Evaluation of Signal in Space Error Bounds to Support Aviation Integrity

Todd Walter, Juan Blanch, and Per Enge,
Stanford University

ABSTRACT

The modernization of GPS and the addition of new GNSS constellations bring forth a multitude of new signals with new capabilities. One of the intended purposes of some of these new signals is the provision of high integrity positioning sufficient for use in aviation. However, one must be extremely cautious, as the integrity requirements for aviation, particularly precision approach, are very strict. There must be less than a one in ten million chance of providing misleading information to the pilot. Further, each individual nation is responsible for approving equipment and procedures in their sovereign airspace. In order to make use of signals over which they have no control, aviation authorities must have a clear understanding of the commitments and capabilities of these new signals. These authorities will also likely want to monitor the signals to ensure that they are conforming to their intended design.

This paper proposes equations to combine information from the satellites together with other error bounds to form appropriate upper bounds on the users' position estimates. The position domain bounds, called protection levels, are matched to proposed data monitoring criteria that ensure consistency between what has been observed and what is assumed by the bounding equations. The protection level equations, their matching monitoring criteria, and their mathematical link are the main topics of this paper.

INTRODUCTION

The Global Positioning System (GPS) is in the process of adding new capabilities. This modernization effort includes new civil signals whose capabilities improve greatly over the currently available signal [1] [2] [3]. In addition, new constellations are being fielded that will offer a much larger number of satellite navigation signals. It is important to study these new signals and capabilities, and plan how to utilize them for aircraft navigation.

It is important for GNSS service providers to clearly describe the performance of their signals. This includes important parameters such as the expected accuracy, probabilities and behaviors of fault modes, time to alert, confidence bounds on the signal in space errors, and how to combine the confidence bounds with other error sources and across multiple satellites. Some of these parameters are more easily specified and evaluated than others. Existing augmentation systems use the concept of Gaussian Bounding to assure that the confidences are correctly combined to produce a confidence bound in the position domain. Unfortunately, Gaussian behavior can be difficult to evaluate. Small samples of Gaussian data may appear to be non-Gaussian and non-Gaussian behavior may not be readily apparent in other sampled data sets. Instead, we propose well-defined tests of sampled data that evaluate specified error quantiles using data sets of fixed length. These quantiles (e.g., within 2-sigma 95% of the time) may correspond to expected Gaussian behavior, but by fixing the sample interval and number of samples we can achieve a common understanding of precisely how to interpret probabilities and error distributions. By specifying several such evaluations at different quantiles and data set lengths, we can bound the full error distributions.

We then show that these evaluations can be rigorously linked to confidence bounds on the signal in space errors through a previously developed technique called paired bounding [4]. Paired bounding allows us to create safe protection level equations that ensure that any combination of signal and user errors has a corresponding confidence bound that meets the required level of integrity. What is unique in this approach is that clear unambiguous evaluations of signal in space errors can be mutually agreed upon by GNSS service providers and integrity service authorities. These evaluations will aid in the approval for use of the high integrity signals. Further, they will enhance the ability to combine integrity from different GNSS service providers to create a more powerful multi-constellation service. This approach is applicable to stand-alone safety of life services or integrity provided through multi-constellation ARAIM.

These evaluations are not intended to serve as a complete integrity analysis. Instead, they are intended to complement design assurance. That is, a system will have integrity because the GNSS service provider designed it into the system, not because it has performed well for a fixed time. Evaluation of the system is merely used to confirm that the integrity design goals continue to be met. A system that is not designed for integrity cannot be assured to be safe into the future no matter how long it has been observed to perform safely in the past. Certain rare fault modes may not be observed, but could present themselves when new conditions arise. Each GNSS service provider must create a complete hazard evaluation on their system and assure that all significant hazards have an adequate mitigation. This analysis combined with evaluation of the actual performance provides the full assurance that the system will meet its integrity specifications.

GPS SPECIFICATIONS

We will concentrate our analysis in this paper on the GPS satellites and use nomenclature typically associated with this constellation [5]. The proposals would apply equally well to any other constellation with only small adjustments to the terminology. The Signal-in-Space (SIS) errors that are under the influence of the GPS ground control and space vehicle segments are referred to as Instantaneous User Ranging Errors (IUREs). These include satellite clock and ephemeris errors, satellite antenna variations, and signal imperfections. Specifically, the IURE does not include ionospheric or tropospheric delay, multipath, or user receiver errors. The IURE is an instantaneous error affecting a particular user at a particular time. The satellites broadcast a parameter called User Ranging Accuracy (URA) that is intended to be a conservative representation of the expected RMS behavior of the IURE at the worst-case location on Earth. The URA is meant to describe the accuracy of the IUREs and indicate an upper limit on their likely magnitude.

The current GPS specification only assures signals to have their IURE no greater than 4.42 x URA with a probability of 10^{-5}/hour [5]. This is not sufficient for high integrity aviation operations without additional protection. However, as part of its modernization, GPS is investigating a significant system design change to support URAs that are assured to bound IUREs to within 5.74 x URA with a probability of 10^{-8}/hour [6]. Satellites that meet this requirement will have a new specific integrity flag set to one. In this event, GPS may be able to support certain aviation applications without augmentation. None of the current GPS satellites have their integrity flag set, nor will they for many years to come. This paper investigates both how future assured URA values may be combined to form an assured position domain bound and how non-assured satellites could be used with RAIM to achieve the same goal.

LESSONS LEARNED FROM PREVIOUS SYSTEMS

One significant difficulty encountered when developing the safety analysis for WAAS and LAAS was the presence of small biases and non-Gaussian behavior observed in data used to validate the analyses. The integrity equations for these systems are based upon zero-mean Gaussian behavior. The sigmas broadcast for use in these equations were inflated to account for worst-case behavior that in turn led to larger protection levels. However, the underlying mathematical assumptions required exactly Gaussian and zero-mean characteristics. Additional analyses were created to ensure that the protection levels were sufficiently large to cover observed imperfect behavior. However, these new analyses imposed additional constraints on the system and further limited performance [7]. It is therefore strongly recommended that future systems use integrity equations that can directly account for biases and non-Gaussian behavior.

Fortunately, both can be accommodated by the inclusion of a bias term in the protection level equation. The handling of biases would then be explicit. A technique called paired bounding [4] is used to bound non-Gaussian behavior. It states that an arbitrary error distribution can be replaced by simple analytic models if the arbitrary distribution lies entirely in between the two models. Because the sign of a bias is not important to the protection level equation, $N(-\mu, \sigma)$ and $N(\mu, \sigma)$ are bounded by the same protection level parameters. Therefore, any distribution whose Cumulative Density Function (CDF) lies between the CDFs of those two distributions can be bounded by that biased Gaussian model.

