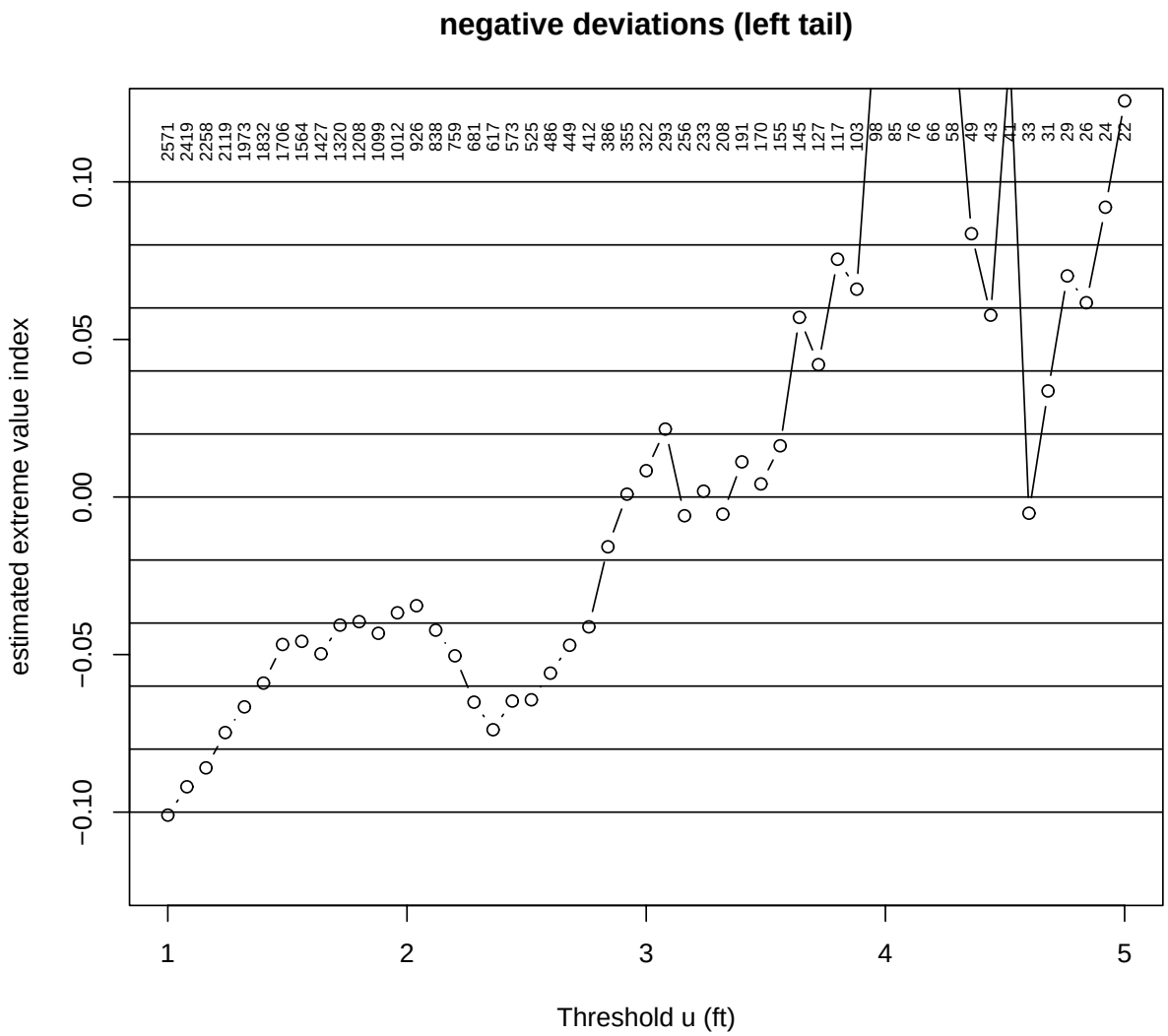
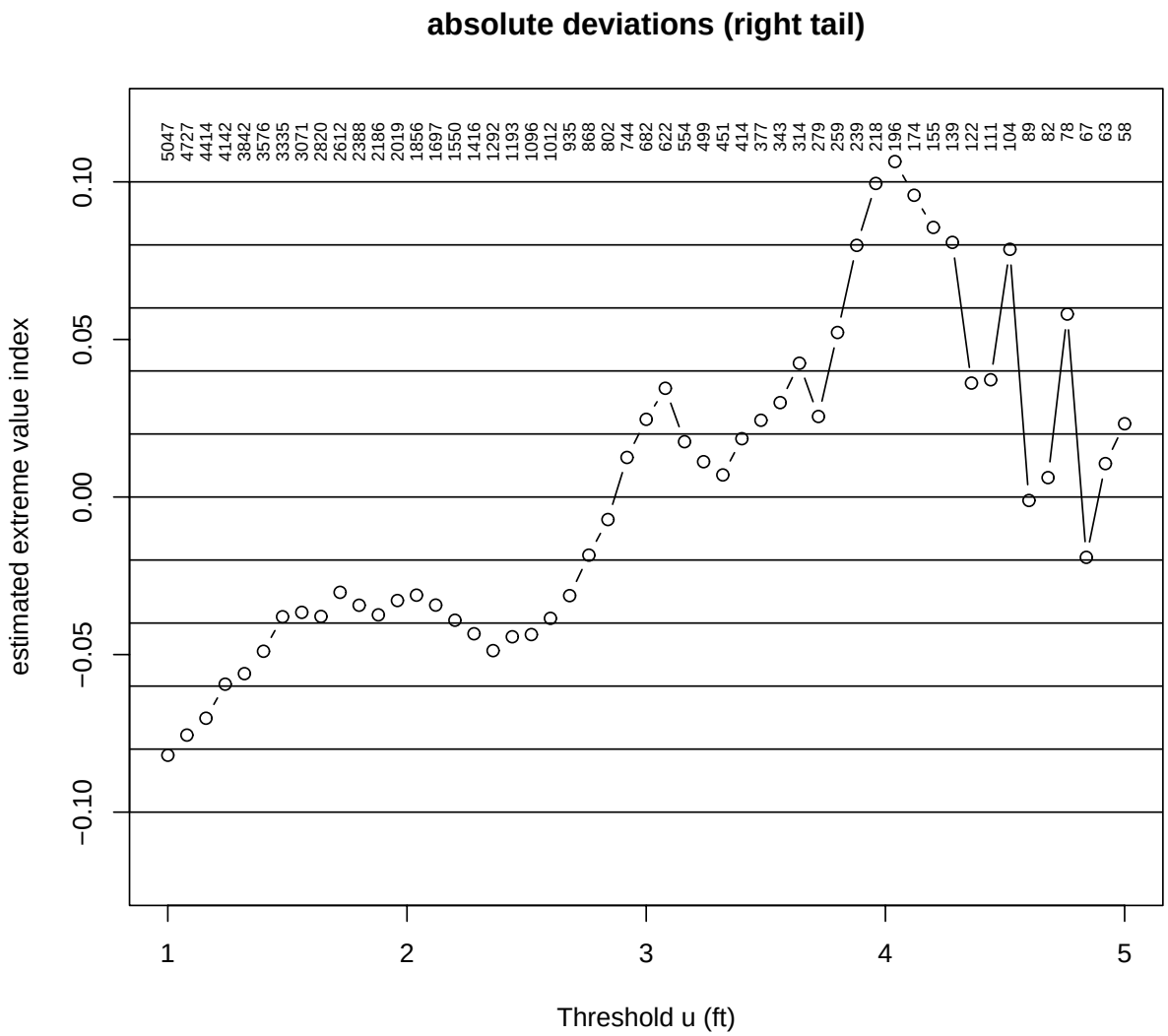


Figure 41: Estimation of Extreme Value Index for Various Thresholds Using Absolute Deviations Over Threshold u



In Figures 36 and 39 there appears to be a level of stability in the estimates for $u \in [3, 5]$ while for $u < 3$ the pattern appears to drift down to a lower level. The level of stability appears to be around .03 (see Figure 39) while the lower level is definitely below zero. Note that $u = 3$ corresponds to $k = 360$, i.e., we are in approximate agreement with our other assessment of the right tail, namely $k = 385$.

