Analysis of Flight Technical Error on Straight, Final Approach Segments

Dr. Benjamin S. Levy, *The MITRE Corporation*
Dr. Pradip Som, *Federal Aviation Administration, AFS-420*
Dr. Richard Greenhaw, *Federal Aviation Administration, AFS-420*

BIOGRAPHY

Dr. Benjamin S. Levy is an analyst for The MITRE Corporation's Center for Advanced Aviation System Development (CAASD). He conducts modeling and simulation, statistical analyses, and algorithm development. His research interests include statistical pattern recognition. His Ph.D. was earned at The University of Maryland, College Park in Civil and Environmental Engineering.

Dr. Pradip Som specializes in Operations Research modeling of problems related to aviation/airlines/defense and works for the Federal Aviation Administration (FAA) Flight Procedure Standards Branch (AFS-420). His areas of interest include Markov modeling, descriptive and prescriptive stochastic modeling, statistical and risk analysis. His work experience includes statistical analysis and forecasting of passenger and equipment data, providing yield management solution, and stochastic modeling of overbooking phenomenon for American Airlines and US Airways. Dr. Som has a Ph.D. in Industrial Engineering and Operations Research from Texas A&M University.

Dr. Richard Greenhaw is an Operations Research analyst with the FAA AFS-420. There he develops and analyzes models for evaluating risk in the National Airspace. His research interests include risk modeling and statistical optimization. He holds a Ph.D. from Oklahoma State University in Mathematics.

ABSTRACT

The Flight Procedure Standards Branch (AFS-420) of the Federal Aviation Administration (FAA) is responsible for developing the criteria to apply to procedures designed and implemented in the National Airspace System (NAS). With the introduction of the Required Navigation Performance (RNP) capability of modern aircraft, development of criteria for the containment widths, or minimum amount of protected airspace needed, requires the accurate statistical analysis of the magnitude of position error for these RNP systems. The position error is represented by the Total System Error (TSE) which is a combination of the Flight Technical Error (FTE) and the Navigation System Error (NSE). The NSE is the error in position due to navigation such as Global Positioning System (GPS), Distance Measuring Equipment (DME)/DME, or Very High Frequency (VHF) Omnidirectional Range (VOR)/DME. FTE is the difference between the position estimated by the Flight Management System (FMS) and the desired aircraft position.

The magnitude of these errors depend upon whether the aircraft is turning, changing speed, flying straight and level, the autopilot mode (e.g., engaged) and navigation mode (e.g., lateral and/or vertical). This paper focuses only on the statistical analysis of FTE for aircraft flying straight, final approach segments.

One of the challenges to the statistical analysis of FTE is limited availability of empirical data that characterizes flights under appropriate conditions. The sample size can be increased by using data from along-track locations on a flight if the errors between locations are statistically independent. Consecutive observations of FTE are, however, highly correlated because the position error data are used by the FMS to control and reduce future position errors. Each flight in the data set can be sampled at intervals at which the errors are statistically independent, to form a plot of lateral versus vertical errors (a billboard). Statistical models are developed based on the billboard data with probability density functions (pdfs). The normal, three-parameter gamma, and Johnson curve pdfs were fitted to the lateral and vertical error data from the final approach segment of the approach.

The decision to fit pdfs to marginal data depends on the demonstration of independence between the lateral and vertical errors. If the non-marginal data can be used, there will be more information available for fitting the pdfs. An important question is how to accurately determine whether the lateral and vertical errors are cross-correlated,

given that there may be some autocorrelation in the lateral and vertical errors.

The autocorrelation in the FTE reduces the number of independent degrees of freedom used in testing the strength of cross-correlation between the lateral and vertical error components of FTE. Failure to consider non-zero autocorrelation in the lateral and vertical errors causes a higher-than-expected Type I error when testing for independence between the lateral and vertical errors. It will be more likely to conclude that there is a relation between the lateral and vertical errors when in fact there is no correlation. Another consequence is that the test on the cross-correlation will have more statistical power than warranted in the detection of a false null hypothesis.

Monte Carlo simulation was used to investigate the effects of non-zero autocorrelation in the lateral and vertical errors on the Pearson's r test statistic for cross-correlation. Sets of random variables of lateral error, y, and vertical error, z, were created which contain specified amounts of autocorrelation (lag-1). Critical values as a function of Type I error, sample size, and ρ_y and ρ_z (i.e., the population lag-1 autocorrelation in the lateral and vertical errors, respectively) were generated that have the correct Type I error. Accurate polynomial estimating equations for the critical values were developed.

Multivariate, lag-1 Markov equations for the generation of random variables containing cross-correlation and lag-1 autocorrelation were derived such that the random variation is correctly partitioned between cross-correlation and autocorrelation. These equations were used to assess the power of the Pearson's r test for cross-correlation between autocorrelated lateral and vertical errors. The multivariate equations are applied to lateral and vertical error and can be used to simulate FTE on complicated routes (e.g., curved, descending). For example, an airplane might be "high" and "right of course" on a descending, curved section of a route. Finally, the accuracy of the algorithm for selecting the Johnson curves for pdf-fitting was evaluated. Johnson curves are four-parameter pdfs that are often used to characterize skewed, tail-heavy data such as FTE.

INTRODUCTION

The AFS-420 of the FAA is developing containment volume criteria for approaches based on RNP equipped aircraft for appropriately trained aircrews. For the current design of these approaches, known as Special Aircrew and Aircraft Approach Requirements (SAAARs), the RNP value changes along the approach, reaching a minimum of 0.3 RNP.

AFS-420 is developing a model of the containment volume with variable physical dimensions to support the process of criteria development. The model will be used to evaluate the likelihood that an aircraft remains within a containment volume (RTCA, 2002) based on lateral, vertical, and along-track position error data.

An important aspect of the containment modeling is the nature of the random variation in TSE, NSE, and FTE. The frequency and magnitude of these errors, measured as the aircraft moves along its route, influences the likelihood estimate that the aircraft remains in the containment volume for a prescribed confidence interval. The errors are measured in three-dimensional space such that there is an along-track error, a horizontal position error taken perpendicular to the track heading (i.e., the lateral error), and a vertical error taken perpendicular to the altitude profile at a position on the route. The FTE for a set of B747 simulations were analyzed, where the FTE is the difference between the indicated position and desired route (RTCA, 2002). The NSE is the difference between the actual and defined routes and the TSE is a combination of the FTE and NSE.

The statistical analysis of the FTE provides important information for the containment surface modeling because FTE is typically larger than NSE. Therefore, it is necessary to accurately characterize the probability density function and the parameter estimates for the FTE. Inappropriate analysis can lead to the overestimation of the required containment surface dimensions, which is inefficient. The underestimation of containment surface dimensions given a stipulated probability causes unsafe conditions.

