A COMPARISON OF LAAS ERROR BOUNDING CONCEPTS^{*}

Curtis A. Shively

Center for Advanced Aviation System Development The MITRE Corporation, McLean VA

BIOGRAPHY

Curtis A. Shively is a member of the Principal Staff at MITRE/CAASD. He has studied the use of satellites for communications and navigation. He received his B.S. and M.S. in Electrical Engineering from the Massachusetts Institute of Technology.

ABSTRACT

The FAA's Local Area Augmentation System (LAAS) broadcasts a parameter $\sigma_{pr_{gnd}}$ to describe the errors in the differential corrections due to the ground facility. An aircraft using LAAS computes an upper bound with high probability on the resulting error in the position domain based on $\sigma_{pr gnd}$ assuming that the errors are Gaussian distributed. This paper compares five bounding methods in terms of assumptions and resulting performance. Assumptions regarding the tails of the error distribution range from a Gaussian model to an exponential model. Performance is compared in terms of the factor by which the estimated value of σ_{pr_gnd} must be inflated before broadcast to ensure the bound is provided with known confidence. For fixed, desired confidence the inflation factor varies with the number of independent samples used to estimate $\sigma_{pr gnd}$. Results show that using the same number of samples, methods assuming Gaussian tails give significantly smaller inflation factors than does the exponential tail method. Between these two extremes, is a more recently conceived method that assumes the error distribution is a mixture of Gaussian distributions with different standard deviations. The Gaussian mixture method gives inflation factors that are smaller than for the exponential tail method and may most closely correspond to the manner in which errors are present in real data.

INTRODUCTION

Background

The FAA's Local Area Augmentation System (LAAS) broadcasts differential corrections for each visible satellite, which are the average of corrections from M reference receivers. A quantity $\sigma_{pr gnd}$ is also broadcast to describe the error in the corrections due to the LAAS Ground Facility (LGF), assuming it is fault free. Since a separate value of $\sigma_{pr_{gnd}}$ is broadcast for each satellite, the correction errors are characterized in the range domain. In the aircraft using LAAS, protection levels are computed to bound the navigation error in the position domain that results from the broadcast corrections and other error sources such as in the avionics and data latency. These protection levels assume that the position error distribution is overbounded by a Gaussian distribution with standard deviation (herein referred to as $\sigma_{vert(M)}$) derived from the broadcast values of $\sigma_{pr gnd}$ and the geometry of the satellites being used for navigation. One such bound known as VPL_{H0} [1] is given for CAT I and M = 3 reference receivers at the LGF by

$$VPL_{H0} = 5.81 \times \sigma_{vert(M)} \tag{1}$$

The probability associated with VPL_{H0} is

Prob{|Fault - Free Vert Error| > VPL_{H0}}
= 2Q(5.81) =
$$6.25 \times 10^{-9}$$
⁽²⁾

where Q is the tail probability (single sided) for a Gaussian distribution with $\sigma = 1.0$. Thus, the fault-free

©2000 The MITRE Corporation

^{*} The contents of this material reflect the views of the author. Neither the Federal Aviation Administration nor the Department of Transportation makes any warranty or guarantee, or promise, expressed or implied concerning the content or accuracy of the views expressed herein.

vertical position error is assumed to exceed VPL_{H0} with an extremely small probability.

Another such bound known as $VPL_{H1,i}$ [1] is used to test the hypothesis that only the ith reference receiver (RR_i) may be faulted. For CAT I and M = 3 reference receivers this bound is given for RR₁ by

$$VPL_{H1,1} = |\text{Estimated Vert Error } RR_1| +2.90 \times \sigma_{vert(M-1)}$$
(3)

The probability associated with $VPL_{H1,1}$ is the probability of not detecting a fault on RR_1 given by

Prob{Fault - Free Vert Error
$$RR_2 \& RR_3$$

> 2.90 $\sigma_{vert(M-1)}$ } = Q(2.90) = 1.9×10⁻³ ⁽⁴⁾

Thus, the probability associated with the $VPL_{\rm H1}$ bound is not nearly as small as that associated with the $VPL_{\rm H0}$ bound.