Another significant lesson is that estimation of the underlying error distribution can depend greatly on the sample set chosen [8]. Because many of the error sources may change behavior over time, the observed characteristics may also change with time. When bounding such errors, it is necessary to bound the worst expected distribution. However, if one only looks at a small population of data, it is unlikely that the largest expected error will be sampled. If one aggregates too much data together, different conditions will all be mixed together creating an average rather than finding the worst-case condition. Therefore, it is best to collect as much data possible, but partition it into the smallest statistically significant subsets as possible. Ideally, these subsets would each contain only like conditions, such that the observed errors are drawn from the same distribution. In practice, this cannot be assured.

For integrity to be maintained, each and every data set must be bounded by the assumed distribution. This ensures that if there is non-stationary behavior, the worst observed cased is overbounded. Each data set must have a sufficient number of independent data points to assure that the distribution is properly sampled. Very small data sets could contain rare normal errors that make them appear to be worse than they really are. By having a sufficient number of independent points, rare normal errors should be appropriately balanced by smaller nominal errors. However, there should not be too many points as rare faults could be completely obscured by the nominal majority of the data. Therefore, it is desirable to partition the data into sets that contain close to the minimum number of independent samples required. How many samples are required depends on what quantity is being evaluated. A mean value can be determined with relatively few samples. However, the tail behavior is being investigated the number required may be in the millions or more.

The final significant lesson discussed in this paper is that small errors can combine to create large position errors. Most integrity analyses focus on single large errors. However, multiple smaller errors can also create a threat if they occur at the same time. This can occur in error distributions that have an excess of one to three sigma errors even if they never have larger errors. Because these moderate errors are occurring too often, there is a greater chance that they are occurring on different satellites at the same time. Thus, the likelihood of large position errors could be greater than implied by independent Gaussian distributions.

Alternately, if the errors are correlated across satellites, they may occur individually with the expected frequency, but when one error is moderate to large, the other errors will be as well. This, too, creates a greater likelihood of unacceptably large positioning errors. Therefore, it is important to ensure that the likelihood of separate errors combining to create a larger positioning error is not higher than expected. To monitor this threat, we propose evaluating the central portion of the error distribution rather than just the tails. Further, we recommend evaluating a chi-square metric to directly observe the sum of the squared errors across multiple satellites.

The recommended actions in response to these three lessons are explained in greater detail in the following sections. We also include some preliminary results of evaluations performed on past GPS data.

VPL EQUATIONS

When the WAAS VPL equation was created, it was believed that WAAS would differentially remove all significant biases. That is, if a bias was known to exist, WAAS would estimate it and remove it. The resulting error would then be unknown and time varying. Since that time, constant unobserved measurement biases have been identified. Therefore the WAAS estimate may have a component that, while unknown, is very slowly time varying. It is not correct to RSS such an error with the other terms when it is not random. WAAS added specific analyses to ensure that these unobserved biases are adequately taken into account despite the lack of a bias term in the VPL equation.

These analyses presume that the number of pseudoranges used in the position solution is limited below a maximum number and that every satellite error bound has a minimum amount of margin to protect the user from the biases. This margin has to exist even on satellites that do not contain the bias [7]. These two constraints make it more difficult to adapt to system changes. Adding new ranging sources increases the maximum number of potential ranging signals in view. This can increase the amount of margin needed in the bounding terms. Also, a slightly larger bias on a particular satellite can lead to requiring additional margin on all other satellites. These effects are due to having to protect against biases that must be added together linearly when there only an RSS term in the VPL equation.

Another early assumption was that the errors could be easily bounded with Gaussian variances. It was recognized that the errors were not purely Gaussian and that conservative estimates would be required for the variances. However, it was felt that the errors were close to Gaussian and that inflating the sigmas to handle the largest expected values would be sufficient. Unfortunately, it was discovered that nearly-Gaussian was a very poorly defined term. It was necessary to determine what was sufficiently Gaussian and what was not. This could be done subjectively, but not everyone would agree on the dividing line. New methods were developed to rigorously define whether the actual observed distributions were sufficiently close. Unfortunately, these methods introduced significant conservatism. Even small deviations from perfect behavior may create large margins in the resulting bound.

This imperfect matching has led to an inflation of the protection level values that may be as much as 20% [7]. As we move forward it is desirable to explicitly include terms to account for non-zero-means and non-Gaussian behavior. We recommend including a bias term. This term is used to bound errors that may appear random, but that affect users in the same way repeatedly. Examples of such biases are antenna biases [9] or nominal signal deformations [10] [11]. These error sources affect a particular geometry identically each time it is encountered. Thus, a maximum bias term is included to bound the effect of these error sources. If we follow the terminology of the SBAS MOPS [12], where K_V is the Gaussian quantile matching the probability of misleading information, $s_{3,i}$ is the projection of the i^{th} pseudorange error into the vertical dimension, and σ_{tropo} and σ_{user} Gaussian bound the tropospheric and dual-frequency user error, respectively, then the recommended VPL is given by:

$$VPL_{GIC} = K_V\sqrt{\sum_{i=1}^{n} s_{3,i}^2 \times \left(\alpha_1^2 \times URA_i^2 + \sigma_{trop,i}^2 + \sigma_{user,i}^2\right)} + \sum_{i=1}^{n}\left|s_{3,i} \times \left(\alpha_2 \times URA_i + \alpha_3\right)\right| \quad (1)$$

where URA is the broadcast confidence factor from the satellite, and the α terms are parameters that can be adjusted to ensure integrity.

If the satellite errors were zero-mean Gaussian and properly overbounded by URA then the α terms could be $\alpha_1 = 1$, $\alpha_2 = \alpha_3 = 0$. If there were small biases independent of the URA value and the URA needed to be inflated 25% to overbound the remaining errors then these terms could be $\alpha_1 = 1.25$, $\alpha_2 = 0$, α_3 = bias overbound. These parameters offer the flexibility to adjust the VPL equation to match currently unknown satellite error characteristics. These terms will be determined later when these characteristics are well known. They could be hardcoded into the MOPS and the receiver, or they could be broadcast dynamically to the user.

If the satellite signals do not have fully assured integrity, then the user will need to perform their own fault detection. Comparison of subset solutions has been shown to be an effective means of detecting and isolating satellite faults. If the probability of multiple faults being present during the required interval is sufficiently small, then only subsets excluding a single satellite need to be investigated.