The statistical analysis is more accurate with an adequate sample size, the correct use of analysis techniques, and suitable probability density functions (pdfs), such as Johnson curves. The Johnson curves can be used to model FTE data, which typically are skewed and which have more mass under the distribution tails than a normal distribution. The selection of the Johnson curve is guided by an algorithm (Hill et al., 1976) based on the sample estimates from the data set. These estimates become more accurate given larger sample sizes. The sample size for analysis can be increased if the lateral and vertical data are uncorrelated (i.e., cross-correlation) and the use of marginal data can be avoided.

The goal of this research is to improve the accuracy of the statistical analysis of FTE data by: 1) applying different pdf models; and, 2) assessing the impact on cross-correlation tests due to autocorrelation in the lateral and vertical FTE. The objectives to meet these goals are: 1) creation of a set of critical values to test the significance of cross-correlation between the lateral and vertical FTE given the presence of autocorrelation in the FTE such that the Type I error is correct; 2) development of polynomial estimating equations for the newly-developed non-zero

autocorrelation critical values; 3) an assessment of the accuracy of the algorithms by which the Johnson curves are selected; and, 4) the derivation and implementation of multivariate, first-order Markov equations for statistical power simulation of the test on cross-correlation.

ANALYSIS METHODOLOGY

The goal of the data analysis of the lateral and vertical flight error data is the ability to infer the probability that an aircraft remains within the lateral and vertical dimensions of the containment volume. The focus of this work is on accuracy, rather than on navigation integrity, which stipulates an hourly failure probability rate. The probability of exceeding a lateral or vertical distance from the route centerline is obtained from a pdf fitted to the error data. Similarly, the containment probability (e.g., $p = 10^{-5}$) can be stipulated, such that the lateral and vertical distances are obtained from the respective error pdfs. Figure 1 shows the overall analysis approach adopted in this work.

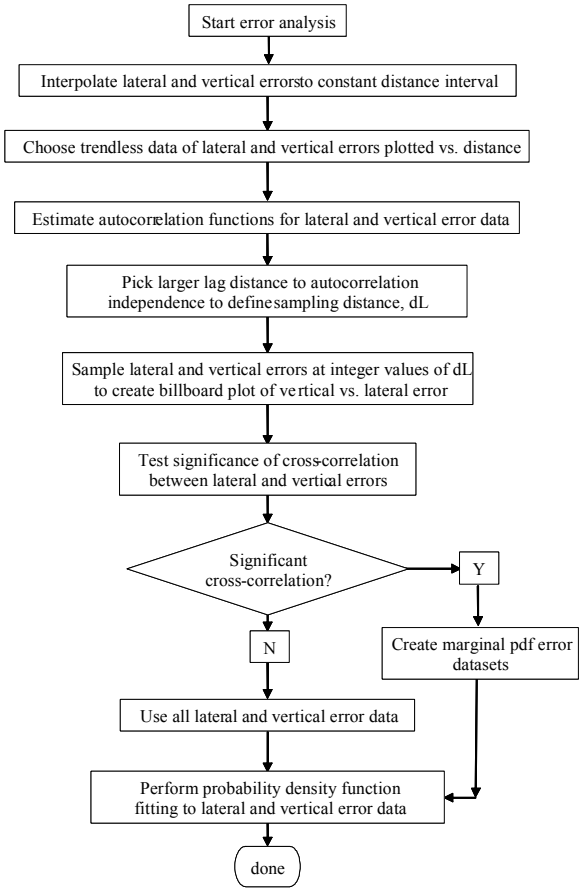

Figure 1. Error Analysis Flowchart

The analysis begins by interpolating the lateral and vertical error data to a constant distance interval on the along-track position data for computation and analysis of the autocorrelation in the lateral and vertical error data. For the error analysis on a constant-heading, steady-descent section of flight, sequences of lateral and vertical error data were selected based on the absence of trends in the data.

Recursive use of lateral and vertical data by the FMS causes autocorrelation in the error data. The sample autocorrelation function (Haan, 1977) was estimated from Equation 1:

$$r_k = \frac{\sum_{i=1}^{n-k} x_i x_{i+k} - \frac{1}{n-k}\sum_{i=1}^{n-k} x_i \sum_{i=k+1}^{n} x_i}{\sqrt{\sum_{i=1}^{n-k} x_i^2 - \frac{1}{n-k}\left(\sum_{i=1}^{n-k} x_i\right)^2}\sqrt{\sum_{i=k+1}^{n} x_i^2 - \frac{1}{n-k}\left(\sum_{i=k+1}^{n} x_i\right)^2}} \quad (1)$$

The implication of the presence of autocorrelation in the lateral and vertical error data is that statistical tests on the correlation between lateral and vertical error which assume data independence will be inaccurately applied. The rectification of this issue is subsequently examined.

The significance of the autocorrelation at the k^{th} lag from Equation 1 is tested to determine the distance at which lateral and vertical error data can be sampled to construct a composite plot of lateral and vertical error data. The null hypothesis that the k^{th} lag autocorrelation value, ρ_k, is not statistically different from zero is given in Equation 2:

$$H_0 : \rho_k = 0 \quad (2)$$

The null hypothesis in Equation 2 is evaluated with the critical values (i.e., one-tailed test) for some Type I error level, α, (Hipel and McLeod, 1994) for the k^{th} lag with Equation 3:

$$r_{\alpha,k} = N^{-1}(\alpha,0,1)\sqrt{\frac{1}{n}\left(1+2\sum_{i=1}^{k} r_i^2\right)} \quad (3)$$

where k = 1..int(n/4) number of lag values and $N^{-1}(\alpha,0,1)$ is the standard normal variate from the inverse cumulative normal distribution at α. The null hypothesis at the k^{th} lag is rejected if $r_k > r_{\alpha,k}$. Testing the autocorrelation in the lateral and vertical errors is performed and the larger distance to statistical independence is the interval, dL, with which the lateral and vertical data are sampled to create the data billboard (Figure 2).

