Anomaly Detection and Improving Predictability of GNSS Timing Signal Quality

Martta-Kaisa Olkkonen[1,*,†], Mikko Kotilainen[1,†] and Sanna Kaasalainen[1]

[1]*Finnish Geospatial Research Institute, Vuorimiehentie 5, Espoo, 02150, Finland*

Abstract

This paper discusses some anomalies that affect the reliability and accuracy of GNSS data. In this paper we study patterns in the Common Generic GNSS Timing Transfer Standard (CGGTTS) data. The pseudorange residuals in this data appear to include patterns that repeat every day. We discuss in this paper more the different factors on these patterns, like satellite age and position on the sky. Ability to detect these anomalies will help to identify the satellites with unreliable timing behavior for an improved time solution on the receiver side.

Keywords

Anomalies, CGGTTS, elevation, ionosphere, satellite aging

1. Introduction

This paper reports work-in-progress in analyzing possible effects in satellite signal quality, especially using CGGTTS data. CGGTTS is currently used by over 70 laboratories [1] to compute the Coordinated Universal Time (UTC) together with Precise Point Positioning [2]. The fundamental method was presented in [3], which was based on REASON (Resilience and security of geospatial data for critical infrastructures) project [4]. REASON was funded by the Research Council of Finland and was focusing on resilience of the timing signal achieved from GNSS. In January 2025, a new project started at Finnish Geospatial Research Institute (FGI) funded by the Research Council of Finland, SURI - Supercomputing GNSS data for Navigation Resilience against Ionospheric Interference [5]. In SURI, our aim is to investigate the effect of ionosphere on the GNSS signal quality, not limited to the timing, but in any application of position, navigation and timing (PNT). We will use in the course of the project the Finnish supercomputer LUMI to have a computationally powerful way to extract from GNSS signal the effect of ionosphere, and separate intentional interference from this. We discuss in this paper some precursors to delving into ionospheric effects on PNT.

We discuss in this paper:

- how bias and slope of the measurements can be used in prediction of satellite reliability?
- effect of satellite position in the sky
- effect of satellite aging on predictability of error

2. Materials and Methods

We use in our study the pseudorange residuals of CGGTTS generated from VTT MIKES time-transfer GNSS receiver (receiver code MI05, Septentrio PolaRx5TR), available in the IDA data storage [6]. CGGTTS files were generated from 24-hour Rinex data using R2CGGTTS v. 8.2 [7]. The pseudorange residuals represent the difference between the satellite clock from a dual frequency L3P solution (P1 & P2) and the MIKES UTC realization UTC(MIKE).

WIPHAL'25: Work-in-Progress in Hardware and Software for Location Computation June 10–12, 2025, Rome, Italy

*Corresponding author.
†These authors contributed equally.
martta-kaisa.olkkonen@nls.fi (M. Olkkonen); mikko.kotilainen2@gmail.com (M. Kotilainen); sanna.kaasalainen@nls.fi (S. Kaasalainen)
0000-0002-5302-081X (M. Olkkonen); 0000-0002-4563-9508 (M. Kotilainen); 0000-0001-6628-418X (S. Kaasalainen)

This chapter discusses the method of processing the measurements by modifying the CGGTTS data (Sec. 2.1). Results of long-term observations are presented in Sec. 2.2, and Sec. 2.3 discusses the satellite age on predictability of its modified data, which is at the core of the presented method. The effect of satellite position in the sky is briefly discussed in Sec. 2.4. We present a result of detecting anomalies by observing the biases and slopes of the cycles in Sec. 2.5.

2.1. Modified measurements

The reported pseudorange residuals were averaged over 16-minute periods, when the satellite was visible. In our method, the data set from each cycle is made zero-mean by allowing each cycle to have their individual bias, and subtracting it. We denote this with modification #1. Secondly, only cycles of most common, mode length are shown, and we denote this with modification #2. The resulting data set, illustrated in [3] was more tightly grouped and showed similar patterns across different days.