The proposed ARAIM VPL equations are very similar in form to the assured integrity equation above. They have several terms starting with the all-in-view solution:

$$VPL_0 = K_0\sqrt{\sum_{i=1}^{n} s_{3,i}^2 \times \left(\alpha_1^2 \times URA_i^2 + \sigma_{trop,i}^2 + \sigma_{user,i}^2\right)} + \sum_{i=1}^{n}\left|s_{3,i} \times \left(\alpha_2 \times URA_i + \alpha_3\right)\right| \quad (2)$$

The subset solution terms are given by:

$$VPL_j = K_j\sqrt{\sum_{i=1, i\neq j}^{n} {}_j s_{3,i}^2 \times \left(\alpha_1^2 \times URA_i^2 + \sigma_{trop,i}^2 + \sigma_{user,i}^2\right)} + \sum_{i=1}^{n}\left|{}_j s_{3,i} \times \left(\alpha_2 \times URA_i + \alpha_3\right)\right| + \Delta_j \quad (3)$$

where Δ_j is the actual or expected difference in the position estimate between the subset excluding the j^{th} satellite and the all-in-view solution, and the K values are related to the probability allocated to each fault mode. The final VPL is determined by taking the maximum over the all-in-view and the subset values,

$$VPL_{ARAIM} = \max_{j=0,n} VPL_j \quad (4)$$

Full details of the algorithm can be found in [13] [14].

MONITORING INDIVIDUAL SATELLITES

In order for the above VPL equations to properly bound the user position error, certain assumptions have to be valid. The CDF of the actual satellite errors must all be bounded to the left by the Gaussian N($-\alpha_2$ x URA - α_3, α_1 x URA) and to the right by the Gaussian N(α_2 x URA + α_3, α_1 x URA). The other errors must be similarly left-right bounded by their respective terms. Unfortunately, we do not know the true CDF of the errors. We can only estimate it from sampled data. This sampling will take place over an extended time, so it may mix many conditions together, yielding not an instantaneous distribution, but an averaged one.

Ideally, we would like the true distribution to be bounded at every instant. If the conditions did not change over time, then the average and instantaneous distributions would be the same. Unfortunately, we know that some conditions do vary with time, the satellites age, the clock and ephemeris estimation accuracy varies with observability, etc. Therefore, it is best to collect data over many sampling periods and compare each set of results to the others, so that we may better understand how conditions may change over time. If we know of changing conditions, we should attempt to partition data sets to group like behaviors together. For example, data from a satellite in the shadow of the Earth may be separated from data taken in direct sunlight. By keeping data sets as small as practical and comparing many of them, we can hope to identify unexpected changes should they occur.

Another issue is the specification of probabilities. The current specification says that the probability of the instantaneous error exceeding 4.42 x URA without a timely alert is less than $1\text{x}10^{-5}$ in any given hour. However, how is this to be interpreted and/or evaluated? Looking at past data one will see that either this condition was met or that it was not. It is not possible to tell what the probability was during any given hour. If the probability were equally likely for any given hour, then one could evaluate 10^5 hours to ensure that no more than one was affected. However, this is longer than 11 years. If a larger error is seen during the first few years of operation, should the satellite continue to be evaluated for the full 11.4 years to make sure that no more occur and then decide that each hour did have a $1\text{x}10^{-5}$ chance? If two separate large errors are seen early on but then nothing for the next 20 years, does the satellite satisfy the specification?

It is impossible to know from the above information whether the satellite met the specification and was safe to use at all times. The user conducting an approach during the time of the failure is not helped by the fact that the satellite performed well at all other times. Nor does this large average probability necessarily reflect the instantaneous likelihood. Unfortunately, it is not possible to evaluate such low probabilities without large amounts of data. But shorter periods than 11 years are possible to evaluate. It is also important to understand how such low probabilities are used to assure the position bound.

A distribution tail probability requirement, such as not to exceed 4.42 x URA with probability greater than $1\text{x}10^{-5}$ per hour, may be used in three different ways: it can become an effective upper limit, particularly for even lower probabilities; it can be used to assure multiple large errors are unlikely, so that combined variances of independent errors can be RSSd; and, for RAIM, it can be used to assure multiple large errors are unlikely, so that only one fault mode at a time need be considered. For each use, it is possible to specify alternate means to ensure these goals. At very low probabilities, such as $1\text{x}10^{-8}$ per hour, such a requirement is best evaluated as a not to exceed number. Thus, even though it is formally acceptable to be exceeded once every 11,000 years, in practice this should never be seen.

In order to assure that errors RSS together as expected, one can monitor their RSS. This is simpler and more direct than requiring Gaussian performance and independence, neither of which is likely true. Instead, by examining the chi-square value, one can directly assure that errors are not all simultaneously becoming large. This is not completely ideal as the users will weight the error sources differently, so the chi-square value does not measure exactly what is being used. Nevertheless, it is a reasonably close metric and still very indicative.

For RAIM, an assurance that there will never be a situation when more than one satellite has an error greater than 4.42 x URA could be used to limit subset analysis to one faulty satellite at a time. This is also a much easier requirement to verify than a $1\text{x}10^{-5}$ probability.

Evaluations of probabilities have an ambiguity. They can be monitored over different time-frames using different data sets. Two observers looking at the same satellite may disagree as to whether a certain probability requirement is met depending on how much conforming data they aggregate together with an observed violation. A specification of probability that can be unambiguously tested must include the length of time for evaluation.

Another important characteristic is the duration of the error. A moderate error may be tolerated for a relatively long time because by itself it is unlikely to create a large positioning error. However, large individual ranging errors can much more easily create hazardous positioning errors. A GPS satellite can be seen by nearly half of the Earth and an hours long error could affect many thousands of aircraft. If a large error were to occur, it is better to have it reduced quickly rather than allow aircraft after aircraft to be affected at each visible location.

Following the discussion above, we propose preliminary monitoring criteria for evaluation. These are based upon expected Gaussian behavior, although the underlying behavior need not be strictly Gaussian. We propose monitoring the mean, RMS, 1 x URA, 2 x URA, 3.29 x URA, 4.42 x URA, and 5.73 x URA values of the maximum projected signal in space error. The last five correspond to Gaussian probabilities of 0.32, 0.05, 0.001, 1 x 10^{-5}, and 1 x 10^{-8}, respectively. Assuming a 15-minute correlation time to match the future GPS data upload rate, the proposed evaluation criteria on each individual satellite are:

- The RMS of IURE/URA over any given day shall not exceed 1,
- The absolute mean value of IURE/URA shall not exceed 0.5 over any given day,
- The absolute value of any IURE shall not exceed the URA for more than 7.7 hours in any given day,
- The absolute value of any IURE shall not exceed 1.96 x URA for more than 1.2 hours in any given day,
- The absolute value of any IURE shall not exceed 3.29 x URA for more than 45 minutes in any given 31 day period,
- If the integrity flag is set, the absolute value of any IURE shall not exceed 4.42 x URA for more than 300 seconds in any given year, otherwise shall not exceed except for major service failures, and
- If the integrity flag is set, the absolute value of any IURE shall not exceed 5.73 x URA for longer than 5.2 seconds at any time or location.