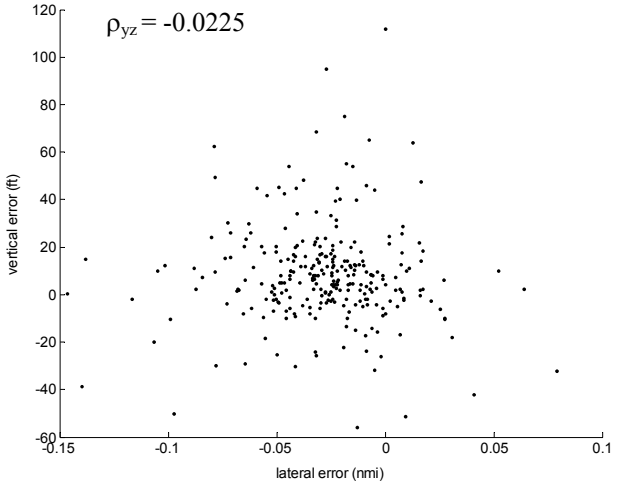

Figure 2. Billboard of Composited Lateral and Vertical Error Data (n = 275)

Given that the lateral and vertical error data from a set of flights have been sampled to create the data billboard of Figure 2, pdfs are fitted to the data and are evaluated. Referring to Figure 1, an important step in the analysis is the assessment of the cross-correlation between the lateral and vertical error data. If the data are found to have a significant cross-correlation, then data sets must be created within discrete ranges of the lateral and vertical error. These restricted datasets are smaller, but control for the presence of statistically significant cross-correlation between the lateral and vertical error, thereby allowing the fitting of marginal pdfs to the data. The implications are poorer pdf fits because of the smaller data sets and the arbitrary construction of discrete data ranges. If the cross-correlation is statistically insignificant, then all the error data may be used in the respective pdf analyses.

The assessment of the cross-correlation strength between the lateral and vertical error data assumes that the data are independent within the lateral and vertical error sequences. Even though the data are sampled at a distance to achieve a statistically independent sequence, correlation between observations for a given flight still exists and may impact the assessment of the cross-correlation by reducing the number of degrees of freedom used in selecting the cross-correlation critical value.

The effect of correlation within the lateral and vertical error data (i.e., autocorrelation) on the cross-correlation inference was studied. Examination of the autocorrelation functions for the sequential lateral and vertical error data suggested that a lag-1 autocorrelated sequence could be used to model the data (see Equation 4):

$$y_i = \rho_y \left(y_{i-1} - \mu_y \right) + \mu_y + u_i \sigma_y \sqrt{1 - \rho_y^2} \quad (4)$$

where y_i is a random variable at its i^{th} observation, ρ_y is the lag-1 autocorrelation between y_i and y_{i-1}, μ_y and σ_y are the mean and standard deviation of y, respectively, and u_i is a standard normal random variate (Haan, 1977).

Given the assessment of the significance of the cross-correlation between the lateral and vertical error data and the data set selection, pdfs were fitted to the data. In this work, the normal distribution, Pearson Type III and Johnson curves were fitted to the data. The choice of Johnson curve, and the accuracy of the algorithm by which the choice is made, is addressed in the subsequent section of this paper.

EVALUATION OF CROSS-CORRELATION IN AUTOCORRELATED LATERAL AND VERTICAL ERROR DATA

Testing of the significance of cross-correlation between the lateral and vertical error is used to evaluate the suitability of the marginal error data for the pdf analyses. If a statistically significant cross-correlation is inferred between the lateral and vertical errors, the pdf analysis is based on the marginal data of the lateral and vertical errors. Absence of a statistically significant cross-correlation allows the analysis of the non-marginal error data set, which has a larger number of observations. Use of the larger non-marginal error data produces a more accurate identification of the pdf and its parameter values. The subjective selection of the marginal data within a range is another source of inaccuracy.

Pearson's r is used to evaluate the strength of cross-correlation between the lateral and vertical errors. The test statistic is given by Equation 5:

$$r_{yz} = \frac{\sum_{i=1}^{n} z_i y_i - \frac{1}{n} \sum_{i=1}^{n} z_i \sum_{i=1}^{n} y_i}{\sqrt{\sum_{i=1}^{n} z_i^2 - \frac{1}{n} \left(\sum_{i=1}^{n} z_i \right)^2} \sqrt{\sum_{i=1}^{n} y_i^2 - \frac{1}{n} \left(\sum_{i=1}^{n} y_i \right)^2}} \quad (5)$$

where r_{yz} is the correlation coefficient between the lateral error, y_i, and the vertical error, z_i, and n is the number of observations in a sample. The null hypothesis (Equation 6) of the statistical test is no statistically significant cross-correlation between y_i and z_i:

$$H_0 : \rho_{yz} = 0 \quad (6)$$

The alternative hypothesis (Equation 7) is significant cross-correlation between y_i and z_i:

$$H_a : \rho_{yz} \neq 0 \quad (7)$$

The null hypothesis is rejected in favor of the alternative hypothesis if the magnitude of the test statistic, $|r_{yz}|$, is larger than the magnitude of the critical value, $|r_{\alpha/2}|$.

Accurate inference of cross-correlation between the lateral and vertical depends on the appropriate use of a statistical test. Failure to adhere to or to acknowledge the assumptions upon which the statistical test is based may cause inaccurate decisions to be made regarding the nature of the cross-correlation. Significantly large autocorrelation in the FTE reduces the number of independent degrees of freedom available for testing the strength of the cross-correlation between the FTE components. Consequently, the decision to reject the null hypothesis of no cross-correlation will be more likely incorrect.

In a hypothetical example, the sample test statistic, r, for Pearson's correlation is calculated as 0.3 for a sample size of 50. Given a Type I error rate of ten percent and the critical value of $|r_{\alpha}| = 0.235$, the decision is to reject the null hypothesis (Equation 6) of no significant cross-correlation in favor of the alternative hypothesis (Equation 7).

The decision to reject the null hypothesis of independence between the lateral and vertical errors is based on the assumption of independence within the error observations. Absence of independence within the lateral and vertical errors is expected because the control laws in the FMS algorithms use past error data in position calculations. The use of past data correlates the successive error measurements (i.e., autocorrelation) and reduces the amount of independent information contained in the data (i.e., smaller effective sample size). The critical value for testing the cross-correlation must be selected on the basis of the smaller effective sample size so that the actual Type I error rate accurately reflects the stipulated Type I error rate for which the critical value is chosen.

Returning to the example developed above, the error data from the FMS may contain autocorrelation between successive observations (i.e., lag-1 autocorrelation). If the autocorrelation is 0.6 for the lateral and vertical errors, then the appropriate cross-correlation critical value is 0.327 for a ten percent Type I rate and a sample size of 50. The critical value was developed by simulation for the presence of non-zero autocorrelation in the error data such that the Type I error rate was preserved. In this instance, the decision would be to fail to reject the null hypothesis of significant cross-correlation (i.e., $|r_{yz}| = 0.3 \leq |r_{\alpha/2}| = 0.327$). Use of the cross-correlation value of $|r_{\alpha/2}| = 0.235$ produced a two percent Type I error rate, instead of the ten percent Type I error rate expected under the condition of non-zero autocorrelation. The implication of using the critical value given non-zero autocorrelation is that the larger set of non-marginal data can be used for the pdf identification and analysis.

