Investigation into Satellite Clock and Ephemeris Errors Bounding Uncertainty and Predictability

Xinwei Liu, Juan Blanch, Todd Walter

ABSTRACT

The Advanced Receiver Autonomous Integrity Monitoring (ARAIM) concept relies on the characterization of conservative error bounds of the clock and ephemeris nominal errors to achieve the required level of integrity. [1]. In this paper, we attempt to investigate the uncertainty inherent in estimating the Gaussian bounding parameters. If the past nominal error bounding parameters are stable, then we can have some confidence that they will stay stable in the future, a conclusion critical for GNSS integrity analysis. We used an error bounding algorithm to examine the GPS and Galileo satellite clock and ephemeris error bounding behavior for the past 12 years. We found that the error distribution overbounding parameter estimation has increased stability after removing the near-fault data points. We evaluated the uncertainty in the bounding process using both the bootstrap and the training-validation methods. We found that a larger data set size significantly reduces the uncertainty, and using 12 years for GPS is enough to characterize the bounding parameter behavior due to its stability, whereas the two years of Galileo's available data set is not.

I. INTRODUCTION

The Advanced Receiver Autonomous Integrity Monitoring (ARAIM) concept relies on the characterization of conservative error bounds of the clock and ephemeris nominal errors to achieve the required level of integrity. [1] To have a conservative bounding, We computed Gaussian overbounds of the sample distribution using the algorithm described in [2] [3] [4], which takes into account the effect of extreme non-faulted events defined as errors lower than the standard 4.42 time of User Range Accuracy (σ_{URA}) for GPS and 4.17 for Galileo. [5] Here, the Gaussian overbound is a Gaussian distribution that bounds over the sample error distribution and can be used to replace the sample error distribution when computing user error bounds. Gaussian bounding is stable through convolution, which means that the convolution of the errors is bounded by the convolution of their Gaussian bounding distributions [4]. Gaussian distributions only rely on the mean and standard deviation parameters. The simplicity of the model is suitable for data transmission and mathematical analysis.

This paper investigates the satellite clock and ephemeris errors uncertainty and predictability of the Gaussian bounding parameter, its mean, and its standard deviation. We refer to those two parameters as bias and σ in the following sections.

We inspect the bounding distribution in section II to characterize the nominal error model. We further characterize the error bounding uncertainty and predictability by exploring the stability of the bounding parameters. If the parameters are stable enough in the past several years, we can conclude that the parameter does not vary much and thus can have some confidence that they will likely stay stable in the future, assuming no drastic changes would occur. We can estimate the uncertainty in the available error data to have some insight into the data behavior in the future. In this sense, service history can characterize the parameter behavior. Taking a step further, we can produce simulations to predict future error bounding models using the existing data. In sections III and IV, we describe the training-validation method for characterizing the bounding uncertainty and prediction simulation and the bootstrap method for characterizing the error bounding parameter uncertainties. We show the preliminary results in the experiment section. [2]

II. ERROR SERVICE HISTORY EXAMINATION

This section examines the bounding parameters variation using twelve years of User Projected Error (UPE) from GPS with three years of the time window for each data point and two years of UPE from Galileo. We applied the standard threshold and later used a lower threshold of three times the σ_{URA} to explore the validity of the standard threshold. The following subsections will explain the data processing method and show the results.

1. Data process

The data we used is the GNSS clock and ephemeris errors normalized by σ_{URA} projected to the line of sight of 200 evenly distributed users around the globe. We have 12 years of data for GPS and two years for Galileo. There are in total 45 satellites for GPS and 22 satellites for Galileo. The data sampling rate for GPS is every 15 minutes, and for Galileo is every 5 minutes.

We obtained the nominal errors by thresholding the projected error normalized by σ_{URA} using the standard definition of 4.42

for GPS and 4.17 for Galileo. However, as long as the faulty probability does not exceed 10^{-5} as defined in [6], we can vary the threshold. We lowered the threshold to 3 to explore the effect of the near-fault data points on the estimated uncertainty.

For GPS, we first selected the error data for one satellite. We computed a common "bias search space" containing possible bias values for all users using the bounding algorithm. We then found the optimal bias and sigma to minimize the error bounds. Not all the biases in the search space can produce a feasible bounding. We binary searched the smallest bias that produces feasible bounding for all the users. Then we took the largest sigma corresponding to that bias across all the users. In this way, we found the optimal bounding parameters for this particular satellite clock and ephemeris error. We applied a similar process to Galileo error data. We represent the algorithm in a flow chart in Figure 1 for the case of GPS. We should expect the majority of the

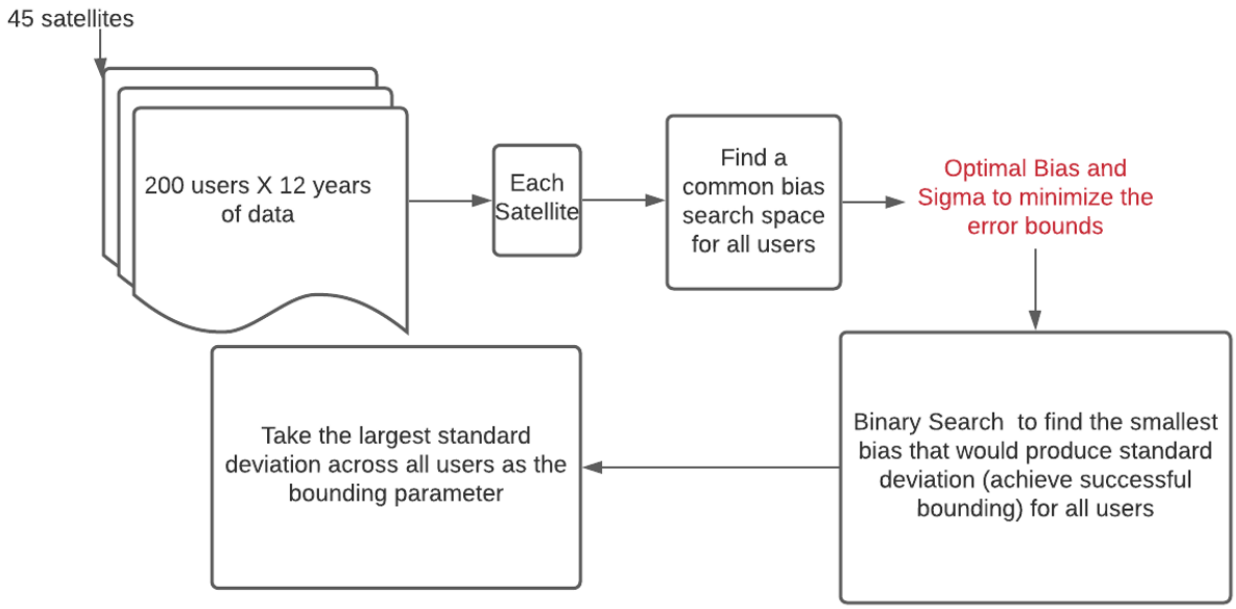

Figure 1: Data process algorithm for GPS. The sampling rate is every 15 minutes, and there are 35040 points in a year.

normalized σ values to be below one since we do not want to nominal error bounding standard deviation to exceed the standard deviation of the URA.

2. Error time history results

We show the results with data normalized by σ_{URA} in Figure 2. We show two years of UPE error bounding parameter histories from Galileo using 0.5 years of a time window in Figure 3 . Each line represents a satellite. Each data point represents the bounding parameter applied to three years of error data for GPS and 0.5 years for Galileo. We slid the three-year window by every six months for GPS and every 0.1 years for Galileo.

The plots show that the bias parameter appears to be stable based on this rough examination. Specifically, all the values are below 0.3 and with slight variations. For σ, we observe more variable behavior. One IIA satellite's σ exceeds the normalization with the standard threshold, and all the satellite bounding parameters stay below the normalization with the lower threshold. In addition, abrupt jumps occur more often in the standard threshold plot, indicating instability.

If we assume that there are two data points per day, and the probability of a faulty event occurring is 10^{-5}, we likely do not have enough independent data to generate meaningful statistical results. This effect could potentially cause the error to exceed the bounding threshold . [5] [6]We mitigated the lack of data issue by aggregating errors from all satellites to obtain enough data to stabilize the bounding parameters. We reasonably assumed that the satellites should have similar error characteristics incentivizing us to aggregate the satellite errors. The results are shown in Figure 4. Here, the red lines represent the error data using the threshold of 3, and the blue lines represent the error data using the standard threshold. The dotted lines are for the σ, and the solid lines are for the bias.

