ANALYSIS OF AIRCRAFT LATERAL PATH TRACKING ACCURACY AND ITS IMPLICATIONS FOR SEPARATION STANDARDS

Michael Cramer, The MITRE Corporation, McLean, VA Laura Rodriguez, The MITRE Corporation, McLean, VA

Abstract

Application of Area Navigation (RNAV) and Required Navigation Performance (RNP) to aircraft separation continues to be a point of difficulty which frequently centers on defining statistical models for Performance-Based Navigation (PBN). Given that questions of the statistics associated with position estimation error are fairly well known, and that path definition error is near zero in modern RNP avionics, this paper will focus on path steering error, which has not been studied using operational data. We propose analysis methods that can be applied to analyzing large volumes of recorded, in-service data to answer the questions of statistical modeling for the path following behavior of these systems. Preliminary analysis results derived from partner airlines' RNAV and RNP equipped aircraft flying in revenue service are presented.

Answers to key questions related to the Flight Management System's (FMS) Path Steering Error (PSE, also known as Flight Technical Error, or FTE) will be presented. This paper will characterize any variations that might exist in the distributions due to flight phase (climb, level or descent), flight path characteristics (turning or straight) and piloting mode, e.g., flight director or autopilot coupled with Lateral Navigation (LNAV). Results are presented as sample statistics, cumulative distributions, and Q-Q plots.

Introduction

In 2009 MITRE was tasked by the Federal Aviation Administration (FAA) to provide an overview of on-board performance monitoring and alerting capabilities in RNP RNAV systems. The resulting paper provided a description of the analytical methods used in such systems to provide performance alerting and it showed the relationship between the alerting capability and possible impacts on separation standards used with RNP RNAV systems [1]. As part of that paper, MITRE provided a review of some working papers from the Separation and Aviation Safety Panel (SASP-15) of the International Civil Aviation Organization (ICAO), from which we found that there were outstanding questions regarding the statistical behavior of the steering algorithms in RNP RNAV systems. In particular, the characterization of Path Steering Error (PSE, also known as cross track error internal to an RNP system) as defined in RTCA DO-236B as a statistical parameter when autoflight systems are engaged was an open question.

Considering that large volumes of flight data might be accessed from revenue service aircraft with RNP RNAV systems, MITRE began a program of working with various operators to attain such data for use in characterizing path steering performance. Initially only one aircraft type was recording the PSE directly out of the RNAV system (FMS), and we are currently receiving initial data from three operators of that aircraft type, Boeing 737. This paper will present the initial results of our analysis of this aircraft type. Other aircraft types will be added as data is received in the future, we are hoping to accumulate data on the full range of aircraft sizes (wide body, narrow body, RJ, business) to eventually complete the analysis to determine if one model might be extended to all the aircraft types when operated in an autoflight mode. Further, the analysis investigates the full range of operation by analyzing the six types of paths that can be defined by the intersection of lateral and vertical elements, I.e., lateral path (turning or straight) and vertical path (climb, level, descent).

We will be attempting to answer the following:

- 1. Are there significant differences between the PSE characteristics
 - a. Based on lateral segment type (turning or straight)?
 - b. Based on the vertical phase (climb, level, descent)?

- c. Based on location during the maneuver, e.g., beginning, middle or end of a turn?
- 2. What theoretical distribution (between Gaussian and Laplace) fits the data best and how well does it fit?
 - a. Can we compensate for the effect of data quantization on the sample statistics?
 - b. What measure should be used for goodness of fit?
- 3. What distribution would we propose in each case?

Data Preparation

In order to report meaningful results, it is necessary to understand the conditions under which the data were obtained. Factors like flight mode (Auto Pilot and Lateral Navigation coupled or Flight Director and Lateral Navigation coupled), flight phase (climb, level or descent), and flight path characteristics (turning or straight) may have an effect on the shape of the PSE probability distribution.

The analysis performed for this paper includes only data captured when the flight mode corresponds to Lateral Navigation (LNAV) and Auto Pilot coupled, so the system is flying in a fully automatic mode. The data was separated in the following groups:

- Turn Climb
- Turn Level
- Turn Descent
- Straight Climb
- Straight Level
- Straight Descent

Smoothing and change detection techniques together with knowledge of the physical limitations of aircraft maneuvers were employed to identify the lateral and vertical profiles of each flight. This identification process does not alter any of the PSE data used in the analysis; it simply tags the data according to the types of segments.