However, for some satellites, subtracting the bias from the data is not sufficient, because each cycle has its individual slope as well. This statement was supported by looking at the measurements of GPS12 during morning cycles [3]. After subtracting the bias, the modified data sets are tightly grouped around middle periods of high elevation.

If the measurements are drifting away or towards the GPS system time, the modified data set spreads out towards the edge periods of low elevation. This was seen as lower standard deviations than 1.5. However, if each cycle is allowed to have its own slope, and it is subtracted, the modified data set is grouped together more tightly. Subtraction of the slope yielded modification #3. The standard deviation of modified data set still increased toward the edge periods, but now less distinctly. This indicates that knowing the bias and slope of the cycle allows us to make more accurate predictions about the pseudorange residuals.

2.2. Long-term observations of bias and slope in improving predictability

Previous day's bias could be hypothetically used in predicting the following day's morning cycle bias. We aim to verify this by plotting a histogram of their differences. We perform a similar analysis in order to predict the same days evening cycle bias based on the morning cycle bias. The results shown in Fig. 1 indicate that in addition to the morning bias time history, prediction of the bias of a next morning cycle could potentially be more accurate if the previous evening cycle bias was set as a precursor. Adding also the evening cycle's time history enables to obtain a more stable estimate. The difference between the cycle biases are centered around 3 ns in Fig. 1. Compared to biases, slopes behave in a different manner, with the longer evening cycles being more centered around zero, whereas the morning cycle slopes have a higher variance. The regression lines in Fig. 2 reveal a positive correlation between the slope of the next evening cycle and that of the known morning cycle. That is, a higher morning slope is associated with a higher evening slope, and vice versa. An outlying evening cycle slope would be easy to spot.

2.3. Effect of satellite age on the predictability of its modified data

The pseudorange residuals of satellites can be predicted more accurately when we shift to the estimated constellation system time. With the goal being able to make a predictive model for tomorrow's data sets based on the data set time history, we can make comparisons between simplistic models and use those to study if the age of the satellite affects its residual error. The simplistic models that we use to study this are

1. no pooling model, where we simply duplicate yesterday's modified data set and use that to make a prediction about the new value,
2. complete pooling model, where we take the mean of all the modified data sets and use that to make a prediction about the new value, and

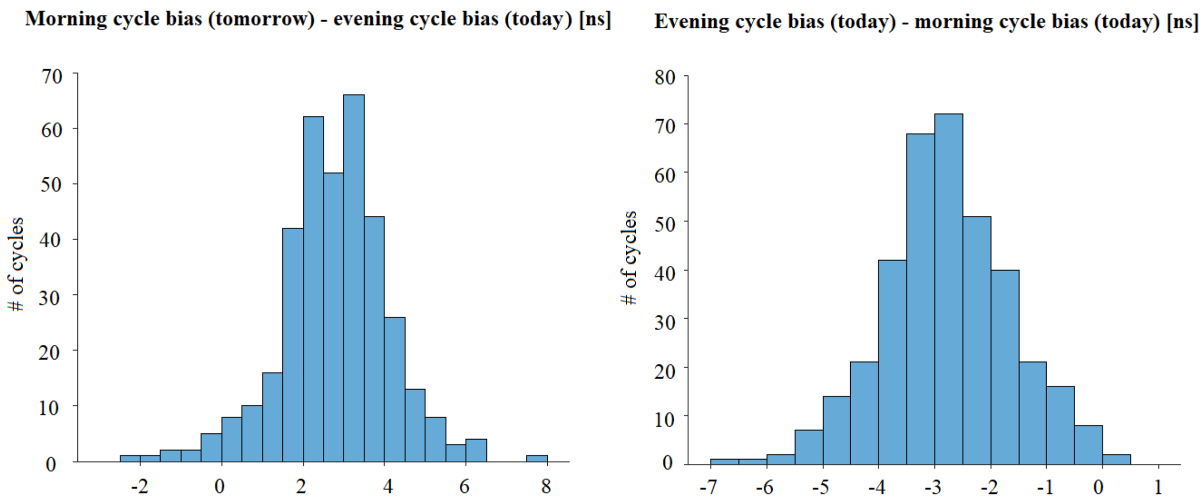

Figure 1: Difference between cycle biases of previous and next cycles. Knowing the morning cycle bias helps predicting the evening cycle bias and vice versa.