The plots show that the bias parameter is highly stable, as represented by the blue and red solid lines. The bias parameter also appears to be stable even with each satellite. For this reason, we focus our study on the σ parameter. Although with the standard

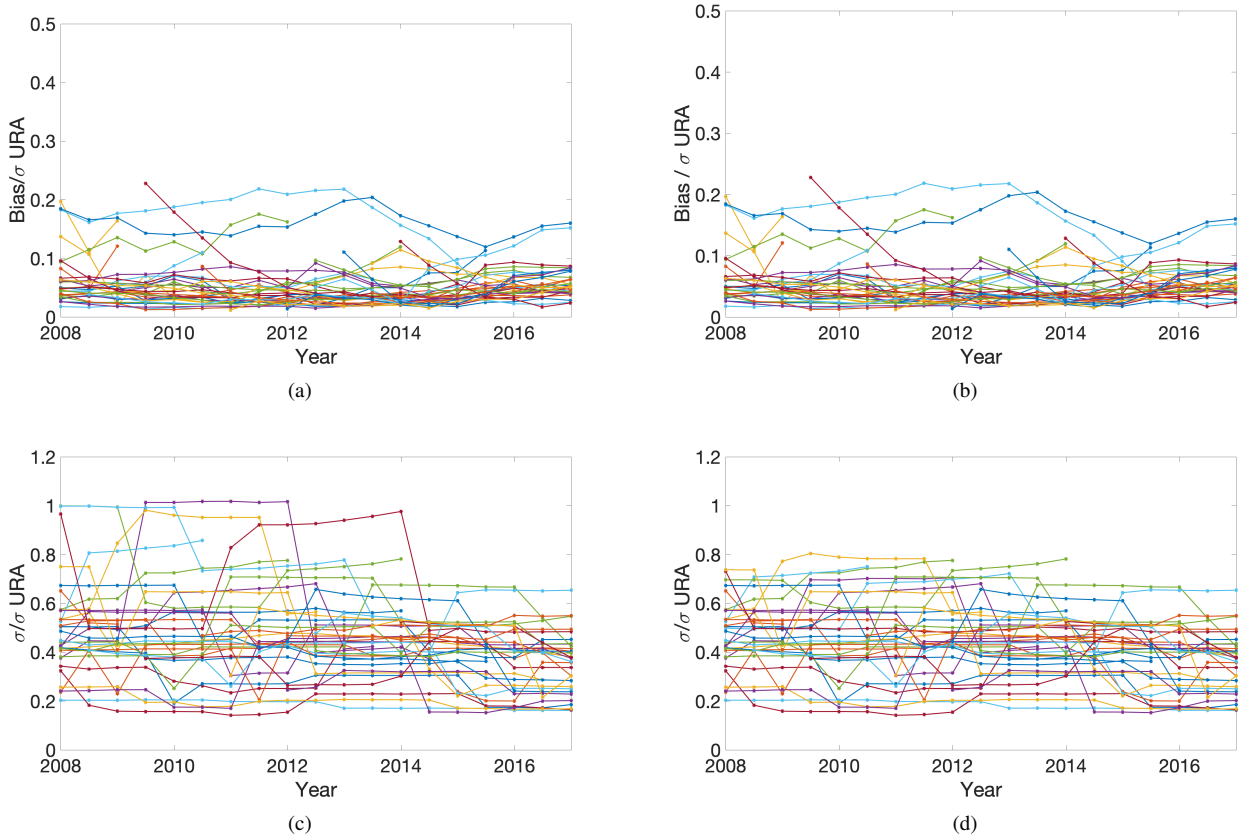

Figure 2: Parameters for GPS clock and ephemeris errors from the year 2008 to the year 2020 using with three years time window (a) Bias with the threshold of 4.42 times σ_{URA} (b) Bias with the threshold of 3 times σ_{URA} (c) σ with the threshold of 4.42 times σ_{URA} (d) σ with the threshold of 3 times σ_{URA}

threshold, normalized σ no longer exceeds one as represented by the dotted blue lines, the instability persists, whereas stability can be observed in the lower threshold plot as represented by the dotted red lines. This result indicates that the near-fault data points are likely responsible for the variability. This finding requires further investigation not covered in this paper.

We employed the training-validation and bootstrap methods introduced in the following two sections to further explore the satellite errors' stability.

III. TRAINING VALIDATION METHOD

To investigate the stability and the predictability in the bounding parameter σ, we applied the training-validation method, an extension of the training-validation method used in the machine learning algorithm. We designated a part of the data set as "training data," computed their bounding parameter, and compared the values to the bounding values generated from the rest of the data set, the validation data. We made those two comparisons to examine how closely the training data bounding parameters resemble the validation's. This method can be viewed as a prediction simulation experiment conducted on the available data, and it gives us insights into what would happen if we use all the available data to predict the future bounding parameters.

We represent the training-validation algorithm in Algorithm 1

1. Training Validation Algorithm

The training-validation algorithm is presented in Algorithm 1 We post-processed the simulated data by plotting the histogram of the ratio between the training and validation parameters and calculating the standard deviations. The results give us an intuition of how well the training data approximate the validation data, or in other words, approximate the predictability of the data set. In addition, the standard deviation will give us an insight into the stability of the parameters. Specifically, this method is applied

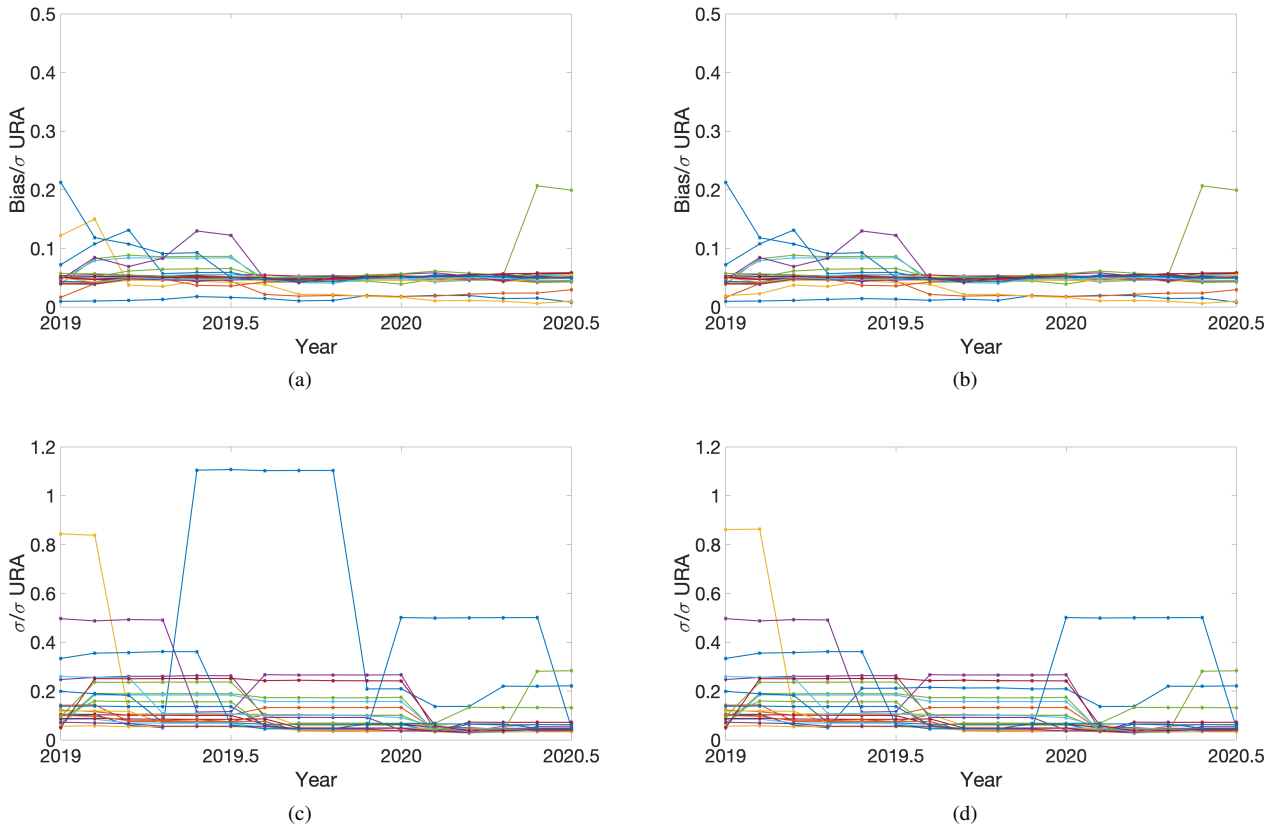

Figure 3: Parameters for Galileo clock and ephemeris errors from the year 2019 to the year 2021 using 0.5 years of the time window(a) Bias with the threshold of 4.17 times σ_{URA} (b) Bias with the threshold of 3 times σ_{URA} (c) σ with the threshold of 4.17 times σ_{URA} (d) σ with the threshold of 3 times σ_{URA}

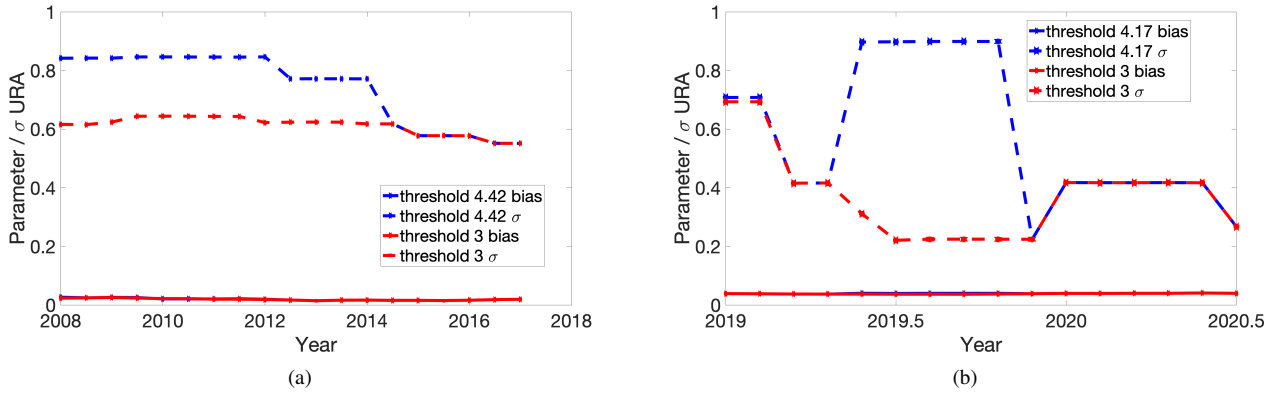

Figure 4: Parameter values plotted against time for aggregated satellite values.(a) Three years time window with the threshold of 4.42 and 3 times σ_{URA} for GPS (b)0.5 years time window with the threshold of 4.17 and 3 times σ_{URA} for Galileo

to investigate the stability and predictability of the satellites over time using the same set of satellites for training and validation. The algorithm divides the data into half and half in time and uses them as training and validation sets. The training set is further divided into units of 6 months to preserve time correlation. We draw the training data from the training set data units. This method can also examine the effect of training time length on the result by varying it.

