GAUSSIAN BOUNDING IMPROVEMENTS AND AN ANALYSIS OF THE BIAS-SIGMA TRADEOFF FOR GNSS INTEGRITY

Juan Blanch, Xinwei Liu, and Todd Walter
Stanford University

ABSTRACT

We describe a method to determine symmetric and unimodal pair overbounding distributions that are key to the determination of strict Gaussian bounds used in GNSS integrity. The method works by casting the search of the symmetric and unimodal bounding distribution as a linear program. We then use the proposed method to compute a set of Gaussian bounds with varying bias (for a given sample distribution) and to determine the approximate optimal choice for a given application. As an application, we apply the method to determine optimal Gaussian bounds for GPS clock and ephemeris errors.

INTRODUCTION

GNSS-based safety-of-life systems like Satellite-based Augmentation Systems (SBAS), Receiver Autonomous Integrity Monitoring (RAIM), and in the near future, Advanced RAIM (ARAIM) are designed to provide very reliable position error bounds. In order to achieve the required level of integrity, it is necessary to, first, characterize the errors that could impact the user position, and, second, to determine the effect of combined errors at the user level. The errors that we are likely to encounter in normal operations, the nominal errors, are characterized by their probability distribution. These probability distributions, whether determined analytically or empirically, are not guaranteed to have a known structure. To simplify both the design of the integrity systems and their analysis, these distributions are modeled using simpler distributions, like Gaussian distributions.

Replacing the expected distribution by a simpler bounding distribution is a necessary step for at least two reasons. First, the expected distribution would likely require a more or less complex description, for example the complete definition of its cdf. This would require large amounts of data, probably too large to be sent through the typically low bandwidth channels available for GNSS integrity. Second, it would be computationally complex to compute the user integrity risk using distributions defined by arbitrary cdfs for each of the error components. The user needs to compute the cdf of the position error (at least for one quantile). In GNSS, the position error is the result of the linear combination of multiple pseudorange errors. The user therefore needs to characterize the convolution of the distributions corresponding to each error source. Computing the convolution of up to a hundred error sources characterized by arbitrary distributions is likely to be prohibitive, even with the potentially increased computational power available to users. That is why replacing these distributions by simpler distributions is essential.

It is important to note that the process outlined in this paper assumes that the actual distribution of the errors is known. In other words, we do not address the uncertainty related to the sampling. So when we refer to the sample distribution, we assume that it represents the actual distribution.

Previous work

There are at least three methods that can be used to generate the bounding distributions: cdf bounding for unimodal and symmetric distributions [1], paired bounding [2], and excess mass bounding [3]. In [4], we presented a method to determine Gaussian bounding distributions that combines these three, and overcomes some of their individual limitations. This method is the basis of a MATLAB toolset that computes strict Gaussian overbounding distributions for any sample distribution [5]. A key step in this method is the determination of a symmetric and unimodal overbounding distribution. In [4], we used an ad hoc method that provided an adequate symmetric unimodal bounding distribution. However, it had two disadvantages: its performance depended on a careful setting of input parameters (controlling the binning and smoothing), and it did not give control on the mean of the resulting bounding distribution. This last point is important, because there is a trade-off between the mean and the sigma of the Gaussian overbounding distribution, and it can be exploited to improve performance [6].

In this work, we first propose an improvement to [4] which will allow us to control the trade-off between mean and standard deviation, second, we propose a method to determine a good balance between the mean and the standard deviation of the overbounding distribution, and finally we apply these results to GPS clock and ephemeris errors.

TWO-STEP BOUNDING METHOD

We start by recalling the method described in [4]. Let us consider a probability distribution defined by its density $f(x)$. The goal is to find a gaussian $f_{ob}(x)$ distribution that can be used instead of $f(x)$ in the integrity calculations and that will result in an upper bound of the integrity risk. This is what is usually called an overbounding distribution. One important feature of the overbounding distribution is that it needs to be stable through convolution (in a certain sense).

