# Validation of the Unfaulted Error Bounds for ARAIM

## Todd Walter, Kazuma Gunning, R. Eric Phelts, and Juan Blanch Stanford University

Advanced Receiver Autonomous Integrity Monitoring (ARAIM) requires accurate modeling of the unfaulted satellite Signal-In-Space (SIS) error distributions in order to properly calculate integrity risk. This error distribution is most commonly described by two terms: the nominal bias,  $b_{nom}$ , and the User Range Accuracy, *URA*. The nominal, or unfaulted, bias describes unchanging errors such as those due to nominal signal deformation or satellite antenna bias. These error sources are expected to remain the same for the user each time they observe the satellite in a similar manner. Fortunately, they are also expected to be small (sub-meter) compared to other error sources. The broadcast *URA* is treated as a 1-sigma parameter (which we will denote  $\sigma_{URA}$ ) and is used to describe the constantly changing error due mainly to inaccuracies in the broadcast satellite clock and ephemeris parameters. These numbers are set to create upper bounds on the true instantaneous values. Thus, a bounding Gaussian distribution is described where negative errors are no more likely to occur than predicted by N(- $b_{nom}$ ,  $\sigma_{URA}^2$ ) and positive errors are no more likely to occur than predicted by N( $b_{nom}$ ,  $\sigma_{URA}^2$ ).

This paper examines how to evaluate the observed instantaneous SIS errors and determine suitable values for  $b_{nom}$  and validate the broadcast  $\sigma_{URA}$ . We compare performance against the commitments and broadcast values from the satellites to determine whether the provided values are sufficient or not. An important aspect is to characterize the errors in light of known or predictable characteristics. Oftentimes errors are grouped together to create a single averaged distribution. However, there may be times and conditions where performance is notably worse. We need to separate out such conditions and evaluate the distributions individually so as not to form overly optimistic estimates of the error bounds. Once the representative data sets have been selected, we need to estimate the appropriate bounding parameters. Further we must ensure that these parameters will continue to bound future fault-free behavior. We will describe the conservative steps taken in the estimation process and the validation effort, both with the real data and versus the stated commitments from the constellation service providers.

## **INTRODUCTION**

An Advanced Receiver Autonomous Integrity Monitoring (ARAIM) user is affected by several error sources [1] [2]. These may be separated into those that arise on the navigation satellites, those that affect the signal during propagation, and those that arise from the receiver or its surrounding environment. All are important and must be accurately described. However, the focus of this paper will be determining an appropriate description of the errors arising from the satellite. These so called Signal-In-Space (SIS) errors may be due to a variety of causes including: satellite clock and position estimation errors, signal generation errors, and antenna characteristics. Inaccuracies in the broadcast values for the satellite location and clock states typically lead to the largest source of SIS errors. However, as these errors are reduced over time, the other SIS error sources grow in importance.

This paper examines the nominal SIS error characteristics. A previous paper [3], examined anomalous behavior and how to determine fault rates. This paper examines how to describe the unfaulted nominal behavior. In the prior paper, faults were defined to be instances when the instantaneous SIS ranging error exceeded a set threshold. This threshold is a fixed multiplier of the broadcast User Range Accuracy (URA). According to the GPS interface specification [4]: "URA provides a conservative RMS estimate of the user range error (URE) in the associated navigation data for the transmitting SV. It includes all errors for which the Space and Control Segments are responsible." As a result we will label this value as  $\sigma_{URA}$  throughout the paper. If a fault is defined to occur whenever the instantaneous URE (IURE) is greater than 4.42 x  $\sigma_{URA}$ , then whenever IURE is below this threshold, must correspond to unfaulted or nominal conditions. We will examine the corresponding data set for this paper.

An important decision is how to model and describe such errors so that a user knows how they may affect their position solution uncertainty. These errors may vary both spatially and temporally, as well as with spacecraft type, or by individual satellite. The error distributions may or may not conform to simple models. Previous studies have examined the error characteristics of the broadcast satellite position and clock errors and found that, in general, smaller errors are much more likely than larger errors and that they are also largely uncorrelated in the longer term. Given the mathematical simplicity afforded when convolving errors using a Gaussian model, it is often the model of choice. A significant advantage of using a Gaussian model to describe error sources is that many overbounding analyses use it as a basis. These analyses describe how an overbound of an individual error source may be combined with other individual overbounds to create an overbound of the combined error (including the final user positioning error).

A Gaussian model contains two terms: a bias and a variance. Prior uses of navigation satellites to support aviation (e.g. Satellite Based Augmentation Systems, SBAS, and Receiver Autonomous Integrity Monitoring, RAIM) tend to set the bias term to zero. This works well when the biases are small compared to the standard deviations and the potential number of combined error sources is small. However, as the relative magnitude of the bias increases and/or the number of potential biases increases, it is better to treat the bias term separately. While the clock and ephemeris errors are frequently changing with little long-term correlation, the signal generation errors and antenna characteristics exhibit quasi-static behavior [5] [6]. Further the clock and ephemeris errors may have bias terms (including the mis-estimations of the inter-frequency term.) The signal generation errors may include effects such as signal deformation [5] [7] and code-carrier divergence. The latter is generally negligible, but its potential effect on L5 needs to be monitored [8] [9]. This paper will focus on clock and orbit, signal deformation, inter-frequency biases, code-carrier incoherence, and antenna effects as the dominant sources of bias.

## **GPS DATA**

In previous works [10] [11], we have described techniques to identify and remove erroneous recordings of broadcast GPS ephemeris parameters. After correctly determining what each satellite broadcast at any given time, we compared the broadcast values to post-processed precise orbit and clock values as determined by the National Geospatial-Intelligence Agency (NGA) [12]. This data allows us to obtain accurate estimates of the instantaneous orbit and clock errors for each satellite every fifteen minutes. We can use such data to identify faulted broadcast ephemerides [3]. In this paper we are examining the nominal or unfaulted error behavior. For this purpose, we have chosen to evaluate the GPS constellation over the last four years so that we may obtain a reasonably up to date view of the constellation performance. Further, there have not been any major service failures (i.e. faults >  $4.42 \times \sigma_{URA}$ ) since June of 2012. We therefore have analyzed all of 2013 through all of 2016 and there was no need to partition the data into nominal and faulted; everything was nominal.





**Figure 1.** Summary of observations for each GPS satellite, where green indicates good observations, blue that the satellite is unhealthy, magenta and yellow indicate missing data

**Figure 2.** Radial, along-track, cross-track, clock, and projected error distributions of the GPS satellites

Figure 1 shows an overview of the data analyzed. The vertical axis identifies each satellite observed by their satellite vehicle number (SVN). The colored squares on either side of the line indicate the colors that will be used to plot individual satellite data in subsequent plots. Each horizontal line indicates a type of measurement for each satellite. Green indicates that the broadcast ephemeris was set healthy and a valid comparison to the precise ephemeris was obtained (the data set contains 4,293,707 such comparisons). Blue indicates that the broadcast ephemeris was set to unhealthy and therefore no comparison was made. Magenta indicates that no broadcast ephemeris was obtained from the IGS database, but that there were precise orbit data. Yellow indicates that broadcast ephemerides were obtained, but no precise orbit data was available. Finally black indicates that the satellite was operational, but we have neither broadcast or precise orbit data. The gray background line indicates that the satellite was not operational (e.g. not yet launched, decommissioned, or not broadcasting).