Algorithm 1: Training-Validation Method

Result: Training Bounding Parameters and Validation Bounding Parameters
Divide the data into training set and validation set;
Compute the bounding parameters from the validation set;
for *m times* **do**
    Divide the training set into chunks of data ;
    Randomly select the data chunks from the training set as training data;
    Compute the bounding parameters correspond to the selected data as training bounding parameters;
    Store the calculated parameters;
end
Construct distribution using the stored training and validation parameters.

IV. BOOTSTRAP METHOD

Bootstrap is a resampling method used to evaluate the uncertainty in the estimation. It assumes that the sample we have is representative of the actual population. We refer to this sample as the original sample in the following paragraphs. The bootstrap method is, in essence, a simulation using sampling with replacement based on the above assumption. The statistics parameter distribution generated from the bootstrap simulation should be similar to the one generated from the actual population. [7] [8] Another way to understand this is that by sampling with replacement, we create alternative history data, and by generating a large number of alternative history, we can explore the inherent uncertainty in the error data. This section will go through a simple explanation of the bootstrap theory, why and how we apply the bootstrap method to our problem, and finally, present the bootstrap algorithm.

1. Bootstrap Theory

This section gives a brief explanation of the mathematics behind the bootstrap method. The bootstrap method computes the probability distribution of a particular parameter and thus can provide its uncertainty. Suppose we are interested in parameter θ. First, let us draw n samples from the population to form the original sample. Then we sample with replacements for M times to generate M new set of sub-samples, each with the length of n. By doing so, we have generated M iid bootstrap samples. For each of those bootstrap samples, we can compute the bootstrap estimator parameter $\hat{\theta}_n^*$. [9]

We represent the distribution of the parameter using the below Equation [9]

$$F_n(t) = \mathbb{P}\left(\sqrt{n}\left(\widehat{\theta}_n - \theta\right) \leq t\right)$$

Here, the $\widehat{\theta}_n$ is the estimator for θ. The bootstrap estimation of the distribution can be represented using the below Equation [9]

$$\hat{F}_n(t) = \mathbb{P}\left(\sqrt{n}\left(\widehat{\theta}_n^* - \widehat{\theta}_n\right) \leq t \middle| X_1, \ldots, X_n\right)$$

Given that we have the original sample as$\{X_1, \ldots, X_n\}$. We can argue that $F_n(t)$ is close to some limiting distribution and that $\hat{F_n}(t)$ is close to another limiting distribution. The two limiting distributions are close to each other; thus, the distribution and the bootstrap estimation are close. Furthermore, we can apply the Monte Carlo approximation using the M samples we have obtained to calculate the distribution represented by the below Equation [9]

$$\bar{F}(t) = \frac{1}{M}\sum_{j=1}^{M} I\left(\sqrt{n}\left(\widehat{\theta}_j^* - \widehat{\theta}_j\right) \leq t\right)$$

This estimation is also close to $\widehat{F}_n(t)$, which is close to $F_n(t)$. Thus, we have obtained the original parameter distribution as long as M is large. [9]A version applied to the mean of the data can be easily proven using the central limit theorem and the law of large numbers. Further expansion on the theory can be found in Asymptotic Statistics by A. W. van der Vaart. In our case, we apply the bootstrap method to the overbounding parameter bias and σ.

2. Bootstrap Application to Bounding Parameters

Bootstrap fits our problem as it is a well-developed method to investigate uncertainty within the data series. [10] In addition, bootstrap does not require a model. Since we do not have an accurate model for the residual bounding parameter distribution, the bootstrap method is a good candidate for evaluating parameter uncertainty. [11] Also, bootstrap does not make many assumptions about the distribution. [7]

3. Bootstrap Algorithm

In this section, we present the bootstrap algorithm. A typical bootstrapping method is executed according to the following Algorithm 2: [7]

Algorithm 2: Bootstrap Method

Result: Statistic Distribution
Draw a sample from population with size n as original sample;
for *m times* **do**
    Draw sub-samples with size n from original sample with replacements and store the sub-samples;
    Calculate the statistics parameter θ for the sub-sample;
    Store θ;
end
Construct distribution using θ

In this case, our statistics parameter θ distribution would reflect the θ distribution if we draw the statistics distribution directly from the population. Our study slightly varies the method by dividing the original sample into smaller units and then drawing the sub-sample from the units with replacements. We elaborate on the incentive for this change in the experiment section. We chose the original sample as the available error history.

There are different advantages for using the training validation and the bootstrap method. The training-validation method is more intuitive than the bootstrap method. It precisely simulates a prediction process of the future error bounding statistics distribution. On the other hand, mathematics is well established for the bootstrap method. It is a commonly accepted algorithm, and it is a valid estimator for quantifying uncertainty. The following section lays out the experimental setup and shows the results.

V. EXPERIMENT

We simulated the training validation and the bootstrap method using the 12 years of clock and ephemeris errors provided by the International GNSS Service, computed using the difference between the precise and the broadcast data. [2] The results are normalized by the σ_{URA}. The following sections discuss the data process and the two experiments.

1. Data Process

In this experiment, we obtained the 12 years of satellite clock and ephemeris maximum projected error (MPE) and URE, normalized by the σ_{URA}. The MPE is the largest possible projected error onto earth calculated using geometry. We first applied a threshold of 4.42 to get the nominal error. We computed the bias and σ directly for the MPE values. For URE, we calculated the smallest bounding bias for all the users and selected the σ that provided bounding for all users as described in the previous section.

2. Training-Validation

The training-validation experiment was applied using the normalized GPS UPE data. After selecting the training and validation data, we aggregated the satellites to avoid instability due to a lack of data.

We selected 2 and 5 years of training data for the simulation, computed the training to validation data ratio, and centered the histogram at 0. We generated a thousand simulation data points and set the thresholds to be 3 and 4.42. The ratios of $\frac{training\sigma}{validation\sigma} - 1$ are plotted in Figure 5

The closer the values are around 0, the better the predictions are. In addition, we want the majority of the data to lie on the negative side to achieve conservative bounding and want to have a small standard deviation for the bounding parameter distribution for better stability.

In the plots, the ratio values are close to 0 and have a low standard deviation, which indicates that the prediction method

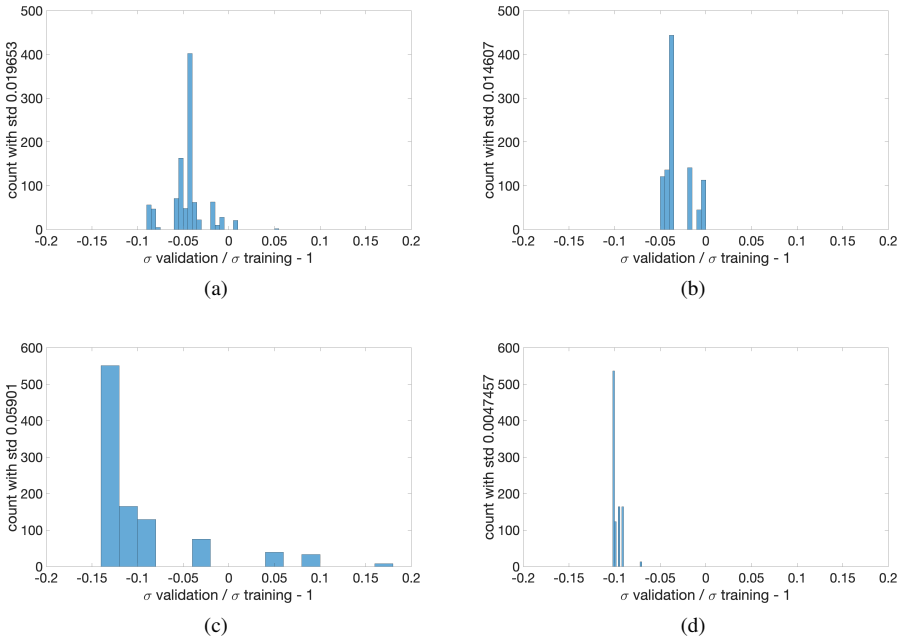

Figure 5: Parameters for GPS clock and ephemeris errors from the year 2008 to the year 2020 using with three years time window for σ with the threshold of 3 times σ_{URA} with (a) 2 years training years and standard deviation of 0.0197(b) 5 years of the training year and standard deviation of with a standard deviation of 0.0146 and with the threshold of 4.42 times σ_{URA} with(c) 2 years of the training year and standard deviation of 0.0590 (d)5 years of the training year and standard deviation of 0.0047

works well with data as little as two years. We can also observe that majority of the data lies in the negative range, implying overestimation, which is desired to achieve conservative bounding.

With the increase of training data size, we have a lower standard deviation, indicating that more data will likely stabilize the parameter. However, we observe contradicting effects after lowering the threshold to 3. For training data size of 2 years, the standard deviation decreased, and for training data size of 5 years, the standard deviation increased. We should not be too alarmed by this effect. Usually, lowering the threshold might eliminate the variation in the bounding parameter. The unexpected standard deviation increasing effect is likely due to the inherent behavior of the data. In Figure 4, we can see that for the first four years, the error data with the threshold of 4.42 seems to behave more stable than error data with the threshold of 3. We draw the training data from the first six years. In other words, it is not surprising that error data with the threshold of 4.42 exhibit more stable bounding behavior for a more extended period.