The behavior of the PSE (also referred to as xTrack in this paper), around flight mode changes was closely studied. In several cases it was observed that xTrack values were unrealistic immediately after both Auto Pilot and LNAV were engaged. Therefore, separation of the transient from the steady state was This requires knowledge of the specific needed. characteristics of the aircraft control systems, which was used to determine where path capture had occurred after engaging the both Auto Pilot and LNAV. Figure 1 shows a very common observed behavior of the xTrack during the transient state. The upper plot of the figure shows xTrack vs. time. The lower plot of the figure shows Flight Mode vs. time, where a value of 1 means that both Auto Pilot and LNAV were engaged, and a value of 0 means otherwise. The green points represent the ones actually used from this flight. From the figure, it could be seen that the applied filter successfully discards the points in the transient state while keeping the rest of the point where the flight mode is engaged.



Figure 1. Transient Behavior of PSE

Despite the effort to filter out undesired points (like the ones from the transient), there are still several cases where the xTrack shows a behavior yet to be fully understood. Figure 2 shows an example. The green points in the figure correspond to those where the flight mode was engaged. From the plot, it can be seen that the flight mode has been engaged for over two minutes during which the xTrack has been behaving as expected. However, suddenly, the

xTrack starts to increase, then jumps back to 0, then increases again until gradually decreasing to the expected values. Due to current reservations about how to interpret this type of behavior, no attempt was done to eliminate such points. The reader should bear this in mind when looking at the reported results, since this type of data will introduce outliers, which may or may not need to be included in the analysis. Until we are sure of the reason for the behavior, we are allowing the points to remain in the data set.



Figure 2. Unexpected behavior of xTrack

In addition to the programmatic script used to screen out transient points, a tool was developed to visually look directly into files that were contributing to large apparent errors and manually remove them if appropriate. See Figure 3 for an example of what was considered bad data. The points at the bottom of the xTrack graph are at full scale for the digital word that contains the PSE data, and obviously, neither the motion of the aircraft or the path would be likely to produce such jumps in the real data, so the file (data) was excluded from further analysis.



Figure 3. Example of bad data

Finally, to standardize the different segment types in time (given that they can be of arbitrary

length), we have divided each segment into 11 sections that represent the percentage of segment completion from 0% to 100% in steps of 10. See figure 4.



Figure 4. Segment Normalization

The "bins" for different segment lengths will likely contain differing numbers of samples. Only one sample is randomly selected from each bin for the statistical analysis. Furthermore, for segments that belong to the same flight, only the first one that took place in time is selected. In this way all the samples in each time bin for a given group (i.e. Straight Level) are assured to be independent. This forms the basis for the sample statistics, which we then compare.

Data Analysis

At this time, the available data is primarily from departures and arrivals leading to a wide range of sample sizes between the 6 groups of data, from only a couple hundred for level turns to nearly 20,000 for descending straights. See Table 1.

Group	Ν
Turn Climb	3290
Turn Level	205
Turn Descent	14010
Straight Climb	4316
Straight Level	2199
Straight Descent	17169

Table 1. Sample Size by Group

Therefore some results are more indicative than statistically significant, which will change as more data is collected. The results in this paper were captured from only one operators data (where PSE is

quantized at a level of 48 feet), and will be expanded to the others although their data is quantized at 96 feet.

The effect of quantization is dramatic when comparing data sample statistics to those from synthetically generated data and when trying to apply any standard tests for normality. Since PSE is likely to be Gaussian in nature due to the design of the control system itself, we have paid special attention to the Gaussian distribution. As part of our analysis, we sampled a number of Gaussian theoretical distributions with different values (within the same order of magnitude than those from the data) of standard deviation and zero mean. We then quantized the results and applied normality tests to the quantized data. It was observed that quantization values coarser than one tenth of the standard deviation led to failure of standard normality tests. This, unfortunately, rules out their use in this analysis.