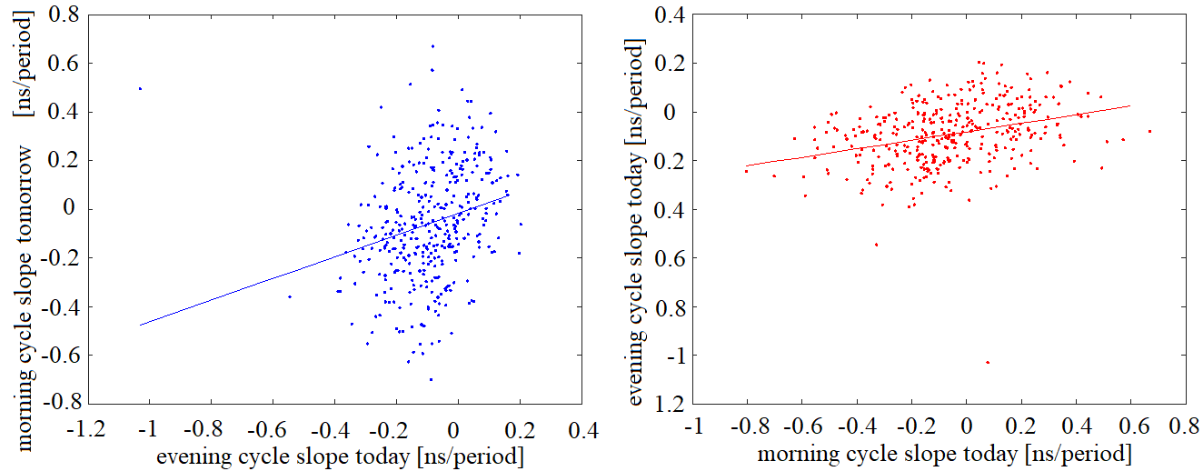

Figure 2: A positive correlation between morning and evening slopes means that accuracy of one's prediction increases if the other (previous one) is known.

3. partial pooling model, where we make a compromise between these two models, and give 70% of the weight to no pooling-model and 30% of the weight to complete pooling-model.

The 70%/30% ratio was selected because its predictions had the smallest standard error of the ratios for each period. The example results are shown in Fig. 3. The partial pooling-model is the best predictor, suggesting some regularization should be used in the prediction models. Only cycles of mode length are analyzed, resulting in some days not present, such as late yellow days in evening cycles. We also see that the predictions are less certain around the edge periods, possibly due to errors from low elevation angles. Both this pattern and the partial pooling being the best predictor hold for all other GPS satellites also.

Next, we take the mean of the residual standard errors across periods for all satellites to see if older satellites have higher residual errors, or in other words, if new satellites have more predictable modified data sets. The results are shown in Fig. 4. Contrary to expectations, the older satellites do not have less predictable modified data sets, and the highest scatter in the modified data sets is for a relatively new GPS8.