3. Bootstrap

We first took every satellite's MPE time history for GPS data and divided it into units of 15 minutes, 0.5 days, one day, and one month. When applying sampling with replacement to the original sample, we used these units as sampling units. Then we took the original sample to be the later six years, nine years, and 12 years to explore the effect of the original sample size on the bootstrapping result. We then sample the units with replacement from the original sample many times.

We gradually relaxed the time independence assumption by increasing the sampling unit size. In the bootstrapping method, we have the underlying assumption that the sampling units are independent of each other. In other words, if we choose every 15 minutes as unit sample size, we assume that every 15 minutes, we have a data set independent of the previous 15 minutes. We incorporated time correlation into the data set by increasing the unit size, thus relaxing the strong time independence assumption.

We chose every 15 minutes because this is the error sampling rate. We chose 0.5 days based on the number of ephemeris uploads to the satellites. We chose one day and one month to explore the effect of expanding the time units. We selected 6 years and 12 years as the original sample size to show the impact of the original sample size. We then applied the same method to the aggregated satellites, expanding the data size by 45.

a). Individual satellite

In Figure 6 and 7, we show the results obtained from satellite with svn 67 and 71. These are normalized histograms. It turns out that the threshold does not matter for these two satellites due to the lack of near-fault data points. Thus we are only showing the results for the threshold of 4.42. In addition, for each satellite, due to the limited operation period, for example, svn 67 and 71's operation data are mostly available in the later six years, it turns out that the result for the later six years and total 12 years are very similar. So we only show the result for 12 years with 1000 samples. The difference between the different original samples size is more obvious for the aggregated error plot since there are always satellites operating, so we do not have the limited operation period issue.

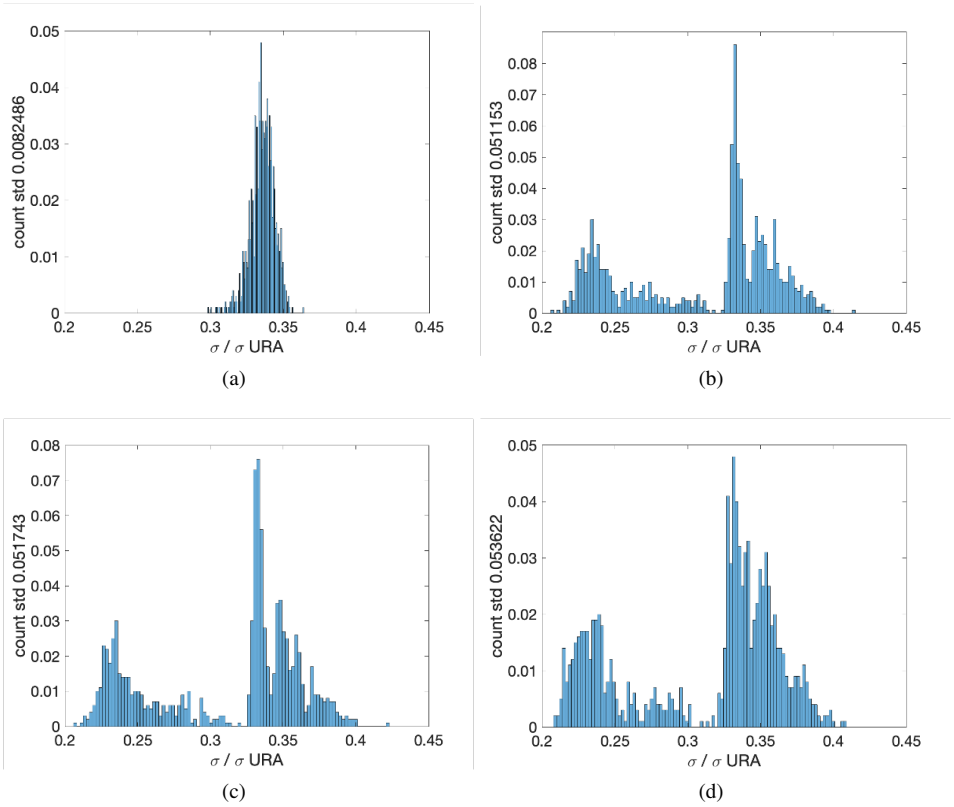

Figure 6: Satellite svn 67 bootstrap result for $\sigma = 0.333$ for GPS clock and ephemeris MPE for 12 years with units of (a) 15 minutes with a standard deviation of 0.00825 (b) 0.5 day with a standard deviation of 0.0512 (c) 1 day with a standard deviation of 0.0517 (d)30 days with a standard deviation of 0.0536

In Figure 6 (a), we observe a mono-peak histogram with a small standard deviation for the bounding parameter distribution. This result indicates stability or small uncertainty in the bounding parameter. However, as we increased the sampling unit's size, the parameter's standard deviation started to increase. Moreover, we see two distinct peaks in the distribution. The two peaks become more evident as the sampling unit size grows. After grouping the time steps into units, we added the time correlation when applying the re-sampling method and added time correlation. We create a second peak, presenting the two modes in the error bounding parameter behavior. The result implies that we could have under or overestimated the parameter with the multi-modal behavior. For example, we could have gotten a value close to the lower peak, whereas the actual result is closer to the higher peak. This bi-modal behavior shows that the error bounding parameter likely has a larger uncertainty when we do not assume that the data points sampled every 15 minutes are independent.

Similar result is shown in Figure 7. As the unit size increased, we increased the time correlation to relax the assumption that error data points sampled every 15 minutes are independent, creating more peaks and a larger standard deviation. This result has the important implication that the uncertainty in the parameter tends to be larger than the result without the data point independence assumption. Even with a small degree of time correlation introduced, taking units every 0.5 days, we could potentially still get large uncertainty in the bounding parameter for a single satellite. Here, by applying the bootstrap method, we created an alternative history of error data that could have happened and can happen in the future, suggesting that the inherent

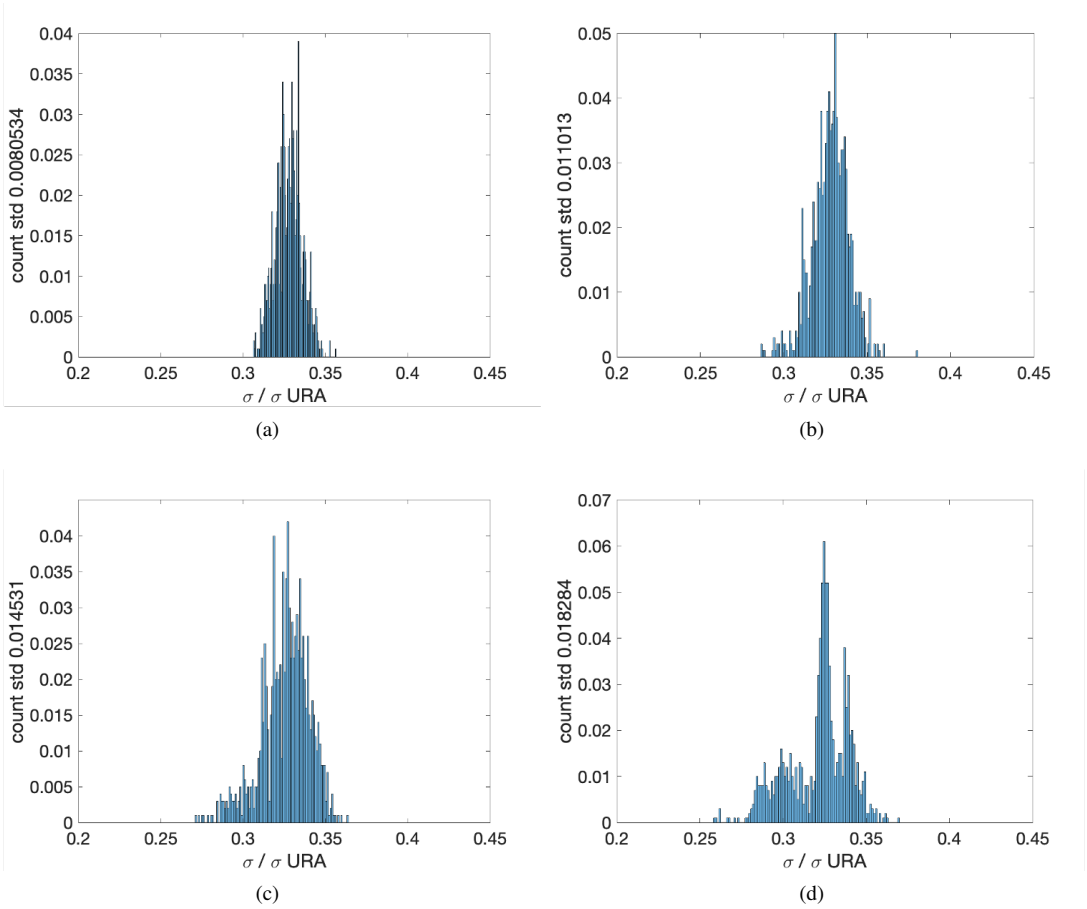

Figure 7: Satellite svn 71 bootstrap result for $\sigma = 0.333$ for GPS clock and ephemeris MPE for 12 years with units of (a) 15 minutes with a standard deviation of 0.00805(b) 0.5 day with a standard deviation of 0.0110(c) 1 day with standard deviation of 0.0145 (d)30 days with a standard deviation of 0.0183

uncertainty might be larger than it appears in the actual time series data.

b). Aggregated satellites

As indicated in the time history plot, we tend to create a more stable parameter with more independent data points by aggregating the satellites together. We examine the bootstrap result from the aggregated satellites with 10000 samples. In Figure 8, we plot the result using a threshold of 4.42 and with the original sample size of 6 years.