There is however, a direct (and apparently linear) relationship between the scaling parameter of the theoretical distribution and the level of quantization. We derived the relationship by sampling various Gaussian distributions with zero mean and different standard deviations, (e.g., $\sigma = 1$, $\sigma = 100$, $\sigma = 200$) and then quantizing each at various levels.

We then, computed the sample standard deviation for each set of quantized data and plotted it against the quantization value.

Since the two most commonly assumed distributions for the PSE are the Gaussian and the Laplace, we experimented with both. We hope to use this linear model to propose a theoretical distribution and compute a goodness of fit parameter to give us an indication of which distribution is closer to the aircraft data behavior.

The first step in the comparison of the sample data sets will be done graphically through use of sample statistics, cumulative distributions from the sample data, and Q-Q plots. This will give a visual comparison of how different or similar the sample data sets seem to be. There are sixty-six data sets given the combination of lateral and vertical path (6 sets of data) and the time bins in the segment (11 subsets for each of the original six sets). The paper will not include all of the sets, but examples will be

shown and conclusions presented based on the complete sixty six sets.

The second step of comparison will be to derive theoretical distributions. Such derivation will be based on the quantization effect on the sample statistics per the above description, followed by computation of a goodness of fit parameter for each of the data sets. The analysis will proceed as follows for each set:

- 1. Compute the sample scaling parameter for the data set,
- 2. Use the linear model to estimate the scaling parameter corresponding to non-quantized data
- 3. Propose a theoretical distribution with the newly found scaling parameter
- 4. Sample the resulting theoretical distribution,
- 5. Quantize the theoretical distribution at 48 feet and create a histogram
- 6. Compute the goodness of fit between the sampled theoretical and the aircraft data set histograms

Once the goodness of fit is scaled appropriately, it will be used to select the best distribution for the data sets.

Initial Results

The mean and standard deviation from each group and time bin are summarized in Figures 5 and 6.



Figure 5. Data Mean Summary



Figure 6. Data Standard Deviation Summary

From the figures it can be observed that the mean for all groups is concentrated between -20 and 20 feet. While the standard deviation ranges between 50 and 160 feet.

The cumulative probability plots for each group are shown in Figures 7 and 8. Each color from the color bar represents a time bin. It can be observed from the plots how the tails differ among the different groups. Comparing segments with the same vertical profile, it can be seen that turning segments in a climb show a longer tail to the right (~3600 feet) compared to straight segments in a climb (~1200 feet). While turning leveled segments show a tail that is one order of magnitude less (~240 feet) than the straight leveled segments. It should be noted that the difference in the order of magnitude could be a result of the small amount of data (205 samples) that contributed to the statistics of this group (Turn Level). Lastly, descending turns show a longer tail to the left (~3600 feet) compared to straight descending segments (~1200 feet), except towards the end of the straight descent segment where both tails are practically the same.

A similar tail comparison could be done among segments with the same lateral profile. Straight climb, level and descent segments appear to have similar tails, except towards the end of the straight descents segments where a longer tail to the left is shown. Turn segments in a climb and turns in a descent exhibit similar tail values, but in the opposite directions. Because of this interesting behavior, the flights contributing to the tails of each group were identified and analyzed. It was observed that most of the flights from Turn Climb causing the long tail to the right were departures with left turns. While most of the flights from Turn Descent producing the long tail to the left were arrivals with right turns. Lastly, comparing segments from Turn Level to segments from Turn Climb and Descent, it could be seen that Turn Level tails a lot shorter. Again, this could be a result of the small number of samples contributing to the Turn Level group.

Overall, the longest tail observed is approximately 3600 feet.



Figure 7. Empirical Cumulative Distribution for Straight Segments



Figure 8. Empirical Cumulative Distribution for Turns

To gain more insight on how the groups compare to each other, Q-Q plots are used. The first sets of pairs were selected to study the effects of the lateral maneuvers. To this end, each pair had the same vertical profile and different lateral profile. This leads to three combinations: Turn Climb vs. Straight Climb, Turn Level vs. Straight Level and Turn Descent vs. Straight Descent.