The key of the method in [4] is the determination of an intermediate symmetric unimodal distribution $f_{su}(x)$ that bounds the original distribution f(x) in the paired bounding sense. That is:

$$\int_{x}^{+\infty} f(t)\,dt \le (1+\varepsilon)\int_{x}^{+\infty} f_{su}(t)\,dt \text{ for any } x \tag{1}$$

(As in paired bounding, we also need a bound for the left hand side). The parameter ε is an excess mass parameter that is tunable (please see [4] for a longer discussion). Once we have determined f_{su}, we determine a gaussian bound. As explained in [4], the gaussian bound can be determined by bounding the cdf to the right of the mean (instead of across the whole range, as it is the case for direct application of paired bounding). Figure (1) illustrates the process: the blue curve is the sample distribution $f(x)$, the red curve is the symmetric overbounding distribution $f_{su}(x)$, and the black curve is the gaussian overbounding distribution $f_{ob}(x)$.

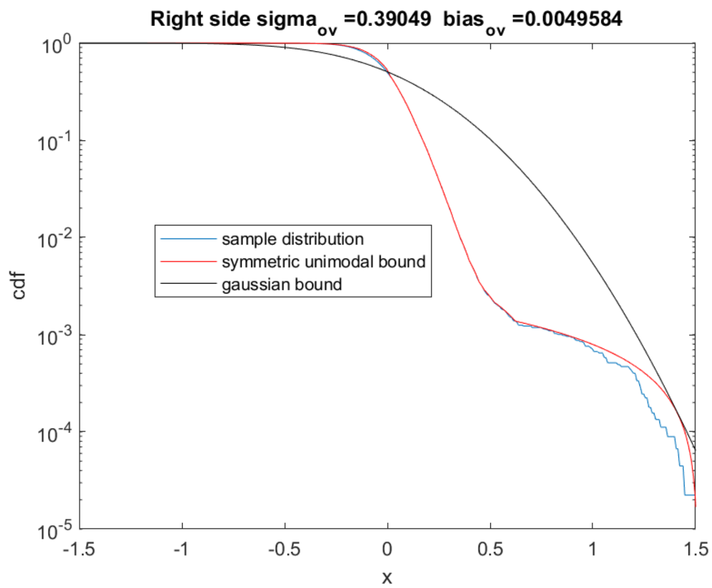

Figure 1. Two-step overbounding method compared to direct paired bounding.

Although it is visually easy to determine a bounding symmetric unimodal (s.u.) bounding distribution, it is not as easy to determine it automatically. This is due partly to the fact that there is an infinite choice of possible s.u. bounding distributions. For example, there is a possible choice for any mean below the median of the sample distribution. The first contribution of this paper is to develop a method where we can control the mean of the s.u. bounding distribution.

DETERMINATION OF A SYMMETRIC UNIMODAL BOUNDING DISTRIBUTION WITH A GIVEN MEAN

Let us write down the constraints that the s.u. bounding distribution must meet. We note F_{su} its cdf:

$$F_{su}(x) = \int_{x}^{+\infty} f_{su}(t)\,dt \tag{2}$$

Symmetry

The symmetry of F_{su} with respect to the median (and mean, since it is symmetric) can be expressed as:

$$F_{su}(x) = 1 - F_{su}(2b - x) \tag{3}$$

Unimodality

Here we define unimodality as the fact that the cdf is convex for *x* larger than *b* and concave for *x* smaller than *b*. This can be expressed as:

$$F_{su}''(x) \geq 0 \text{ for } x \geq b$$

$$F_{su}''(x) \leq 0 \text{ for } x \leq b \tag{4}$$

Bounding constraint

The bounding constraint is written in Equation (1). In the cdf notation it becomes:

$$F(x) \le (1+\varepsilon) F_{su}(x) \tag{5}$$

Objective function

For a given bias *b*, the objective is to minimize the standard deviation of the Gaussian distribution that bounds the F_{su} to the right of the mean. The gaussian bounding distribution defined by $N(b,\sigma_{ob})$ must be such that:

$$F_{su}(x) \le Q\left(\frac{x-b}{\sigma_{ob}}\right) \text{ for } x \ge b \tag{6}$$