In Figures 37 and 40 the situation is again somewhat less compelling as was the case with the modified Hill estimator. Definitely one can perceive the drifting to a lower level for $u < 3$. Within $[3, 5]$ one sees several possible levels of stability (see Figure 40), respectively, ending roughly at $k = 98, 145, 256$, and 322, in relatively close agreement with our previous choices.

In Figures 38 and 41 the thresholds $u = 3$ seems to be the best choice for separating fluctuations around a stable level from the drifting off behavior. This corresponds to $k = 682$ again agreeing roughly with our earlier assessment of $k = 700$.

5.4 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 and for those we have nonparametric confidence bounds which are 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 that are shown in Figures 42-44 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 reasonable.

5.5 Results Using EXTRAP

Based on the above examinations we decided to use the EXTRAP program (described in [8]) to extrapolate the positive tail of the deviation data using $k = 385$ extremes, the negative tail of the deviation data using $k = 98, 145, 270$, and 325 extremes, and the right tail of the absolute deviations using $k = 700$ extremes. The four choices of k for the negative tail data should give us some appreciation of the sensitivity of the extrapolations, especially for extremely small exceedance risks.

Figure 42: Estimated Median Excess over Threshold for Various Thresholds Using Positive Deviations Over Threshold u

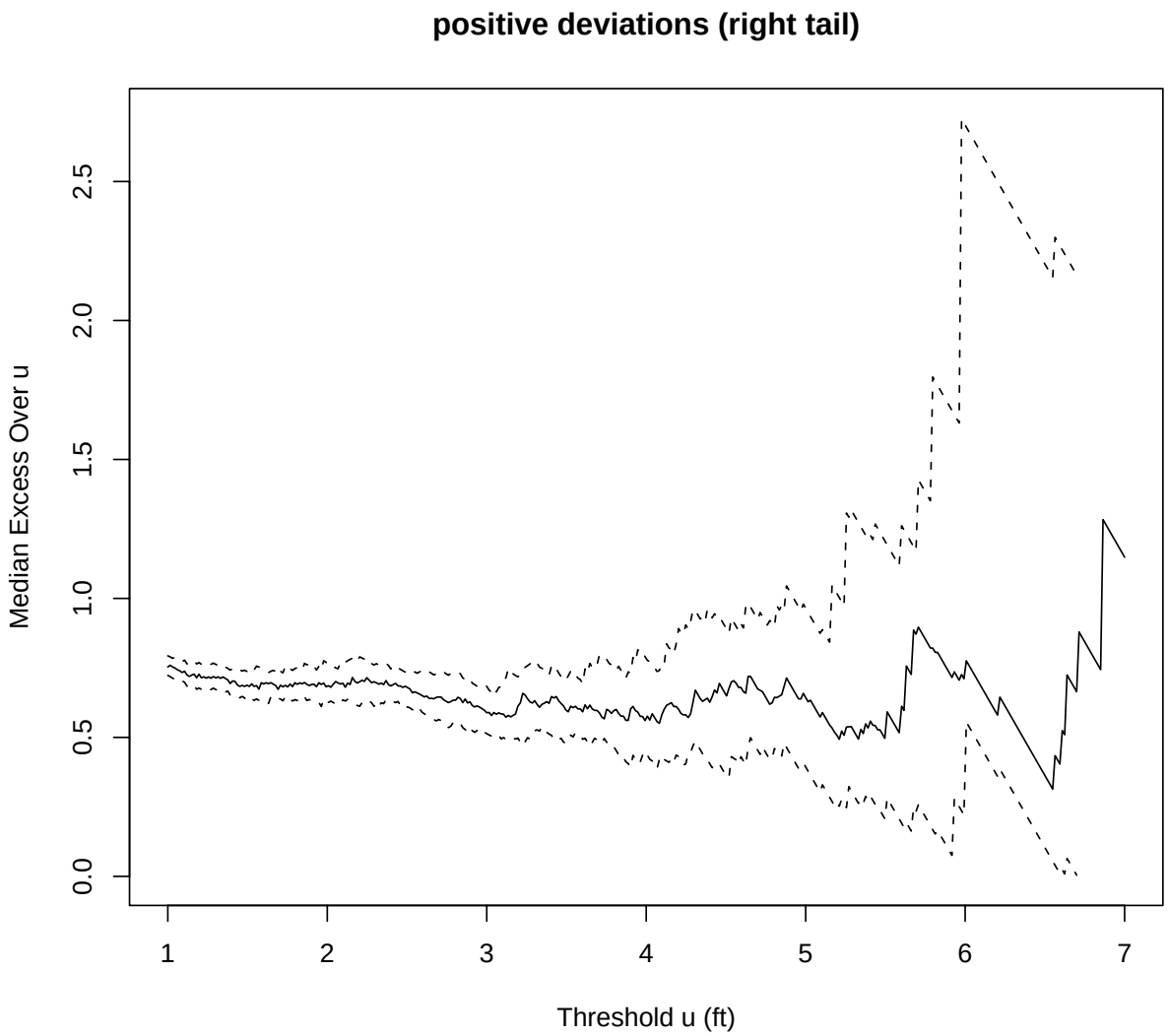


Figure 43: Estimated Median Excess over Threshold for Various Thresholds Using (−) Negative Deviations Over Threshold u