For each lateral and vertical profile combination three pairs were selected corresponding to start, middle and end of the segment. For example, for the combination Turn Climb vs. Straight Climb, the following pairs are used:

- Turn climb at 0 % time bin vs. straight climb at 0 % time bin.
- Turn climb at 50 % time bin vs. straight climb at 50 % time bin.
- Turn climb at 100 % time bin vs. straight climb at 100 % time bin.

The complete set of pairs across all groups total nine. The Q-Q plots for the nine pairs are shown in Figure 9, 10 and 11.



Figure 9. Q-Q Plot for Turn vs. Straight Climbs



Figure 10. Q-Q Plot for Turn vs. Straight Levels



Figure 11. Q-Q Plot for Turn vs. Straight Descents

The points in the Q-Q plots for turn versus straight segments in a climb are clearly not on the 45° reference line (y=x). This is an indication that the underlying distribution for each population is different. However, the Q-Q plots for level turn and level straight segments show high proximity to the 45° reference line (magenta line). This behavior is an indication that each pair could probably belong to the same distribution. Lastly, the Q-Q plot for turn and straight descending segments at 50% and 100% exhibit strong departure from the reference line. While the O-O plot for the pair corresponding to 0% has more points on the reference line, there are still several others that are significantly far. Therefore, these pairs are likely to come from different distributions.

From the above analysis, it appears that the statistical behavior of the PSE is not affected by the lateral maneuvers during leveled segments. For this reason, the Turn Level and Straight Level groups are merged into one. As for the other groups, the statistics of the PSE seem be sensitive to lateral maneuvers when climbing and descending. Therefore, these groups stay separate, reducing the number of groups to 15 out of the original 18.

After analyzing the effects of the lateral maneuvers, we proceed to determine whether the vertical maneuvers affect the statistics of the PSE. A similar approach is followed where a number of comparisons of pairs with the same lateral but different vertical profile are performed. For example: Straight Climb at 0% time bin vs. Straight Descent at 0% time bin. Again, three time bins are selected 0%, 50% and 100% for each profile combination. The complete set of pairs across all combinations total 18. The Q-Q plots for the 18 pairs are shown in Figure 12 through 17.





Figure 13. Q-Q Plot for Turn Climb vs. Descents



Figure 14. Q-Q Plot for Straight Climb vs. Level Turn and Straight



Figure 15. Q-Q Figure X1. Q-Q Plot for Turn Climb vs. Level Turn and Straight



Figure 16. Q-Q Plot for Straight Descent vs. Level Turn and Straight



Figure 17. Q-Q Plot for Turn Descent vs. Level Turn and Straight

Upon analyzing the Q-Q plots above, it is determined that the samples from Descent Straight are likely to come from the same distribution as the samples from the already merged group Level Turn and Straight. Therefore, these two groups are merged into one. The rest of the groups stay separate since the current data does not support the hypothesis that they come from the same distribution.

The next step in this study is to find theoretical distributions to model each of the 12 resulting groups. As mentioned earlier the available path steering error data is quantized to approximately 48 feet. The effects of the quantization are such that attempting to directly fit a Gaussian or even a Laplace distribution to the sample data is impossible. However, before ruling out the possibility of modeling the path steering error using either of these two distributions, we investigate the effects of quantization on each.

The following experiment was designed to understand the effects of quantization on the standard deviation.

- 1. Set an initial standard deviation and mean
- 2. Generate random data from a Gaussian distribution with the initial values of standard deviation and mean
- 3. Quantize the data to different resolutions starting with a resolution of $1/10^{\text{th}}$ of the initial standard deviation and ending with a value equal to the initial standard deviation.
- 4. Compute the sample standard deviation of the quantized data at each of the resolutions.
- 5. Plot the sample standard deviation vs. resolution

For example, selecting an initial value for the standard deviation of 100 ft and a mean of 0 ft; then performing steps 2 through 5, results in the plot shown in Figure 18.



Figure 18. Standard Deviation vs. Quantization

After repeating this experiment for different initial values of standard deviation, it was found that there was always a linear relationship between the standard deviation and the quantization resolution. Furthermore, it was found that the slope of the best fitting line changed very little from case to case. The experiment was also conducted using Laplace distribution instead of Gaussian and similar results were found. Table 2 shows the values of the slope for each case.