Figure 2 shows the observed probability density functions (PDFs) for the radial, along-track, cross-track, clock, and projected user ranging errors. The vast majority of the errors appear to be nearly Gaussian. There is evidence of mixing starting around the  $10^{-3}$  level, particularly for the clock. By mixing we mean that small subset of the data appears to have a larger variance than the majority. Below the  $10^{-5}$  level there is clear indication of some data with significantly larger errors. However, the data also clearly indicates that a Gaussian model appears to be a reasonable descriptor for the vast majority of the data. The increased mass at the tails can be conservatively handled by inflating the assumed variance. In order to determine appropriate overbounding parameters for the data, we must first decide on more precise definitions of the errors and then decide on how to aggregate the data. The next two subsections address these points.

## **IURE Definition**

Two different definitions for projected error have been used in previous analyses: maximum projected error (MPE) and user projected error (UPE). MPE is the maximum satellite orbit and clock error projected onto Earth at a particular time [11]. There is one value per healthy satellite at each time epoch. The MPE can only take on the value of zero if the three orbital errors and the clock error are all simultaneously zero. The MPE can sometimes switch rapidly between positive and negative as the corresponding projections change. As a result the MPE distribution is bimodal with a notch at zero. This error distribution is not expected to be Gaussian even if all the underlying distributions were Gaussian. The UPE is the projected satellite orbit and clock error at a specific user location. Each user can see many, but not every healthy satellite at any given epoch. Unlike MPE, there will be multiple UPE values per satellite at each epoch (one for each user that has the satellite in view). If the underlying errors are Gaussian, the UPE distributions will also be Gaussian, both at each individual user location and aggregated across all user locations. For this reason, we advocate using UPE as the error to be used to evaluate the core of the error distributions. Both MPE and UPE are well suited to describe the tail behavior. We have selected 200 evenly distributed user locations around the globe [11]. This density has been found to be sufficient such that a value within 2 cm of the MPE will be observed at one or more of the user locations. Other users may see much smaller errors at the same time. In evaluating performance, we will label whether we are using MPE or UPE and in some cases take the maximum value across the two options.

## **GPS Data Partitioning**

Another important question is: how to aggregate data for evaluation? Infrequent hazardous conditions run the risk of being hidden by more common, benign data. Thus, it is best to try to form partitions that only contain similar underlying conditions. However, if a partitioned data set is too small, it will not be statistically significant. Further, how do we separate and identify the more hazardous conditions? We cannot simply select data that has larger magnitude errors as they may not have any underlying common characteristic. Instead, we should partition our data by independent means that reflect expected performance variations. The IURE magnitudes and/or signs should not be included as a basis for partitioning data (although investigating patterns of larger errors may be useful for determining which external conditions are most hazardous). Various partitions have been suggested. These include:

- Individual satellites
- Satellite block type (including clock type)
- Time (by year, by season, by month, or by day)
- Satellite age
- Age of navigation data
- Broadcast URA value
- Combinations of above



Figure 3. Monthly mean UPE for each GPS satellite

**Figure 4.** Monthly 95% bound of UPE on each satellite and across all GPS satellites (heavy black line)

As these criteria are combined, the sample size can become very small. Unfortunately it is not clear at what number a data set becomes too small. Further, it is not merely a matter of sample size, 100 data points collected over two minutes can have very different properties than 100 samples collected over several days. Neither will likely provide very much useful information and certainly not about low probability event (e.g. below 10% likelihood). A recent DLR study [13] concluded that MPEs had a correlation time on the order of 12 hours. Our own investigations reveal that the errors on the broadcast orbital positions contain 12-hour periodic components that take many days to decorrelate. The errors on the broadcast clock terms depend on the clock type. The error terms for the cesium clocks are larger and have a correlation time on the order of ten hours, and the error terms for the rubidium clocks are smaller, but also contain 12-hour periodic components that take many days to decorrelate. Similar findings have been previously reported [9]. Thus, there may be as few as two independent samples per satellite per day. Creating a monthly partition for an individual satellite may have fewer than 100 independent samples. It is not necessarily appropriate to draw conclusions about tail behavior from such small datasets. Even core parameters, such as the sample mean and sample standard deviation may be subject to noticeable variations due to the small sample size. If the 12-hour correlation time is correct, it will require a minimum of  $\sim 40$  satellite years worth of data to approach significance for events that have probabilities of 10<sup>-4</sup> or below. This is longer than the lifetime of the individual satellites. Even if all of the satellites in a constellation could be aggregated together, it would require more than a full year's worth of data to match this level of significance.

Although the nominal errors have long correlation times, faults can be comparatively short. Many of the faults observed in the last nine years lasted about fifteen minutes each. A fault is a concern even if it lasts for only seven seconds (faults that remove themselves within six seconds can be ignored). Clock errors can grow very rapidly and navigation data errors can produce extremely large step errors. Therefore, even though the nominal errors have a correlation time of ten hours or longer, the sampling interval ideally should be set to six seconds or below. To date we have evaluated performance using fifteen-minute sampling intervals, but we are developing tools to process the data with a five-second interval. Higher sampling rates do not affect most of the metrics used here to evaluate performance. Instead, the long correlation times lead to a particular nominal error value being sampled many times. One should not mistake the increased sample count as providing information about the error distribution to smaller probabilities. Even though the five second sampling will lead to 180 times as many data points, the effective sample size will not be changed. Instead, the benefit is that we can be certain we have not missed any significant transient anomalies.

Figures 3 and 4 show monthly mean and monthly 95% bounds on UPE (separate colors for each SVN and a heavy black line for the 95% bound across all satellites). The most obvious feature is that SVN 64 (PRN 30) had a  $\sim$ 1.3 m average bias during its first month of operation. Fortunately, it quickly reduced to absolute values below 0.1 m for the remainder of the period of investigation. A few other satellites (SVNs 68, 70, and 71) also exhibited increased biases during their first month of operation, but all were below 0.4 m. For the most part, the satellite monthly means appear to exhibit a random pattern of values with most having absolute values below 0.1 m. The 95% error bounds exhibit much more variation. Again SVN 64 has a large value ( $\sim$ 3 m) during its first month of operation, but it settles down close to its mean value of 0.71 m by its third

month of operation. SVN 70 (PRN 32) and SVN 71 (PRN 26) also showed increased initial 95% bounds (~1.75 m) before settling down below 0.7 m. It is hard to declare either a clear temporal pattern or lack of temporal pattern from this data. There is definitely temporal variation, but no clear trends. All 95% error values are well below 4.8 m, which corresponds to a two-sigma bound given the minimum broadcast value for  $\sigma_{URA}$  of 2.4 m.

The variation of the monthly means can also be used to examine the effective sample size. For a stationary distribution, the variance of the means is equal to the variance of the errors divided by the effective number of samples. Therefore, we can take the variance of the errors and divide it by the variance of the monthly means to determine the effective sample number contained in each month. For the error variance we used the average of the monthly 95% bound divided by two and then squared. We also excluded the first two months of operation for SVNs 64, 68, 70, and 71. We found that the estimated values ranged from 20 to 200, but with an average value of 58. This corresponds to about two independent points a day, which is in line with the 12-hour correlation time. We also found that the square root of the variance of the means was about 15 cm. The mean values seen in Figure 3 are mostly within two sigma of the expected deviation. Aside from the first months bias for SVN 64, these biases are nearly all within the expected range that would correspond to a zero mean distribution. At worst, the clock and ephemeris error biases would be on order 20 cm.