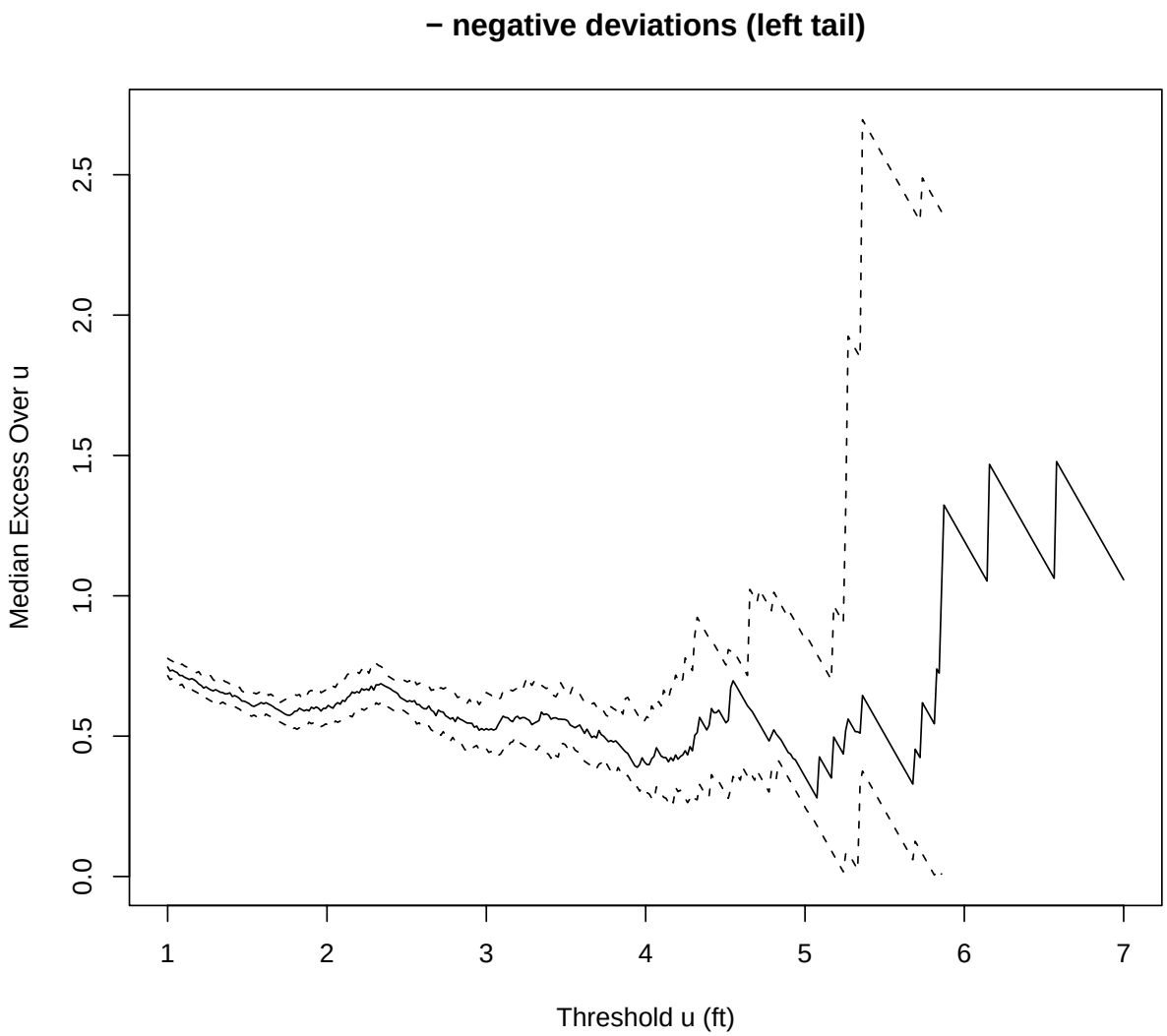
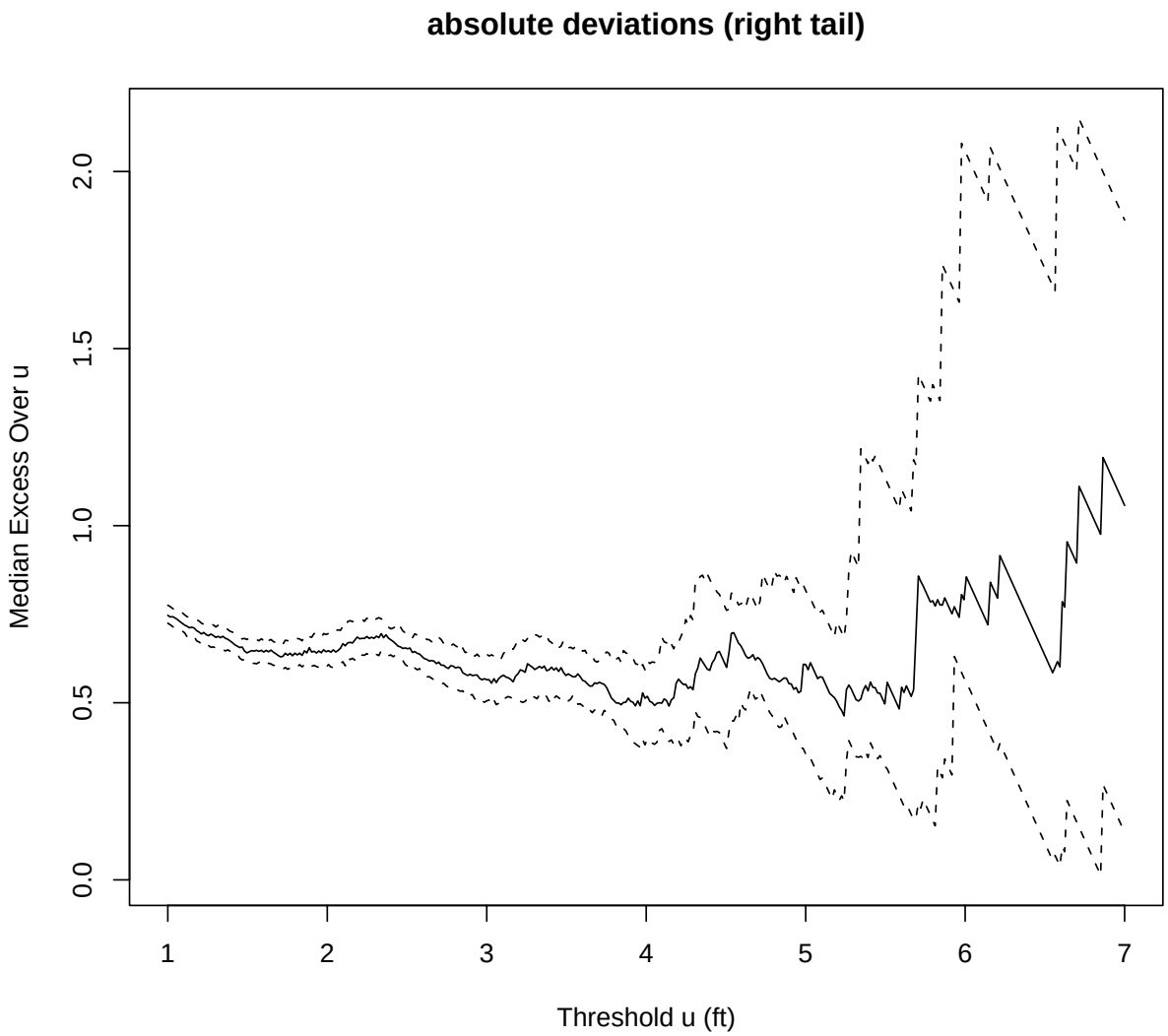


Figure 44: Estimated Median Excess over Threshold for Various Thresholds Using Absolute Deviations Over Threshold u



The resulting estimates and 95% upper bounds are shown in Figures 45-50 on pages 64-69 (explained below) and the risk thresholds are tabulated in Tables 2 and 3 on page 63, respectively. The reason for doubling the exceedance risks in the case of absolute deviations is to make the deviation thresholds comparable, since for $x \geq 0$

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

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

Note that the abscissas of the plots in Figures 45-50 have two scales, a distorted probability scale (indicating the probability of exceeding a certain taxiway centerline deviation level as indicated on the ordinate) and a linear $h(p)$-scale (labeled "transformed p(i)") which relates to the exceedance probabilities through the following transformation formula

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

with ln denoting the natural logarithm, $n = 9796$ denoting the total sample size, and c the estimated extreme value coefficient or index as given in the plot.

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 ith-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 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 and 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 value c employed in the abscissa scale transformation $h(p)$ is known. The fact that it is 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 exceedance risk or upper confidence bound for that risk for a specified exceedance threshold. This inversion can be implemented via the following formula

$$p = 1 - \exp\left(-\frac{1}{n}\left[1 + c\left(\frac{y-\lambda}{\delta}\right)\right]^{-1/c}\right) , \tag{2}$$

where y is the specified threshold, p is the risk estimate or upper bound on 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 line. 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. For the negative extremes (for which we reversed signs in extrapolation plots) we give the results for four different values of k.

Within the range of the data (corresponding to exceedance risks $\geq 10^{-4}$) the tabulated values don't differ much for these four choices. However, for very small exceedance risk at 10^{-7} one encounters differences of about 11 ft (for $k = 98$ and 270). These differences in results mostly derive from the different estimates of the extreme value index and in the case of $k = 98$ one would deal with great uncertainty. Thus the 11 ft may be a result of that uncertainty. There is a possibility that this different behavior for the negative deviations is related to the centerlight offset from the centerline. However, we have not been able to throw further light on this.

When going to absolute deviations one appears to end up in a compromise position. Given that one would not want to have different risk assessments for the two sides of a taxiway we let the results for the extrapolation on the absolute values stand as the final answer.