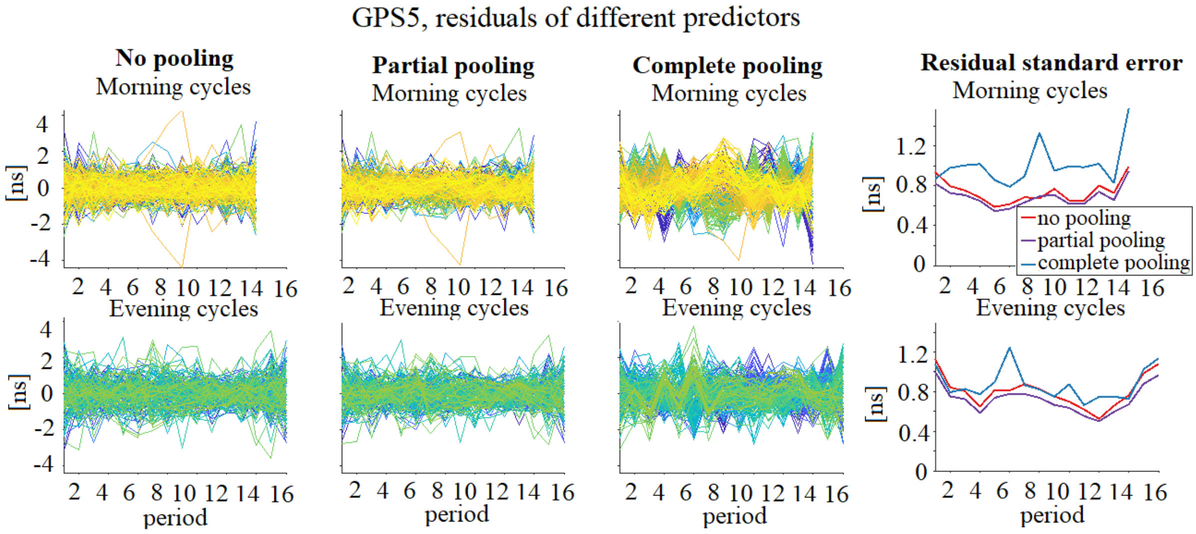

Figure 3: Errors of simplistic prediction models for cycles of mode length. Partial pooling model is the best at predicting future modified data sets in almost all epochs.

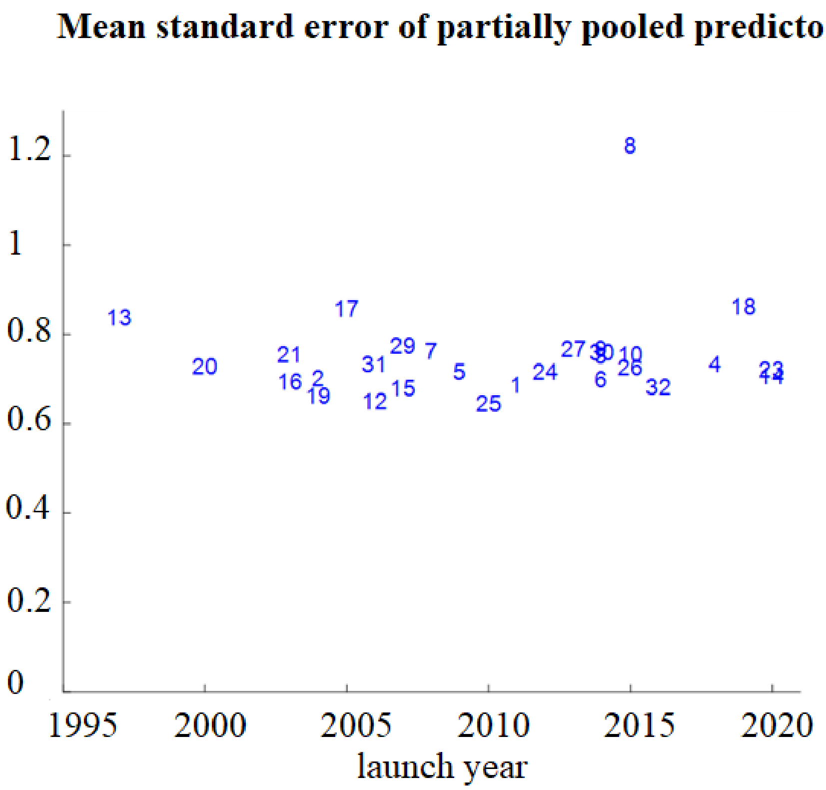

Figure 4: The average prediction error for satellites based on the year they have been launched. The error seems relatively stable across years with one outlier.