We therefore need:

$$\sigma_{ob} \ge \frac{x-b}{Q^{-1}\left(F_{su}(x)\right)} \tag{7}$$

For a given F_{su}, the smallest possible σ_{ob} is given by:

$$\sigma_{ob} = \max_{x \ge m} \frac{x-b}{Q^{-1}\left(F_{su}(x)\right)} \tag{8}$$

The objective is to minimize σ_{ob}. This is not a simple objective function. In order to simplify it, we consider the lowest possible bound σ_{ob}. Because of the constraint expressed by Equation (5), we have:

$$\sigma_{ob} \ge \max_{x \le m} \frac{x-b}{Q^{-1}\left(\frac{F(x)}{1+\varepsilon}\right)} \tag{9}$$

If we want to approach this bound, we need to minimize the difference between *F* and F_{su}. For practical reasons (as will be see in the next section), we chose a linear objective function. More precisely, we minimize:

$$L(F_{su}) = \int_{-\infty}^{+\infty} w(t)\left(F_{su}(t) - F(t)\right) dt \tag{10}$$

where *w(t)* is an arbitrary weighing function. The following form was found to work well:

$$w(t)=\frac{1}{F(t)+\delta} \tag{11}$$

Linearity of the constraints and discretization

Both the constraints and the objective function are linear. As a consequence, we can use a linear program to find the optimal solution (in the sense of the minimum objective function). For the practical application of this approach, we first need to discretize the problem. We consider *n* points x_1 to x_n . At each of these points, the cdf *F* is represented by y_k. The discretization is carried out such that the piecewise linear function defined by the points y_k is an upper bound of *F(x)* over the whole range:

$$\forall x \in [x_k, x_{k+1}] \quad F(x) \le y_k + \frac{y_{k+1}-y_k}{x_{k+1}-x_k}(x-x_k)$$

Let us now label *z* the discretized version of the s.u. cdf. The symmetry and bounding constraints (Equations (3) and (5)) are straightforward, so we will not write down the corresponding equations. For the unimodality constraint, we need to discretize Equation (4). We express the fact that the derivative must be an increasing function above the median:

$$\frac{z_k - z_{k+1}}{x_k - x_{k+1}} \le \frac{z_{k-1} - z_k}{x_{k-1} - x_k}$$

The objective function in the discretized form is given by:

$$L(z)=\sum_k w_k (z_k - y_k)=\sum_k w_k z_k + \text{constant}$$

Once we have expressed the constraints and the objective function, we can see that the search of the optimal s.u. bounding distribution is a linear program. It can be expressed as:

$$\begin{aligned}&\min\ c^T z\\ &Az \le d\\ &A_{eq} z = d_{eq}\end{aligned}$$

with the appropriate definitions of A, A_{eq}, c, d, and d_{eq}. In this work, we used the MATLAB® function linprog.m to compute the solution. In some cases, there will not be a feasible solution. This means that there is no s.u. bounding distribution with the requested median *b* and requested excess mass parameter ε (and therefore a larger *b* must be used).

Gaussian bounding distribution

Once we have the s.u. bounding distribution, we determine σ_{ob} as indicated in Equation (8). The condition at the points x_k is written:

$$\sigma_{ob} \geq \max_{x_k \geq m} \frac{x_k - b}{Q^{-1}\left(\frac{z_k}{1+\varepsilon}\right)} \tag{12}$$

In order to account for the fact that the cdf is assumed to be piecewise linear between the points x_k, we used the method described in [4] to compute a σ_{ob} value that strictly bounds F_{su}.

The result of the process can be observed in Figure 1. With the first step, we obtain the red curve. In this case, the median b has been set at the value 0.00459. In the second step, we compute the σ_{ob} value, which in this example is 0.39. The black curve corresponds to the distribution N(b, σ_{ob}). In the rest of the paper, we will call b the bias of the bounding distribution.

Bounding both sides

As described in [4], the final bound is obtained by repeating the above process for the left hand side. This can be done using the same tools by considering the distribution obtained by a symmetry around the origin.