Table 2: Estimated Thresholds by Exceedance Risk
for Adjusted 747 Centerline Deviations from Taxiway Centerline

	exceedance risk				
	10^{-7}	10^{-6}	10^{-5}	10^{-4}	10^{-3}
positive right tail $k = 385$	17.02 ft	13.95 ft	11.15 ft	8.60 ft	6.26 ft
negative left tail $k = 98$	21.73 ft	15.50 ft	11.05 ft	7.87 ft	5.59 ft
negative left tail $k = 145$	15.86 ft	12.71 ft	10.00 ft	7.67 ft	5.65 ft
negative left tail $k = 270$	12.80 ft	11.02 ft	9.25 ft	7.47 ft	5.69 ft
negative left tail $k = 325$	13.74 ft	11.58 ft	9.53 ft	7.57 ft	5.69 ft
	$2 \cdot 10^{-7}$	$2 \cdot 10^{-6}$	$2 \cdot 10^{-5}$	$2 \cdot 10^{-4}$	$2 \cdot 10^{-3}$
absolute deviations $k = 700$	15.84 ft	13.04 ft	10.48 ft	8.13 ft	5.99 ft

Table 3: Threshold 95% Upper Bounds by Exceedance Risk
for Adjusted 747 Centerline Deviations from Taxiway Centerline

	exceedance risk				
	10^{-7}	10^{-6}	10^{-5}	10^{-4}	10^{-3}
positive right tail $k = 385$	18.23 ft	14.92 ft	11.90 ft	9.13 ft	6.61 ft
negative left tail $k = 98$	24.98 ft	17.65 ft	12.41 ft	8.67 ft	5.99 ft
negative left tail $k = 145$	17.57 ft	14.01 ft	10.94 ft	8.30 ft	6.02 ft
negative left tail $k = 270$	13.74 ft	11.81 ft	9.88 ft	7.94 ft	6.01 ft
negative left tail $k = 325$	14.71 ft	12.38 ft	10.15 ft	8.02 ft	5.99 ft
	$2 \cdot 10^{-7}$	$2 \cdot 10^{-6}$	$2 \cdot 10^{-5}$	$2 \cdot 10^{-4}$	$2 \cdot 10^{-3}$
absolute deviations $k = 700$	16.66 ft	13.69 ft	10.98 ft	8.49 ft	6.22 ft

Figure 45: Extrapolation for Positive Deviations
Using 385 Most Extreme Positive Deviations

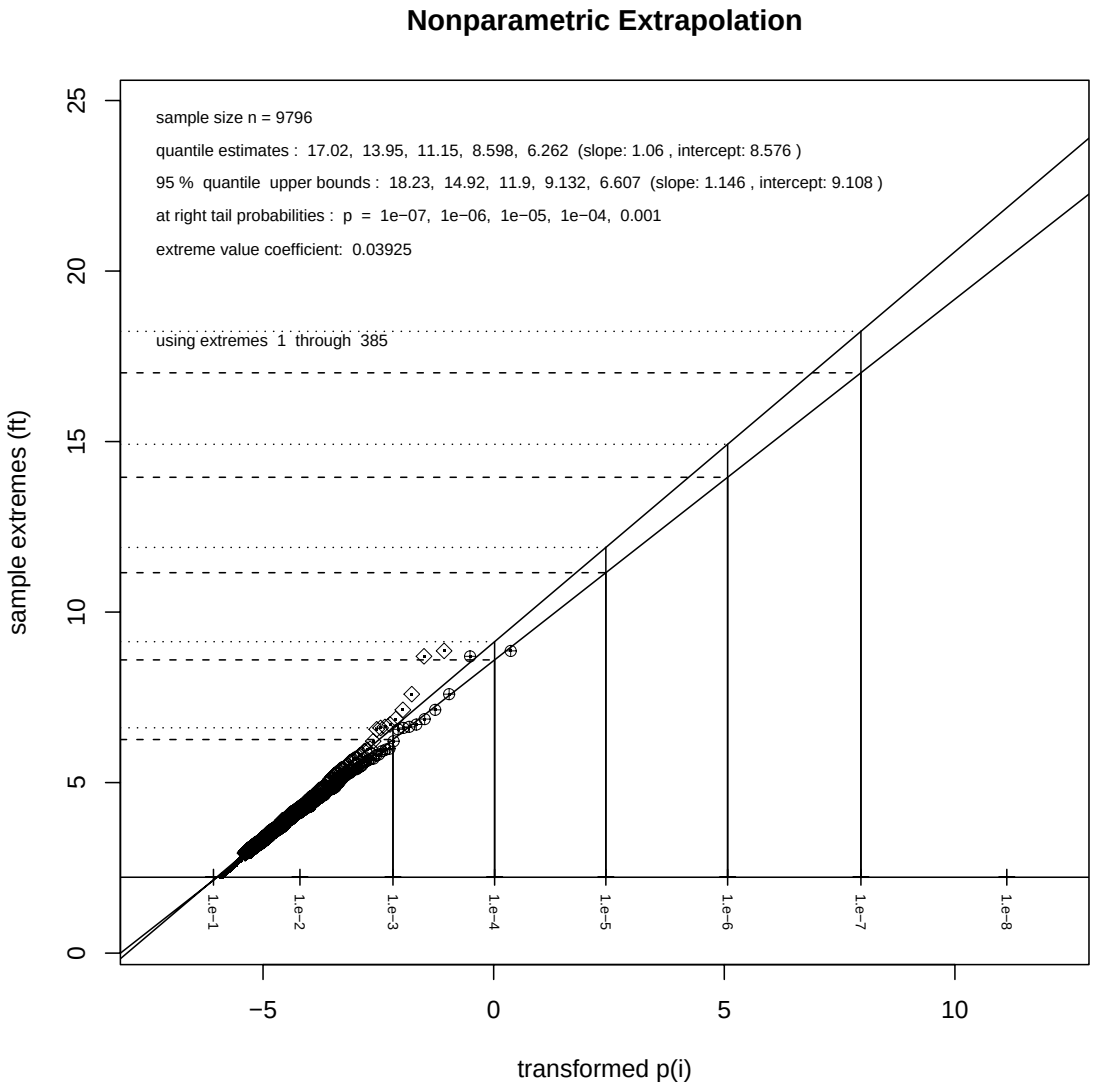


Figure 46: Extrapolation for Negative Deviations
Using 98 Most Extreme Negative Deviations (sign reversed)

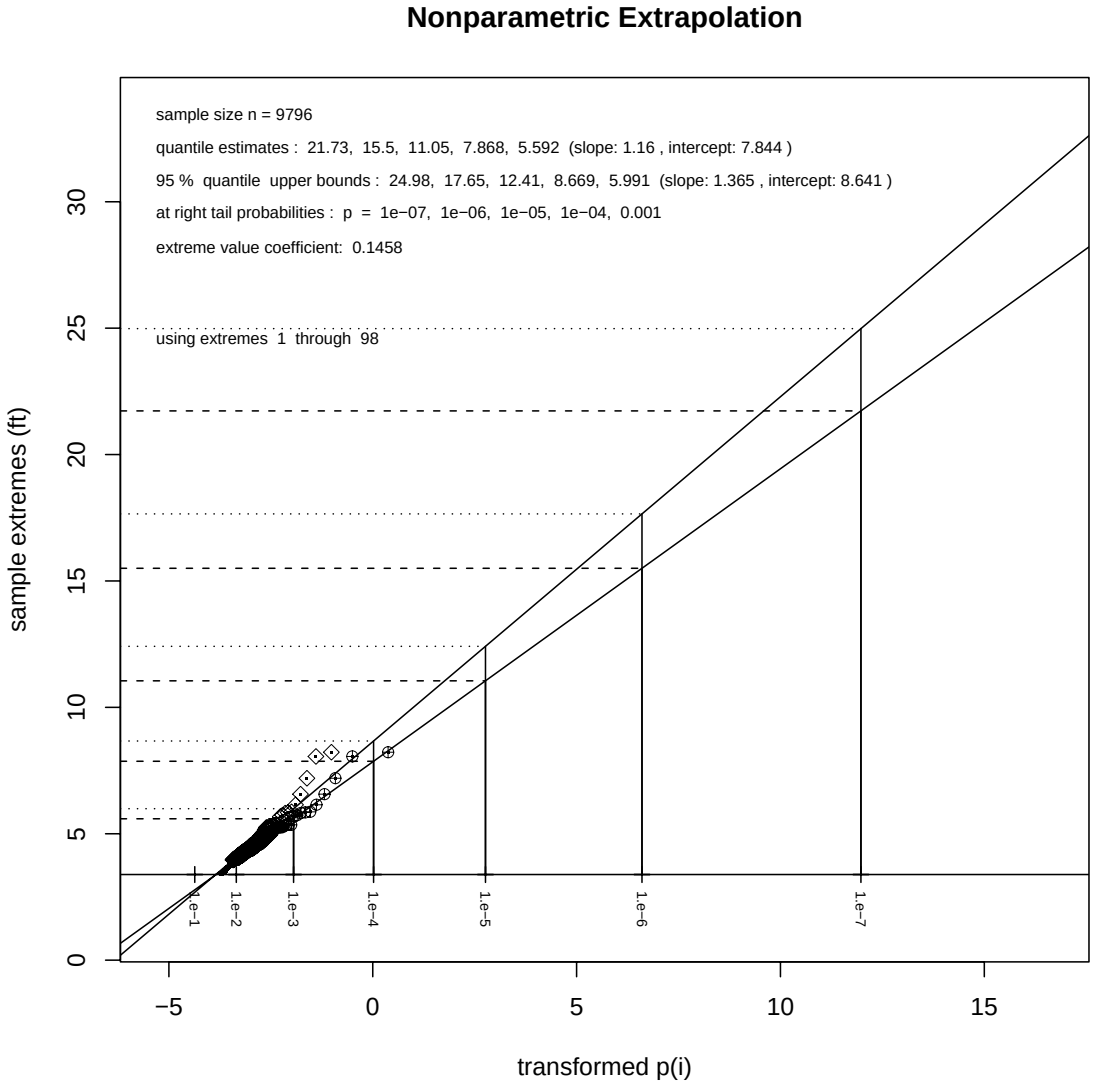


Figure 47: Extrapolation for Negative Deviations
Using 145 Most Extreme Negative Deviations (sign reversed)