Table 2. Slope of Linear Fit for StandardDeviation vs. Quantization

Initial	Slope (m _o)	Slope (m _o)
Standard	(Gaussian Data)	(Laplace Data)
Deviation		
1	-0.30566	-0.26579
100	-0.3125	-0.27242
200	-0.31573	-0.27233
300	-0.31422	-0.27209
400	-0.31226	-0.27274

This allows creating a linear model to find an approximate value of the standard deviation of a potential theoretical distribution based on the sample statistics. From Table 2, it is reasonable to select - 0.31 as the value for the slope (m_o) of our linear model for the Gaussian data, and 0.27 for Laplace data. Thus, having a value for the standard deviation from the data, the quantization resolution of the data, and the slope from the linear model allows us to solve for the intercept of that line. This will be our approximation of the true standard deviation of our quantized data. Figure 19 shows an example of the true standard deviation.

- Example:
 - $\sigma_{true} = \sigma_{data} m_o * quantization$
- Where
 - $\sigma_{data} = 74 ft$
 - $m_o = -0.31$
 - Quantization = 48 ft
 - $\sigma_{true} = 74 (-0.31) * 48 = 88.8 ft$

Figure 19. Example of Derivation of the True Standard Deviation

With the derived value of the true standard deviation, and a selected mean, a theoretical distribution could be proposed. It should be noted that, due to the small variation (within 40 feet) of the sample mean across all groups and the little impact it has on both the Q-Q plot and the Chi Square distance, the true mean was assumed to be zero for all proposed theoretical distributions.

To check whether the found distribution is a good fit, we sampled it, then quantized it to 48 feet, and compared it against the data. For comparing, we use Q-Q plots and the χ^2 (Chi Square) distance between the histograms. Figure 20 shows the Q-Q plot and the histogram of the data belonging to

Straight Climb group at 100% (green) and the synthetically generated Gaussian data quantized to 48 feet (blue). The sample standard deviation of the data is 74 feet, the derived true standard deviation for the Gaussian data is 89 feet, and the computed standard deviation of the quantized Gaussian data is 72 feet. The Q-Q plot in the figure shows the majority of the points on the 45° reference line. This is an indication of a good match at least in the core of the distribution. However, there are a few points that do not line up with the reference line. Currently, there are reservations as to whether these points could be outliers or indicators of legitimate heavy tails in the data.



Figure 20. Example of Theoretical Model

Two theoretical distributions were computed for each of the remaining 12 groups, one Gaussian and one Laplace. The χ^2 distance was computed in each case. Except for two groups, the model proposed for each case corresponds to the one where χ^2 is the smallest. The two exceptions correspond to group Straight Climb at 0% and group Straight Descent merged with Straight and Turn Level at 50%. In those cases the Q-Q plot looked more linear for the Laplace Model than for the Gaussian, giving initial indication that the Laplace was the better fit, while the corresponding χ^2 distance was smaller for Gaussian than for Laplace, giving the opposite indication. Table 3 contains a summary of χ^2 distance between histograms for each group. In the table, the numbers with one asterisk indicate the best fitting model (Gaussian or Laplace) based on the χ^2 distance, and the ones with two asterisks indicate the two that produce inconsistent results (between Q-Q plot and χ^2 distance). It should be noted that, in general, all Q-Q plots showed strong match to the reference line towards the core of the distribution. However, this was not the case towards the tails.

Table 3. χ² Distance for Each Group and Theoretical Model

Group	χ ² Distance	χ^2 Distance	
	Gaussian	Laplace	
	Model	Model	
Straight Climb 0%	1.38	1.82**	
Straight Climb 50%	2.73*	5.07	
Straight Climb 100%	1.51*	2.79	
Turn Climb 0%	1.64*	5.01	
Turn Climb 50%	3.81	3.19*	
Turn Climb 100%	5.9	4.51*	
Turn Descent 0%	10.79	4.76*	
Turn Descent 50%	7.47	4.09*	
Turn Descent 100%	4.52	2.97*	
0% Straight Descent and	3.42	1.15*	
Level Turn & Straight			
50% Straight Descent	1.1	1.46**	
and Level Turn &			
Straight			
100% Straight Descent	6.19	3.69*	
and Level Turn &			
Straight			

Finally, Table 4 summarizes the proposed best fit distribution type and standard deviation for each of the 12 groups. In the table σ_o represents the proposed theoretical standard deviation, σ_1 represents the standard deviation of the sampled and quantized proposed theoretical distribution, and σ_d represents the data sample standard deviation. All units in table 4 are in feet.