A significant challenge is associated with verifying that the broadcast values of σ_{pr_gnd} do in fact provide the assumed bounds, particularly the VPL_{H0} bound. Although the core of the error distribution appears to be Gaussian, its tails may not be Gaussian. An extremely large amount of data would be needed to observe and precisely characterize the tails of the error distribution. A general approach to this challenge has been proposed and investigated by a number of researchers. In this approach, an estimate of the standard deviation (herein referred to as $\sigma_{pr_gnd_estimate}$) is computed from observed values of the error. An inflation factor (herein referred to as INF) is then applied to give the broadcast value as

$$\sigma_{\text{pr}_gnd} = \text{INF} \times \sigma_{\text{pr}_gnd_estimate}$$
(5)

The value of INF is determined so that the assumed bounds are theoretically provided. A previous paper [2] analyzed one particular method of computing INF which involves assuming the tail is exponential beyond an analysis point.

Purpose and Organization of Paper

This paper compares five overbounding methods including the previously analyzed exponential tail method. The paper begins with general assumptions, considerations and an initial observation. Then a very brief description of each method is presented. Thereafter, each method is discussed in turn, including underlying assumptions regarding the postulated error distribution, development, resulting values of INF and observations. The paper concludes with a comparison of all methods, a summary and recommendation.

Considerations and Assumptions

Research is currently underway to develop theoretical characterizations for LAAS errors that may not be Gaussian, particularly those due to multipath. However, such theories that can be verified from data have not yet been developed. Consequently, the emphasis in this paper is on what can actually be estimated or confirmed from observed data in conjunction with additional a priori assumptions that must be taken on judgment.

An important consideration when characterizing the performance of estimation techniques is the confidence associated with the result. Therefore, a procedure for quantifying confidence is developed for each method. A confidence value of 0.90 was used in the analysis. While this value may seem rather small, it was chosen because inflation factors are difficult to compute for Method 5 (see list below) for confidence higher than 0.90 and it was desired to compare all methods using the same value of confidence.

For simplicity, several other factors that may be relevant are not considered in this paper. Theory for quantifying confidence assumes that samples are independent. Carrier smoothing of data with a 100 s time constant introduces autocorrelation of data samples. This analysis assumes that independence of data has been ensured by procedures such as taking samples at least 200 s apart during any single day. Multipath that repeats from day to day is assumed to be avoided by skipping an appropriate number of days between data gathering sessions. Correlation of errors can also be present between reference receivers due to specular ground multipath. Possible ways of addressing this factor have been treated in [3]. Crosscorrelation of errors between reference receivers is assumed negligible in the present analysis.

Initial Observations

The characteristics of LAAS errors that can be verified depend on the number of independent samples that can be collected in practice. Based on assumptions similar to those mentioned above, it was observed in [2] that roughly 25,000 independent samples could be collected per satellite elevation bin per year. The probability $P_{any>K_{\sigma}}$ of observing any samples greater than $K\sigma$ can easily be computed for the Gaussian distribution as a function of the number of samples taken (refer to Figure 1 in [2]). Assuming 25,000 samples, for K = 3.72, $P_{any>3.72\sigma} \sim 0.9$, while for K = 4.27, $P_{any>4.27\sigma} \sim 0.2$. Thus, even for a year's worth of data, the observations will be limited to roughly $\pm 4\sigma$. Thus, only characteristics relevant to the VPL_{H1} bound, which uses 2.90σ can be verified from data. For the VPL_{H0} bound, which uses 5.81σ , some a priori assumptions must be made regarding the tail of the error distribution beyond what can actually be observed from data.

Methods Analyzed

The following brief list is intended to suggest only the essence of the five methods analyzed.

Method 1: Gaussian tail, inflation factor determined from $\sigma_{pr \ gnd \ estimate}$ alone

Method 2: Gaussian tail, inflation factor determined from tail observed at $K_L \sigma_{pr \ gnd \ estimate}$

Method 3: Gaussian tail, inflation factor determined from largest observed value

Method 4: Overbounding at $5.8INF\sigma_{pr_gnd_estimate}$, tail from mixed Gaussians beyond $K_L\sigma_{pr_gnd_estimate}$

Method 5: Overbounding at $5.8INF\sigma_{pr_gnd_estimate}$, tail is exponential beyond $K_L\sigma_{pr_gnd_estimate}$