Here, we observe the same multi-modal behavior in the plots. The peaks become more distinct than before, and the standard deviation grows with the increasing unit size from 15 minutes to 12 hours, consistent with each satellite's result. The time correlation effect becomes less evident among the unit size of 12 hours, one day, and one month. The peaks and the large standard deviation are likely caused due to the introduced time correlation described in the above sections. Notice that we should not compare the standard deviation of the aggregated satellite bounding parameter distribution with the single satellite since the satellites we selected are newer and thus are likely to produce a better result than the other satellites.

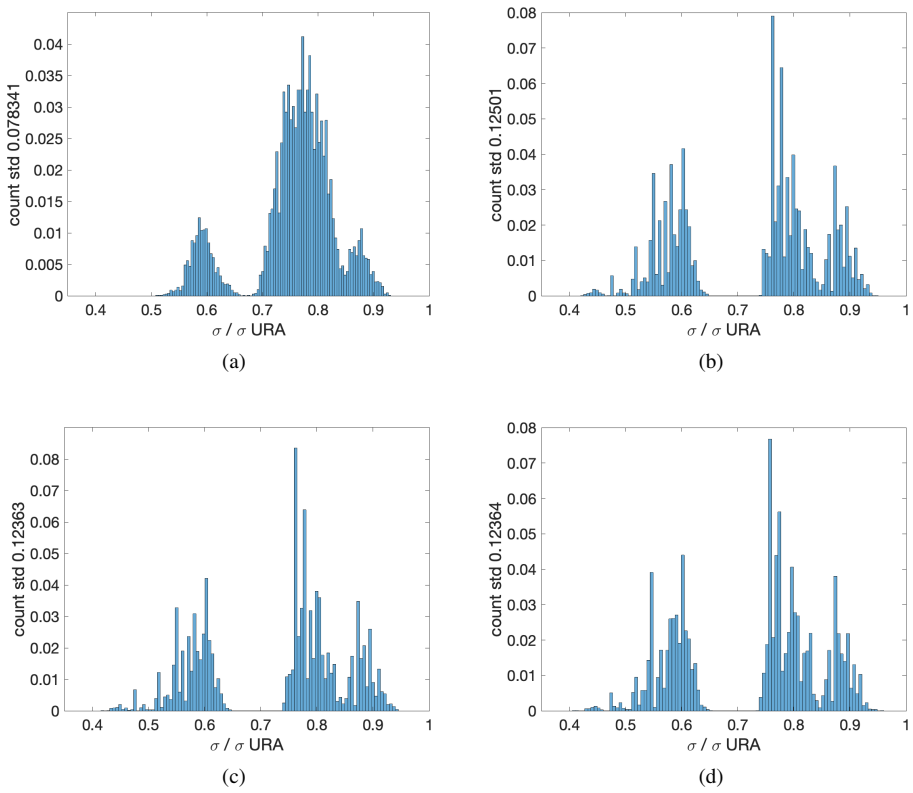

Figure 8: Satellite aggregated bootstrap result for $\sigma = 0.776$ for GPS clock and ephemeris MPE for six years with the threshold of 4.42 units of (a) 15 minutes with a standard deviation of 0.0783 (b) 0.5 day with a standard deviation of 0.125(c) 1 day with standard deviation of 0.124(d)30 days with a standard deviation of 0.124

One important factor we wish to explore is the effect of the original sample size on the bootstrapping result. We took to expand the original sample size to the later nine years and plotted it in Figure 9

As we can observe, the standard deviation of the σ parameter distribution is significantly reduced with a larger original sample size, which is more representative of the whole population, in our case, the entire time of service of the satellites. In addition, we observe that the multi-modal behaviors are less noticeable. Both results indicate a potential decrease in the parameter uncertainty as we increase the original sample size.

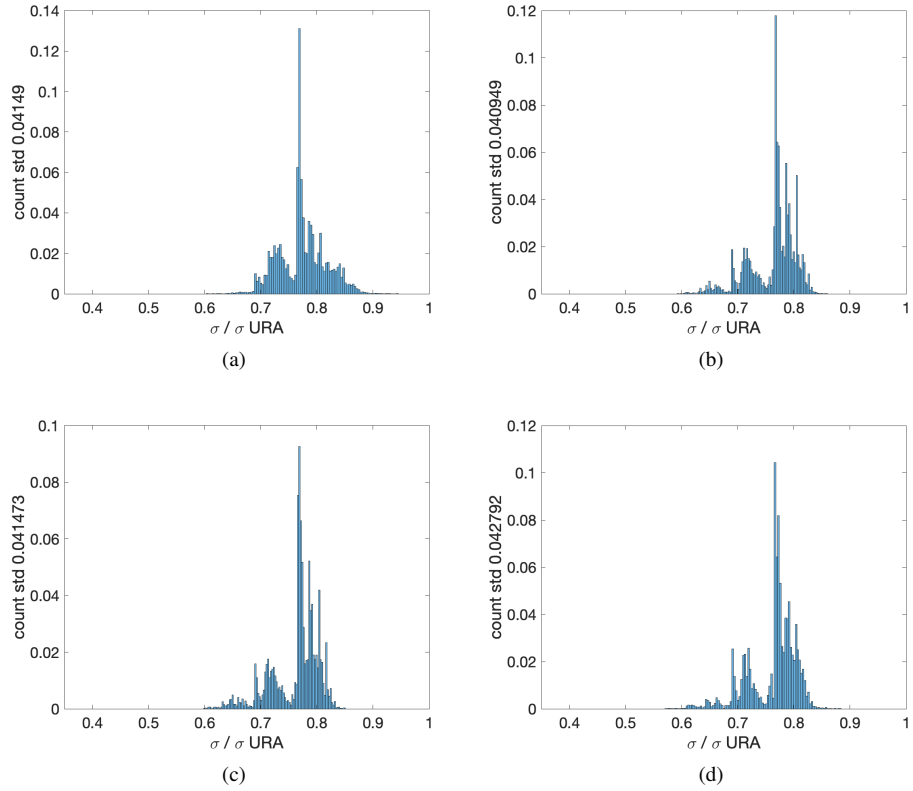

Figure 9: Satellite aggregated bootstrap result for $\sigma = 0.768$ for GPS clock and ephemeris MPE for nine years with the threshold of 4.42 units of (a) 15 minutes with a standard deviation of 0.0415 (b) 0.5 day with a standard deviation of 0.0409(c) 1 day with standard deviation of 0.0415(d)30 days with a standard deviation of 0.0428

Now we can further expand the original sample size to 12 years of data and apply bootstrap to the aggregated satellite MPE with units of 15 minutes, 0.5 days, one day, and 30 days. The results are plotted in Figure 10. This bootstrap result should represent the whole population more since we expand the bootstrap sample size.

The plot shows that the multi-modal behavior becomes less noticeable with more data available. The standard deviations also significantly decreased compared to the results obtained using a smaller original sample. This result indicates that the bounding parameter for the aggregated satellite error is stable and will likely stay stable in the future.

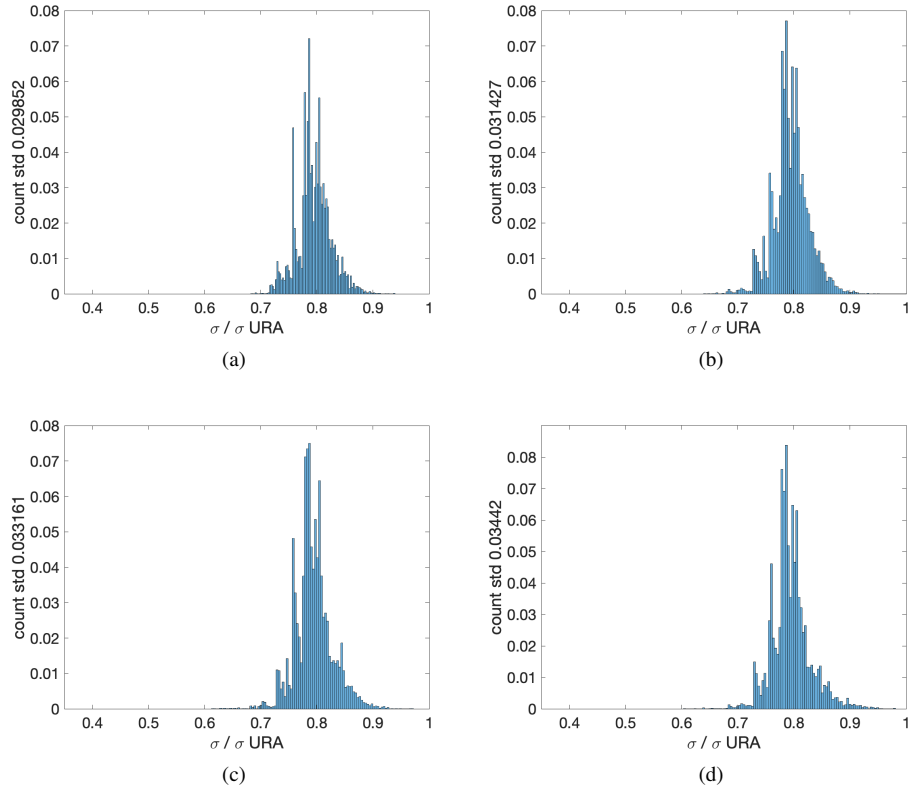

Figure 10: Satellite aggregated bootstrap result for σ for GPS clock and ephemeris MPE for 12 years with the threshold of 3 units of (a) 15 minutes with a standard deviation of 0.03041(b) 0.5 day with a standard deviation of 0.04124(c) 1 day with standard deviation of 0.03352(d)30 days with a standard deviation of 0.03503

As suggested by the time history study, the near-fault data points increase the bounding parameter variability. We thus choose to eliminate these bad players in the bootstrapping process by lowering the threshold to 3. The results are shown in Figure 11.