Table 4. Proposed Theoretical Distribution forEach Group

Group		PDF Type	σ	σ1	σ_{d}
Straight 0%	Climb	Laplace	143	129	130

Straight Climb	Gaussian	105	87	90
50%				
Straight Climb	Gaussian	89	72	74
100%				
Turn Climb 0%	Gaussian	101	84	86
Turn Climb 50%	Laplace	143	126	130
Turn Climb 100%	Laplace	173	154	160
Turn Descent 0%	Laplace		96	97
Turn Descent	Laplace	124	110	111
50%				
Turn Descent	Laplace	123	108	110
100%				
0% Straight	Laplace	83	67	70
Descent and				
Level Turn &				
Straight				
50% Straight	Laplace	62	49	49
Descent and				
Level Turn &				
Straight				
100% Straight	Laplace	81	67	68
Descent and				
Level Turn &				
Straight				

Conclusions

The authors wish to point out that any conclusions of this paper must be treated as preliminary, given the relatively low numbers of data samples for some flight segment types, and the limitation to only arrivals and departures. As we expand the amount of data we process, we expect results to become more consistent, particularly if investigation shows that the remaining "tail" behavior can be traced to system operation in a known transient condition, rather than a valid "normal" operational tail.

At this point, we feel that the following observations are consistent with the data so far:

- Based on the direction of the largest tails for climbing and descending turns, and the turn directions being flown in the arrivals and departures, it appears that turn direction could affect the PSE statistics in climbs or descents.
- 2) The analysis of the data using Q-Q and Chisquared tests leaves the question of the underlying theoretical distributions still open at this time, although the Laplace seemed to fit better in a majority of cases. (Table 3)
- 3) The sample statistics can provide some indication of overall performance (Figure 5), where straight level and descent, as well as turn level have similar standard deviations of around 50 feet. Straight climb and turn descent are similar at around 115', and turning climb is bounded by 165 feet standard deviation. That is not to say that we can use this for theoretical prediction of probability in the extreme tails, but it does give a bound on the normal behavior.
- 4) The fact that the 18 separate possible distributions reduced to 12 after comparison of the Q-Q plots may be an indication that more reduction may be found as the data set

increases in size and we understand more of the underlying operation related to outliers.

Work is on-going to expand the analysis to other aircraft types and sizes. We are currently working toward A380 and B777, B767 and B757, and Gulfstream. We are also working to accumulate more data from B737 to augment this papers result.

References

[1] Cramer, Michael, September 2009, On-board Performance Monitoring and Alerting (OPMA), McLean VA, MITRE Corporation

[2] Cramer, Michael, Sam Miller, Laura Rodriguez, June 2010, Performance Based Navigation (PBN) Assessments, On-board Performance Monitoring and Alerting, McLean VA, MITRE Corporation, MTR100179

[3] Cramer, Michael, Sam Miller, Laura Rodriguez, August 2011, Performance Based Navigation (PBN) Assessments; On-board Performance Monitoring and Alerting, McLean VA, MITRE Corporation, MTR110324

Acknowledgements

The authors would like to acknowledge the contributions of Sheila Conway, Michael Ulrey and Janna Young of the Boeing Company for providing a portion of the data and critique of our methods.

We would also like to thank Sam Miller of MITRE for his continuing efforts to expand the dataset to other aircraft types and other airlines.

Disclaimer

The contents of this material reflect the views of the author. Neither the Federal Aviation Administration nor the Department of Transportation makes any warranty, guarantee, or promise, either expressed or implied, concerning the content or accuracy of the views expressed herein. Approved for public release 13-1647.

2013 Integrated Communications Navigation and Surveillance (ICNS) Conference April 23-25, 2013