BIAS-SIGMA GAUSSIAN BOUNDING TRADE-OFF

As pointed out above, there are many possible choices for the bias b. More precisely, there is always a value b_l, such that for any m above b_l there will be a finite σ_{ob} such that N(b, σ_{ob}) is a Gaussian bound. In order to illustrate this point, we will use a sample distribution corresponding to normalized GPS clock and ephemeris errors computed as described in [7]. The points defining the sample are unitless because they have been normalized by the broadcast URA. (Note that the data shown here is not necessarily representative of the nominal GPS behavior, as it is only a subset of the data).

Figure 2 illustrates shows the bounding process for three different values of the bias b for a given sample distribution.

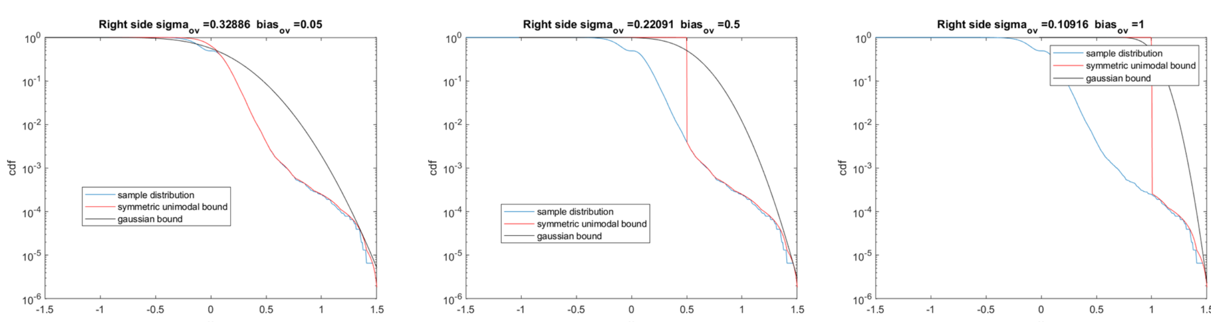

Figure 2. Bounding process for three different values of the bias b

We can repeat this process for a range of values of b. In Figure 3, we show the resulting σ_{ob} for the same sample distribution. As expected, the value of σ_{ob} decreases as the bias b increases.

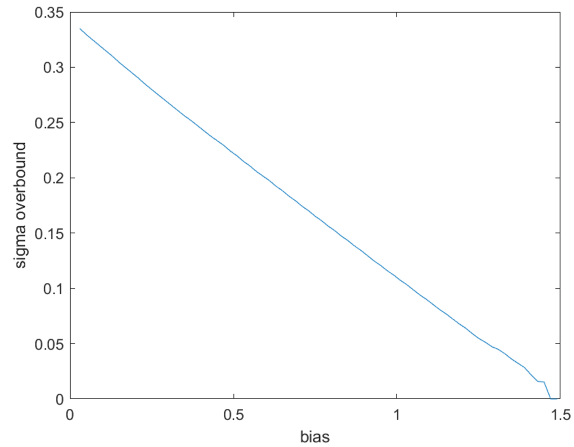

Figure 3. σ_{ob} as a function of the bias m.

The overall purpose of this process is to determine one set (b, σ_{ob}) that bounds the distribution of interest. The question is therefore how to choose the best set. To answer it, we need to evaluate the impact on performance.

User error bound dependence on b and σ_{ob}

The user error bound will be dependent on many different error sources, and on the coefficients that are used to weigh them. However, all user error bounds based on a linear estimator *s* will be very close to the following form:

$$\text{User error bound} = \sum_{i=1}^{n} \left|s_i b_i\right| + K\sqrt{\sum_{i=1}^{n} s_i^{\,2} \sigma_i^{\,2}} \tag{13}$$

where *n* is the number of measurements and *K* is related to the integrity risk (usually between 2 and 6). Let us further assume that all the biases and sigmas are the same (or what is equivalent, given the coefficients, that each pair has the same ratio). We have:

$$\text{User error bound} = b\sum_{i=1}^{n} \left|s_i\right| + \sigma K\sqrt{\sum_{i=1}^{n} s_i^{\,2}} = \left(\sqrt{\gamma}b + \sigma K\right)\sqrt{\sum_{i=1}^{n} s_i^{\,2}}$$

where:

$$\gamma = \frac{\left(\sum_{i=1}^{n} \left|s_i\right|\right)^2}{\sum_{i=1}^{n} s_i^{\,2}}$$

