GLOBAL INTEGRITY MONITORING FOR PRECISE SATELLITE NAVIGATION

A DISSERTATION
SUBMITTED TO THE DEPARTMENT OF
AERONAUTICS AND ASTRONAUTICS
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Yu-Fang Lai
July 2026

© 2026 by Yu Fang Lai. All Rights Reserved.
Re-distributed by Stanford University under license with the author.



This work is licensed under a Creative Commons Attribution-Noncommercial 3.0 United States License.
https://creativecommons.org/licenses/by-nc/3.0/legalcode

This dissertation is online at: https://purl.stanford.edu/jm332cz8952

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Todd Walter, Primary Advisor

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Sigrid Elschot

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Grace Gao

Approved for the Stanford University Committee on Graduate Studies.

Kenneth Goodson, Vice Provost for Graduate Education

This signature page was generated electronically upon submission of this dissertation in electronic format.

Abstract

The Global Navigation Satellite System (GNSS) provides absolute positioning services to users based on radio signals transmitted by GNSS satellites. One of the most prominent and the pioneering GNSS service is the Global Positioning System (GPS) owned and operated by the United States government. While the typical GNSS service can provide meter level positioning accuracy, Precise Point Positioning (PPP), derived from utilizing corrections to these same measurements, can achieve centimeter level accuracy for static receivers and decimeter level accuracy for kinematic receivers. For any applications of PPP with safety-critical requirements, the integrity of the positioning result is crucial and thus must also be delivered for PPP to be trusted. This dissertation develops a ground-based integrity monitor for precise GNSS services such as PPP. I demonstrate that the integrity monitor can track any type and any number of satellite clock and ephemeris faults in real-time while maintaining decimeter level accuracy and protecting users anywhere on Earth. The satellite clock faults and ephemeris faults are two of the most common and relevant satellite faults observed over the operation history of GPS. The monitoring results for several real satellite fault events are evaluated and shown to be bounded by the computed confidence bounds. The performance of the integrity monitor primarily relies on the geometric relations of the ground reference receivers, so the design of the network geometry is of importance. In particular, the placement and selection of the receivers for the integrity monitor are formulated as a convex optimization problem that can optimally determine the best network design for a given number of monitoring receivers. I address the questions about where to place the receivers and how many of the receivers are required to obtain the desired level of performance, for users

at arbitrary locations globally. Together, these results describe the first complete global high-integrity ground-based monitoring system to protect users against satellite clock and ephemeris faults while providing decimeter level accuracy and meter level confidence bounds. This is an order of magnitude lower than similar systems such as the Wide Area Augmentation System (WAAS).

Acknowledgements

When I was admitted to the Master's program in the Department of Aeronautics and Astronautics at Stanford in January 2020, I had never thought about pursuing a Ph.D. In January 2022, I was doubtful about whether I should become a Ph.D. student or a software engineer. I can proudly say to four-years-ago me: you made the right choice to stay at Stanford. I have learned so many things from the world of academia, and I am grateful that some of my ideas, building upon the great work of numerous predecessors, can be contributed to this world.

I want to thank my advisor, Professor Todd Walter, for giving me the opportunity to join the GPS Lab. You guided me through all the difficulties and uncertainties of doing research and led me to the world of academia. I really enjoyed the time spent in your office discussing all the possibilities and potential solutions to problems. Your incredible knowledge of GNSS, and incredible patience and support to the students, helped me navigate through all the technical and mental challenges, especially as an international student. This Ph.D. work could not be completed without your supervision.

Special thanks to Fiona Walter for proofreading this dissertation.

I want to thank Dr. Juan Blanch. From you I learned how to mathematically and rigorously examine claims, equations, and manuscripts. The solution to a problem does not always have to be complicated. My work, and this dissertation, became mathematically reliable thanks to your scrutinous evaluation. I also want to thank you for reviewing this dissertation.