2.4. Effect of satellite position in the sky

Looking at the modified data set for consecutive days in Fig. 3 (of similar colors, for example yellow) reveals that the similar colors are more grouped than being random draws from the distribution based on where the satellite is in the sky. Fig. 5 demonstrates a more subtle pattern and adds to the prediction accuracy of the modified data set as a function of where the satellite is positioned in the sky. Thus, we can obtain a prediction with an even smaller standard deviation based on how the modified data set

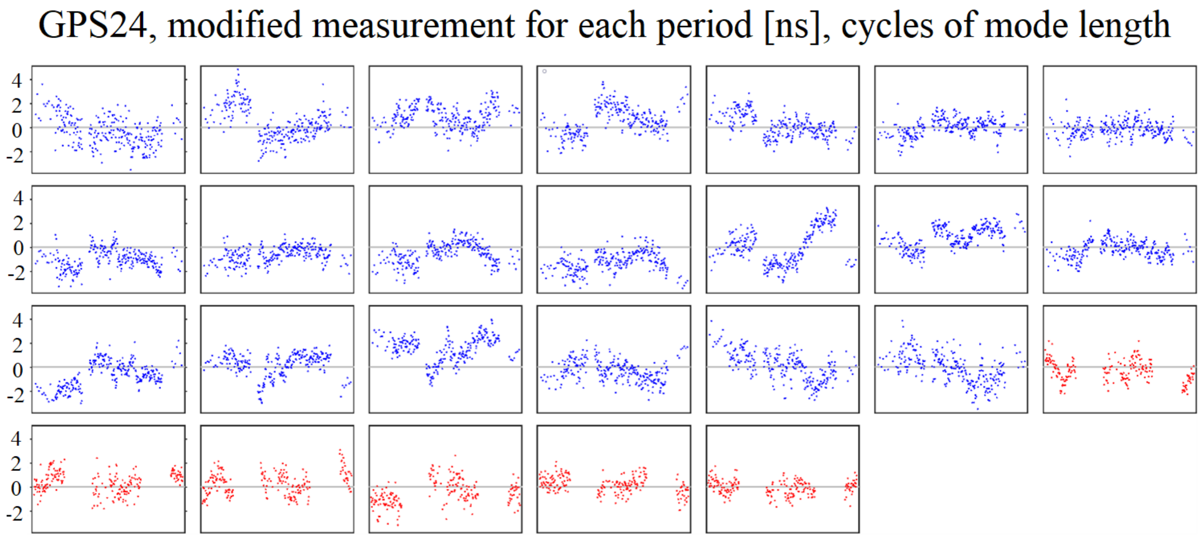

Figure 5: Modified data of each period as a function of their sidereal day, for only cycles of mode length. The top left panel shows how the modified data from first period of morning cycles evolves throughout the year, whereas the first panel with the red dots represents the first period of evening cycles. There are patterns evolving from one day to next. Also note that for morning cycles, the modified data set at edge periods of low elevation have a higher scatter than the middle periods of high elevation.

has behaved in the previous several days. This latter observation is also better visible with another perspective on the data, where we plot the modified data set for each period separately as a function of their sidereal day. The drawback of this graph shown in Fig. 5 is that the cycles must be restricted to be of a specific length, leading the modified data from cycles of other lengths to be removed. The residual error appears to be around 1 ns for the central periods and higher for the edge periods.

2.5. Method for detecting anomalies

If we assume that the bias and slope of the cycle are changing slowly across different days, we can detect anomalies by looking at the biases and slopes of the cycles. Two cases were presented in [3]. The third case was found by looking at the modified data set of Galileo 1. To get more data for the cycles repeating every 17th cycle, we extended the analyzed days to Modified Julian Dates 59300-59805. The modified data set in Fig. 6 are divided into cycles so that every subplot contains modified data from similar cycles that are of mode length. The modified data set for one cycle, marked with red in the first subplot appear different when compared to others. Neither of the suggested anomalous GPS measurement cycles (starting June 6th 2022, 06:46 UTC and April 7th 2022, 15:02 UTC, respectively) coincided with the reported NANU [8] or with the GPS problem report status [9].