METHOD 1

Method 1 is based on accepting the a priori assumption that the entire error distribution is Gaussian. The standard deviation $\sigma_{pr_gnd_estimate}$ is estimated in the usual manner from the square root of the sample variance computed from N independent samples. To achieve the desired confidence, the estimate thus derived need only be inflated due to the finite number of samples used. The value of INF to give an upper confidence limit $\sigma_{est_conf} =$ INF $\sigma_{pr_gnd_estimate}$ is based on the well-known Chi-Square distribution with N-1 degrees of freedom. Method 1 provides a bound for all values of error, even beyond $5.81\sigma_{pr_gnd}$ needed for VPL_{H0}.

Figure 1 shows values of INF versus N for Method 1. At the first point plotted (N = 100), INF ~ 1.1. For larger values of N, INF decreases rapidly and is negligible (< 1.02) for N > 2,000. Method 1 makes the most idealized assumption and thus gives by far the smallest values of INF of all methods. The goal of current research is to establish the plausibility of the assumption underlying Method 1 that the entire error distribution is Gaussian. However, as of this writing no data has produced an observed distribution that can be overbounded without inflating σ_{pr_gnd} considerably more than this method would predict.



Figure 1. Method 1 Inflation Factor vs. Number of Samples (Confidence = 0.90)

METHOD 2

The next two methods assume the tail of the error distribution is Gaussian, but tie the broadcast value of $\sigma_{pr gnd}$ to some point on the observed distribution rather than just the value of $\sigma_{pr_gnd_estimate}$ computed from the sample variance. Method 2 is based on accepting the a priori assumption that the distribution is Gaussian beyond some analysis point $E_L = K_L \sigma_{pr_gnd_estimate}$. This method is equivalent to finding the σ (larger than $\sigma_{pr gnd estimate}$) of a Gaussian distribution that has the tail probability actually observed (with desired confidence) at E_L. The tail probability is estimated from the number of samples n (out of all N gathered) whose values are greater than E_L . The binomial distribution is then used to compute an upper confidence limit P_{T_conf} on the observed tail probability at E_L. The value of INF is then determined so that $P_{T_Gaussian}(K_L / INF) = P_{T_conf}$. By the a priori

assumption bounding is provided beyond E_L , and thus at $5.8\sigma_{pr_gnd}$. The analysis was done assuming $K_L = 2.9$ for two reasons. First, K = 2.9 in VPL_H1. Second, since $P_{T_Gaussian}(2.9) \sim 0.002$, it is anticipated that the tail probability can be estimated with reasonable confidence for N on the order of a few thousand samples.

For Method 2 the value of INF that would be computed from real data depends on the value of n actually observed. While n is a random variable having a binomial distribution, it is assumed in this analysis that n takes on its expected value, i.e., n = 0.002N. Figure 2 shows these "expected" values of INF for Method 2. A comparison with Figure 1 indicates that Method 2 gives slightly larger values of INF than does Method 1 for the same N. While any use of Method 2 is unknown to this author, it is included here for completeness as a plausible technique.



Figure 2. Method 2 Inflation Factors (Expected) Assuming Gaussian Tail Beyond 2.9 $\sigma_{pr \ gnd \ estimate}$ (Confidence = 0.90)

METHOD 3

Method 3 ties the broadcast value of σ_{pr_gnd} to the largest error value observed, $E_{max} = K_{max}\sigma_{pr_gnd_estimate}$, rather than a selected value E_L . This method is based on the a priori assumption that the tail of the error distribution is Gaussian beyond E_{max} . Method 3 is equivalent to finding the σ (larger than $\sigma_{pr_gnd_estimate}$) of a Gaussian distribution that has the tail probability actually observed (i.e., zero plus upper confidence limit) beyond E_{max} . The upper confidence limit P_{T_conf} on the tail probability is determined from the binomial distribution when 0 of N are observed. A value of K_g is then determined such that $P_{T_Gaussian}(K_g) = P_{T_conf}$. Then, since $E_{max} = K_g\sigma_g =$ $K_{max}\sigma_{pr_gnd_estimate}$, $INF = \sigma_g / \sigma_{pr_gnd_estimate} = K_{max} / K_g$. By the a priori assumption bounding is provided beyond E_{max} and thus at 5.8 σ_{pr_gnd} .