(Because, we have $\sum_{i=1}^{n} \left|s_i\right| \le \sqrt{n}\sqrt{\sum_{i=1}^{n} s_i^{\,2}}$, we expect that $\gamma \sim n$). Using these approximations, we see that to minimize the error bound, we need to minimize the expression $\sqrt{\gamma}b + K\sigma$. We have seen above that for a given distribution, we can determine the set of pairs *(b,σ(b))* that bound the error distribution (where *σ(b)* is the minimum sigma for a given bias b). If we note *b** the optimum bias, we have:

$$b^* = \text{argmin}\sqrt{\gamma}b + K\sigma(b)$$

In Figure 4, we show the bound $\sqrt{\gamma}b + K\sigma$ for the distribution bounded in Figure 3 for different values of *γ* and *K*.

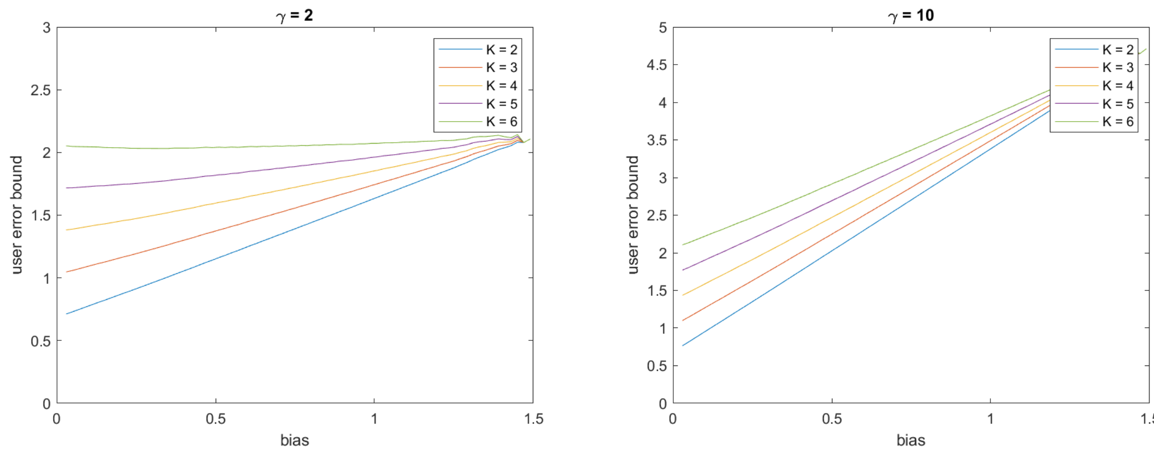

Figure 4. User error bound variation with different biases for the sample distribution used in Figure 3

We can see that for *γ* = 10, it is always better to choose the smallest possible bias. For *γ* = 2, it is only for large values of *K* that larger biases could become advantageous. This result is of course very dependent on the actual sample distribution. A different *σ(b)* curve could lead to a different conclusion. It will also be dependent on the user geometries. In Figure 5 we show the ratio *γ* for a set of users over a period of 24 hours for GPS. For this set of users, the ratio *γ* is well above 2.