DEVELOPMENT OF CROSS-CORRELATION CRITICAL VALUES GIVEN NON-ZERO AUTOCORRELATION

The preceding section identified the need for cross-correlation critical values that preserve the stipulated Type I error rate under the conditions of non-zero autocorrelation, (Objective 1 in the final paragraph of the Introduction). Non-zero autocorrelation in the lateral and vertical error data reduces the independent degrees of freedom for testing the significance of the strength of the cross-correlation between the lateral and vertical error data. This problem can be either resolved by the identification of the effective number of independent degrees of freedom or by simulation of the critical values given non-zero autocorrelation and sample size such that the correct Type I decision error is preserved.

Monte Carlo simulation was used to create a set of new critical values, r_{α}, that meet these requirements. Standard normal random variates were created with the polar method (Press et al., 1992). In the Monte Carlo simulation, the test statistic, r_{yz}, was repeatedly calculated from the random variates z_i and y_i drawn from standard normal distributions (i.e., mean = 0, standard deviation = 1). The sample size, n, and lag-1 population autocorrelation in the vertical error (ρ_z) and in the lateral error (ρ_y) were specified. A cumulative frequency histogram of the test statistic was created given 3×10^6 Monte Carlo repetitions for each set of simulation conditions (i.e., sample size, Type I error rate, autocorrelation values). The critical values were taken from the cumulative frequency histogram at specified Type I error rate values.

Critical values were estimated for sample size values of 25 to 105, step 5 and from 150 to 500, step 50. The Type I error rate values for which the critical values were taken were 0.9, 0.95, 0.975, 0.99, and 0.995. The autocorrelation values were 0, 0.2, 0.4, 0.6, 0.8, and 0.9. As a validation check, simulated critical values for zero autocorrelation in z_i and in y_i were compared to the published critical values as a function of sample size and Type I error rate and were found to be accurate to $\leq 10^{-3}$.

The permutation of the simulation parameter values created a large number of critical values. A fourth-order polynomial equation was fitted to the new set of critical values to facilitate usage and to remove or reduce reliance on the extensive set of tabulated critical values (Objective 2 in the final paragraph of the Introduction). The polynomial equation accurately estimates the critical values, r_{α}, as a function of sample size, n, and the lag-1 autocorrelation in the y_i and z_i random variables (i.e., ρ_y

and ρ_z) for a given Type I error (i.e., α) value. The polynomial estimating equation is given in Equation 8:

$$r_\alpha = r_{\alpha 0} + \left(\sum_{i=1}^{3} b_i \rho_y^i \right) \left(\sum_{i=1}^{3} b_i \rho_z^i \right) \quad (8)$$

where $r_{\alpha 0}$ is the critical value at $\rho_y = \rho_z = 0$ given some Type I error level, α, and sample size, n. Equations for estimating $r_{\alpha 0}$ and the remaining polynomial coefficients were determined graphically as functions of sample size. The $r_{\alpha 0}$ term was estimated by Equation 9:

$$r_{\alpha 0} = a_0 n^{a_1} \quad (9)$$

The coefficients b_1, b_2, and b_3 were estimated by Equations 10, 11, and 12:

$$b_1 = c_1 + \frac{c_2 c_3}{c_3 + n^2} \quad (10)$$

$$b_2 = \frac{d_1}{n} + \frac{d_2 d_3}{d_3 + n} + d_4 \quad (11)$$

$$b_3 = \frac{e_1}{n} + \frac{e_2 e_3}{e_3 + n} + e_4 \quad (12)$$

Table 1 gives the estimates of the polynomial coefficients for Type I error rates of 0.9, 0.95, 0.975, 0.99, and 0.995. Table 1 also gives the goodness-of-fit statistics for each polynomial approximation of the critical values at each Type I error rate. The polynomial coefficient values were found by minimization of the squared sum of errors between the observed and predicted critical values, for 901 critical values.

Referring to Table 1, the standard error of the estimate, S_e, is less than 0.005. The relative error (S_e/S_y) measures are less than 3.5 percent, further indicating that the polynomial estimating equations are very accurate. The maximum error between the simulated critical values and the corresponding values estimated from the polynomial equations was 0.015.

An example calculation with the polynomial estimating equations is given. Following the example given in the preceding section, the critical value is estimated for a sample size of 50, a ten percent Type I error rate, and lateral and vertical autocorrelation values of 0.6. From Table 1 at 1-α = 0.95 (upper tail), the coefficients a_0 = 1.7410 and a_1 = -0.5090 are used in Equation 9 to estimate $r_{\alpha 0}$ = 0.2373, which is very close to the tabulated value of 0.237. With the remaining coefficient values in the same column of Table 1, Equation 10 through Equation 12 were used to estimate b_1 = 0.34861, b_2 = 0.03786, and b_3 = 0.30302, respectively. Given the values of b_1, b_2, b_3, and $r_{\alpha 0}$, the polynomial estimate of r_α is 0.321 from Equation 6 which is close to the simulated value of 0.327.

Table 1. Polynomial Coefficient Values for Critical Value Estimation, with Goodness-of-Fit Statistics

Type I Error Level	α	0.9	0.95	0.975	0.99	0.995
Standard error of estimate	S_e	0.0032	0.0036	0.0039	0.0042	0.0043
Critical value standard deviation	S_y	0.0910	0.1130	0.1280	0.1440	0.1533
Relative error	S_e/S_y	0.0346	0.0314	0.0301	0.0288	0.0277
Polynomial Coefficient Values	a_0	1.3890	1.7410	2.0280	2.3330	2.5220
	a_1	-0.5140	-0.5090	-0.5050	-0.5000	-0.4960
	c_1	0.3050	0.3440	0.3760	0.4140	0.4380
	c_2	-4.4700	-3.7390	-10.4320	-9.1810	-11.3310
	c_3	-4.9780	-3.0380	1.1380	5.4880	7.3570
	d_1	8.8740	11.9320	17.0700	22.4270	26.0940
	d_2	0.5710	0.6400	0.6120	0.6440	0.6700
	d_3	109.4640	109.3790	139.4670	148.0110	151.0340
	d_4	-0.5780	-0.6410	-0.7040	-0.7780	-0.8220
	e_1	-20.2870	-35.1210	-46.3930	-67.8740	-100.4550
	e_2	-12.4560	-11.6080	-16.3490	-16.0540	-30.2750
	e_3	-0.4350	-1.2860	-1.2980	-2.1610	-2.0220
	e_4	0.6280	0.6990	0.7480	0.8170	0.8530

FITTING OF JOHNSON CURVES TO LATERAL AND VERTICAL ERROR DATA

The objective of the analysis is the characterization of the statistical distributions of the lateral and vertical error data. The identification of the pdf and the estimates of its parameters is performed on the lateral and vertical error data after the assessment of the strength of the cross-correlation and the data set selection (see Figure 1).