There are clearly performance differences among the satellites. Most obviously, the older SVNs (lines in Figures 3 & 4 that tend toward blue) tend to have larger biases and sigmas than do the newer satellites (lines that tend towards red). This could be due to satellite age, but is most likely due to satellite design (as well as to the onboard clocks). The original Block IIA satellites were never as accurate as the current IIR and IIF rubidium satellites currently are. The notable exceptions among the newest satellites are SVN 65 (PRN 24) and SVN 72 (PRN 08). These IIF satellites are operating cesium clocks that have significantly larger associated errors than the IIF satellites with rubidium clocks. In looking at longer data sets (going back to 2008 [10] or even prior [14]) we do not see much impact from aging. The black line in Figure 4 steadily decreases over time. This is primarily due to the retirement of the older blocks of satellites that are then replaced with newer better ones. This set of partitions has indicated that the most important considerations are satellite clock type and satellite block. However, it remains important to continue to look for hidden patterns within the data.

Another way to partition the data is by age of data. The broadcast time of ephemeris ( $t_{oe}$ ) values can be used to identify when a new upload occurs [4]. Typically, each GPS satellite navigation data is uploaded once per day. The data ages over the next 24 hours or so and then is replaced with new information. This upload is distinct from the ephemeris data set changes that generally occur every two hours. Each ephemeris set represents a curve fit that is valid for a limited time (most often four hours). These sets are spaced two hours apart and are preloaded onto the satellite for then next day. The relevant set is broadcast when its corresponding curve fit interval begins. The first set after an upload has a corresponding age of data ranging from approximately zero to two hours, the age of data for the second set ranges from approximately two to four hours, etc. Therefore, the first set after upload should be more accurate than the later sets. Figure 5 shows the 95% UPE error bound as a function of time since last upload. The data was aggregated over each two hour segment, so the first



**Figure 5.** 95% UPE bound for each GPS satellite as a function of time since the last upload.

**Figure 6.** *Minimum overbounding*  $\alpha$  *value for MPE for each GPS satellite as a function of time since the last upload* 

ephemeris set with an age between zero and two hours is plotted at the two hour mark, the second set with the age between two and four hours is plotted at the four hour mark and so on.

The error clearly grows as the age of data increases. There is also an obvious difference between satellite blocks and clock types. The four Block IIA satellites with cesium clocks (the four highest lines in blue) have the largest initial error and most rapid initial error growth. Interestingly their error growth slows after about six hours. The two Block IIF satellites with cesium clocks (uppermost red and orange lines) have the next largest initial error and similarly large error growth. The four remaining Block IIA satellites with rubidium clocks are in the middle (we have excluded the data from SVN 35 (PRN 30) as it was at the end of its life with a very poorly performing rubidium clock and was never set to the lowest  $\sigma_{URA}$  value of 2.4 m). The Block IIR satellites (all with rubidium clocks) have generally good performance (eight of the best ten performers are Block IIRs). However, some, like SVN 44 (PRN 28), have worse much worse performance (light blue line that ends near the top). SVNs 46 (PRN 11), 47 (PRN 22), and 53 (PRN 17) all are above 2 m 95% when more than 25 hours after upload. The Block IIFs with rubidium clocks perform generally well, but several exhibit larger errors at around six and eighteen hours after upload and reduced error twelve hours after upload. This is a result of the periodic error terms previously mentioned.

Figure 6 shows overbounding results on the MPE data. The data in this figure, unlike for Figure 5, is limited to times when the corresponding  $\sigma_{URA}$  value was set to 2.4 m. As can be seen some satellites have no data beyond certain ages, this is because, for those satellites, after a certain age, the  $\sigma_{URA}$  was increased to a larger value. The worst performing satellites in Figure 5 were most likely to see this  $\sigma_{URA}$  increase. Data overbounding is described in the next section. The main take away from Figure 6 is that all values are below one, which indicates that the  $\sigma_{URA}$  value of 2.4 m was a safe descriptor of the error distribution. Error bounding is much more sensitive to outliers than the 95% error bound, so the trends in Figure 6 are less clear than in Figure 5. It appears that outliers may be more common shortly after an upload for some satellites. Indeed, sometimes there are uploads in relatively quick succession (within a couple of hours of the previous on) that appear to correct a less accurate earlier upload. Otherwise the general trend does match Figure 5 in that there is reduced margin in the 2.4 m error bound as the age of the data increases. MPE was used for this analysis because partitioning by satellite and into twohour windows for age of data results in sparse data sets. MPE has data for every healthy satellite epoch while UPE data only has data when the satellite is in view of the user. The next section will discuss further issues surrounding overbounding with sparse data sets.

## **GPS Data Overbounding**

There are many overbounding theories and approaches [15][16][17] that have been applied to satellite navigation. The earliest and simplest [15], computes an overbounding sigma using the cumulative distribution function (CDF) bounding. The general idea behind CDF bounding is that the probability of having an error magnitude greater than x for the bounding distribution is at least as large as for the observed data for all values of x. In practice, there are some considerations to take into account when using this method:

- 1. Bounding cannot strictly succeed if the data has a non-zero mean
- 2. The underlying distribution must be symmetric
- 3. The underlying distribution must be unimodal

Various methods have been proposed to separately account for biases [18] or to avoid the requirements on symmetry and unimodality [16][17]. However, there is usually a penalty in terms of added conservatism (i.e. the overbounding sigmas are excessively inflated) and limits are imposed on the number of distributions that may be convolved. A simple approach is to declare that the true distributions are symmetric and unimodal (even if the sample histograms are not) and/or to include a separate bias term in the protection level equation. The simple approach is often used, but one must take care when overbounding the CDFs, because when the sample histograms are not zero-mean, symmetric and unimodal, they cannot necessarily be CDF bounded at all points. When looking at Figure 2, it seems obvious that the tails will imply the largest sigma values. Therefore, in practice, the CDF bound is applied only to the tails and not to the core values (i.e. the central 68 - 95% of the data is not subjected to the CDF bounding). Counterintuitively, the central portion creates issues if not excluded from this tail bounding process. To see this better, we will present some examples.

Figures 7 and 8 show standard Quantile - Quantile (Q - Q) plots for the normalized errors from individual satellites. Each normalized error data set (range error divided by  $\sigma_{URA}$ ) is sorted from smallest to largest and then plotted on the y-axis versus the inverse of the normal CDF of 0.5, 1.5, ..., N-0.5 divided by N, where N is the number of data points. Gaussian





**Figure 7.** *Q-Q plot for SVN 52 (PRN 31) showing UPE/\sigma\_{URA} data for 200 users (colored circles), MPE/\sigma\_{URA} data (black plus sign), and the overbounding Gaussian (red line)* 

**Figure 8.** *Q-Q plot for SVN 61 (PRN 02) showing UPE/\sigma\_{URA} data for 200 users (colored circles), MPE/\sigma\_{URA} data (black plus signs), and the overbounding Gaussian (red line)* 

distributions are represented by straight lines on a Q-Q plot, a zero-mean, unity-variance Gaussian would be plotted as a  $45^{\circ}$  line going upward along the diagonal, passing through (-1, -1), (0, 0), etc. CDF overbounding is achieved if the assumed Gaussian overbound is below the sampled distribution for all values to the left of zero and above the sampled distribution for all values to the right. That would indicate that the overbounding distribution always predicts that non-zero errors are more likely to occur than actually observed at all probability levels. The colored circles represent UPE data from each one of the 200 evenly spaced users and the black plus signs represent the MPE data.