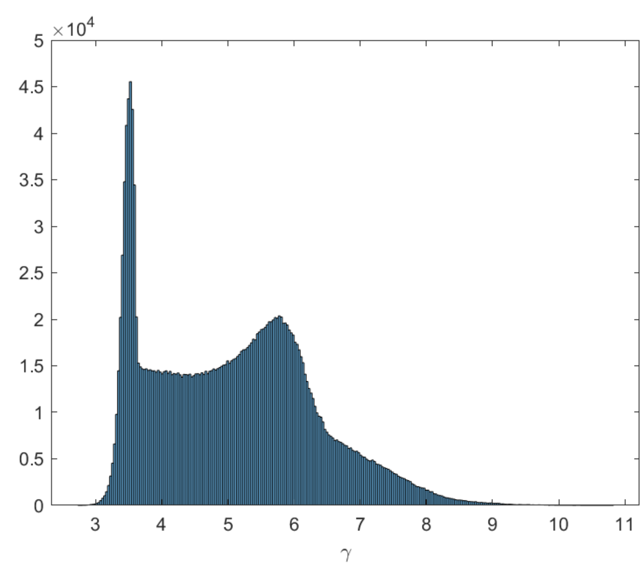

Figure 5. γ ratio for a typical set of GNSS geometries

EXAMPLE: APPLICATION TO GPS CLOCK AND EPHEMERIS ERRORS

In this section we show the sigma-bias curve obtained with the proposed method for a few GPS satellites. As with the examples before, the data corresponds to GPS clock and ephemeris residuals normalized by the URA [7] for 12 years, sampled every 15 minutes. For each satellite, the clock and ephemeris errors were computed over a grid of 200 users evenly spaced over the globe. This process resulted in 200 sample distributions per satellite. Each line in Figure 6 corresponds to the Gaussian bounding for one of the users. The curve defined by the maximum of all user curves corresponds to a *(b,σ(b))* family of Gaussian distributions that bounds the observed behavior at each of the locations.