The Johnson curves (Johnson, 1965; FAA, 2003) are often applied to FTE and TSE data. The preference for the Johnson curves over the normal pdf is based on the ability to characterize data that are skewed and that have more mass in the extremes (i.e., thick tailed) relative to the normal pdf. The Johnson curves are based on transformed normal distributions and have three major types: the SL, SB, and SU curves.

A random variable from the Johnson SL (i.e., log-normal) distribution, x_{SL}, is related to a standard normal variate, x_{SNV}, by Equation 13:

$$x_{SNV} = \gamma + \eta \ln\left(\frac{x_{SL} - \delta}{\lambda}\right) \tag{13}$$

where γ, η, ε, and λ are the parameters of the Johnson SL curve to be estimated (Hahn and Shapiro, 1994). The γ, η, δ, and λ parameters are estimated by an algorithm developed by (Hill et al., 1976). The parameter ranges for the Johnson SL curve are $x_{SL} \geq \delta$, $\eta > 0$, $-\infty < \gamma < \infty$, $\lambda > 0$, and $-\infty < x_{SL} < \infty$ (Hahn and Shapiro, 1994).
The Johnson SB curve is given by Equation 14:

$$x_{SNV} = \gamma + \eta \ln\left(\frac{x_{SB} - \delta}{\delta + \lambda - x_{SB}}\right) \tag{14}$$

for the range of $\delta \leq x_{SB} \leq \delta + \lambda$ for a Johnson SB distributed random variable, x_{SB}.

Finally, the Johnson SU curve is given by Equation 15:

$$x_{SNV} = \gamma + \eta \sinh^{-1}\left(\frac{x_{SU} - \delta}{\lambda}\right) \tag{15}$$

for the range of $-\infty < x_{SU} < \infty$ for a Johnson SU distributed random variable, x_{SU}.

The results of fitting the normal, three-parameter gamma, and Johnson curve distributions to the billboard of lateral error data (see Figure 2) are shown in Figure 3. Referring to Figure 3, it can be seen that the normal distribution was fitted to the lateral error data with $\mu \approx -0.028$ nmi and $\sigma \approx 0.03$ nmi for 275 data points. The three-parameter gamma pdf (i.e., Pearson Type III) was also fitted to the lateral error data with standardized skew, $\gamma_s \approx -0.516$. Finally, the Johnson SU curve was selected and fitted to the data with $\beta_1 \approx 0.243$ and $\beta_2 \approx 5.107$, where β_1 and β_2 are the relative skew and relative kurtosis estimates, respectively, that are used to select the Johnson curve type are given in Equation 16 and 17:

$$\beta_1 \approx \left(\frac{n}{(n-1)(n-2)}\right)\left(\frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^3}{s_x^3}\right) \tag{16}$$

and

$$\beta_2 \approx \left(\frac{n}{s_x^4 (n-1)(n-2)(n-3)}\right)\times\ldots \times\left((n+1)\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^4 - 3\left(s_x^4 \frac{(n-1)^3}{n}\right)\right) \tag{17}$$

where s_x is the sample standard deviation of a random variable x_i of size n.

Johnson curves allow the characterization of skewed data with "thick tails". Referring to Figure 3, it can be seen that the normal distribution (dotted curve) underpredicts the frequencies in the tails of the data histogram. By contrast, the Johnson SU curve is a better match (solid line) to the data histogram.

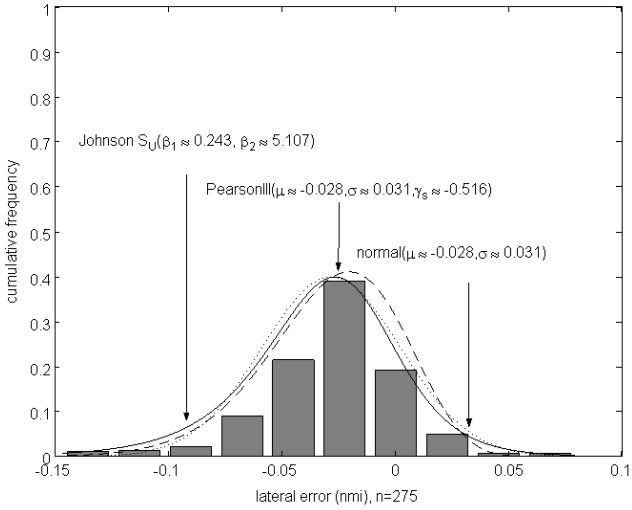

Figure 3. Lateral Error Data and PDF Fits

The implication of relying on the normal pdf to characterize the error data, instead of using the Johnson curves, is an underprediction of the probability of transgressing from the containment surface. For example, in the data of Figure 3, the distances estimated from the

normal pdf at a 10^{-5} probability are -0.161 and 0.105 nmi. These distances correspond to a 10^{-3} exceedence probability according to the Johnson SU pdf. In this example, the reliance on the distances derived from the normal pdf at a 10^{-5} probability would underestimate the true risk of transgressing the containment surface by a hundredfold factor.

EVALUATION OF THE ACCURACY OF THE JOHNSON CURVE SELECTION ALGORITHM

The selection of one of the Johnson curves (i.e., SL, SB, or SU) is based on the sample estimates of the relative skew, β_1, and the relative kurtosis, β_2, with the β_1-β_2 surface of Figure 4. Figure 4 is redrawn and adapted from Hahn and Shapiro (1994) and the algorithms in Hill et al. (1976). The procedure for the graphical curve selection entails finding the region in Figure 4 where the sample estimates of β_1 and β_2 plot. For example, if $\beta_1 \approx 1$ and $\beta_2 \approx 6$, the Johnson SU curve is indicated. The SB curve is selected if $\beta_1 \approx 2$ and $\beta_2 \approx 4$. The line separating the SU region from the SB region describes data characterized by the SL (i.e., log-normal) distribution.