If the integrity flag is not set, then the satellite has 1 x 10^{-5} per hour probability of a major service failure occurring. Here a major service failure is defined to be an error of 4.42 x URA or greater and its duration could be as long as 6 hours. Clearly, a satellite that does not have its integrity flag set cannot provide assured position domain integrity without some additional augmentation. However, such a satellite may be used with RAIM. This specification should be recast to make the interpretation of 1 x 10^{-5} per hour clear. If there are 32 satellites without the integrity flags set, an average of 3 per year would experience major service failures. Therefore, the above requirements should be amended to allow up to 3 separate major service failures per year to be excluded. Each service failure must affect only one satellite at a time and up to six contiguous hours may be removed per failure. Events more than six hours apart must be counted as separate, as must events affecting separate satellites. It is also not acceptable for two satellites to have overlapping major service failures (one must end or be alerted to the user before another may start).

Figure 1 provides an illustration of several of these checks and their implication on the CDF of the errors. This plot is a normal probability plot also referred to as a quantile-quantile (q-q) plot. The y-axis shows the ordered, observed errors and the x-axis corresponds to the quantile (fraction) of the observed data that is below each value.

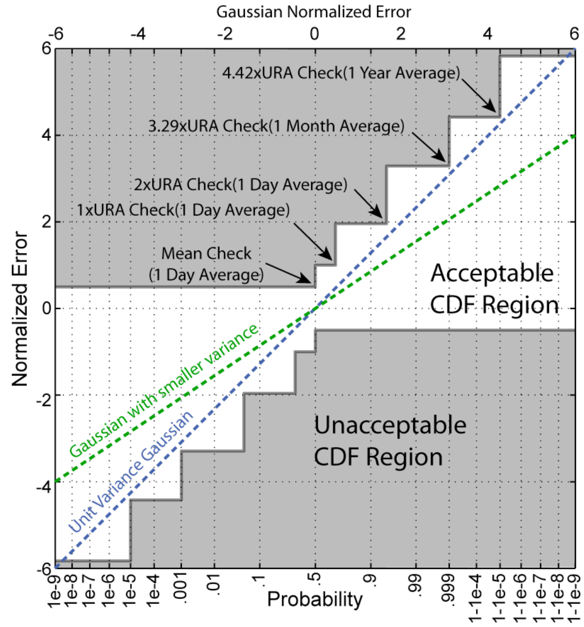

Figure 1. *The normal probability plot is shown with the proposed evaluations highlighted. The actual CDF must not enter the shaded region. A standard, unit-variance Gaussian distribution is shown as the diagonal line. A Gaussian with a smaller variance would have a smaller slope as shown.*

The bottom shows the probabilities and the top axis shows the corresponding Gaussian values. If the observed distribution were a zero-mean, unit-variance Gaussian, the data would fall along the diagonal line shown. If the data were zero-mean, but with a smaller variance, it would still follow a straight line but it would have a smaller slope as shown. Non-Gaussian distributions will have curved lines. The proposed evaluations are shown as corners in a staircase boundary on the plot. Data that passes these evaluations will lie in the white unshaded region. Data that fails will cross the boundary into the shaded region.

Note that a typical q-q plot is usually made up of a single set of data rather than the different time periods noted. Further, the tests are to apply to any such period, for example any 24 hour period, not merely ones that begin and end at midnight. Thus, Figure 1 is more illustrative of how the tests restrict the CDF rather than specifically how they are implemented.

Figure 2 focuses on the upper right-hand quadrant of this plot. It shows that a CDF that avoids the shaded region can also be bounded by a Gaussian with a mean of 0.9 and a sigma of 1.33. Thus, these discrete evaluations can be replaced with a single Gaussian form. The concept of paired bounding [4] can be applied. If the actual distribution is always to the right of N(μ,σ) and to the left of N(-μ,σ), then it can be said to be bounded by N(μ,σ). It will also be bounded by N(μ,σ) if it can be right-left bounded by a Gaussian with the same means, but a smaller sigma. Thus, a nearly arbitrary distribution in the acceptable region of Figures 1 and 2 can be Gaussian bounded with known attributes.

Figure 2 also suggests that a tighter Gaussian bound may be possible if the evaluation points are trusted and one is not concerned that the actual CDF may have the worst-case properties in between each evaluation point. In this case, the Gaussian N(.5, 1) bounds the evaluation points and likely behavior in between. Bounding with this smaller Gaussian may translate into a significant availability increase. Here we begin to see the linkage between the evaluation points and the Gaussian bound used for analysis. It is possible to add more evaluations such that even tighter bounds are possible, or it may be possible to achieve the same level of performance with fewer well chosen evaluations.

The corresponding Gaussian bound from paired bounding directly connects back to the α parameters in the VPL equations completing the link between them and the monitoring. The mean parameter corresponds to α_2 and the sigma parameter to α_1. Thus for the conservative bound in Figure 2, we would have $\alpha_1 = 1.33$ and $\alpha_2 = 0.9$. For the less conservative bound we could use $\alpha_1 = 1$ and $\alpha_2 = 0.5$. The difference in availability between these two is likely significant. The smaller values may be used if one may safely assume that the expected behavior is unlikely to be worst-case. If this is not an acceptable assumption, additional evaluations may reduce the conservatism of the worst-case overbound.

MULTIPLE SATELLITE MONITORING

One of the key aspects of having low probability of large errors is that, should a larger error occur on one satellite, similarly large errors on other satellites or from other error sources are unlikely to occur at the same time. Another means of evaluating whether or not one or more errors are combining to form a very large position error is to look at the sum of the square of the normalized errors. If the errors are close to Gaussian and independent, then this sum will be close to chi-square distributed. By formally evaluating the chi-square, it is possible to ensure that the RSS of the errors is close to independent Gaussian expectations.

Therefore we propose an evaluation of the form:

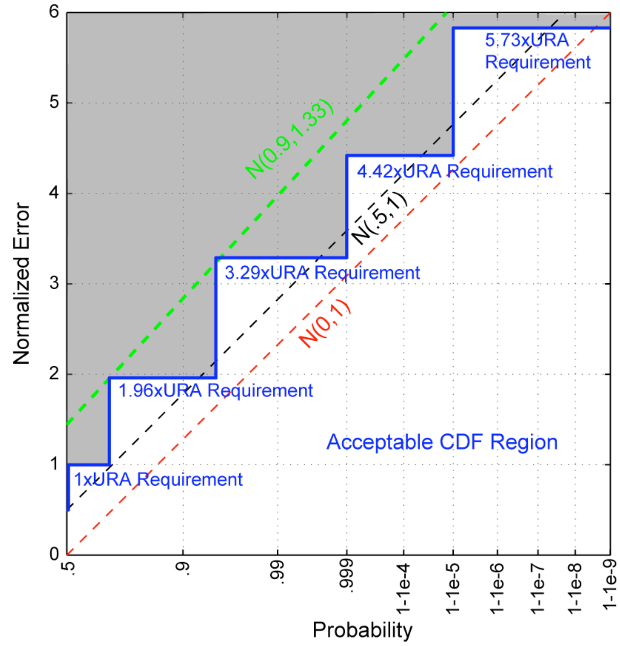

Figure 2. *This figure shows how Gaussian overbounds may be determined from the evaluation. A smooth distribution that stays inside the acceptable region may look similar to the zero-mean, unit-variance Gaussian shown on the diagonal. Such a distribution stays within the right-most corners of the boundary. This boundary can be represented by a half-unit mean, unit-variance Gaussian. A worse distribution that only just stays within the boundary can be bounded by a more conservative Gaussian model.*

$$\sum_{i=1}^{n}\left(\frac{IURE_j^i(t) - \langle IURE_j(t)\rangle}{URA^i(t)}\right)^2 \le K_{prob}^2 \qquad (5)$$

where a common mode clock term, $\langle IURE_j(t)\rangle$, is removed from the evaluation. The next question is at what probabilities and time periods should such evaluations be made? Following chi-square statistics, an upper bound on the expected value, assuming nine degrees of freedom and 10^{-7} probability, leads to a value of ~7.1 for K_{prob}.

We propose adding the following monitoring requirement:

- For satellites with the integrity flag set, the sum of the squared ratios (IURE minus a common clock term divided by URA) shall not exceed 50.2 for longer than 5.2 seconds at any time or location.

To understand the potential application of this monitor limit we look at the RSS component of the VPL equation coming only from the SIS terms:

$$VPL_{SIS} = K_V \sqrt{\sum_{i=1}^{n} s_{3,i}^2 \times \alpha_1^2 \times URA_i^2} \quad (6)$$

This is to be compared to the Vertical Position Error (VPE) coming solely from the SIS errors:

$$VPE_{SIS} = \left| \sum_{i=1}^{n} s_{3,i} \times IURE_i \right| \quad (7)$$

This can be rewritten as:

$$VPE_{SIS} = \left| \sum_{i=1}^{n} s_{3,i} \times \frac{IURE_i}{URA_i} \times URA_i \right| \quad (8)$$

According to the Cauchy-Schwartz inequality:

$$VPE_{SIS} \le \sqrt{\sum_{i=1}^{n} s_{3,i}^2 \times URA_i^2} \times \sqrt{\sum_{i=1}^{n} \frac{IURE_i^2}{URA_i^2}} \quad (9)$$

From (5) the second radical can be replaced with K_{prob} leading to:

$$VPE_{SIS} \le K_{prob} \times \sqrt{\sum_{i=1}^{n} s_{3,i}^2 \times URA_i^2} \quad (9)$$

Therefore, provided $K_V \times \alpha_1 \ge K_{prob}$ we can be assured that $VPE_{SIS} \le VPL_{SIS}$ to the required probability. This is only rigorously demonstrated when the other error terms are neglected, but because these other error terms are independent of the SIS errors and if both are close to Gaussian, the full VPL equations (1)-(4) are expected to hold as well.

This leads to the requirement that if $K_V = 5.33$ as in SBAS and $K_{prob} = 7.1$ as above, then $\alpha_1 \ge 1.33$. This coincidentally matches the conservative overbounding value from the previous section. Smaller values may be possible by exploiting other expected properties of the $s_{3,i}$ factors or by exploiting existing conservatism in the URAs to set K_{prob} lower. In the next section we will see that historical data has much lower observed chi-square values.

INDIVIDUAL SATELLITE RESULTS

To begin to see whether these proposed monitor limits are reasonable, we examined historical GPS data from the year 2008. IUREs were approximated by comparing the broadcast satellite clock and ephemeris data to precise values as determined by NGIA [15]. These errors were then projected onto the surface of the Earth to determine the maximum projected error at a given time [16].

These errors only approximate the actual IURE as they only include the satellite clock and ephemeris errors. Other possible SIS errors such as signal deformation [10] [11], antenna group delay variation [9], and others are not included in this formulation. Further, we only have precise orbit information available every 15 minutes and even so, they are not always available for every satellite at every 15-minute epoch. Thus, some errors may be missed if they are shorter in duration than 15 minutes or precise estimates are not available at the time. If an anomaly makes it difficult to determine a precise ephemeris estimate, then that data may be missing and the anomaly will not appear in this record.

Some errors may be alerted to the user in ways that are not evident in the recorded broadcast ephemeris data. If the satellite switches to non-standard data, it may not be reflected in the data investigated. That is, precise orbits may still be calculated, but new ephemerides that fail parity checks are not recorded. Consequently, some of the anomalies observed may not have really affected users as they were alerted by alternate means than the health bits in the ephemeris. These are still being investigated to determine whether users were actually affected. For this analysis, we will assume that until the health bits were set unhealthy, users would incorporate that signal.

The most significant limitation of evaluation with current or previous data is that the broadcast ephemeris information is typically uploaded to the satellite only once per day. Thus, errors are strongly correlated over many hours instead of the 15-minute period assumed in selecting the proposed evaluation periods. This is unfortunate because the evaluation periods would need to be extended by a factor between 10 and 100. However, even the lower end of this range starts to make the evaluation periods much too long to be practical. Therefore, we will start by using the proposed evaluation periods, although we would not necessarily expect even well-behaved satellites to pass all tests.

For the existing GPS interface, the minimum possible broadcast value of URA is 2.4 m. It appears that much of the time the maximum projected IURE is much smaller, leading to excess margin in the URA. It is not clear if this margin would be maintained if it were possible to broadcast smaller values. The future signal on L5 does allow smaller values to be broadcast and values of 0.7 m

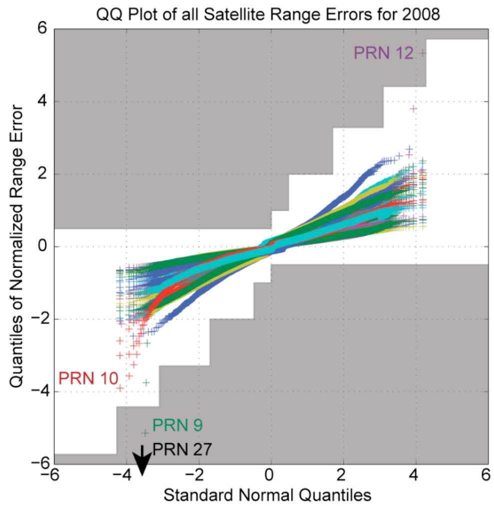

Figure 3. *This figure shows the q-q plot for the full year's worth of data for all satellites. In this plot, all but three satellites appear to lie within the acceptable CDF region. However, several of the satellites fail the shorter time period evaluations.*

are being targeted. For 2008, the URAs appear to be fairly conservative and many satellites pass all of the evaluations proposed despite the large difference in correlation time and update rate (24 hours vs. 15 minutes). Some of the satellites fail a few daily tests (RMS, mean, 1 x URA, and/or 2 x URA). Three of the satellites have apparent major service failures (errors greater than 4.42 x URA). Again, these are unconfirmed failures as they may have been alerted by a different means than the ephemeris health bits. The affected satellites are PRN 12, on April 2, PRN 9 on June 7, and PRN 27 on November 14. There are unscheduled outage NANUs associated with each event.