For Method 3 the value of INF that would be computed depends on the value of E_{max} actually observed. While E_{max} is a random variable, it is assumed in this analysis

that E_{max} takes on its expected value. Figure 3 shows the expected value of E_{max} in multiples of σ (i.e., K_{max}) as a function of N. Note from the figure that for N = 1,000 $K_{max} \sim 3.5$ and increases to roughly 4.0 for N = 10,000. Beyond that point the expected value of E_{max} increases slowly due to the very small probabilities involved.

Figure 4 shows values of INF for Method 3 assuming the expected value of E_{max} is actually observed. Note that INF does not vary as much with N, and for large N stabilizes at a higher value, than for Method 2 (Figure 2). A procedure similar to Method 3 has often been used to justify an overbound by graphical presentation. In that case, an error distribution from data is plotted along with a so-called "overbounding" Gaussian curve based on an inflated σ . As typically illustrated, the Gaussian tail probability is larger than the probability from the data for the largest observed value. However, such a method does not quantify the confidence in the overbound unless the probabilities for the observed distribution have already been appropriately increased.



Figure 3. Expected Value of Largest Observed Magnitude



Figure 4. Method 3 Inflation Factors (Expected) Assuming Gaussian Tail Beyond Largest Observed Value (Confidence = 0.90)

METHOD 4

Method 4 is based on the a priori assumption that the tail of the error distribution is from a mixture of Gaussian distributions. This could be the case for the following reasons. Data are often pooled from different azimuths and even elevations for the same satellite. Moreover, data from more than one satellite are sometimes pooled in order to obtain enough independent samples for the analysis to have any meaning at all. If the data pooled in this manner were all Gaussian distributed with the same mean and standard deviation, the result would obviously be Gaussian distributed. However, the multipath error characteristics are likely to be a function of satellite azimuth and elevation. If elevations are combined there is often an attempt to normalize the data using the standard deviation for each elevation bin. The imperfection in the bin standard deviation estimates can lead to the pooling producing a mixture of Gaussians with standard deviations different from unity. Even if just azimuths are combined, there is insufficient data or theory to attempt any kind of normalization. Thus, the actual distribution of the data to be analyzed might at best be a mixture of Gaussians with different standard deviations.

Method 4 attempts to estimate parameters that describe a mixed Gaussian model probability density function (PDF)

A model which represents pooling of sigmas uniformly distributed between two extremes might be appropriate. Such a model was investigated and found to require numerical solution for estimating the parameters. Therefore, to be more informative and to illustrate the potential of this method, this paper analyzes the case assuming the mixture combines errors in equal amounts from distributions with just two distinct values of σ , i.e., σ_1 and $\sigma_2 > \sigma_1$.

The overbounding model for this case simply assumes a Gaussian distribution with $\sigma = \sigma_2$. The overbounding provided by this model is illustrated in Figure 5. The top curve is the tail of a Gaussian distribution with $\sigma = \sigma_2$. The next curve is the tail for the actual distribution from the mixture of data. Note that the model tail closely bounds the actual distribution in this example. The lowest curve is the tail of a Gaussian with the expected value of the σ that would be observed from the pooled data.



Figure 5. Method 4 Model Overbounding Illustration (Tail Probability $\sigma_1 = 0.5$, $\sigma_2 = 1.5$)

Since

$$\sigma_{\rm obs} = \sqrt{\frac{\sigma_1^2 + \sigma_2^2}{2}} = \sqrt{\frac{0.5^2 + 1.5^2}{2}} = 1.1 \quad (6)$$

the tail based on σ_{obs} underestimates the probability by roughly five orders of magnitude at the 5.8 σ point of interest. INF in this case is given by

INF =
$$\frac{\sigma_2}{\sigma_{obs}} = \frac{1.5}{1.1} = 1.34$$
 (7)

The values of σ_1 and σ_2 can be estimated from the data using the method of moments [4]. The second and fourth moments (designated m_2 and m_4 , respectively) are expressed in terms of σ_1 and σ_2 by

$$\mathbf{m}_2 = \overline{\mathbf{E}^2} = \frac{\sigma_1^2 + \sigma_2^2}{2} \tag{8}$$

$$m_4 = \overline{E^4} = \frac{3\left(\sigma_1^4 + \sigma_2^4\right)}{2} \tag{9}$$