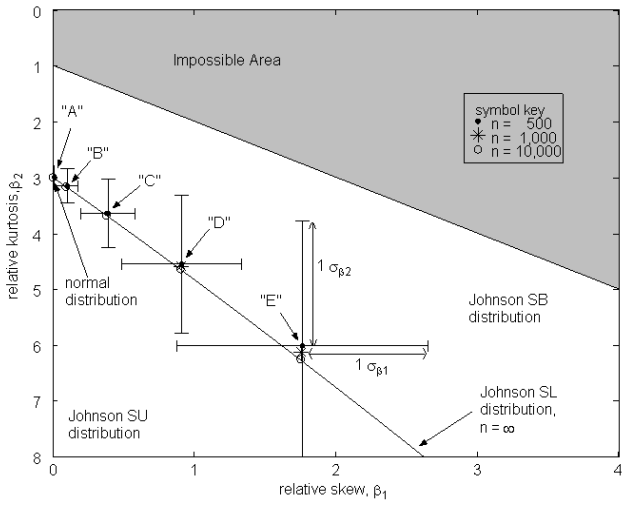

Figure 4. The β_1- β_2 Surface for Johnson Curve Selection

A logical question with regard to the use of Figure 4 is: "How far from the SL line can a sample distribution be, given the sample estimates of β_1 and β_2, and still be considered characterized by the SL distribution instead of another curve?" The answer to this question begins with the evaluation of the accuracy of the Johnson curve selection procedure given by Hill et al. (1976) (Objective 3 in the final paragraph of the Introduction). This algorithm provides for the selection and parameter estimation of the Johnson curves from data based on sample estimates of β_1 and β_2. Figure 5 shows the curve selection logic from Hill et al. (1976).

Johnson curve selection occurs at the decision node in Figure 5 (Hill et al., 1976):

$$\left|U - \beta_2\right| < \varepsilon \tag{18}$$

where U is defined by:

$$U = W^2\left(3 + W\left(2 + W\right)\right) - 3 \tag{19}$$

and

$$W = \sum_{i=-1}^{1} \left(-1\right)^{|i|-1} Z^i \tag{20}$$

and

$$Z = \left\{ \frac{\beta_1}{2} + 1 + \sqrt{\beta_1\left(\frac{\beta_1}{4} + 1\right)} \right\}^{\frac{1}{3}} \tag{21}$$

If Equation 18 is true, the SL distribution is selected. If $U - B_2 > \varepsilon$, then the SB distribution is selected; otherwise, the SU distribution is chosen. An important decision parameter in the selection procedure of Figure 5 is the scalar quantity ε, which is a tolerance set at 10^{-2} in the algorithm.

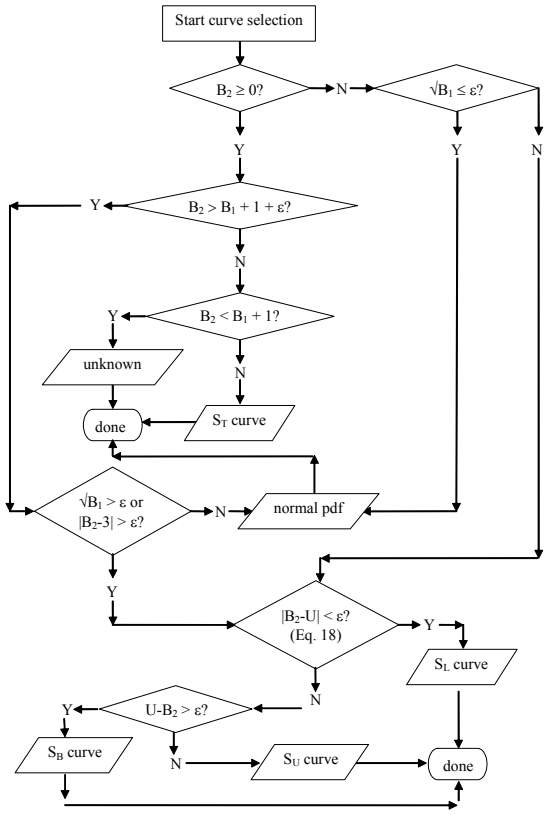

Figure 5. Procedure for Choosing Johnson Curves

A simulation study assessed the ability of the Hill et al. (1976) algorithm to correctly identify a data set as Johnson SL distributed, given that the data created from a SL distribution. Monte Carlo simulation was used to count the number of correct and incorrect identifications of data sets of size 500, 10^3, and 10^4 drawn from the SL distribution. For each simulation experiment, $3x10^4$ realizations were used. The tolerance, ε, was varied to inspect its effect on the accuracy (i.e., $\varepsilon = \{0.01, 0.05, 0.1, 0.5\}$).

Johnson SL distributed random variates, x_{SL}, were generated by Equation 22:

$$x_{SL} = \exp\left(\sigma_L x_{SNV} + \mu_L\right) \tag{22}$$

where x_{SNV} is a random variate drawn from N(0,1) and μ_L and σ_L are the population mean and standard deviation, respectively, for the log-normal (or SL) parent distribution. Equation 13 degenerates to Equation 22 for $\gamma = 0$, $\eta = 1/\sigma_L$, $\mu_L = 0$, $\delta = 0$, $\lambda = 1$. In this simulation study, $\mu_L = 0$ and σ_L was varied (i.e., $\sigma_L = \{0.01, 0.1, 0.2, 0.3, 0.4, 0.5\}$) to inspect the effect of increased variability on the accuracy of the algorithm. For an infinite sample size, increasing σ_L from zero moves the β_1,β_2 coordinate out along the Johnson SL line in Figure 5 and away from $\beta_1 = 0$, $\beta_2 = 3$. The Johnson SL line is described by Equation 19 through Equation 21.

Referring to Table 2, the simulation results are presented as the percent correct and incorrect identifications of the SL-distributed data sets for n = 500, for four of the curve types in Figure 5 (i.e., SL, SU, SB, and normal). The results for each curve type are organized by tolerance (i.e., rows) and by the SL population standard deviation (i.e., column). It is clear for $\sigma_L > 0.1$ (i.e., beyond $\beta_1 \approx 0$, $\beta_2 \approx 3$ in Figure 4), that the algorithm misidentifies the SL distribution with a high frequency.

This occurs because the sample data have estimates of β_1 and β_2 that are near the parent values for the normal distribution (i.e., $\beta_1 = 0$, $\beta_2 = 3$). The ability of the algorithm to correctly identify the SL curve improves as σ_L increases and then diminishes as explored in the following paragraph.

Table 3 gives the sample estimates of the mean and standard deviation of β_1 and β_2 as a function of σ_L for n = 500. As the value of σ_L increases from 0.01 to 0.5, the simulated average estimates of β_1 and β_2 create a trajectory that parallels the SL line on Figure 4, as described by the solid dotted symbols for sample size of 500.