3. Discussion on the results and scalability of the method

We have studied satellite aging, elevation as well as slope and bias of measurements as a precursor for further studies in extracting atmospheric effects from the data sets, in particular discerning ionospheric effects from other intentional interference like jamming and spoofing. More accurate bias and slope estimation will enable improved prediction of the measurements. As we gather more measurements from the cycle, the accuracy of the bias and slope estimates improves. At the final periods of the cycle, we have a very accurate estimate of both the bias and the slope and the uncertainty only comes from the uncertainty of the modified data set. This uncertainty comes from the filtering step that excludes data from some cycles. The method includes subtracting the bias from the CGGTTS data, but each cycle has its individual slope as well. The modified data set is constrained in that it might not be comparable to a different data set obtained at a different operational condition: the method is affected by measurement

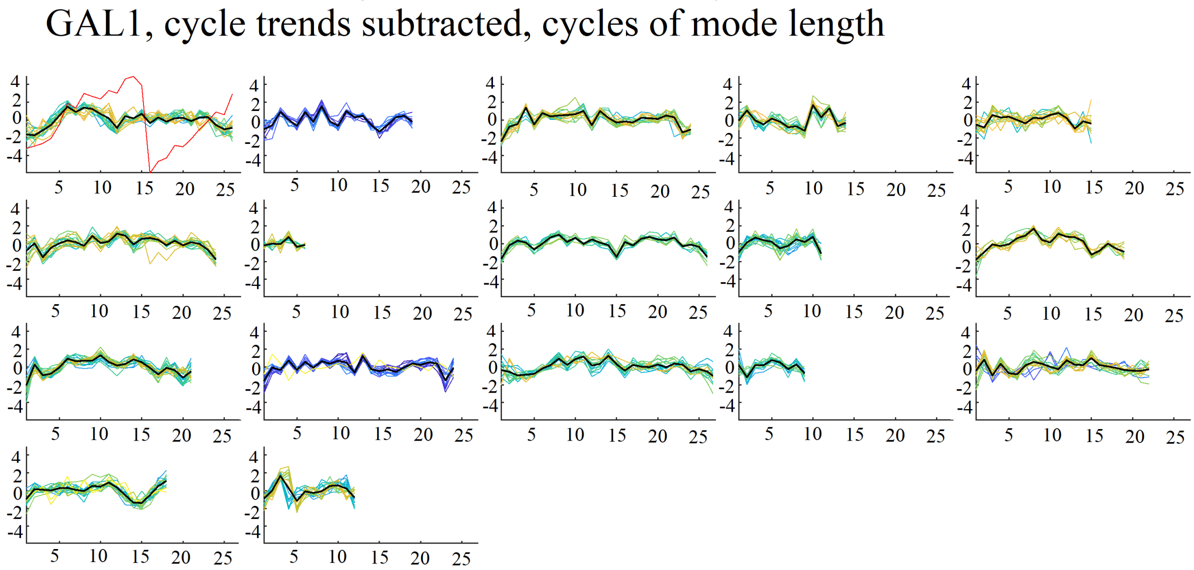

Figure 6: Modified data set for Galileo 1 grouped so that each subplot contains similar cycles of mode length. The anomalous modified data set are depicted with a red curve in the first subplot.

uncertainties when gathering the CGGTTS data. Most importantly, the stability of the receiver is an essential factor. If the antenna is moved intentionally or unintentionally, or some changes occur around, say, a new tall building is built, the effect on the CGGTTS data is significant. Namely, the system would need recalibration from time to time. In addition, the method is not really scalable to be comparable to different locations without a suitable calibration method, which has not yet been discussed in the framework of this work. Therefore, developing a suitable calibration method in order to improve the scalability of the method would be imperative.