Values of σ_1 and σ_2 may then be estimated using the moments observed from the data

$$\sigma_{1_est} = \sqrt{m_2 - \sqrt{\frac{m_4}{3} - m_2^2}}$$
(10)

$$\sigma_{2_est} = \sqrt{m_2 + \sqrt{\frac{m_4}{3} - m_2^2}}$$
(11)

In order to find an upper confidence limit on σ_{2_est} it is necessary to find the variance of σ_{2_est} . Use of methods in [4] gives

$$\operatorname{var}(\sigma_{2_est}) = \operatorname{var}(m_{2}) \left(\frac{\partial \sigma_{2_est}}{\partial m_{2}}\right)^{2} + 2\operatorname{cov}(m_{2}, m_{4}) \frac{\partial \sigma_{2_est}}{\partial m_{2}} \frac{\partial \sigma_{2_est}}{\partial m_{4}}$$
(12)
$$+ \operatorname{var}(m_{4}) \left(\frac{\partial \sigma_{2_est}}{\partial m_{4}}\right)^{2}$$

Evaluating the variance and covariance terms gives

$$\operatorname{var}(\mathbf{m}_2) = \frac{\left(\sigma_1^4 + \sigma_2^4\right)}{n} \tag{13}$$

$$cov(m_2, m_4) = \frac{6(\sigma_1^6 + \sigma_2^6)}{n}$$
 (14)

$$\operatorname{var}(\mathbf{m}_4) = \frac{48\left(\sigma_1^8 + \sigma_2^8\right)}{n} \tag{15}$$

Evaluating the partial derivatives gives

$$\frac{\partial \sigma_{2_est}}{\partial m_2} = \frac{-\sigma_1^2}{\sigma_2 \left(\sigma_2^2 - \sigma_1^2\right)}$$
(16)

$$\frac{\partial \sigma_{2_est}}{\partial m_4} = \frac{1}{6\sigma_2 \left(\sigma_2^2 - \sigma_1^2\right)}$$
(17)

Substituting equations (13) through (17) into equation (12) gives

$$sigma(\sigma_{2_est}) = \frac{\sqrt{(\sigma_{1}^{8} + 3\sigma_{1}^{4}\sigma_{2}^{4} - 6\sigma_{1}^{2}\sigma_{2}^{6} + 4\sigma_{2}^{8})}}{\sqrt{n}\sqrt{3}\sigma_{2}(\sigma_{2}^{2} - \sigma_{1}^{2})}$$
(18)

According to [4] $\sigma_{2_{est}}$ is asymptotically Gaussian (with respect to sample size) with expected value equal to σ_2 . If the observed estimate is denoted $\sigma_{2_{observed}}$, the upper confidence bound $\sigma_{2_{est_{conf}}}$ satisfies

$$\Pr ob \begin{cases} \sigma_{2_est} \le \sigma_{2_observed} \\ \text{if } \sigma_2 = \sigma_{2_est_conf} \end{cases} = 1 - \text{conf (19)}$$

The upper confidence limit $\sigma_{2_est_conf}$ is the largest that σ_2 could be and still have probability 1-conf that the estimated value σ_{2_est} would be less than the actually observed value $\sigma_{2_observed}$. This is equivalent to

$$\sigma_{2_est_conf}$$

$$-CF \times sigma \begin{pmatrix} \sigma_{2_est} \text{ Given } \sigma_2 \\ = \sigma_{2_est_conf} \end{pmatrix}$$
(20)

$$= \sigma_{2_{observed}}$$

$$\int_{-CF}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = conf$$
(21)

The inflation factor is then given by

$$INF = \frac{\sigma_{2_est_conf}}{\sigma_{pr_gnd_estimate}}$$
(22)

The value computed for INF depends on the value of $\sigma_{2 est}$ that results from the data analysis. Values of INF are shown in Figure 6, under the assumption that $\sigma_{2_{est}}$ takes on its expected value, i.e., $\sigma_{2 \text{ est}} = \sigma_2$. Several curves are shown corresponding to different spreads between the values of σ_1 and σ_2 . The top curve corresponds to the bounding example given in Figure 5 with $\sigma_1 = 0.5$ and σ_2 = 1.5, or a spread of ± 0.5 relative to 1.0. The bottom two curves correspond to smaller spreads of ± 0.1 and ± 0.05 . Spreads of this smaller size might correspond to the residual difference in standard deviation resulting from normalization of two groups of data by their own standard deviation estimates. Note that all curves flatten rapidly with increasing number of samples, as was the case for Method 3. Note, however, that for the lowest curves, the inflation factors are at best somewhat larger than for Methods 1 through 3.