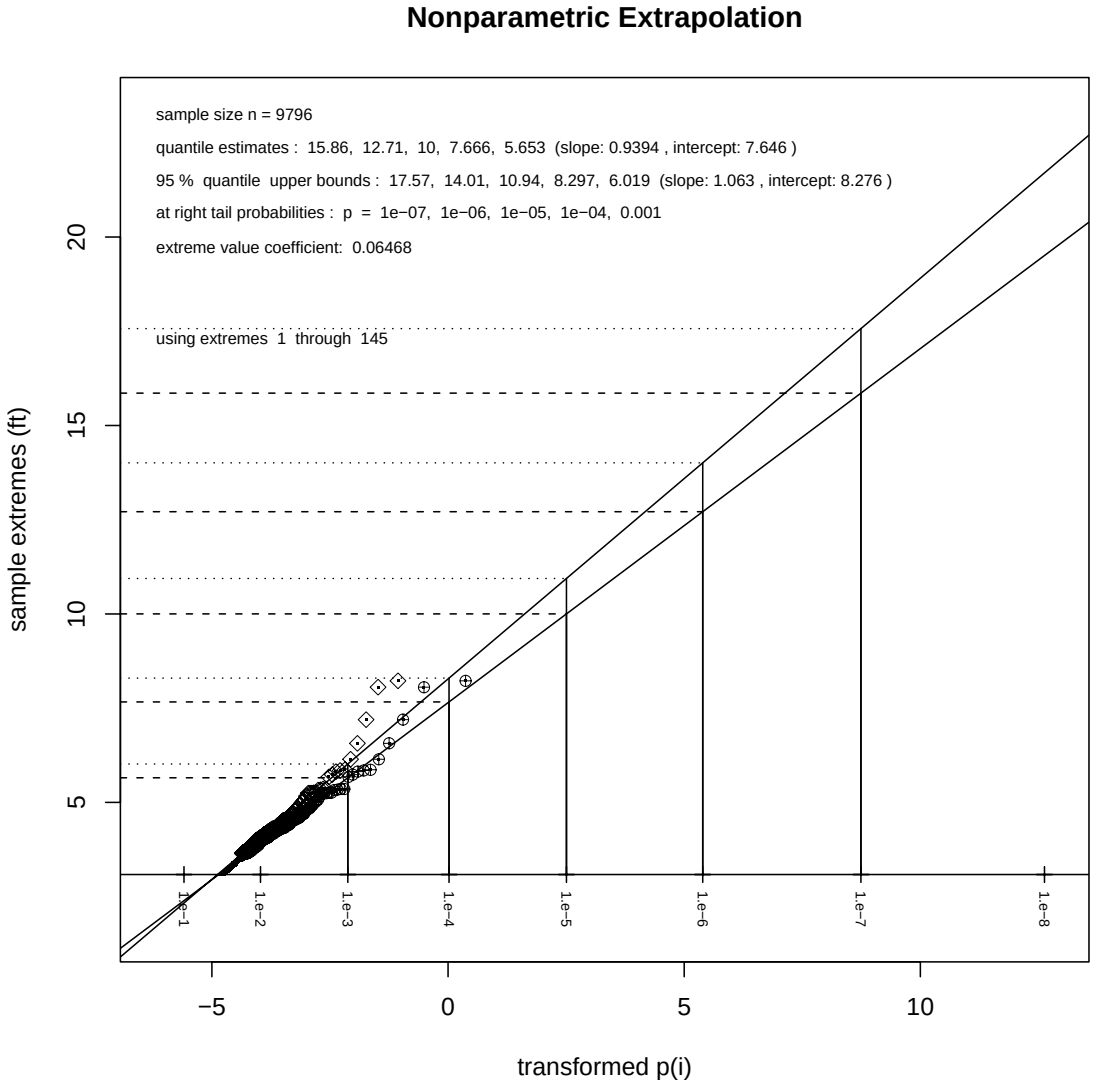


Figure 48: Extrapolation for Negative Deviations
Using 270 Most Extreme Negative Deviations (sign reversed)

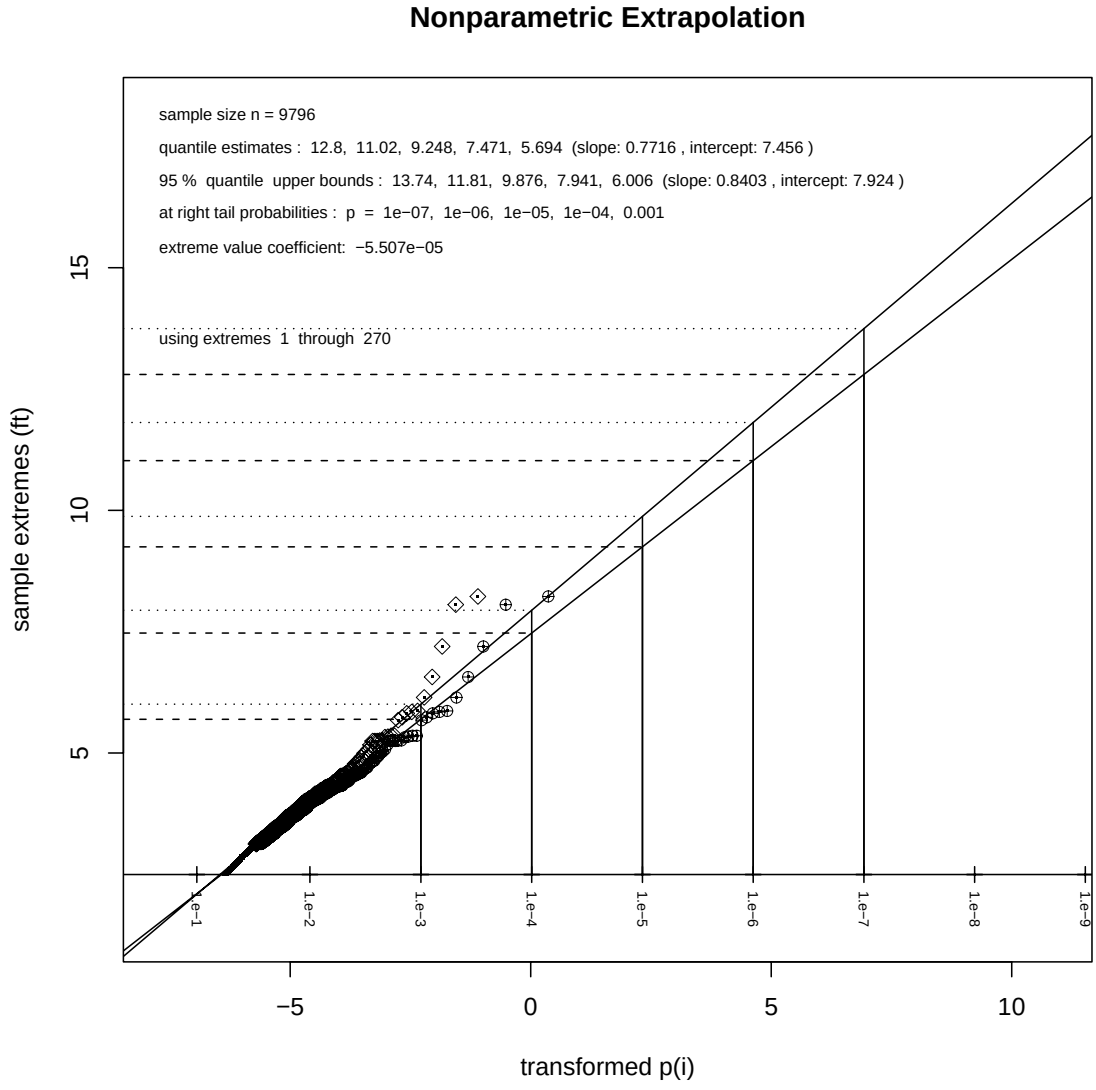


Figure 49: Extrapolation for Negative Deviations
Using 325 Most Extreme Negative Deviations (sign reversed)