As shown in Figure 11, the multi-modal behavior becomes less noticeable compared to the bootstrap result obtained before near-fault data point elimination. This result is a good indication that the multi-modal behavior is likely caused by the few near-fault data points in the aggregated satellite error data. After eliminating the near-fault data points, the bounding parameter tends to exhibit more stable behavior.

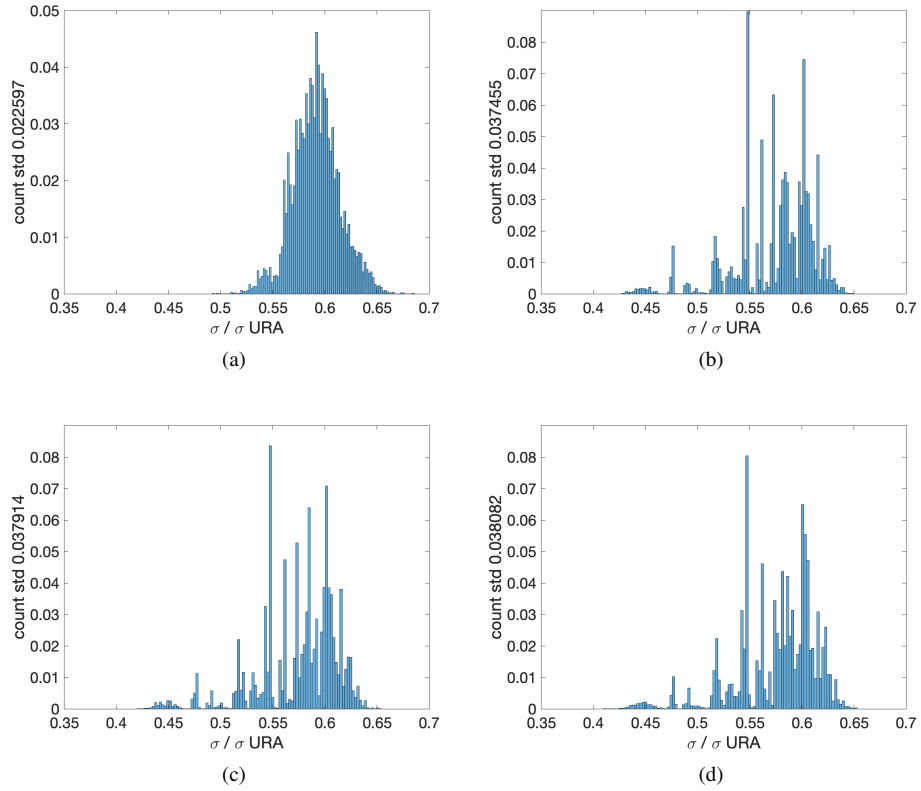

Figure 11: Satellite aggregated bootstrap result for σ for GPS clock and ephemeris MPE for six years with the threshold of 3 units of (a) 15 minutes with a standard deviation of 0.0226(b) 0.5 day with a standard deviation of 0.0375(c) 1 day with standard deviation of 0.0379(d)30 days with a standard deviation of 0.0380

Now, we increase the original sample size as we did before. The results are shown in Figure 12

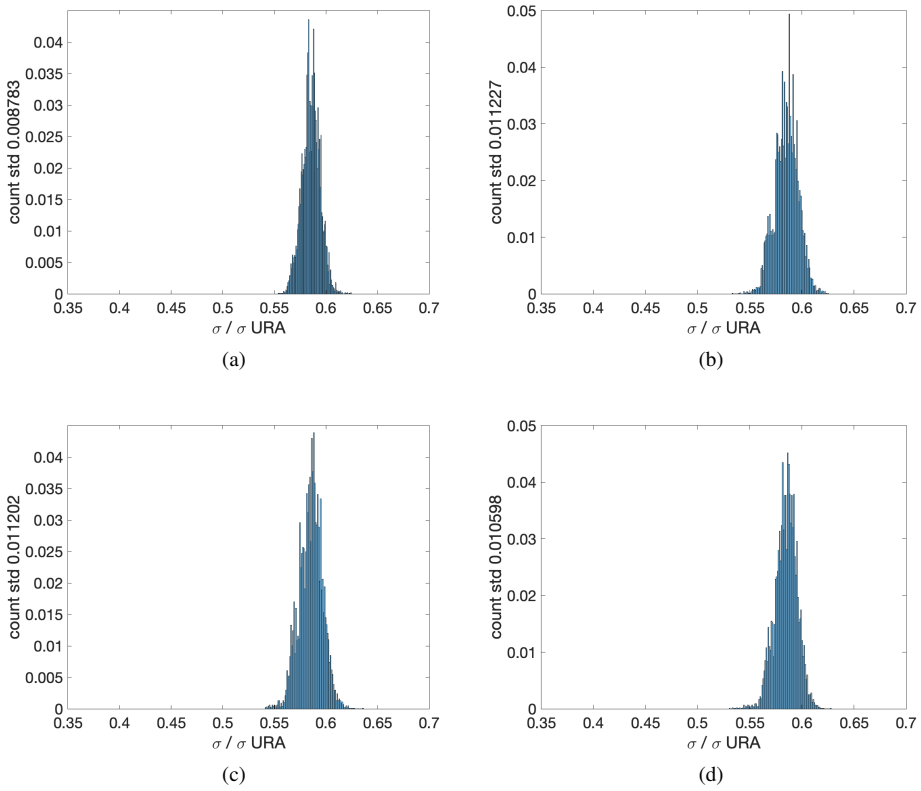

Figure 12: Satellite aggregated bootstrap result for σ for GPS clock and ephemeris MPE for nine years with the threshold of 3 units of (a) 15 minutes with a standard deviation of 0.00878(b) 0.5 day with a standard deviation of 0.0112(c) 1 day with standard deviation of 0.0112(d)30 days with a standard deviation of 0.0106

As we can observe from the result, the multi-modal behavior nearly vanishes, and we are left with highly stable parameter statistics. Finally, we increased the original size to 12 years with the threshold of 3. The result is shown in Figure 13

We can barely observe the multi-modal behavior after we lower the threshold to 3 and increase the original sample size to 12 years. The standard deviation of the parameter distribution significantly decreased. This result again shows that the near-fault data points likely contribute to the uncertainties in the bounding parameter, and by expanding the original sample size, the parameter is stabilized. In Figure 13, we see little variation in the error bonding parameter. This result represents the inherent uncertainty of the Gaussian bounding estimates since it is not affected by the few near-fault data points and uses all the error data available. The plot shows that different sampling units produce similar distribution and standard deviation, suggesting that the time correlation no longer has a significant impact. The σ, as shown in the plot, has a small variation ranging from 0.542 to 0.628. The standard deviation is lower than 0.0095.

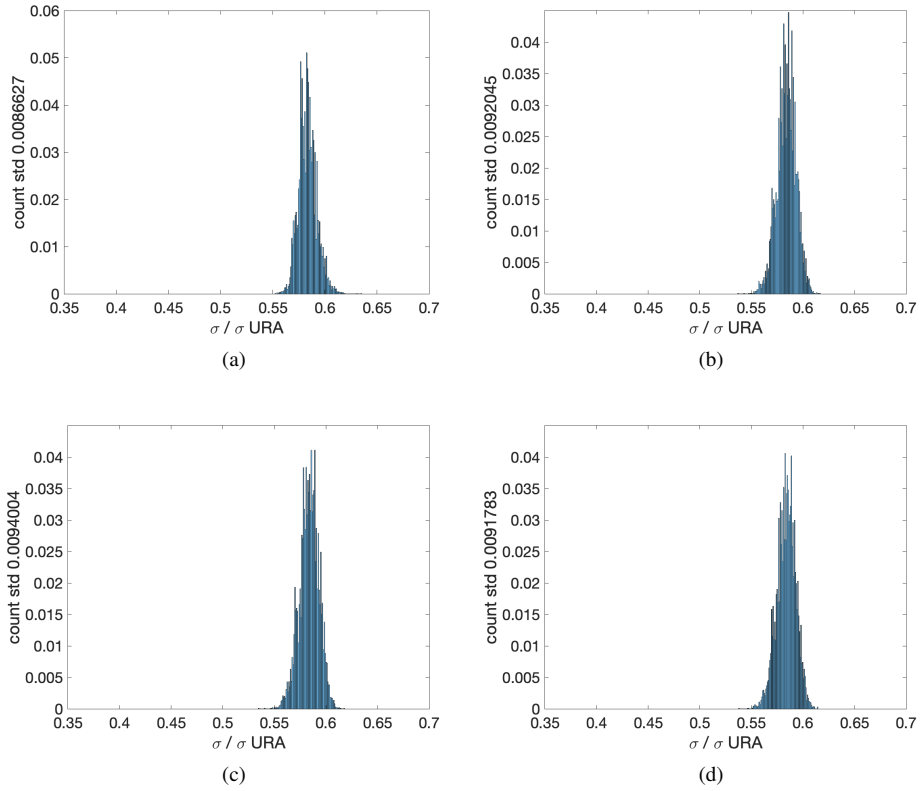

Figure 13: Satellite aggregated bootstrap result for σ for GPS clock and ephemeris MPE for 12 years with the threshold of 3 units of (a) 15 minutes with a standard deviation of 0.00866(b) 0.5 day with standard deviation of0.00920(c) 1 day with standard deviation of 0.0094(d)30 days with a standard deviation of 0.00918

c). Galileo

For Galileo, we have two years of data taken every 5 minutes as indicated in the time history study. We applied the same comparison metric to the Galileo bootstrapping results. In Figure 14 we show the bootstrap result with sampling units of 5 minutes, 12 hours, 1 day, and 1 month for svn 205 using 2 years of data.