The patterns for each of the satellites are dependent on the environment around the antenna of the GNSS receiver used in generating the CGGTTS files. Collecting the data from the entire track enables us to find the anomalous satellites sooner than the geodetic time transfer method [7]. Limitation of this approach is that if the antenna is moved or the surrounding environment is significantly changed (for example, a new tall building is built nearby), the above parameters need to be re-estimated. In general, the standard deviations of modified data sets are higher for lower elevation periods than for higher elevation periods. This indicates that the atmospheric effects are not entirely mitigated by using the dual frequency solution and subtracting the cycle slope. This is because the dual frequency solution only mitigates the first order of the ionospheric error, and not that of the tropospheric error, which could be included in the future model. Subtracting the slope allows us to more accurately predict the modified data sets down to the elevation angles of 10 degrees. If the data sets would contain other than linear terms, they would be showing in the modified data sets. For example, parabolic modified data sets would be indicated by the modified data sets at both edge periods being higher or lower than in the middle periods. There does not seem to be any clear indications of these.

4. Conclusion and future work

This paper presented results of ongoing work to detect anomalies in satellites' timing solutions. In the future work, a predictive model should be built. Utilizing past data in the manner described in this paper, the expected behavior of each satellite can be modeled. The model can predict plausible intervals for the bias and slope based on historical data and integrate that uncertainty with the modified data set uncertainty, allowing us to quantify the now descriptive anomaly identification. If the actual data doesn't agree with the prediction interval at any period, it would be possible to warn the user from using the time solutions from this satellite. In SURI project, we intend to gather large data sets of

GNSS data that contains both ionospheric scintillation and intentional interference, and apply machine learning methods in order to be able to gain a situational awareness of quality of GNSS signals in a larger scale. We will use the supercomputer LUMI for modeling the effect of ionosphere very accurately on the PNT solution.

Acknowledgments

This work was supported by the Research Council of Finland under Grant 338042 (REASON) and Grant 364761 (SURI).

Declaration on Generative AI

The author(s) have not employed any Generative AI tools.

References

[1] G. Panfilo, F. Arias, The coordinated universal time (utc), Metrologia 56 (2019) 042001. URL: https://dx.doi.org/10.1088/1681-7575/ab1e68.

[2] J. Ray, K. Senior, Geodetic techniques for time and frequency comparisons using gps phase and code measurements, Metrologia 42 (2005) 215. URL: https://dx.doi.org/10.1088/0026-1394/42/4/005.

[3] M. Kotilainen, M. Mäkelä, K. Hanhijärvi, M.-K. Olkkonen, A. Wallin, T. Fordell, S. Kaasalainen, Detecting consistent patterns in pseudorange residuals in gnss timing data, in: 2023 International Conference on Localization and GNSS (ICL-GNSS), 2023, pp. 1–5.

[4] S. Kaasalainen, M. Mäkelä, L. Ruotsalainen, T. Malmivirta, T. Fordell, K. Hanhijärvi, A. Wallin, T. Lindvall, S. Nikolskiy, Reason-resilience and security of geospatial data for critical infrastructures, in: A. Ometov, J. Nurmi, E. S. Lohan, J. Torres-Sospedra, H. Kuusniemi (Eds.), WiP Proceedings of the International Conference on Localization and GNSS (ICL-GNSS 2021) (CEUR Workshop Proceedings), volume 2880, 2021.

[5] Finnish geospatial research institute, suri project, 2025. URL: https://www.maanmittauslaitos.fi/en/research/suri-supercomputing-gnss-data-navigation-resilience-against-ionospheric-interference.

[6] Fairdata, Ida data storage, 2023. URL: https://ida.fairdata.fi/apps/files/?dir=/&fileid=74397824.

[7] P. Defraigne, C. Bruyninx, On the link between gps pseudorange noise and day-boundary discontinuities in geodetic time transfer solutions, GPS Solutions 11 (2007) 239–249.

[8] T. S. Kelso, Gps nanu messages 2022, 2023. URL: https://celestrak.org/GPS/NANU/2022/.

[9] Navigation center, united states coast guard u.s. department of homeland security, gps user issue detection evaluation (guide) tool, 2023. URL: https://www.navcen.uscg.gov/gps-problem-report-status.