I want to thank Dr. Yu-Hsuan Chen. Thank you for the snacks you brought from Taiwan, and thank you for being there all the time, listening to my complaints and

lecturing me about the hardware.

I want to thank other members of this dissertation committee, Professor Grace Gao and Professor Sigrid Elschot. Thank you for providing your valuable comments and suggestions to this dissertation.

I want to thank other research engineers in the GPS Lab, Dr. Sherman Lo, Dr. Sam Pullen, and Dr. Eric Phelts, for helping me prepare my presentations and providing valuable advice on how to improve them.

I want to thank the students in the GPS Lab. Rebecca Wang: Thank you for letting me know that there was an opportunity to join the GPS Lab. We prepared for the qualification exam together and made it through. I believe you will be the next one to graduate from the lab. Dr. Zixi Liu: I enjoyed the lunch we had together and all the snacks you shared. Argyris Kriezis and Jade Babcock-Chi: Good luck with your continued Ph.D. research. I really enjoyed the trips we took to Hawaii and all the dinners we had at Greek restaurants.

I want to thank Professor Shau-Shiun Jan. The first time I heard of Stanford University and learned about the existence of the GPS Lab was in your class during my undergraduate studies at NCKU. It is hard to believe that I would eventually join the same lab as yours and also become a Ph.D. from the GPS Lab. I really appreciate your humor, conversations, and dinners during conferences and whenever I was back at NCKU.

Finally, I want to thank my parents. Thank you, mom, for continuously and unconditionally supporting the decision I made. Thank you, dad, for worrying about me but still supporting my decision. Moving abroad from Taiwan and pursing a higher degree across the continent has never been easy, physically and mentally. Your words and well-wishes are the reasons I was able to accomplish my Ph.D. degree. Thank you.

Contents

Abstract v

Acknowledgements vii

1 Introduction 1

1.1 Precise Point Positioning . . . 3

1.1.1 Measurement Model . . . 4

1.1.2 Observation Modeling . . . 5

1.1.3 Estimator . . . 18

1.1.4 PPP User . . . 22

1.2 Integrity Requirement . . . 24

1.3 Continuity Requirement . . . 28

1.4 Contributions . . . 30

1.4.1 Integrity Monitor . . . 30

1.4.2 Network Design . . . 30

1.4.3 Tropospheric Error Analysis . . . 31

1.4.4 Publications . . . 31

1.5 Outline . . . 33

2 Ground-Based Integrity Network 35

2.1 Fault-Tolerant Filter . . . 35

2.1.1 Single Receiver Self Monitoring . . . 35

2.1.2 Multi Receiver Monitor Network . . . 45

2.1.3 Calibration of Receiver Differential Code Biases . . . 65

2.2 Global Integrity Monitor . . . 69
2.2.1 GPS Clock Fault at PRN-22 on October 2nd, 2022 . . . 70
2.2.2 GPS Clock Fault at PRN-1 on January 25th, 2023 . . . 71
2.2.3 GPS Clock Fault at PRN-1 on July 10th, 2023 . . . 72
2.2.4 GPS Clock Fault at PNR-7 on May 25th, 2024 . . . 73
2.3 Regional Integrity Monitor . . . 73
2.3.1 Limitations of a Regional Network . . . 75
2.4 Summary . . . 75

3 User Integrity from the Monitor **78**
3.1 User Integrity and Performance . . . 79
3.1.1 Dataset and Monitor Network . . . 80
3.1.2 Test Setup . . . 81
3.1.3 Test Results . . . 84
3.2 Summary . . . 86