Figure 3 shows the results for all satellites evaluated for the full year. Note that most satellites are very well behaved, that is, they are reasonably linear indicating Gaussian behavior and their slopes are noticeably smaller than 1, indicating margin in the URA. All pass nearly through the intercept indicating small mean values. For this long time period all but the three previously identified satellites appear to pass the evaluations and lie in the acceptable region. However, when evaluated over shorter periods of time, several of the satellites fail an occasional daily test. These smaller data set tests are an indication of potentially non-stationary behavior.

To better understand the type of problem that these daily tests may identify, we examined PRN 14. PRN 14 is a Block IIR satellite launched in late 2000. It failed the RMS, mean, 1 x URA, and 2 x URA tests around day 45 of 2008. Figure 4 shows the RMS test evaluated for overlapping 24-hour periods every 15 minutes over the course of the year. Around day 45, the RMS value increases to more than triple its typical upper values. This indicates that the satellite behavior at this time may be unlike its behavior during the rest of the year.

Figure 5 looks at the time history of the maximum projected error around the day in question. Up through day 44, the behavior is extremely good. However, at the

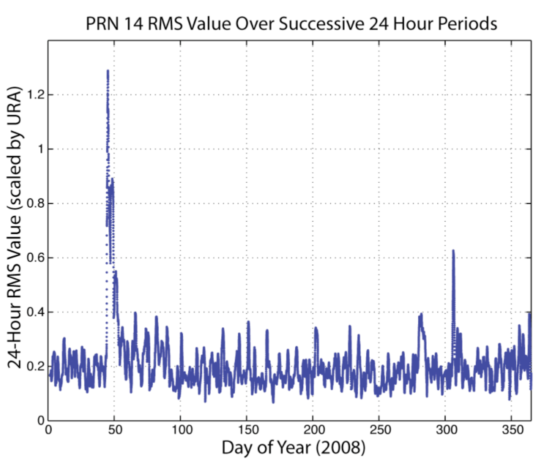

Figure 4. *The RMS of the maximum projected error divided by the URA is shown for PRN 14 for all of 2008. These are overlapping data sets accumulating 24 hours of data evaluated at 15 minute steps. Near data 45 a short but significant increase is observed indicating non-stationary behavior.*

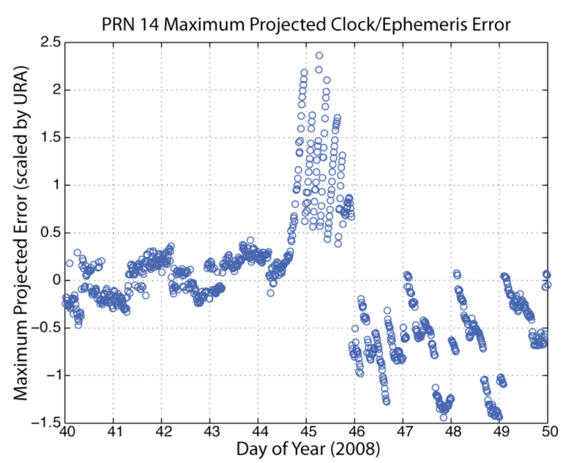

Figure 5. *The maximum projected clock and ephemeris error is shown for PRN 14 around the time of the anomalous behavior. Clearly, there is a change of behavior late on day 45 with a more quickly changing error that is not fully compensated by the broadcast parameters. Starting on day 46 another change in behavior is observed.*

end of day 44 and into day 45 we see that the broadcast parameters are not describing the actual satellite performance nearly as well, and after day 45 we see another change in behavior. During this time, the error never went above 2.5 x URA, so it is hard to say for certain what harm this effect causes. Around day 45 the statistics for PRN 14 are different from other times, but because the event only lasts for a day, it is hard to fully characterize its new behavior. It does appear that the probability of large error was much higher during this period. Figure 4 indicates that at a minimum the expected RMS behavior was at least three times worse during this event. Thus, any effort to exploit the margin seen during the rest of the year, would be limited by such behavior. Such events make it hard to place too much confidence in the future performance of the satellite. Other Block IIR satellites that were launched near the same time do not exhibit this same behavior. Does this anomaly indicate that all satellites are susceptible to such behavior even if they did not experience it in 2008? Satellites that pass all tests provide much greater confidence that their operation will continue to be well behaved into the future.

Figure 6 shows the q-q plot for PRN 14. The CDF is distinctly non-linear and hence non-Gaussian. A year's worth of good data does not hide the bad day's results in this case, although the daily tests can be made to pass by extending evaluation period to just 3 days. This is largely due to the excess margin in the URA value. Thus, such tests still do not guarantee finding all non-stationary behavior. Although there is too little data here to say for certain that this CDF is unsafe, it does not firmly indicate that it is safe either. Aviation integrity works more from the principle of guilty until proven innocent, which means that this behavior is a cause for concern. Precisely what happened to PRN 14 on day 45 of 2008 should be further investigated to understand how likely it is to happen again, to learn what are the appropriate statistics during such an event, and to determine if are other satellites likely to be similarly affected.

Contrast the behavior of PRN 14 with the performances of PRNs 15, 19, 21, and 23 shown in Figure 7. These are the best performing satellites of 2008. Their behavior is very linear, with no indication of significant non-Gaussian behavior. The observed maximum projected error never exceeds 1 x URA. Thus, there is much excess margin in their performance that could be exploited. Further, these pass all daily and other tests and closer inspection reveals no evidence of significant variability from day to day. PRNs 19, 21, and 23 are Block IIR satellites launched in 2003 and 2004. PRN 15 is a Block IIR-M satellite launched in 2007. It is not clear why these four are much better performing than many others from the same blocks with similar or more recent launch dates. This aspect also needs to be further investigated as one would expect identically designed satellites to perform similarly barring failures.