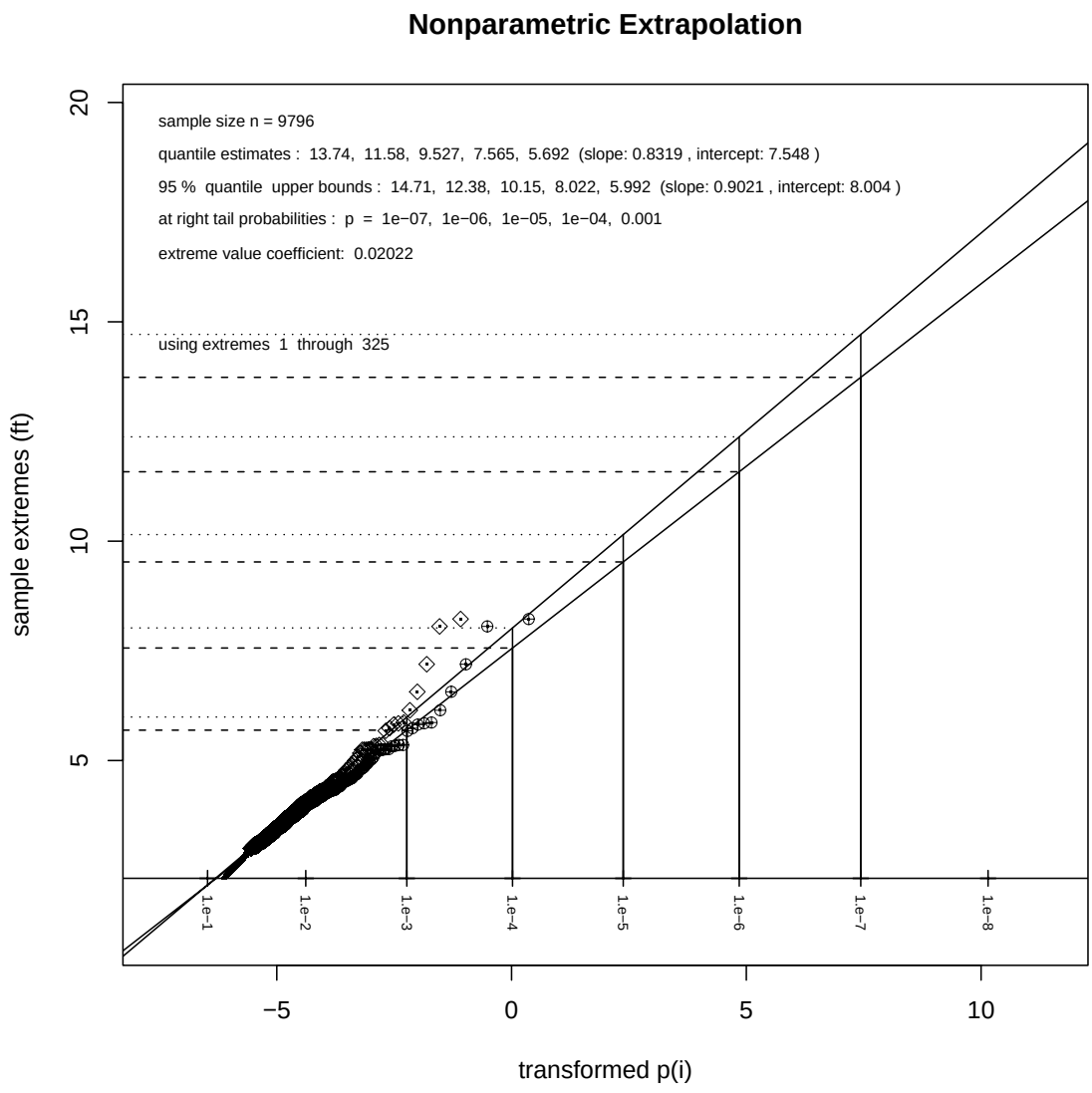
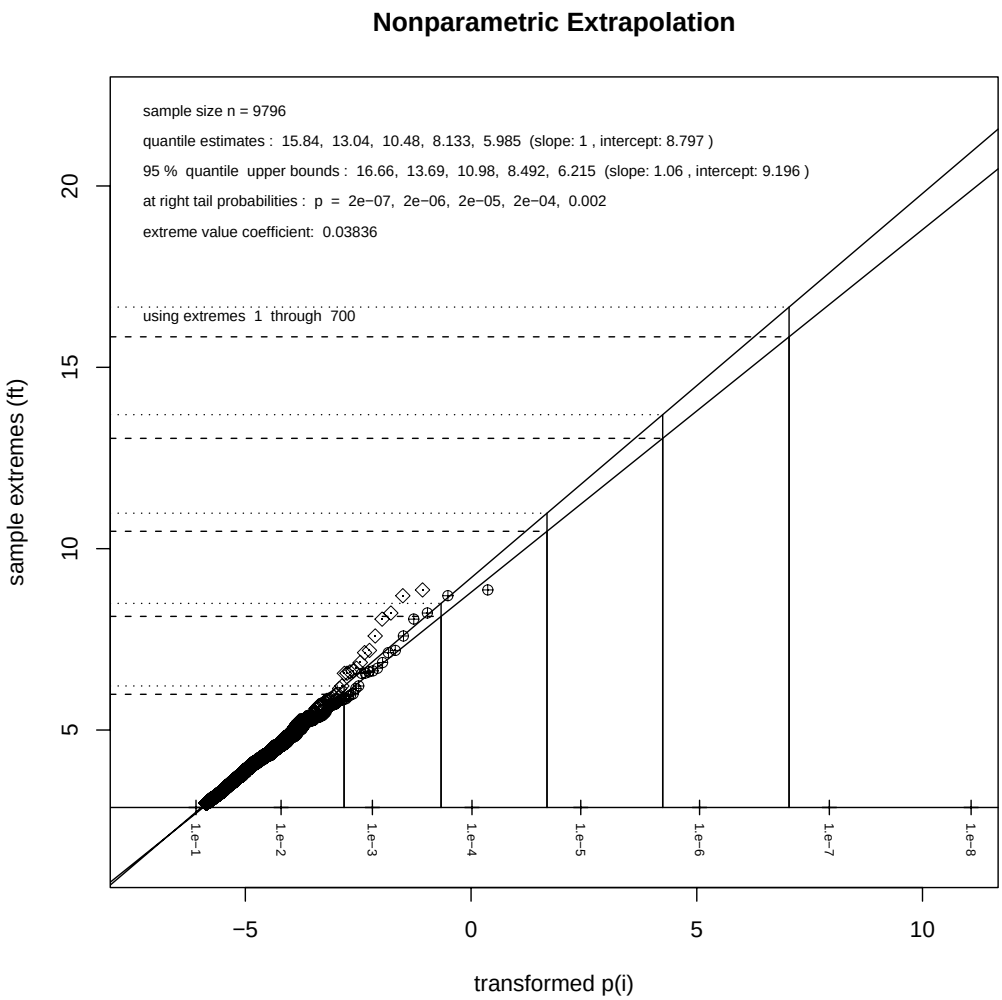


Figure 50: Extrapolation for Absolute Deviations
Using 700 Most Extreme Absolute Deviations



6 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 both taxiways. 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 2 and 3, are based on the bias adjusted deviation data. These values are meaningful only if we pretend that there is no bias of 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 either of the form $\hat{\mu}_O + \hat{\mu}_P$ or $\hat{\mu}_O - \hat{\mu}_P$ for an eastbound (northbound) or westbound (southbound) heading, respectively.