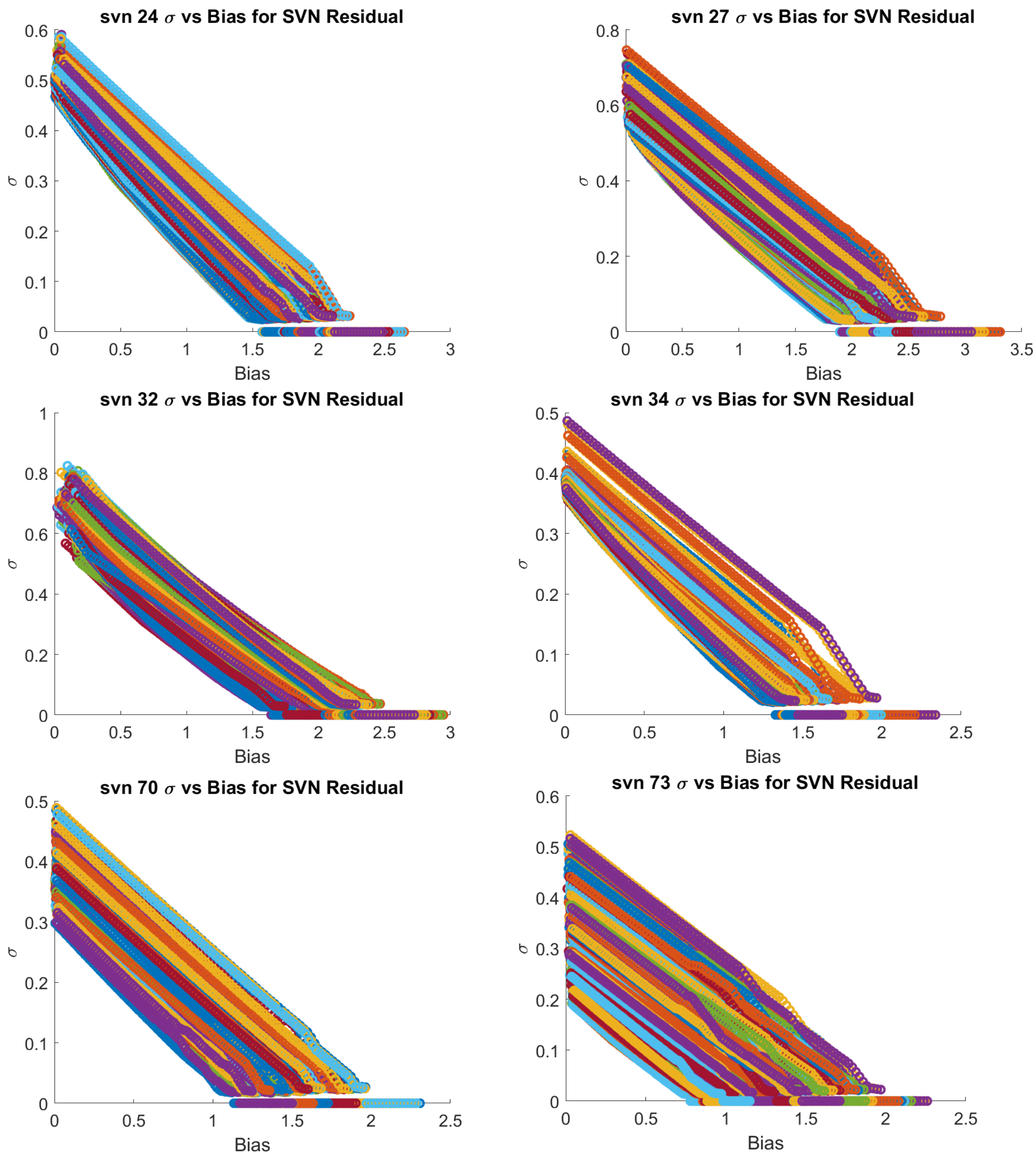

Figure 6. Sigma-bias bounding curve for 6 GPS satellites and 200 users

For these 6 satellites, the Gaussian bounds are such that they can approximately be modeled as:

$$\sigma(b) = \sigma_{\max} - \frac{b}{4}$$

We now plug this expression in the user bound factor developed earlier, and we obtain:

$$\text{User error bound factor} = \sqrt{\gamma} b + \sigma K = \left(\sqrt{\gamma} - \frac{K}{4} \right) b + K \sigma_{\max}$$

For $\gamma = 3$ and $K = 6$, we can see that the user bound is minimum for the smallest bias. (It should be noted that there might be other constraints that place a lower bound on σ [7]. For example, we might not want to use error bounds that below what is promised by the constellation providers).

SUMMARY

The main contribution of this paper is a method to determine symmetric and unimodal pair overbounding distributions with a fixed bias. This determination, which works by casting the search of as a linear program, is key to the computation of strict Gaussian bounds used in GNSS integrity. For a given sample distribution, this method enables the computation of a family of Gaussian bounds N(b,σ) parametrized by the bias b. We then propose a method to choose the best Gaussian bound in this family (for a given application). We apply this approach to the determination of overbounding distributions of GPS clock and ephemeris errors and find that the best choice is generally to minimize the bias.

The MATLAB tools developed for this paper will be made publicly available (as an evolution of [5]).

ACKNOWLEDGEMENTS

This work was funded by the Federal Aviation Administration under the memorandum of agreement DTFAWA-16-A-00005.

REFERENCES

[1] DeCleene, Bruce, "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.

[2] Rife, J., Pullen, S., Pervan, B., and Enge, P. Paired Overbounding for Nonideal LAAS and WAAS Error Distributions. IEEE Transactions on Aerospace and Electronic Systems, 2006, 42, 4, 1386 -1395.

[3] Rife J, Walter T, and Blanch J, “Overbounding SBAS and GBAS error distributions with excess-mass functions,” *Proceedings of the 2004 International Symposium on GPS/GNSS*, Sydney, Australia, 6-8 December, 2004.

[4] Blanch J., Walter, T., and Enge, P., “Gaussian Bounds of Sample Distributions for Integrity Analysis”. IEEE Transactions on Aerospace and Electronic Systems.” PP. 1-1. 10.1109/TAES.2018.2876583.

[5] https://gps.stanford.edu/resources/software-tools/Gaussian-overbound

[6] Martini,I., Sgammini, M., Boyero, J.P.” Error Bounding for ARAIM Integrity Support Message Generation”. *Proceedings of ION GNSS + 2020*, September 2020.

[7] Walter, T., 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.