Table 2. Johnson SL Curve Identification Percentages by the Hill et al. (1976) Algorithm, Sample Size of 500

		Log-Normal Standard Deviation					
Curve Type	**Tolerance**	**0.01**	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**
Johnson SL	0.01	91.04%	97.53%	57.12%	3.29%	0.09%	0%
	0.05	61.11%	94.48%	56.04%	2.48%	0.04%	0.03%
	0.1	32.24%	88.93%	56.49%	2.93%	0.22%	0.17%
	0.5	0.00%	0.02%	6.03%	4.31%	1.18%	0.65%
	Tolerance	**0.01**	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**
Johnson SU	0.01	0.00%	0.00%	0.34%	1.05%	1.08%	0.58%
	0.05	0.00%	0.00%	0.09%	1.15%	0.92%	0.41%
	0.1	0.00%	0.00%	0.01%	1.38%	1.06%	0.38%
	0.5	0.00%	0.00%	0.21%	0.51%	1.08%	0.24%
	Tolerance	**0.01**	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**
Johnson SB	0.01	0.10%	2.00%	41.46%	94.53%	98.08%	98.85%
	0.05	0.03%	2.50%	40.06%	89.58%	94.31%	96.69%
	0.1	0.04%	1.15%	35.42%	80.80%	88.93%	93.96%
	0.5	0.05%	0.32%	12.58%	44.14%	58.61%	72.95%
	Tolerance	**0.01**	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**
Normal	0.01	8.84%	0.41%	0.00%	0.00%	0.00%	0.00%
	0.05	38.86%	2.68%	0.00%	0.00%	0.00%	0.00%
	0.1	67.70%	9.44%	0.00%	0.00%	0.00%	0.00%
	0.5	99.81%	97.98%	48.86%	1.34%	0.00%	0.00%

Note: Balance in ST curve comprises 100 percent for each entry.

Table 3. Estimates of β_1 and β_2 for Johnson SL Distributed Data, Sample Size = 500

	Sample Relative Skew		Sample Relative Kurtosis	
Standard Deviation, σ_L	**Average**	**Standard Deviation**	**Average**	**Standard Deviation**
0.01 (point "A")	0.0128	0.0185	3.0031	0.2174
0.1 (point "B")	0.1025	0.0753	3.1576	0.3068
0.2 (point "C")	0.3894	0.1939	3.6459	0.6155
0.3 (point "D")	0.9098	0.4243	4.5482	1.2296
0.4 (point "E")	1.7606	0.8849	6.0078	2.2252
0.5 (point "F")	3.0426	1.9298	8.1877	4.3424

Shown also on Figure 4 are the simulated sample averages of β_1 and β_2 for $n = 10^3$ (asterisks) and $n = 10^4$ (open circles). As the sample size increases, the sample averages of β_1 and β_2 approach the population β_1-β_2 line for the SL distribution as defined by Equation 19 through Equation 21. The implication of this shift with increasing sample size is that the decision rule of Equation 18 from Hill et al. (1976) will be inaccurate for samples sizes less than 10^3, as indicated by the results in Table 2 for $n = 500$.

The estimates of the standard deviation of β_1 and β_2 increase with σ_L (see Table 3). The increase in variability of β_1 and β_2 with σ_L reduces the accuracy of the Johnson SL curve selection algorithm. The loss of accuracy as the data depart from $\beta_1 = 0$, $\beta_2 = 3$ along the SL line is caused by the constant tolerance, ε, used in selecting either the SL, SB, or SU curve. A more rational curve selection algorithm should be developed which does not rely on the use of a constant tolerance. An improved algorithm should use a tolerance that depends on sample size and position along the SL line for curve selection.

POWER SIMULATION WITH AUTOCORRELATED AND CROSS-CORRELATED ERROR DATA

Statistical inference and decision-making based on finite-size data sets is error-prone. The more common error type is the chance of rejecting a true null hypothesis (i.e., the Type I error). The Type I error level is controlled by the selection of the critical value at some α level. For example, if the test statistic exceeds the critical at $\alpha = 0.05$, there is a five percent chance of rejecting the true null hypothesis. That is, there is a small chance that the test statistic is very large (or very small) due only to random variation and finite-size data and not due to an attributable influence.

Harder to evaluate is the likelihood of failing to reject the false null hypothesis (i.e., the Type II error). In this case, the risk is that the test statistic is statistically different from the criterion value under the null hypothesis, but the decision is to incorrectly accept the null hypothesis. The Type II error risk increases as the Type I error level decreases. The Type II error risk is also large if there are an insufficient number of observations or the data have a large variability. A quantity related to the Type II error is the statistical power which is defined as 1 – Type II error.

Simulation of statistical power defines the data requirements for a statistical test, validates the critical values, and defines the reliability of a statistical test to discriminate between the true and false null hypothesis (Objective 4 in the final paragraph of the Introduction). Power simulation is accomplished with the Monte Carlo technique. Data sets of sample size n are constructed such that the null hypothesis is deliberately falsified. The test statistic is computed and evaluated against the critical value specified by the Type I error rate and the sample size. A Type II error occurs if the decision is to incorrectly fail to reject the false null hypothesis. This process is repeated for nSim realizations to tally the number of Type II errors. The statistical power is found as Power = 1 - # Type II errors/nSim. Typically, the value of nSim is large (e.g., 10^6) in order to obtain an accurate power estimate.

The null hypothesis is falsified by generating random data by choosing non-zero values of the population test statistic. For example, a power simulation of Pearson's r applied to independent (i.e., zero autocorrelation) data would generate n random values for the independent variable, x_i. The dependent variable vector, y_i, is then calculated as $y_i = \rho_{xy}\, x_i$, where $\rho_{xy} \neq 0$. As $|\rho_{xy}|$ approaches 1, the statistical power approaches 100 percent. As $|\rho_{xy}|$ approaches zero, the statistical power approaches α. Therefore, a validation of the critical values is the satisfaction of the condition that Power = 100α percent given $|\rho_{xy}| = 0$.