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 2 and 3, i.e., add $|\hat{\mu}_P| + |\hat{\mu}_O|$.

Previously we identified two sets of bias corrections, one set for each taxiway. These are slightly different from each other, presumably just due to sampling variation. It would make sense to make a bias correction that is independent of the taxiway. One could simply average the two values for $\mu_{P,K}$ and $\mu_{P,R}$ and the two values for $\mu_{O,K}$ and $\mu_{O,R}$, but since different sample sizes were involved at KILO and ROMEO, namely $n = 3520$ for KILO and $m = 6276$ for ROMEO, 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 bias corrections

$$\mu_P = \frac{n}{m+n}\mu_{P,K} + \frac{m}{m+n}\mu_{P,R} = .27 \quad \text{and} \quad \mu_O = \frac{n}{m+n}\mu_{O,K} + \frac{m}{m+n}\mu_{O,R} = .48$$

and we will add $.27 + .48 = .75$ ft to any threshold given in the Tables 2 and 3.

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 51 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.

The middle density represent the distribution of centerline deviations with maximal bias of $.48 + .27 = .75$ 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 $2.25 \cdot 10^{-5}$ because the center of the distribution is $.75 + .75 = 1.5$ ft further to the right of the back-adjusted threshold on the left.

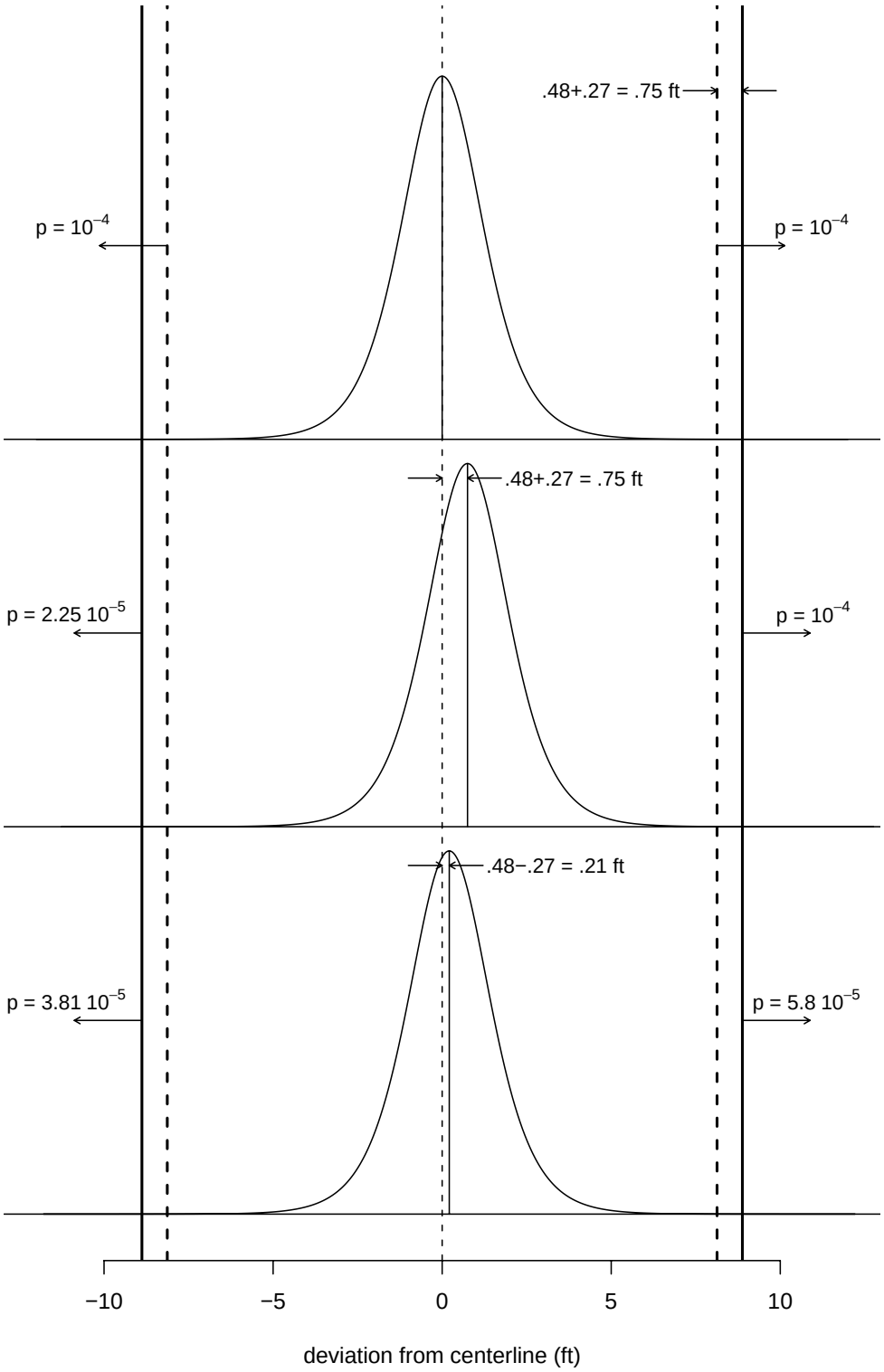
The bottom density represent the distribution of centerline deviations with the smaller bias of $.48 - .27 = .21$ ft. Here the other bias and the parallax bias cancel to some extent. Note that the risk of exceeding the back-adjusted threshold on the right now is $5.8 \cdot 10^{-5}$, while on the left it is only $3.81 \cdot 10^{-5}$ because the center of the distribution is $.75 + .21 = .96$ 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 2 on page 62.

For example, when dealing with a risk of 10^{-6} of the 747 centerline at main gear location exceeding a deviation threshold to a particular side of the taxiway, we find from Table 2 the estimated threshold 13.04 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 $.27 + .48 = .75$ ft and obtain $13.04 + .75 = 13.79$ 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 (KILO or ROMEO) and a particular direction perpendicular to the taxiway centerline. Then 13.79 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) only about once in 10^6 passes. The conservative nature is the same as was discussed in the context of Figure 51, 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 4.

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

	two-sided exceedance risk				
	$2 \cdot 10^{-7}$	$2 \cdot 10^{-6}$	$2 \cdot 10^{-5}$	$2 \cdot 10^{-4}$	$2 \cdot 10^{-3}$
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

Figure 51: Exceedance Probabilities for Various Bias Scenarios



7 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 width 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 52.

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

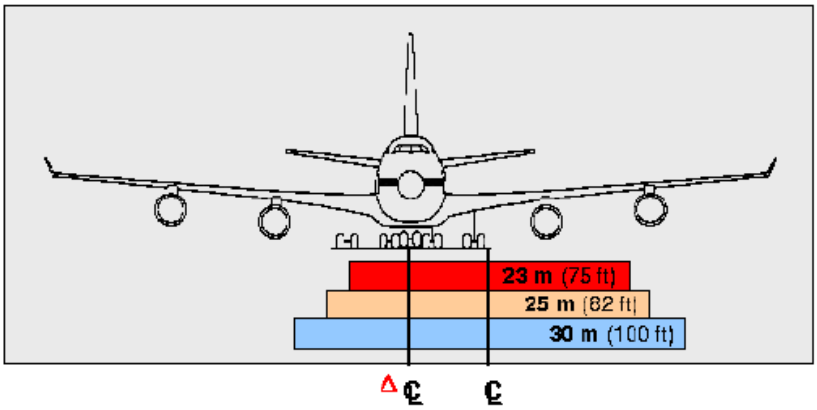
The previously derived back-adjusted threshold of 13.79 ft corresponding conservatively to a 10^{-6} one-sided risk of exceedance translates into a $13.79 + 20.67 = 34.46$ ft deviation of the outside edge of the 747 landing gear relative to the taxiway centerline. Here we assume that the outside edge of the main landing gear is 20.67 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 68.92 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 $2 \times (13.69 + .75 + 20.67) = 70.22$ 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 20.67 and .75 from 75/2 ft we arrive at a threshold of $37.5 - 20.67 - .75 = 16.08$ for the bias corrected centerline deviations. Using equation 2 from page 62 we get the estimated two-sided risk of exceeding this threshold as

$$p = 1 - \exp\left(-\frac{1}{9796}\left[1 + .03836\left(\frac{16.08 - 8.797}{1}\right)\right]^{-1/.03836}\right) = 1.66 \cdot 10^{-7} \ ,$$

Figure 52: Taxiway Centerline Deviation in Relation to Standards



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

FAA DG5 ICAO CODE E	ICAO CODE F	FAA DG 6

where the relevant numbers were taken from Figure 50 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. Figure 53 shows a comparison of the landing gear geometry for 747 and A380.