METHOD 5

Since Method 5 has been described extensively elsewhere by the author [2], the discussion in this paper is limited to a brief summary of the method concept and presentation of the resulting inflation factors for comparison to the Method 5 assumes that the error other methods. distribution for a single reference receiver has a Gaussian core and an exponential (Laplacian) tail beyond the analysis point $E_L = K_L \sigma_{pr_gnd_{est}}$. The two parameters needed to fully describe the exponential tail are verified through hypothesis tests on the data. One of these tests is conducted on the value of the PDF at E_L and the other test is conducted on the total value of the tail beyond E_L , i.e., the probability that E exceeds E_L. Computation of these two parameters considers the confidence that is associated with the limited number of data samples analyzed. The error corresponding to averaging differential corrections is modeled by convolution of the individual model PDFs for three reference receivers. The inflation factor is determined so as to provide bounding of the resulting average error at 5.8INF $\sigma_{pr_gnd_est}$. For further details on the method, the reader is referred to [2].

Figure 7 shows values of inflation factors for Method 5 versus the number of independent samples. Three curves are shown corresponding to different values of K_L. Note that the lowest inflation factors are associated with the smallest value of $K_L = 1.96$. This occurs because the probabilities to be confirmed by hypothesis testing are largest in that case. Recall that the model in Method 5 assumes an exponential tail for an individual reference The averaging process reduces this tail receiver. somewhat. Even with this advantage of averaging after applying the model, Method 5 still has significantly larger inflation factors than Methods 1 through 4. Method 5 is believed to provide a limit for Method 4, particularly for the case of pooling data from Gaussian distributions with uniformly distributed σ . Further work beyond the scope of this paper is needed to quantify the relationship between the bounds provided by Methods 4 and 5.



Figure 6. Method 4 Inflation Factors (Expected) 50-50 Mixture of Two Gaussians (Confidence = 0.90)



Figure 7. Method 5 Inflation Factors (Exponential Tail Before Averaging References, Confidence = 0.90)

SUMMARY

A convenient comparison of inflation factors for all methods is presented in Figure 8. Representative curves with mid-range parameter values are shown for Methods 4 and 5. Note how much smaller the inflation factors are for Methods 1 through 3. These methods require the strongest a priori assumption, that the error distribution has a Gaussian tail. Note how much more conservative the inflation factors of Method 5 are, particularly for smaller numbers of independent samples. This method requires the least restrictive assumption, that the error distribution has an exponential tail (before averaging reference receivers). Method 4 gives somewhat moderate inflation factors and requires perhaps the most plausible a priori assumption, that the data pools Gaussian distributed errors with different standard deviations. Consequently, it is recommended that Method 4 be developed further and applied to field data.

ACKNOWLEDGMENT

The author wishes to thank Ron Braff of MITRE for reviewing the paper.

REFERENCES

- 1. RTCA, Minimum Aviation System Performance Standards for the Local Area Augmentation System (LAAS), RTCA/DO-245, September 28, 1998.
- 2. Shively, C., "An Overbound Concept for Pseudorange Error from the LAAS Ground Facility", *Proceedings of Institute of Navigation Annual Meeting*, San Diego, CA, 26–28 June 2000.
- 3. Pervan, B. and I. Sayim, "Issues and Results Concerning the LAAS σ_{pr_gnd} Overbound," presented at IEEE Position Location and Navigation Symposium (PLANS) 2000, San Diego, CA, 13–16 March 2000.
- 4. Cramer, H., *Mathematical Methods of Statistics*, Princeton University Press, 1946.



Figure 8. Inflation Factor Comparison (Confidence = 0.90)