Figure 7 shows the data for SVN 52 (PRN 31). Here the real distributions are close to the minimum overbounding Gaussian indicated by the heavy red line. Note that between -5 and -0.5 and the red line is below all of the actual distribution lines and that between 0.5 and 5 and the red line is above all of the actual distribution lines. This indicates the true distributions are all CDF bounded at the tails. The mean of the data was removed beforehand and the slope of this minimum overbound corresponds to a one-sigma value of 0.322 (the MPE behavior was the limiting factor). That is to say, the tails of the data could be overbounded by a zero-mean Gaussian with a sigma value of  $0.322 \times 2.4 \text{ m} = .773 \text{ m}$ . Larger sigma values would also overbound the data. Note that the distribution is actually fairly close to Gaussian throughout and does not contain large tails. However, there is some evident asymmetry as the slope of the data to the left of zero is a little smaller than the slope of the data to the right. It is not clear how much asymmetry would present a problem, some of the other methodologies can account for this and we are continuing to investigate in order to provide clearer guidelines.

Figure 8 shows the results for SVN 61 (PRN 02). Note that the tails are much larger than would be expected given the core behavior. Given the core behavior and volume of data, there should not be any normalized errors larger than one. Instead there are values nearly as large as three. The minimum overbounding sigma for SVN 61 is  $0.686 \times 2.4 \text{ m} = 1.65 \text{ m}$  (set by a worst case user's UPE.). This also helps to illustrate an issue with smaller partitions. The UPE is worse than the MPE here due to data volume. Both the UPE and the MPE have the same magnitude of error (just below three sigma). However the MPE has more data points in the set. MPE can be calculated for every epoch that there are valid broadcast and precise ephemerides. UPE is only calculated for a particular user when the satellite is in view. Therefore, the UPE has roughly one third as many data points (since each satellite is in view only about one third of the time). The UPE and MPE have the same maximum *y* value, but having more points in the data set means that the *x* position will be larger for MPE resulting in a lower slope. If the data set is even further partitioned, the three sigma error will be in some data set with even fewer points leading to an even greater slope.

It is important to use a data set that is as large and representative as possible. If there are too few data points, it does not make sense to claim knowledge of the tail behavior. Roughly speaking, an effective sample size of N allows us to probe probabilities down to at best 1/N. Thus, with four years of data with 31 satellites at each epoch and one independent sample approximately every twelve hours, we have a net effective sample size of about 90,000 points. We may be able to probe





**Figure 9.** Distribution of the square root of the horizontal chi-squares value at each of the 200 user locations

**Figure 10.** Distribution of the square root of the vertical chisquare values at each of the 200 user locations

down to nearly  $10^{-5}$  assuming all satellites and all times are roughly comparable. As we partition our total set by satellite and by age of data, we divide our data set by 31 and by 12 to end up with about 244 independent points. It is hard to confirm behavior below  $10^{-2}$  with this effective sample size. Therefore, we should not be too surprised if the occasional partition were to fail. If we were to simulate one million data points with a perfect zero-mean, unity-variance Gaussian distribution, we would expect a few thousand of the data points to be bigger than three-sigma, about sixty to be bigger than four-sigma, and perhaps one as big as five-sigma. If any such points were to be included in a partition with 244 total samples, the partition would appear fail an overbounding test relative to the true Gaussian. Even though this simulated data is truly Gaussian, many sub partitions of it will indicate the overbounding variance should be larger than the true value. As the sample sizes are made smaller, the indicated overbound of some subsets will be larger than necessary. Therefore, we argue against drawing a strong conclusion from a data set with a relatively low effective sample size.

Previously, we have addressed the convolution of errors across multiple satellites by examining the sum of the squares of the normalized residuals [19] [20]. We have shown that as long as these values remains below thresholds determined by the protection levels, the errors do not convolve together in a harmful way [11] [20]. Figures 9 and 10 show the 1 - CDF of the square root of the chi-square values for both the horizontal and vertical calculations [20]. All distributions are below the Gaussian 1 – CDF for all probabilities below 0.6. This indicates that the probability of a subset position error is correctly bounded by the Gaussian assumption between the values 0.6 and  $\sim 10^{-4}$ . The bound is likely also good beyond both those limits but the method is not well suited for the larger probability values and we do not have enough independent data points to go lower probability with confidence. A simulated set of Gaussian data with zero-means and unity variance would not lie below the Gaussian either. Being above the line does not indicate underbounding or excessive correlation. It is simply a limitation of the methodology. Being below the line, however, definitely indicates overbounding.

Figure 11 shows the daily maximum of the square root of each chi-square value across the 200 users. As can be seen there is quite a lot of variation from day to day. There are only a handful of spikes above 2.5 and they are widely distributed from each other. Each spike is usually due to a small clock runoff that gets as large as two to three times the  $\sigma_{URA}$  value before the satellite is set unhealthy. The runoff is usually corrected within an hour and the satellite returned to healthy status with a new error below one times the  $\sigma_{URA}$ . The overall chi-square trends show a decrease in both the horizontal and vertical values. This is consistent with Figure 4, which shows older noisier satellites being retired and newer less noisy ones taking their place. The floor value for  $\sigma_{URA}$  of 2.4 remains fixed, so the resulting chi-square values continue to get smaller.

Figure 12 shows the sample means and the overbounding sigma values (for both UPE and MPE) for all of the satellites in the data set. The means are all small (below 10% of the broadcast  $\sigma_{URA}$ , which is generally 2.4 m) and all overbounding sigmas are below one. Sometimes the UPE implied bound is larger than the MPE bound and sometimes it is the other way around. To be conservative, we require both to be below one in order to claim that the broadcast  $\sigma_{URA}$  bounds the nominal performance. The values are determined using the full data sets for each satellite rather than also partitioning by age of data. We could use Figure 5 and 6 to infer an inflation that could be applied to Figure 12 to account for age of data. It would



**Figure 11.** Daily maximums of the square root of the chisquare values.

**Figure 12.** Sample means and overbounding sigma values for each of the satellites in the data set

appear that the average end-point value is perhaps 20% - 30% larger than the mid-point value. All but SVN 40 (which has since been retired) would continue to be bound with that inflation applied. As Figure 6 shows, the MPE bound did pass for all satellites and all age of data partitions (including SVN 40). However, a few of the UPE were slightly above one. As previously mentioned, this is not surprising given the small effective sample sizes. We state that the  $\sigma_{URA}$  as broadcast were adequate to bound the actual orbit and clock errors over the last four years.