As was seen in equation (1) on page 14 this distance L impacts the amount of dampening that the main gear amplitude experiences relative to the nose gear amplitude. If we denote by $h(L)$ the dampened amplitude of the main gear as a function of L, i.e.,

$$h(L) = A' = A \Big/ \sqrt{\left(\frac{2\pi L}{TV}\right)^2 + 1}$$

we see that its derivative is

$$h'(L) = -h(L)\frac{L\,(2\pi/TV)^2}{(2\pi L/TV)^2 + 1}$$

and by a one term Taylor expansion none obtains

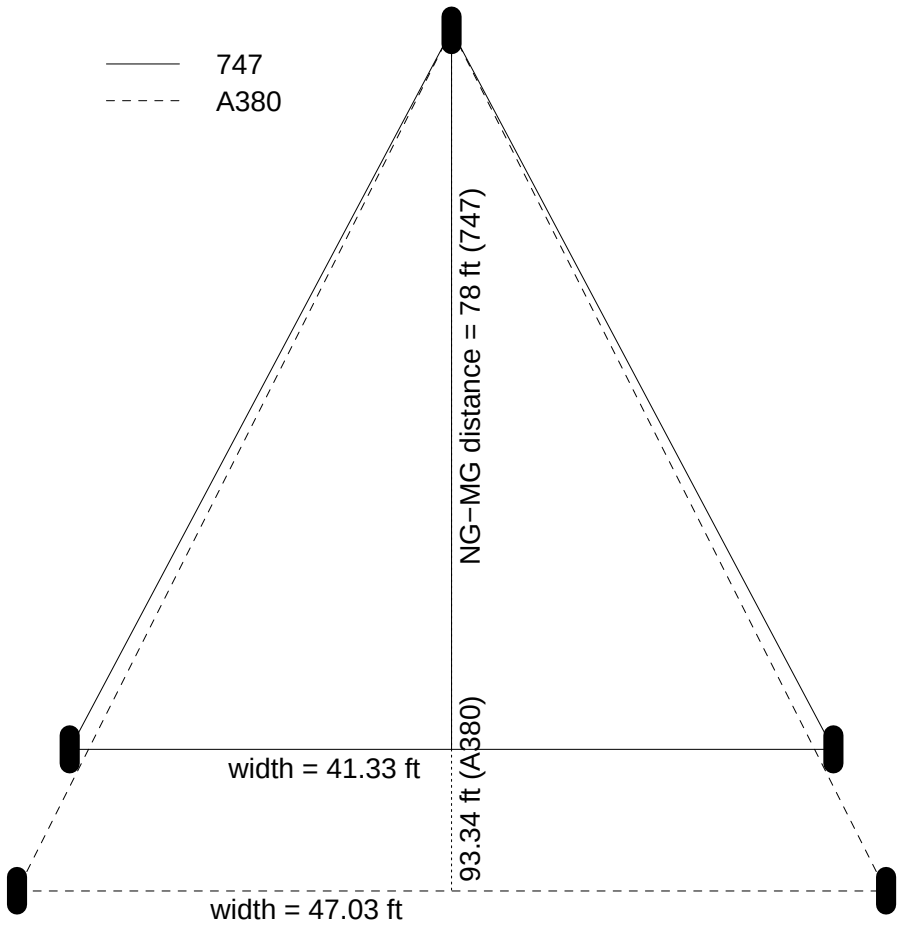
$$h(L_2) - h(L_1) \approx h'(L_1)(L_2 - L_1) = -h(L_1)(L_2 - L_1)\frac{L_1\,(2\pi/TV)^2}{(2\pi L_1/TV)^2 + 1}$$

or in terms of relative change in amplitude dampening

$$\frac{h(L_2) - h(L_1)}{h(L_1)} \approx -\frac{L_2 - L_1}{L_1}\frac{(2\pi L_1/TV)^2}{(2\pi L_1/TV)^2 + 1}\ .$$

Since $(L_2 - L_1)/L_1 = (93.34 - 78)/78 = .1967 \approx .2$ and since it seems reasonable to assume that $L_1 \gg TV$ (the distance traveled during one wave period is much larger than L_1) we can set $(2\pi L_1/TV)^2/[(2\pi L_1/TV)^2 + 1]$ to approximately 1 and conclude that the main gear amplitude dampening experienced by an A380 gear geometry is about 20% more than that experienced by a 747 gear geometry. This dampening refers to the path traversed by the main gear centroid and does not take into acount the outer-to-outer main gear width which is wider for the A380 geometry. Of course the stronger dampening effect for the A380 geometry does not reflect the whole picture since a sinusoidal path parallel to the taxiway centerline has two components, the parallel offset and the sinusoidal wave around the offset line. As

Figure 53: Landing Gear Geometry of 747 & A380



far as the parallel offset is concerned there is no dampening effect and only the outer-to-out main gear width comes into play. Thus a clear comparison between 747 and A380 gear geometry is not an easy proposition. One would need some understanding of how much the deviations reflect parallel offsets and how much they represent sinusoidal paths parallel to the centerline. So far there is no data to throw light on this issue. Similar considerations for the NLA were not possible since the corresponding value for L_3 was not available.

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	$1.66\ 10^{-7}$	$1.23\ 10^{-8}$	$4.13\ 10^{-11}$
	95% Upper Bound on Risk	$3.09\ 10^{-7}$	$2.50\ 10^{-8}$	$9.81\ 10^{-11}$
from A380	Estimated Risk	$1.69\ 10^{-6}$	$9.99\ 10^{-8}$	$2.19\ 10^{-10}$
	95% Upper Bound on Risk	$2.91\ 10^{-6}$	$1.90\ 10^{-7}$	$4.99\ 10^{-10}$
from NLA	Estimated Risk	$1.95\ 10^{-5}$	$8.85\ 10^{-7}$	$1.21\ 10^{-9}$
	95% Upper Bound on Risk	$3.05\ 10^{-5}$	$1.56\ 10^{-6}$	$2.63\ 10^{-9}$

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 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 2 from page 62 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 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.

8 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 (ANC) for which the data were collected. Although we found that the deviation data from the two taxiways (KILO and ROMEO) showed comparable behavior once we corrected for any biases one cannot make any conclusions concerning similar behavior once the common factors (ANC, 747, 75 ft wide straight segment with shoulders) 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 analysis of the JFK data, although this is still for 747s and for 75 ft wide straight segment 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 JFK is sufficiently similar to that at ANC. Thus one might hope for similar deviation behavior at 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.

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.

9 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 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 ANC with a different collection of $\approx 10,000$ 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.

10 Additional ANC Data

The data set analyzed here was collected by the FAA and is sometimes referred to as the Phase I data as opposed to another data set called the Phase II data. The latter was collected during subsequent time spans of 7/2/2001-7/2/2002 for ROMEO and 9/27/2001-7/2/2002 for KILO through the University of Alaska. This phase II data set captured 14318 events of a 747 passing the lasers. It is planned to analyze this data set as well and compare findings. If the data are reasonably comparable an analysis of the pooled data would also be feasible.

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. (1995), "Nonparametric Tail Extrapolation," *ISSTECH-95-014*, ISS Technology, Boeing Information & Support Services.

[9] Steiner, David (2003), personal communication.

[10] Taxiway Centerline Deviation System at Anchorage International Airport, Users Manual, September 2001.