Under the case that $\rho_{xy} = 0$, the power simulations were accomplished with Equation 4. As demonstrated in the section on critical value development, however, the presence of non-zero autocorrelation and non-zero cross-correlation can be expected to partition variability between the independent and dependent variables. This identifies a need for a set of equations that may be used to

synthesize simultaneous sets of random variables y_i and z_i that contain lag-1 autocorrelation and have a non-zero cross-correlation between y_i and z_i. Levy (1999) derived Equation 23, which is used in the power simulations.

$$x_{i,j} = \begin{matrix} \mu_{x_j} + u_{i,j}\sigma_{x_j}\sqrt{\left(1-\rho_{x_j}^2\right)-\left(1-\rho_{x_j}\rho_{x_k}\right) \times \left|\rho_{x_j x_k}\right|}+\ldots \\ +\rho_{x_y}\left(x_{i-1,j}-\mu_{x_j}\right)+u_{i,3}\sigma_{x_j}\sqrt{\left(1-\rho_{x_j}\rho_{x_k}\right) \times \left|\rho_{x_j x_k}\right|} \end{matrix} \quad (23)$$

Referring to Equation 23, $x_{i,j}$ is a generic random variable. The subscript j refers to the x_j j^{th} generic random variable; the subscript k refers to the x_k k^{th} random variable with which the random variable x_j covaries. For example, if j = 1 represents the lateral errors, y_i, then k = 2 represents the vertical errors, z_i. The subscript i represents the i^{th} sequential observation of y_i or z_i. The mean and standard deviation of the j^{th} variable are μ_j and σ_j, respectively. The autocorrelation coefficients for the j^{th} and k^{th} variables are ρ_{xj} and ρ_{xk}, respectively, and $\rho_{x(j),x(k)}$ is the cross-correlation between x_j and x_k. Finally, the $u_{i,j}$ represents the random variation from the standard normal distribution with x_j and the $u_{i,3}$ is the variation to be partitioned between x_j and x_k.

Figure 6 shows the contour lines of equal statistical power from the simulation of Equation 4 and 23 from the newly-developed critical values (solid lines) for a five percent Type I error level. The contours of equal power for the zero-autocorrelation critical values are shown as dashed lines for the five percent Type I error level. It can be seen that at $\rho_{xy} = 0$, the statistical power is five percent for all sample size values for the newly-developed critical values. This satisfies the previously stated validation condition on the critical values for any level of autocorrelation. By contrast, the power is 15 percent for the condition of $\rho_{xy} = 0$ for the zero-autocorrelation critical values, which is incorrect. Figure 6 also shows that the power increases to 1 as the ρ_{xy} increases to 1.

Finally, the power simulations show that the statistical power from using the newly developed critical values is less (e.g., about ten percent) than expected when using the zero-autocorrelation critical values. This is rational given the expected increase of the Type I error level to its correct, stipulated value when using the newly-developed critical values in the presence of non-zero autocorrelation.

CONCLUSIONS

The development of containment surface criteria depends, in part, on the accurate analysis of the statistics of flight technical error. Failure to properly apply the appropriate statistical analysis techniques may result in unreliable lateral and vertical distance estimates for containment modeling. Another possible unsatisfactory outcome may be the over-estimation of containment distances, resulting in the inefficient use of airspace or unnecessarily restrictive conditions on the route use and deployment.

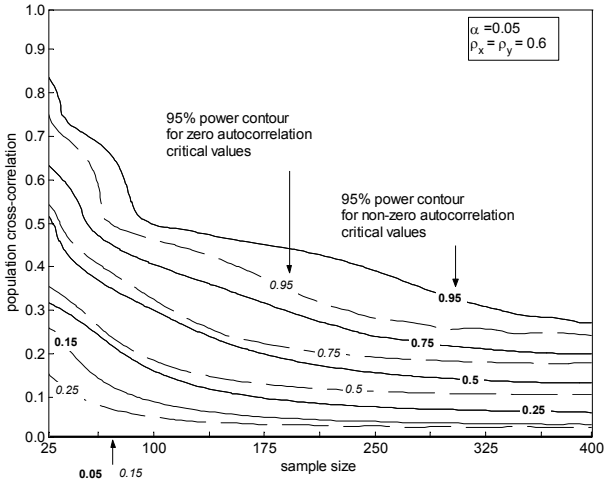

Figure 6. Contours of Statistical Power for Non-Zero Autocorrelation and Zero-Autocorrelation Critical Values ($\alpha = 0.05$, $\rho_x = \rho_y = 0.6$)

Part of accurate modeling of flight technical error involves the assessment of the strength of cross-correlation between the lateral and vertical errors. This assessment may be made inaccurate by the presence of autocorrelation in the errors, leading to an increased likelihood of rejecting the true null (i.e., no cross-correlation) hypothesis. The result of such a decision is to reduce the data set size, which leads to a more inaccurate identification and estimation of pdf parameters. This inaccuracy is accompanied by a higher than warranted level of statistical power. Furthermore, the accuracy of the algorithm for Johnson curve selection is not known, despite the attractiveness of the Johnson curves for pdf-fitting to the error data.

This work has addressed the deficiencies addressed above by use of Monte Carlo simulation. A set of critical values for testing cross-correlation with Pearson's r were developed that depend on sample size, Type I error rate, and lag-1 autocorrelation in the lateral and vertical errors. The critical values may be accurately estimated by a fourth-order polynomial equation for ease of use. The work herein also assesses the power of the Pearson's r test given the newly-developed critical values for non-zero autocorrelation and multivariate, lag-1 Markov equations. The multivariate Markov used in the power simulation may be useful for modeling of flight technical error because of the capability to partition variability between cross-correlation and autocorrelation. Finally, a preliminary study of the Johnson curve identification algorithm indicates that it is inaccurate for sample sizes less than 10^3. Improvement in the accuracy of identifiability could be made to the algorithm by allowing

for different sample sizes and changes in variability with population pdf parameters.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the reviews of Thomas Becher and Charlotte Laqui of CAASD. The lead author also acknowledges a special debt of gratitude to Dr. R. McCuen and Dr. K. Brubaker at the University of Maryland, College Park for their past guidance.

REFERENCES

Federal Aviation Administration, 2003, AFS-420, Personal Communication.

Haan, C. T., 1977, *Statistical Methods in Hydrology*, The Iowa State University Press, 378 pp.

Hahn, G. and S. S. Shapiro, 1994, *Statistical Models in Engineering*, John Wiley & Sons, Inc., 376 pp.

Hill, I. D., R. Hill, and R. L. Holder, 1976, Algorithm *99: Fitting Johnson Curves by Moments*, Applied Statistics, vol. 25, pp. 180-189.

Hipel, K. W. and A. I. McLeod, 1994, *Time Series Modeling of Water Resources and Environmental Systems*, Developments in Water Science, Ch. 45, Elsevier, 1013 pp.

Johnson, N. L., 1965, *Tables to Facilitate Fitting SU Frequency Curves*, Biometrika, Vol. 52, Issue 3 and 4, pp. 547-558.

Levy, B. S., 1999, *Multiple Trend Detection in Hydrologic Time Series*, Ph.D. Dissertation, University of Maryland, Department of Civil and Environmental Engineering, 510 pp.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, 1992, *Numerical Recipes in FORTRAN*, Cambridge University Press, pp. 279-280.

RTCA, 2002, *Minimum Aviation System Performance Standards: Required Navigation Performance for Area Navigation*, RTCA DO-236A, 165 pp.