The σ for this satellite is 0.17122. We observe similar behavior as the ones we saw in GPS satellites. The uncertainty significantly increased, jumping from 5 minutes unit size to 12 hours unit size. We also observe the apparent multi-model behavior. Thus, we can reach the same conclusion as we did in the GPS case.

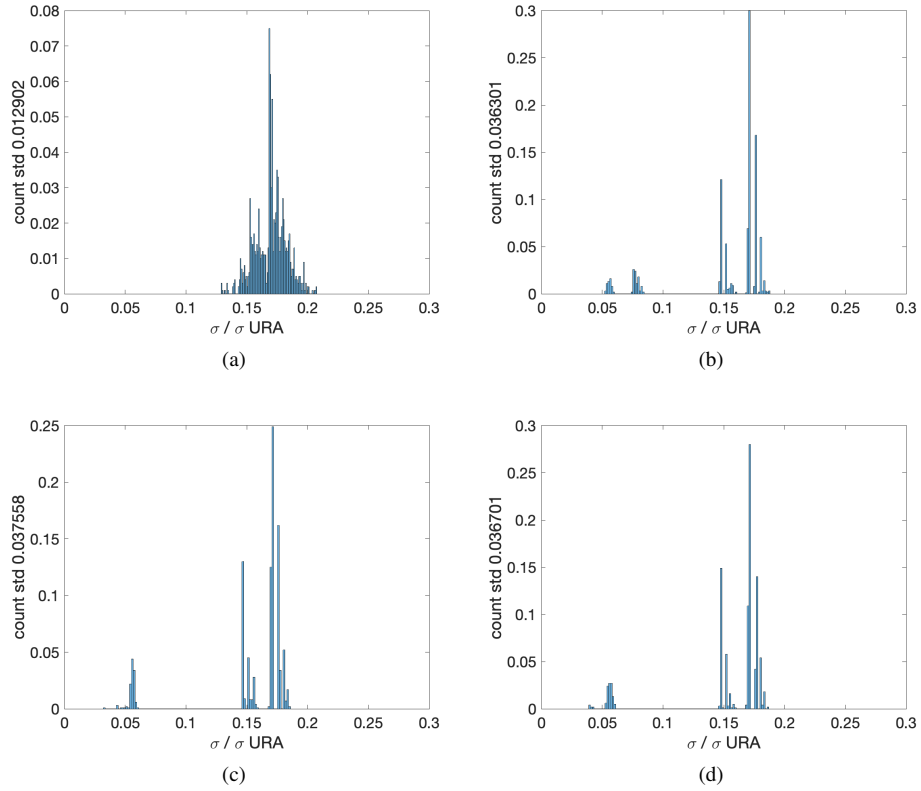

Figure 14: svn 205 bootstrap result for $\sigma = 0.171$ for Galileo clock and ephemeris MPE for 2 years with the threshold of 4.17 units of (a) 5 minutes with a standard deviation of 0.0129(b) 0.5 day with standard deviation of 0.0363(c) 1 day with standard deviation of 0.0375(d)30 days with a standard deviation of 0.0367

Then we plotted the bootstrap result for aggregated Galileo satellite errors. In Figure 15 and 16, we plotted the bootstrap results using the 4.17 threshold and 1.5 years and 2 years of original sample sizes, respectively, with the different unit sizes.

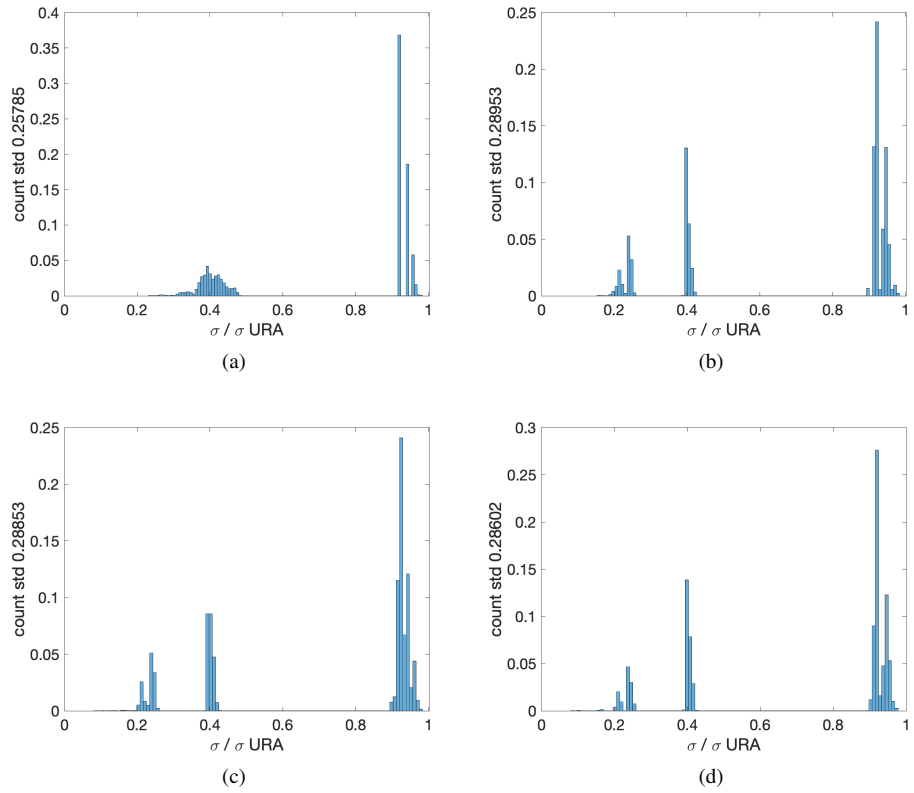

Figure 15: Aggregated bootstrap result for $\sigma = 0.921$ for Galileo clock and ephemeris MPE for 1.5 years with the threshold of 4.17 units of (a) 5 minutes with a standard deviation of 0.258(b) 0.5 day with standard deviation of 0.290(c) 1 day with standard deviation of 0.290(d)30 days with a standard deviation of 0.286

As we can see, the uncertainty decreased with a larger original sample size. The uncertainty increased when we increased the unit size from 5 minutes to 12 hours. However, unlike the GPS case, using the largest original sample size for Galileo did not eliminate the multi-modal behavior. The lack of total data likely caused this behavior. Even after we lower the threshold and eliminate the near-fault data points, we do not observe stable behavior as shown in Figure 17 and 18.

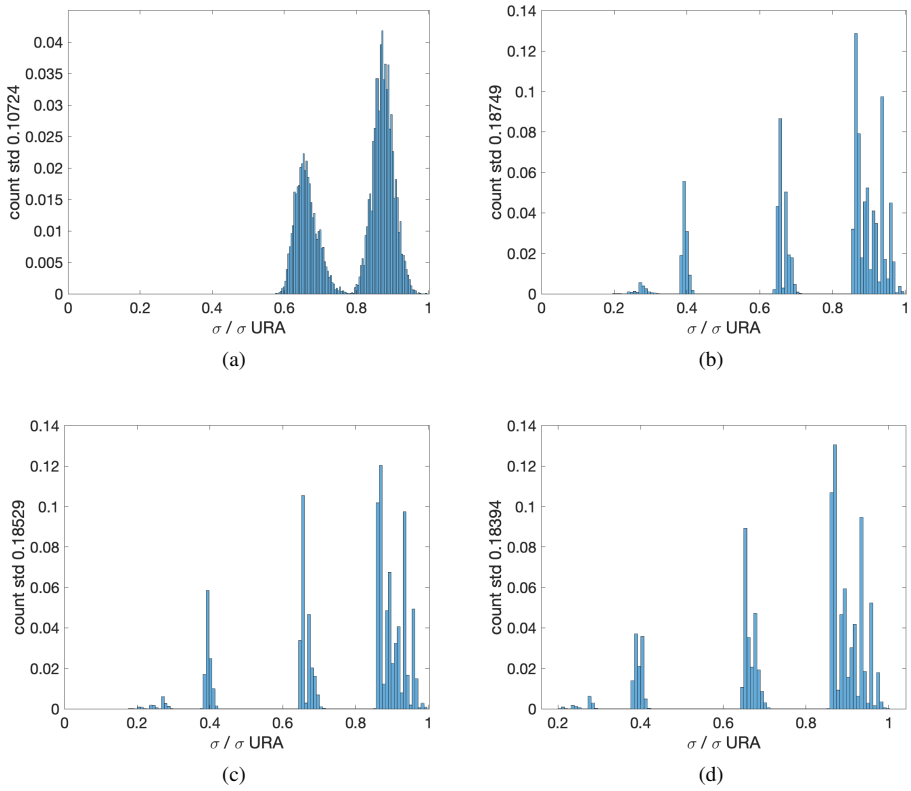

Figure 16: Aggregated bootstrap result for $\sigma = 0.867$ for Galileo clock and ephemeris MPE for 1.5 years with the threshold of 4.17 units of (a) 5 minutes with a standard deviation of 0.107(b) 0.5 day with standard deviation of 0.187(c) 1 day with standard deviation of 0.185(d)30 days with a standard deviation of 0.184

We observe a significant drop in the standard deviation from the plots after eliminating the near-fault data points. This result again shows that the near-fault data points are likely responsible for the instability in the bounding parameter. However, possibly due to the small original sample size, we do not observe a highly stable behavior as we did in the GPS case since in this case, we were only using two years of data.