## **GPS SIS BIASES**

The prior sections addressed determining sigma overbound results from the observed GPS error distributions. However, there is a concern that some errors may not be represented in the above error distributions. The specific method of comparing broadcast and precise ephemerides does not necessarily capture effects such as signal deformation, inter-frequency bias error for different frequency combinations (i.e. L1/L5 vs. L1/L2), code-carrier incoherence, and/or satellite antenna biases. Signal deformation in particular is very hard to capture with data, as it requires evaluating different user receiver implementations. The following sections address determining sufficient values for  $b_{nom}$ .

#### **GPS Clock and Orbit Biases**

As we saw in Figure 3 there can be large biases in the clock and orbit estimation terms. However, we saw that such large biases were transient and already well covered by the broadcast  $\sigma_{URA}$ . This was further validated by the chi-square evaluations in Figures 9 - 11. We also found that the uncertainty on the monthly bias estimates (~15 cm) was on par with observed distribution of means. We therefore do not recommend increasing  $b_{nom}$  to account for any potential bias components from the clock and orbit errors as we feel that the current  $\sigma_{URA}$  terms are sufficient to describe this error component.

#### **GPS Nominal Signal Deformation Biases**

Differences in receiver tracking loop implementations lead to different, constant delay values on the ranging measurements made by these receivers. These delays are due to the combined effect of the satellite Radio Frequency (RF) filter, the user RF filter, the specific tracking loop implementation, and the receiver correlator spacing. Aviation receivers are constrained to operate within a certain space that limits their choice of filter bandwidth, correlator spacing and total group delay [21]. The existing L1 SBAS service allows a fairly wide set of values for bandwidth and correlator spacing. Unfortunately, this has led to potentially large bias differences between the reference receiver and the user receiver. The next generation of SBAS receiver requirements for L1/L5 SBAS is being developed. The allowed user space for this next generation of receivers has been substantially reduced to lower the magnitudes of the potential biases due to nominal deformation

Figures 13 and 14 below show the differences in receiver discriminator correlator spacings between the currently allowed user space and proposed values for the next generation of SBAS/ARAIM L1/L5 receivers. The modeled error for each of these configuration spaces is shown in Figures 15 and 16. The green highlighted areas show the user receiver configurations and indicate the origins of the signal deformation bias numbers. The reduction in this allowed user design space alone has lowered the maximum expected biases from 90 cm to 15 cm for L1 and from 40 cm to 10 cm for L5.



**Figure 13.** Previous (current) and proposed dual-frequency user receiver discriminator spacings and filter bandwidths for L1 C/A and E1.



Figure 14. Proposed dual-frequency MOPS user receiver discriminator spacings and filter bandwidths for L5 and E5a.



**Figure 15.** Previous (current) and proposed estimated user receiver signal deformation biases for L1 C/A. (Bias estimates based on analysis in [7].)

For L5, the proposed MOPS configuration limitations have produced similar levels in assumed bias magnitudes. The left side of Figure16 shows that the assumed bias can be reduced from 21 cm (for "all" users) to 10 cm for "typical" users to approximately 8 cm for the current proposed MOPS users. (The largest bias for the latter occurs at a bandwidth of 12 MHz and an early-minus-late correlator spacing of 1.1 L5-chips.)



**Figure 16.** Proposed MOPS user receiver signal deformation biases for L5. (Bias estimates based on analysis in 2014 paper.) and signal deformation biases for L5 using high-resolution dish data from MITRE.

The right side of Figure 16 shows analysis results from data obtained from MITRE in late 2016 taken form the 12 GPS L5capable SVs. Using this data (as opposed to the analysis of the 2014 paper), a maximum GPS-L5 signal deformation bias of approximately 7.4 cm is obtained. (This somewhat validates the previous analysis methodology for estimating the biases.) However, to account for more than 12 SVs and provides some margin, we continue to assume a worst-case signal deformation bias on L5 of 10 cm.

We therefore advocate having an L1 signal deformation contribution to  $b_{nom}$  of 15 cm and an L5 contribution of 10 cm. The resulting dual-frequency iono-free contribution would be ~35 cm.

#### **GPS Interfrequency Biases**

Interfrequency errors will be very similar to clock errors in that there is no spatial variation. The difference is that their effect is specific to the frequency combination used by the user. The satellite clock is in reference to a specific combination. Currently for GPS, the broadcast clock is in reference to the L1/L2 iono-free code combination. The L1-only clock is offset from this reference by a value called the group delay time offset or  $T_{GD}$ . The future L1/L5 iono combination will be offset by a combination of  $T_{GD}$  and an inter-signal correction or ISC. Estimates for these values will be broadcast to the user as part of the navigation data. Any errors in these values will appear as a clock difference to the user. Provided that hardware components used to generate the signals are sufficiently stable, these values are not expected to vary quickly with time. Ideally the  $T_{GD}$  and ISC values are selected to minimize any biases. Future effort should be devoted to studying both the overall magnitude and time-varying aspects of these terms. For now we will assume that such errors are in line with the prior evaluation of the satellite clock and ephemeris parameters and assume no additional contribution to  $b_{nom}$ . However we will continue to refine this analysis in the future.

## **GPS Code-Carrier Incoherence Biases**

If the code and carrier are not fully coherent, the process of carrier smoothing can introduce a bias to the smoothed estimate of pseudorange. To date, the GPS satellites have not exhibited any signs of incoherence between the code and carrier signals on L1 or on L2. Recent studies have seen some clock variations on L5 that may have a differential effect on the code and carrier (as well as the ISC between L1 and L5) [8][9]. However, such variations are of order tens of cm with a periodicity of 4 hours or longer. These should have very little effect on the 100-second carrier smoothing that will be performed. The five-second data analysis that we are developing should be able to better investigate the effects of any true variations on the smoothing. In the meantime we feel that no contribution to  $b_{nom}$  is necessary at this time.

#### **GPS Nominal Satellite Antenna Bias**

Look-angle dependent biases in the code phase and carrier phases on both L1/E1 and L5/E5a are present on L-band satellite antennas. For the GPS satellites, these biases may be up to 40 cm peak-to-peak [6]. Given that the spacecraft position and attitude relative to a fixed user repeat every sidereal day for GPS spacecraft (excluding long-term drifts) and approximately every 10 days for the Galileo spacecraft, the effect of the antenna biases could be considered as a periodic systematic error for a fixed user. Therefore, there might be some points in the service volume where the biases tend to more consistent across multiple satellites. This systematic error could be accounted for in  $b_{nom}$ . These biases depend on the look angle of the signal through the antenna and may be different for each frequency and for code and carrier. It is likely that that are already present in some form in the above analysis of clock and orbit errors. We want to further investigate this possibility to avoid double counting the effects of any such error.

## **GPS SIS Bias Bound**

There is still much work to be done in characterizing the effects of these biases and how such upper bounds should be combined. It is unlikely that a single user will suffer the maximum possible bias from each contributor on a single satellite at the same time. However, for the moment, we do consider that the biases can all take on their maximum values and the worst-case signs. Table 1 shows the potential range of values for these various contributors and the corresponding ranges for  $b_{nom}$ .