4 Network Integrity of the Monitor **88**
4.1 Receiver-wise Solution Separation Statistics . . . 88
4.1.1 Probability of Receiver Failure . . . 89
4.1.2 Receiver Fault Detection . . . 91
4.1.3 Receiver Fault Exclusion . . . 93
4.1.4 Robust Integrity Information . . . 94
4.2 Evaluation Using Global Data . . . 95
4.2.1 AIV UREs and Sigmas . . . 97
4.2.2 Detection of Receiver Failure . . . 99
4.2.3 Exclusion of Receiver Failure . . . 101
4.2.4 Robust Integrity Information . . . 104
4.2.5 Suspected Ephemeris Anomaly . . . 106
4.2.6 Global Projection . . . 108
4.2.7 Integrity of the Estimator . . . 112
4.3 User Integrity from the Robust Integrity Information . . . 115
4.4 Summary . . . 117

5 Network Performance Evaluation **119**
5.1 Defining Performance of an Integrity Network . . . 120
5.2 Steady-State Kalman and Information Filter . . . 121
5.2.1 Steady-State Covariance from Discrete-Time Algebraic Riccati Equation (DARE) . . . 121
5.2.2 Evaluation of the Protection Levels . . . 124
5.3 Steady-State Information Approximation of Satellite Errors . . . 127
5.3.1 Heterogeneous Receiver Priors . . . 130
5.4 Full Simulation Via Almanac . . . 130
5.5 Summary . . . 135

6 Network Geometry Design **136**
6.1 Introduction . . . 136
6.2 Heuristic Receiver Selection Via Monte-Carlo . . . 138
6.3 Receiver Location Selection Via Convex Optimization . . . 140
6.3.1 Problem Formulation . . . 140
6.3.2 Selection Result . . . 145
6.3.3 Analysis of Coverage . . . 149
6.4 Receiver Number Selection . . . 151
6.5 Summary . . . 152

7 Troposphere Error Analysis **154**
7.1 Introduction . . . 154
7.2 Troposphere Delay Models: UNB3, GPT2W, and GPT3 . . . 155
7.3 The Dataset and Residuals . . . 157
7.4 Residuals Analysis . . . 158
7.4.1 The Grouping Criteria . . . 158
7.4.2 Quantile-Quantile Plot . . . 159
7.4.3 Residual Distribution . . . 160
7.5 Paired-Gaussian Bounding . . . 165
7.6 Bounding the Residuals . . . 167
7.6.1 Group Paired-Bound . . . 168

7.6.2 Overall Paired-Bound . . . 169
7.7 Alternate SBAS VPL . . . 171
7.7.1 Current SBAS VPL . . . 171
7.7.2 Alternate SBAS VPL . . . 174
7.7.3 Sigma-Bias Value for Alternate VPL Equation . . . 176
7.8 Summary . . . 178

8 Conclusions **180**
8.1 Integrity Monitor . . . 180
8.2 Network Design . . . 182
8.3 Tropospheric Error Analysis . . . 182

A Temporal Modeling of User Latency **183**
A.1 Gaussian Processes . . . 183
A.2 Choice of Design . . . 186
A.2.1 N: Total Time Steps of Past Information . . . 186
A.2.2 $\hat{f}(t_p|\cdot)$: The Prediction Model . . . 186
A.2.3 $\hat{\sigma}_{t0}$, ν_{t_0}: Priori of URE . . . 186
A.2.4 $\rho(t, t')$: Correlation of Time . . . 187
A.3 Test Results . . . 188
A.3.1 Latency: 30 seconds . . . 188
A.3.2 Latency: 60 seconds . . . 190
A.3.3 Latency: 90 seconds . . . 192
A.3.4 Latency: 120 seconds . . . 194

B Horizontal Protection Levels **195**

C Paired Bounds Ad-hoc Algorithm **199**
C.1 Minimum Bias . . . 199
C.2 Constant σ . . . 200

BIBLIOGRAPHY **206**

List of Tables

1.1 Deterministic Propagation Models 19
1.2 Initial Noise 20
1.3 Process Noise 20
1.4 Measurement Noise 21

2.1 Process Noise of Satellite Errors 49

List of Figures

1.1 Receiver clock bias (in meters) of STK2 station from IGS network. . 7
1.2 Receiver clock bias (in meters) of STFU station from IGS network. . 8
1.3 Region where MW and PIR cycle slip detectors cannot detect. 17
1.4 Position error estimates in East-North-Up direction. 23