Figure 8 shows both the nominal and yearly overbounding sigma values as a fraction of the broadcast URA for each

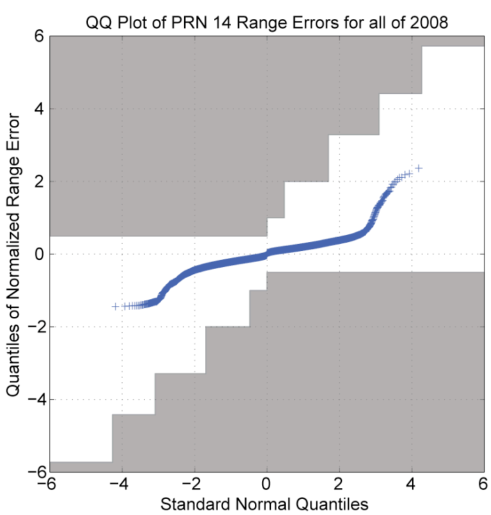

Figure 6. *This figure shows the q-q plot for the full years worth of data for PRN14. Although the data appears to pass all requirements, its behavior is distinctly non-Gaussian. This is also indicated in the daily evaluations. It is hard to be confident that future behavior will remain within the acceptable region.*

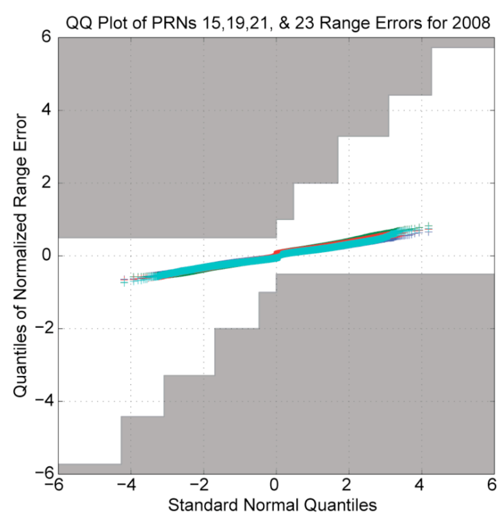

Figure 7. *This figure shows the q-q plot for the full year's worth of data for PRNs 15, 19, 21, and 23. All satellites here easily pass all evaluations. In fact the maximum error never exceeds 1 x URA and the overall behavior is exceedingly Gaussian.*

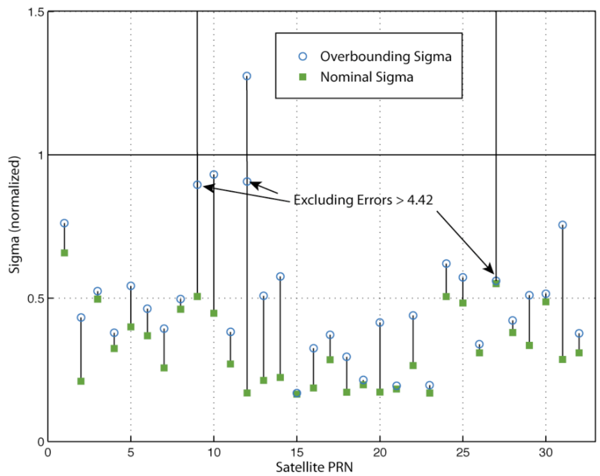

Figure 8. *This figure shows the nominal and overbounding sigma for each satellite. The green square is the nominal value (majority of data) and the open circle bounds the tails. For PRNs 9, 12 and 27 the open circle is calculated both with and without the errors greater than 4.42 x URA.*

PRN. The green square is the result of a linear fit to the q-q CDFs and represents the nominal value or sigma of the majority of the data. The open circles are calculated by finding the minimum value that would keep the q-q curve no worse than a unit-variance Gaussian. It is generally driven by the tails. A large difference between the two indicates non-Gaussian behavior. The best four PRNs have both the lowest overall sigma values and the open circles lie right on top of the green squares. PRN 10 by contrast has a very large difference between its nominal and tail behavior.

PRNs 9, 12, and 27 have open circles calculated both including and excluding the possible major service failure points. As can be seen many of the satellites could have safely used URA values close to one third of what was broadcast. The best four satellites, could divide their URA by nearly four. However, some satellites were only just covered by the existing URA. In order to exploit the margin in the best performing satellites, there would need to be a way for the control segment to distinguish between their performance in real time and send larger URA values for the worse performing satellites.

CORRELATION RESULTS

The previous section examined the behavior of individual satellites against the Gaussian ideal. This section looks at the potential for correlation among multiple satellite errors. Here we evaluate the chi-square value across all satellites in view. Instead of looking at the maximum projected error, each satellite error is projected to a five-degree by five-degree grid of users on the surface of the Earth. A common clock term is removed from each user error at each epoch, and the remaining error is divided by the URA. These normalized residuals are then squared and added together to form the chi-square value.

Figure 9 shows the histogram of values where we have excluded the three major service faults from the calculation, but included all other errors. As can be seen performance is quite good with a maximum observed value of 25.4. The corresponding K_{prob} value, if this were an upper bound, is 5.04. This value works very well with the K value used in the VPL equation. However, the result is not quite as optimistic as it initially appears. The average nominal sigma value from Figure 8 is close to one third. Therefore, the expected chi-square values should be reduced by approximately a value of nine. Indeed, this is nearly the case for the average chi-square value. However, as already observed, the tail extends out to 25.4 which is very large compared to the mean. For this volume of data, we would expect about a five to one ratio between the maximum and average. Instead we observe a value greater than 17. Thus, large chi-square values appear out of proportion to smaller values. We saw that this was often the case with the individual distributions in Figure 8.

We also observed that, for the most part, each satellite was well behaved although some exhibited non-Gaussian behavior and increased tails relative to the nominal. The concern in this section is the possibility of correlated behavior. To assess the effect of a single error versus a combination of several smaller ones, we calculated the chi-square value removing the single largest residual error at each location and at each time step. Thus, the effects of individual errors would be eliminated, but multiple smaller ones would still be included. Figure 10 shows that in this case, the largest observed value is 5.18. This is much more in line with the expected five to one ratio to the mean value and broadcast URAs that are often three times larger than required.