Error Source	Lower Range	Upper Range	Expected Value
Clock & Orbit	0 cm	20 cm	0 cm
Signal Deformation	30 cm	50 cm	35 cm
Inter-frequency Bias	0 cm	20 cm	0 cm
Code-Carrier Incoherence	0 cm	10 cm	0 cm
Antenna Bias	20 cm	50 cm	40 cm
Total:	50 cm	150 cm	75 cm

Table 1 Expected contributions to  $b_{nom}$  for the GPS satellites

## **GLONASS DATA**

We have performed a similar analysis on the GLONASS constellation [22]. As was done for GPS, we have analyzed all of 2013 through all of 2016. Unlike GPS, there were many faults in the data set that had to be removed in order to evaluate the nominal performance. GPS has long published performance commitments. Its most recent edition was published in 2008 GLONASS recently submitted a paper describing proposed commitments to the International Civil Aviation [23]. Organization (ICAO) [24]. In this paper, they defined a major service failure to be any error greater than 70 m. Therefore, we removed all errors greater than 70 m from the data set and analyzed the remainder as part of the nominal operation. GLONASS broadcasts a parameter called F<sub>T</sub>, which is similar to the GPS URA value. However, F<sub>T</sub> does not provide an upper bound on the error sigma, instead F<sub>T</sub> corresponds maximum likelihood RMS value. The document further states that faults will occur with a probability below  $10^{-4}$  per satellite. The probability  $10^{-4}$  corresponds to a 3.89  $\sigma$  for a zero-mean Gaussian, i.e. 99.99% of the distribution will be between plus and minus 3.89  $\sigma$ . Therefore any error greater than 3.89  $\sigma$  also occurs with less than  $10^{-4}$  probability. The commitment that any error greater than 70 m will occur with a probability below  $10^{-4}$ implies that  $\sigma \ge 18$  m. To use a value lower than 18 m is equivalent to assuming that errors smaller than 70 m also occur with a probability below  $10^{-4}$  per satellite. Since such an assumption goes beyond the stated commitment, we will assume a floor value for  $\sigma_{URA}$  of 18 m. Using this fixed value of  $\sigma_{URA}$  is convenient, as  $F_T$  is not recorded in RINEX navigation files for GLONASS [24] and we therefore do not have access to the broadcast values.

Figure 17 shows the overview of the GLONASS data analyzed. The color-coding is the same as for the GPS data in Figure 1. A noticeable difference from the GPS plot is the presence of faults as indicated by the red markers (circles for clock faults, squares for ephemeris faults, and asterisks for uncategorized faults). There were many faults observed during this period [22] including a constellation wide fault on April 1<sup>st</sup> 2014, where all satellites experienced large ephemeris errors. As mentioned above, all errors greater than 70 m were removed from subsequent processing. Note that the GLONASS constellation is smaller than the GPS constellation with no more than 24 healthy satellites at any given time. Figure 18 shows the observed PDFs for the radial, along-track, cross-track, clock, and projected user ranging errors. There are similar behaviors as with GPS. The errors are roughly twice as large, but otherwise appear to be largely Gaussian (particularly near the core) with some evidence of mixing below 10<sup>-4</sup>.

Figures 19 and 20 show monthly mean and monthly 95% bounds on UPE (separate colors for each SVN and a heavy black line for the 95% bound across all satellites). There are many differences compared to GPS. The mean errors are much larger and remain present throughout the four-year period. In November of 2016, the biases all changed and increased by about 50%. We do not yet understand if this is a fundamental change in the satellites or due to some other part of our process. Its cause is being actively investigated. There is some uncertainty over the biases in general, as different receiver manufacturers have their own methods for removing inter-frequency biases (recall that GLONASS satellites use different frequencies to





**Figure 17.** Summary of observations for each GLONASS satellite, where green indicates good observations, blue that the satellite is unhealthy, and red indicates faults

**Figure 18.** *Radial, along-track, cross-track, clock, and projected error distributions* 





Figure 19. Monthly mean UPE for each GLONASS satellite

**Figure 20.** Monthly 95% bound of UPE on each satellite and across all GLONASS satellites (heavy black line)

distinguish themselves rather than using different codes). More study is required to determine whether or not there are valid methods for the receivers to estimate and reduce these biases. For now, we will treat these biases as real and part of the clock and ephemeris error. The 95% errors in Figure 20 show some significant variation over time and there is indication of aging effects for a few of the satellites. There is no clear trend as was seen in Figure 4, because the new satellites are roughly equal in performance to the ones they replace. Overall biases of order 3 m appear to affect the ranges (compared to 15 cm for GPS) and the average 95% error range is about 4.5 m (compared to below 1.5 m for GPS). Figure 20 shows less satellite-to-satellite variation than Figure 4, perhaps indicating that the satellite designs are more similar to each other than the different Blocks and clocks on the GPS satellites.

The variation of the monthly means is not a useful method to examine the effective sample size for GLONASS clock and orbit errors. We obtained very low values, on the order of 25 independent samples per month. This is because we see more temporal variability in the error and bias behavior, so approximating the distribution as stationary does not appear to be very accurate. We have performed a separate time correlation analysis of the errors and found that, like GPS, the correlation time appears to be in the neighborhood of 12 hours. Therefore the effective sample size per satellite appears to be comparable to GPS.



**Figure 21.** Distribution of the square root of the horizontal chi-squares value at each of the 200 user locations



**Figure 22.** Distribution of the square root of the vertical chisquare values at each of the 200 user locations



Figure 23. Daily maximums of the square root of the chisquare values.

**Figure 24.** Sample means and overbounding sigma values for each of the GLONASS satellites in the data set

Unfortunately, there is no indication for the age of data in the GLONASS broadcast navigation data, so we are unable to perform a similar analysis as was done for GPS in Figures 5 and 6. However, previous research [14] has observed worse accuracy in satellites that were farther from Russia, where the GLONASS upload stations are located [22]. GLONASS is adding more observation and upload sites, so that behavior may not hold true after these sites are incorporated.

Figures 21 and 22 show the 1 - CDF of the square root of the chi-square values for both the horizontal and vertical calculations [20]. The  $\sigma_{URA}$  values used were 18 m for all satellites at all times. All distributions are below the Gaussian 1 – CDF for all probabilities below 0.7. Figure 23 shows the daily maximum of the square root of each chi-square value across the 200 users. As can be seen, there is quite a lot of variation from day to day. There are only a handful of spikes above 1.5 and they are widely distributed from each other. We have not yet investigated the causes of these spikes in greater detail. Figure 24 shows the sample means and overbounding sigma values (for both UPE and MPE) for all of the satellites in the data set. Unlike for the GPS data, this plot is not normalized by broadcast  $\sigma_{URA}$  (or by  $F_T$ ). Instead, it is displayed in meters. The means are much larger than for GPS, with many values around two meters and one value at three meters. The overbounding sigmas range from ~2 - 17 m. It appears that 18 m is perhaps unnecessarily large for a great number of the GLONASS satellites.

In reviewing the data, we found relatively few GLONASS errors whose magnitudes were between 25 and 70 m [22]. If the commitment were changed such that errors greater than 25 m were claimed to occur with a probability below  $10^{-4}$  per satellite, then a floor value for  $\sigma_{URA}$  of just ~ 6.5 m could be applied. However, we do not advocate basing the  $\sigma_{URA}$  on empirical data. We do not know where the GLONASS operators have set their threshold for taking action. If they do not view errors between 25 and 70 m as faulted behavior, then they may not take immediate action to set such satellites unhealthy. If the operators of GLONASS are unwilling to claim that the  $10^{-4}$  number applies to error magnitudes smaller than 70 m, the ISM should not make that claim for them. Instead the  $\sigma_{URA}$  should be based on the commitments and the data should be only be used to contribute to the validation of the commitments. Communication with the CSP is another key contributor to the overall validation.