2.1 Fault injection types. 39
2.2 Position error in East-North-Up direction with 30 m step fault. 40
2.3 Position error in East-North-Up direction with 5 cm/sec ramp fault. . 41
2.4 Satellite clock error with 30 m step fault injection 42
2.5 Position error in East-North-Up direction with satellite clock error state and 30 m step fault. 42
2.6 Satellite clock error state with 5 cm/sec ramp fault injection 43
2.7 Position error in East-North-Up direction with satellite clock error state and 5 cm/sec fault. 43
2.8 Satellite clock error state value on January 25th. 44
2.9 56 selected IGS stations. 58
2.10 UREs of 56-station network on a normal day at JPLM. 59
2.11 URAs of 56-station network on a normal day at JPLM. 59
2.12 UREs of 56-station network with 60 m artificial step fault on GPS-PRN-1 at JPLM. 61
2.13 Centered UREs of 56-station network with 60 m artificial step fault on GPS-PRN-1 at JPLM. 63
2.14 UREs of 56-station network with 60 m artificial step fault on PRN-1 and -1 cm/sec ramp fault on PRN-19 at JPLM. 64

2.15 UREs on January 25th from multi-receiver network. 65
2.16 Nominal UREs without difference code bias corrections. 66
2.17 Fitting result of the first 15 RDCBs of AC23 station. 68
2.18 Nominal UREs with difference code bias corrections. 68
2.19 UREs on October 2nd, 2022 from multi-receiver network. 70
2.20 UREs on July 10th, 2023 from multi-receiver network. 72
2.21 UREs on May 25th, 2023 from multi-receiver network. 73
2.22 38 WAAS stations. 74

3.1 UREs and Sigmas of 30 m Step Fault at JPLM. 81
3.2 UREs of $\frac{1}{120}$ m/s ramp fault. 82
3.3 UREs of ephemeris faults with 30 m step offset on y and z directions. 83
3.4 Position Errors and Sigmas at JPLM with integrity risk. 84
3.5 Position Errors and protection levels at JPLM with integrity information and zero latency. 85

4.1 Probability of Number of Receiver Faults Greater than Number of Receiver Exclusions. 90
4.2 Estimation results of the UREs from the 56-station network with data from April 15th to May 20th in 2023. 97
4.3 SS statistics of AIV set with 5.74 sigma threshold. 99
4.4 SS statistics of the correct and incorrect exclusions at JPLM. 101
4.5 Positioning result from MIKL on May 19th, 2023, using RTKLIB as PPP user. 103
4.6 Integrity Information. 105
4.7 Normalized Integrity Information. 106
4.8 Subset UREs of PRN-20 during the anomaly. 107
4.9 120 uniformly distributed global users. 109
4.10 Max. SS statistics of AIV set with 6.5 sigma threshold at all 120 locations. 110
4.11 SS statistics of the correct and incorrect exclusions at all 120 locations. 111

4.12 Histogram of the normalized integrity information. The ephemeris anomaly event is excluded from the counts. 112
4.13 QQ-plot of the Empirical CDF of the Normalized UREs and the corresponding overbound. 114
4.14 Position errors of CPVG in East-North-Up direction under the ephemeris anomaly at PRN-20 on May 19th. 115
4.15 Position errors and protection levels of CPVG in East-North-Up direction under the ephemeris anomaly at PRN-20 on May 19th, with robust integrity information. 116
4.16 Position errors and protection levels of CPVG in East-North-Up direction under the ephemeris anomaly at PRN-20 on May 19th, with AIV integrity information. 116