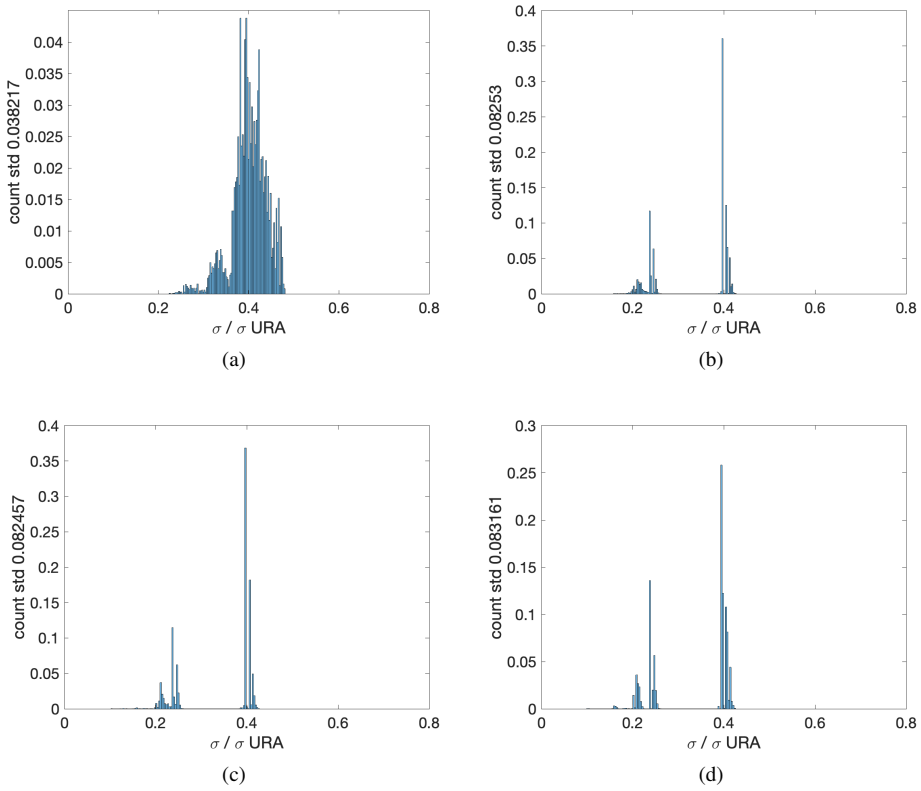

Figure 17: Aggregated bootstrap result for $\sigma = 0.921$ for Galileo clock and ephemeris MPE for 1.5 years with the threshold of 4.17 units of (a) 5 minutes with a standard deviation of 0.03828(b) 0.5 day with standard deviation of 0.0825(c) 1 day with standard deviation of 0.0825(d)30 days with a standard deviation of 0.0831

VI. CONCLUSION

Our study shows that three years of data for each satellite likely provides too few independent data points for time series analysis. Thus, we should use the aggregated satellite error for the time history error analysis of the bounding parameter. The time history plots show that the bias parameter is highly stable, and the parameter is less stable. This result is likely caused due to the near-fault data points.

We then explore using the training-validation method to simulate the data prediction process and find that the past aggregated UPE data for GPS has good stability and predictability and will likely produce conservative bounds for the future errors using as little as two years of data.

Finally, we used the bootstrap to evaluate the uncertainty in our bounding estimate. By varying the sampling units, we include the temporal correlation. The result shows that the past error estimates have acceptable uncertainty even with a high degree of time correlation assumption. The standard deviation of the Gaussian bounding distribution ranges from 0.542 to 0.627, with high stability for GPS. This result shows that the GPS clock and ephemeris error bounding parameter is highly stable and can be well characterized with the available data. For Galileo, on the other hand, with the currently available data, the bounding parameters do not appear to be stable. We suspect that this is due to the lack of data.

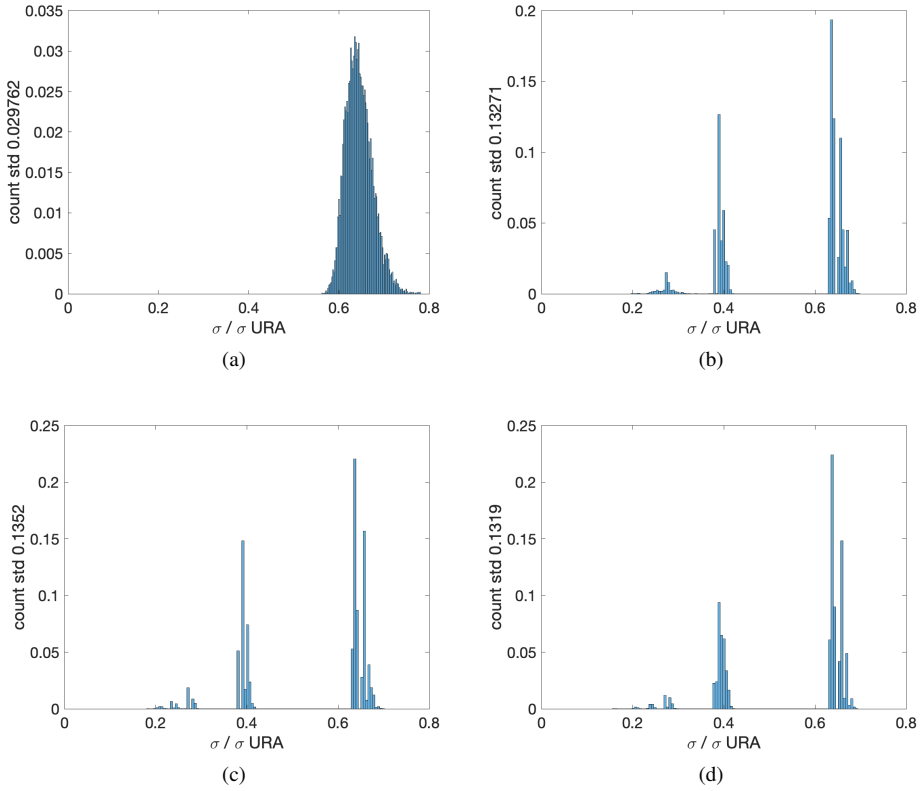

Figure 18: Aggregated bootstrap result for $\sigma = 0.867$ for Galileo clock and ephemeris MPE for 1.5 years with the threshold of 4.17 units of (a) 5 minutes with a standard deviation of 0.0298(b) 0.5 day with standard deviation of 0.133(c) 1 day with standard deviation of 0.135(d)30 days with a standard deviation of 0.132

ACKNOWLEDGEMENTS

We would especially like to thank the FAA Satellite Navigation Team for funding this effort under MOA 693KA8-19-N-00015.

REFERENCES

[1] Gssc.esa.int. 2021. Navipedia.

[2] Walter, T., Gunning, K., Eric Phelts, R., and Blanch, J. (2018) Validation of the Unfaulted Error Bounds for ARAIM. J Inst Navig, 65: 117– 133. doi: 10.1002/navi.214.

[3] Blanch, Juan, Liu, Xinwei, Walter, Todd, "Gaussian Bounding Improvements and an Analysis of the Bias-sigma Tradeoff for GNSS Integrity," Proceedings of the 2021 International Technical Meeting of The Institute of Navigation, , January 2021, pp. 703-713.

[4] J. Blanch, T. Walter and P. Enge, "Gaussian Bounds of Sample Distributions for Integrity Analysis," in IEEE Transactions on Aerospace and Electronic Systems, vol. 55, no. 4, pp. 1806-1815, Aug. 2019, doi: 10.1109/TAES.2018.2876583.

[5] Pullen, Sam, Lo, Sherman, Katz, Alec, Blanch, Juan, Walter, Todd, Katronick, Andrew, Crews, Mark, Jackson, Robert, "Ground Monitoring to Support ARAIM for Military Users: Alternatives for Rapid and Rare Update Rates," Proceedings of the 34th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS+ 2021), St. Louis, Missouri, September 2021, pp. 1481-1507.

[6] 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.

[7] Hesterberg, Tim Monaghan, Shaun Moore, David Clipson, Ashley Epstein, Rachel Freeman, W York, Company. (2005). Bootstrap Methods and Permutation Tests. Introduction to the Practice of Statistics. 14.

[8] Victor Chernozhukov. Econometrics. Spring 2017. Massachusetts Institute of Technology: MIT OpenCouseWare, https://ocw.mit.edu/. License: Creative Commons BY-NC-SA.

[9] L. Wasserman, Class Lecture, Topic: "Statistical Methods for Machine Learning." POS 152, the Machine Learning Department, Carnegie Mellon University, Pittsburgh, Spring, 2018.

[10] Efron and Gong (February 1983), A Leisurely Look at the Bootstrap, the Jackknife, and Cross Validation, The American Statistician.

[11] H. Barreto and F. M. Howland, Introductory Econometrics: Using Monte Carlo simulation with Microsoft Excel. Cambridge: Cambridge Univ. Press, 2013.

[12] J. Blanch, T. Walter and P. Enge, "Gaussian Bounds of Sample Distributions for Integrity Analysis," in IEEE Transactions on Aerospace and Electronic Systems, vol. 55, no. 4, pp. 1806-1815, Aug. 2019, doi: 10.1109/TAES.2018.2876583.