## **GLONASS SIS BIASES**

The observed monthly clock and orbit biases for GLONASS sometimes exceeded four meters. The average bias on at least one satellite was three meters over the entire four-year period and several others were around two meters. These biases could have arisen from erroneous clock and orbit broadcast values, inter-frequency bias values, signal deformations, or antenna biases. We have not studied the GLONASS signals closely enough so as to be able to separate the observed bias by the points of origin. Much work has been done on estimating and removing the interfrequency biases on GLONASS [26]. We believe that some of the observed biases may be reduced with better processing at the receiver or by improving the broadcast

navigation parameters. For the data observed, we would recommend setting  $b_{nom}$  to at least three meters, but this value is subject to change as the underlying causes are better understood.

#### DISCUSSION

We have evaluated four years worth of data for both GPS and GLONASS and determined minimum safe error bounds on each. However, it is important to remember that our goal is not describe error bounds that would have been safe in the past. Instead we wish to describe parameters that will be safe in the future. In this regard we are dependent on the performance of the Constellation Service Providers (CSPs). The constellations are not stationary, new satellites are being launched, new control software is being developed, and new operators will take the place of past operators. What is most important is that each CSP makes a set of trustworthy commitments. We must be confident that the CSPs have tests and procedures in place to ensure that new equipment, software, and personal will perform as well as prior iterations (or at least to within the commitments). We must trust that if something goes wrong, that the fault will be quickly prevented from affecting the user and that the faulted component will not be returned to service until it is sufficiently unlikely to fault again. Our data analysis is really an evaluation of CSP performance.

The ARAIM Working Group C Milestone reports [1][2] describe a parameter to be broadcast as part of the Integrity Support Message (ISM) called  $\alpha_{URA}$ . This parameter is intended to multiply the broadcast  $\sigma_{URA}$  in order to increase the value if needed. In this paper, we advocate that this value should always be set to one and therefore does not need to be part of the ISM. Our rationale is that the data analysis should only be used to validate the CSP commitments. If  $\alpha_{URA}$  needs to be larger than one in order to properly bound the error, then the CSP is not meeting its commitments (or they have been misinterpreted by the aviation community) and that constellation should not be used for ARAIM. There needs to be assurance from the CSP as to what value of  $\sigma_{URA}$  is safe. If the ISM needs to inflate this value to properly bound the error, then the commitment is not trustworthy. Similarly, after looking at our data, it may be tempting to use  $\alpha_{URA}$  values that are below one. However, this makes the assumption that future operations will continue exactly as they have in the past. If the CSP is not willing to commit to lower  $\sigma_{URA}$  values, then neither should the ISM.

GPS has developed a new civil navigation (CNAV) message that will allow it to send smaller URA values. Unlike the currently used legacy navigation (LNAV) messages, the CNAV messages can send URA values that are smaller than 2.4 m. Further, the new URA values have terms to allow spatial and temporal variation. Thus, even though most GPS satellites would be overbounded with  $\alpha_{URA} = 0.5$  today, that margin may well be removed when the new CNAV messages become operational. GPS will also be launching new Block III satellites shortly. These satellites have been developed according to the current commitments. They most likely will perform as well or better than previous generations, or like the Block IIF cesium clocks, their performance may be a little worse than the prior generation, but still be within the commitments.

We have evaluated a substantial set of GPS data and had frequent discussions with the Air Force on how they operate GPS and deal with anomalies. We can see first hand that larger errors are quickly identified and removed. We have seen the beginnings of clock errors such as were more common in the past. However, now we see that action is taken about the time such errors get to 3  $\sigma_{URA}$  rather than after they have crossed 4.42  $\sigma_{URA}$ . The Air Force has provided information about past faults and described what actions they have taken to prevent those faults from occurring again. We see that the accuracy of the system is continually improving and that the Air Force is committed to maintaining and improving this level of performance. We have had considerably less interaction with the operators of GLONASS (and Beidou and Galileo). GLONASS has had more variability in its performance over time. In between 1998 and 2009 the constellation was well below full size [22]. In 2015, just before the paper describing expected performance was presented to ICAO [24], the number of observed GLONASS faults was decreasing. However, in 2016 the number of faults increased again. We have also seen that once a GLONASS satellite has a fault, it is much more likely to have a subsequent fault after initially returning to nominal performance. This behavior is not seen on GPS. A faulty GPS satellite does not appear any more likely to have a future fault than any of the other GPS satellites.

The performance paper [24] describes empirical behavior rather than providing a commitment as to when action would be taken. It is too soon to consider GLONASS for higher integrity applications like vertical guidance. Much more discussion needs to take place about when they do, and do not, have observability to the satellites. The GLONASS CSP needs to communicate the threshold where they will immediately take action to set the satellite unhealthy and how long such an action will take to complete. Indeed such dialog is required for all of the constellations before they can be considered suitable for

aviation use. We had previously investigated using GLONASS for horizontal navigation using what we believed would be conservative values [27]. The numbers used were close, but not necessarily quite sufficient given our current analysis. However, that paper did show the value of adding more satellites even when the constellation performance worse.

## CONCLUSIONS

We have analyzed the performance of two constellations: GPS and GLONASS. GPS performance was found to be consistent with its official commitments. We can confirm that the broadcast URA value bound the error distributions for all of the satellites and that we found no patterns or conditions where the URAs appear to be too small. We analyzed the error distributions of the individual satellites and of the combined errors, as they would affect position estimates. The chi-square metrics show that the position error bounds formed from the individual satellite ranging error bounds do indeed bound any resulting positioning error. The bias terms still require further investigation. We examined five potential sources of nearly constant error and placed preliminary limits on each. The commitment on the GPS URA is intended to cover all SIS errors so it may be argued that a  $b_{nom}$  term is not strictly necessary for GPS. However, our preliminary values provide additional protection that may cover a wider range of user configurations than considered by the Air Force. Further, we may choose to be more conservative in how we treat the error convolution across multiple satellites. We propose a range of values between 50 and 150 cm for  $b_{nom}$ , with a nominally expected bound of ~75 cm. These values will continue to be investigated to refine the minimum safe upper bound value.

We assert that the GLONASS commitment of less than a  $10^{-4}$  probability of there being errors greater than 70 m, corresponds to a floor value on the  $\sigma_{URA}$  of 18 m. This value could be further refined should more information be provided as part of the commitment or should the operators be willing to lower the threshold for a fault. When using this 18 m floor value, we found that it bounds the observed satellite ranging errors for all GLONASS satellites. Although it appears that smaller values could be successfully used for most of the satellites, we do not advocate using values that imply better satellite performance than are supported by the official commitment. Neither do we advocate using values larger than the official commitment as such would only be required if we do not trust the official commitment. Instead, the aviation community and the operators of GLONASS should have further discussions to determine if the commitments can be strengthened, as well as to gain as much information about the operation of GLONASS as possible. Until that dialog occurs, the 18 m value should be used as a floor on  $\sigma_{URA}$ . The bias values for GLONASS still require further analysis, the initial values we have observed are roughly bounded by a value of 3 m. However, we are hopeful that these biases can be reduced for aviation users.