The chi-square data in Figure 10 indicates that correlated SIS errors were not an issue during 2008. There were individual satellite errors that grew large compared to the URA and, in three cases, may have led to major service failures. However, at any given time there was no more than one such large error. The N-1 chi-square evaluation in Figure 10 is consistent with the observed individual performance in Figure 8 and near independence of the errors. Together these can be used to demonstrate that

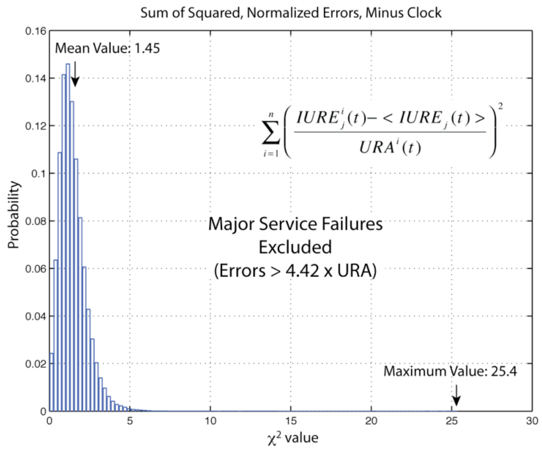

Figure 9. *The chi-square values is shown for a grid of terrestrial users is shown excluding the suspected major service failures. Although the largest value is well within expectations, it is very large compared to the main distribution.*

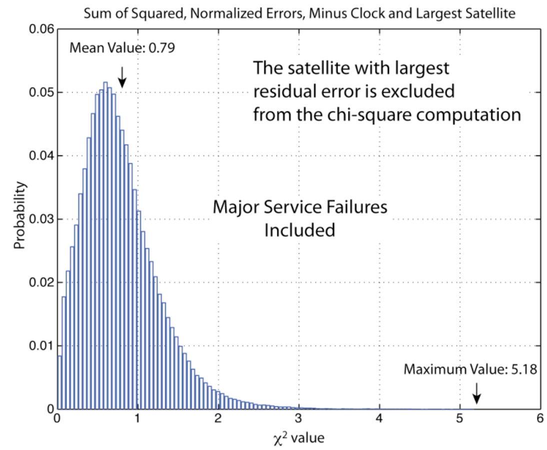

Figure 10. *Here the chi-square values are calculated excluding the single largest error at each location. Now the main distribution and tail are much more consistent indicating that the residuals typically only ever experience a single large error at a given time.*

RSSing the errors, as in the ARAIM VPL equations, would have been safe with the existing data.

CONCLUSIONS

We have proposed specific data monitoring evaluations that are both unambiguous and directly linked to requirements that support VPL equations. We have proposed evaluations at both the core and tails of the error distributions as well as on the sum of the squared errors. These evaluations have been preliminarily evaluated on actual data from 2008. It has been shown that the tests are effective at identifying behavior that requires further investigation to determine its impact. These tests are linked to Gaussian models that in turn set the alpha parameters in the VPL equations. Tighter monitor limits can be directly translated to smaller VPLs. Once these monitor limits are established, the alpha values can be determined and provided to the users. Each sovereign state would have the ability to determine their own alpha values and indicate which satellites may be safely used. Thus, they would have much more confidence in and control over integrity determination within their airspace.

Although the vast majority of GPS data from 2008 indicates excellent behavior, there are some subsets of data that behave very differently from others. Further, different satellites have very different levels of performance, some having large margins compared to the broadcast URA while others have very little margin. Ideally, the satellites would be much more consistent in their behavior with regard to the margin in the URA.

It is important to remember that the monitoring does not assure integrity on its own. Rather, it is a properly designed system, carefully analyzed to ensure that it is capable of meeting such monitoring requirements, that assures safety. The monitoring requirements merely provide confidence that the system continues to meet its design goals.

ACKNOWLEDGMENTS

The authors would like to gratefully acknowledge the contributions from Professor Jiyun Lee of the Korean Advanced Institute for Science and Technology (KAIST) for her initial evaluation of GPS data, and Dr Grace Gao and Mr. Liang Heng of Stanford University for their preparation of the 2008 data. The work for this paper was supported by the FAA Satellite Product Team under research grant 95-G-005.

REFERENCES

[1] McDonald, K. D. and Hegarty, C., "Post-Modernization GPS Performance Capabilities," Proceedings of ION annual Meeting, San Diego, CA, 2000.

[2] Van Dierendonck, A. J., Hegarty, C., Scales, W., and Ericson, S., "Signal Specification for the Future GPS Civil Signal at L5," Proceedings of ION annual Meeting, San Diego, CA, 2000.

[3] Hegarty, C. J. and Chatre E., "Evolution of the Global Navigation Satellite System (GNSS)," in Proceedings of the IEEE Vol. 96, Issue 12, Dec. 2008.

[4] Rife, J., Pullen, S., Pervan, B., and Enge, P. "Paired Overbounding and Application to GPS Augmentation" Proceedings of the IEEE Position Location and Navigation Symposium, Monterey, CA, April 2004.

[5] GLOBAL POSITIONING SYSTEM STANDARD POSITIONING SERVICE PERFORMANCE STANDARD, 2001 available at http://www.navcen.uscg.gov/gps/geninfo/2001SPSPerformanceStandardFINAL.pdf.

[6] Kovach, K., Dobyne, J., Crews, M., and Miles, C., "GPS III Integrity Concept," Proceedings of ION GNSS 2008, Savannah, GA, 2008.

[7] Walter, T., Blanch, J., and Enge, P., "L5 Satellite Based Augmentation Systems Protection Level Equations," in Proceedings of the International GNSS conference, Sydney, Australia, December, 2007.

[8] Walter, T., Enge, P., and DeCleene, B., "Integrity Lessons from the WAAS Integrity Performance Panel (WIPP)," in proceedings of ION NTM, Anaheim, CA, January 2003.

[9] Shallberg, K. and Grabowski, J., "Considerations for Characterizing Antenna Induced Range Errors," Proceedings of ION GNSS 2002, Portland, OR, 2002.

[10] Mitelman, A. M., Phelts, R. E., Akos, D. M., Pullen S. P., and Enge, P. K., "Signal Deformations On Nominally Healthy GPS Satellites," Proceedings of the National Technical Meeting of the ION, San Diego, CA, 2004.

[11] Phelts, R.E., Walter, T., and Enge, P., "Characterizing Nominal Analog Signal Deformation on GNSS Signals," Proceedings of ION GNSS 2009, Savannah, GA, 2009.

[12] RTCA, "Minimum Operational Performance Standards for Global Positioning System/Wide Area Augmentation System Airborne Equipment," RTCA publication DO-229D, 2006.

[13] Walter, T., Enge, P., Blanch, J., and Pervan, B., "Worldwide Vertical Guidance of Aircraft Based on Modernized GPS and New Integrity Augmentations," in Proceedings of the IEEE Vol. 96, Issue 12, Dec. 2008.

[14] Blanch, J., Ene, A., Walter, T., Enge, P. "An Optimized Multiple Solution Separation RAIM Algorithm for Vertical Guidance". Proceedings of the ION GNSS 2007, Fort Worth, TX, September 2007.

[15] Wiley, B., Craig, D. Manning, D., Novak, J., Taylor, R., and Weingarth, L. "NGA's Role in GPS," Proceedings of the ION GNSS 2006, Fort Worth, TX, September 2006.

[16] Gao G. X., Tang, H., Blanch, J., Lee, J., Walter. T., and Enge, P., "Methodology and Case Studies of Signal-in-Space Error Calculation -- Top-down Meets Bottom-up," Proceedings of ION GNSS 2009, Savannah, GA, 2009.