This paper describes various analyses to verify whether the observed SIS error ranges are consistent with performance commitments from CSPs. The data is in no way a substitute for the commitments themselves. The validation effort should be viewed strictly as pass/fail. Data that passes all tests is can be used to indicate that service commitments are mutually understood and have been met in the past. However, only the service commitments provide assurances for the behavior of the errors for the future. Communication and trust must be established between the aviation community and the CSPs. Data verification is a necessary but not sufficient condition in order to establish the ISM parameters. Data that fails the tests indicates that either the commitments are not understood or are not being honored. We further recommend that  $\alpha_{URA}$  not be included in the ISM as its value should either be one or the constellation should not be trusted.

#### REFERENCES

[1] Working Group C, ARAIM Technical Subgroup, Milestone 2 Report, February 11, 2015. Available at: http://www.gps.gov/policy/cooperation/europe/2015/working-group-c/

[2] Working Group C, ARAIM Technical Subgroup, Milestone 3 Report, February 26, 2016. Available at: http://www.gps.gov/policy/cooperation/europe/2016/working-group-c/ http://ec.europa.eu/growth/tools-databases/newsroom/cf/itemdetail.cfm?item\_id=8690

[3] Walter, T., Blanch, J., Joerger, M., Pervan, B., "Determination of Fault Probabilities for ARAIM," Proceedings of IEEE/ION PLANS 2016, Savannah, GA, April 2016, pp. 451-461.

[4] GPS Interface Specification, Navstar GPS Space Segment/Navigation User Segment Interfaces (IS- GPS-200H), GPS Directorate, Revision H, September 2013.

[5] Wong, G., Phelts, R.E., Walter, T., Enge, P., "Characterization of Signal Deformations for GPS and WAAS Satellites," Proceedings of the 23rd International Technical Meeting of The Satellite Division of the Institute of Navigation (ION GNSS 2010), Portland, OR, September 2010, pp. 3143-3151

[6] Haines, B., Bertiger, W., Desai, S., Harvey, N., Sibois, A., and Weiss, J. "Characterizing the GPS Satellite Antenna Phase and Group Delay Variations Using Data from Low-Earth Orbiters: Latest Results," 2012 IGS Workshop

[7] Phelts, R.E., Blanch, J., Walter, T., Enge, P., "The Effect of Nominal Signal Deformation Biases on ARAIM Users," Proceedings of the 2014 International Technical Meeting of The Institute of Navigation, San Diego, California, January 2014, pp. 56-67.

[8] Montenbruck, O., Hauschild, A., Steigenberger, P., and Langley, R. B., "Three's the Challenge," GPS World, July, 2010.

[9] Montenbruck, O., Hugentobler, U., Dach, R., Steigenberger, P. and Hauschild, A., "Apparent clock variations of the Block IIF-1 (SVN62) GPS satellite," published in GPS Solutions, Vol. 16, No. 3, pp. 303-313, 2012.

[10] Heng, L., Gao, G. X., Walter, T., Enge, P., "GPS Signal-in-Space Anomalies in the Last Decade: Data Mining of 400,000,000 GPS Navigation Messages," *Proceedings of the 23rd International Technical Meeting of The Satellite Division of the Institute of Navigation (ION GNSS 2010)*, Portland, OR, September 2010, pp. 3115-3122.

[11] Walter, T. and Blanch, J., "Characterization of GNSS Clock and Ephemeris Errors to Support ARAIM," Proceedings of the ION 2015 Pacific PNT Meeting, Honolulu, Hawaii, April 2015, pp. 920-931.

[12] http://earth-info.nga.mil/GandG/sathtml/ephemeris.html

[13] Diaz, S. P., Meurer, M., Rippl, M., Belabbas, B. Joerger, M., Pervan, B., "URA/SISA Analysis for GPS-Galileo ARAIM Integrity Support Message," Proceedings of the 28th International Technical Meeting of The Satellite Division of the Institute of Navigation (ION GNSS+ 2015), Tampa, Florida, September 2015, pp. 735-745.

[14] Heng, L., "Safe Satellite Navigation with Multiple Constellations: Global Monitoring of GPS and GLONASS Signal-In-Space Anomalies," Ph.D. Dissertation, Stanford University, December 2012 available at: http://web.stanford.edu/group/scpnt/gpslab/pubs/theses/LHengThesisFinalSignedSecured.pdf

[15] DeCleene, B., "Defining Pseudorange Integrity - Overbounding," *Proceedings of the 13th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GPS 2000)*, Salt Lake City, UT, September 2000, pp. 1916-1924.

[16] Rife, J. H., Pullen, S. P., Enge, P. K., and Pervan, B. S., "Paired Overbounding for Nonideal LAAS and WAAS Error Distributions," Published in IEEE Transactions on Aerospace and Electronic System, Vol. 42, No.4, October 2006.

[17] Rife, J. H., Walter, T. F., and Blanch, J. A., "Overbounding SBAS and GBAS Error Distributions with Excess-Mass Functions," Presented December 2004 at the International Symposium on GNSS/GPS in Sydney, Australia.

[18] Walter, T., Rife, J., and Blanch, J., "Treatment of Biased Error Distributions in SBAS," *Proceedings of GNSS 2004*, Sydney, Australia, December 2004.

[19] Walter, T., Blanch, J., and Enge, P., "Evaluation of Signal in Space Error Bounds to Support Aviation Integrity", *NAVIGATION, Journal of The Institute of Navigation*, Vol. 57, No. 2, Summer 2010, pp. 101-113.

[20] Walter, T. and Blanch, J., "Improved User Position Monitor for WAAS", *NAVIGATION, Journal of The Institute of Navigation*, Vol. 64, No. 1, Spring 2017, pp. 165-175.

[21] RTCA, DO-229E, "Minimum Operational Performance Standards for Global Positioning System/Satellite-Based Augmentation System Airborne Equipment (SBAS MOPS)," prepared by RTCA SC-159, December, 2016.

[22] Gunning, K., Walter, T., and Enge, P., "Characterization of GLONASS Broadcast Clock and Ephemeris: Nominal Performance and Fault Trends for ARAIM," Proceedings of the 2017 International Technical Meeting of The Institute of Navigation, Monterey, California, January 2017, pp. 170-183.

[23] Global Positioning System Standard Position Service Performance Standard, 4th Edition, September 2008

[24] Kaplev, S. Bolkunov, A., "GLONASS Status," presentation to ICAO NSP Joint Working Group, June 2016

[25] Walter, T., Gunning, K., and Blanch, J., "Improved Ephemeris Monitoring for GNSS," Proceedings of the 2016 International Technical Meeting of The Institute of Navigation, Monterey, California, January 2016, pp. 600-608.

[26] Sleewaegen, J.M., Simsky, A., De Wilde, W., Boon F., and Willems, T., "Demystifying GLONASS Inter-Frequency Carrier Phase Biases," Inside GNSS, May/June 2012.

[27] Walter, T., Blanch, J., Choi, M. J., Reid, T., and Enge, P., "Incorporating GLONASS into Aviation RAIM Receivers," Proceedings of the 2013 International Technical Meeting of The Institute of Navigation, San Diego, California, January 2013, pp. 239-249.