5.1 Actual and predicted steady-state protection levels. 123
5.2 Steady-State snapshot VPL from 56-station network using real integrity information. 125
5.3 99% Availability Steady-State VPL from real network. 126
5.4 Ephemeris simulation from almanac. 131
5.5 True and steady-state protection levels. 132
5.6 Larger true and steady-state protection levels. 133
5.7 99% Availability Steady-State VPL from Almanac. 134

6.1 5°×5° Global Grids. 137
6.2 519 IGS stations. 139
6.3 200 evenly spread-out candidate receivers. 146
6.4 Evenly spread-out Receiver Selections 147
6.5 IGS Receiver Selections 149
6.6 Coverage vs. Availability for 3 different networks with 56 receivers. . 150
6.7 Coverage vs. Availability for evenly spread-out network with receiver number ranges 40 ∼ 60. 151

7.1 IGS Map and stations in 2022. [1] 157

7.2 Normal CDF and QQ-plot . . . 159
7.3 Median residuals of UNB3 model in 2021. . . . 160
7.4 Maximal and Minimal residuals of UNB3 model in 2021. . . . 161
7.5 Maximal and Minimal residuals of UNB3 model in 2021 after removal of data. . . . 163
7.6 Median residuals of GPT2W model in 2021. . . . 164
7.7 Maximal and Minimal residuals of GPT2W model in 2021. . . . 164
7.8 Gaussian Paired-Bound . . . 166
7.9 Group Minimal Bias Gaussian Paired-Bound in 2021. . . . 168
7.10 Overall residual paired-bound in 2021. . . . 169
7.11 Overall Paired-Bound in 2021. . . . 170
7.12 Overall residual in 2012∼2021. . . . 170
7.13 SBAS Tropospheric Error vs. VPL . . . 173
7.14 SBAS Tropospheric Error vs. Corrected VPL . . . 175
7.15 SBAS Tropospheric Error vs. Corrected VPL without Non-tropo errors 176
7.16 SBAS Tropospheric Error vs. Corrected VPL with Non-tropo errors . 178

A.1 Position errors and protection levels at JPLM with integrity information under 30-second latency. . . . 188
A.2 Position errors and protection levels at JPLM with integrity information under 60-second latency. . . . 190
A.3 Position errors and protection levels at JPLM with integrity information under 90-second latency. . . . 192
A.4 Position errors and protection levels at JPLM with integrity information under 120-second latency. . . . 194

B.1 99% Availability steady-state HPL from partial simulation. . . . 195
B.2 99% Availability steady-state HPL from full simulation. . . . 196
B.3 99% Availability steady-state HPL from 200 evenly spread-out candidate receivers. . . . 197
B.4 99% Availability steady-state HPL from IGS candidate receivers. . . . 198

Chapter 1

Introduction

The Global Navigation Satellite System (GNSS) provides position, navigation, and timing (PNT) services that are irreplaceable to modern society [2]. Aviation [3], maritime [4, 5], automotive [6], agriculture [7], mobile phones [8] and infrastructure such as telecommunication, internet, commerce and banking industries [9], to name a few, all demand PNT services for proper operation. The widespread reliance on PNT services requires GNSS to provide continuous and reliable service across a wide range of applications. However, satellite faults inevitably occur for various reasons, such as clock anomalies [10, 11], accidental thruster firing [12], faulty ephemeris error upload [13], harsh space environment [14], and other unknown reasons. The Global Positioning System (GPS) is the pioneering GNSS service developed by the U.S., and has been successfully operated for more than 40 years. It is the most widely used and stable system on earth, but still has experienced several satellite clock anomalies/faults in recent years [11, 15].

Techniques such as Receiver Autonomous Integrity Monitoring (RAIM) [16, 17] and Advanced RAIM (ARAIM) [18] are developed to preserve the integrity of the user positioning results using redundant measurements to perform consistency checks that allow users to obtain probabilistic bounds on the position errors given the possibility of the satellite faults and priors. RAIM and ARAIM focus on acquiring integrity without the assistance from other systems. This is beneficial for users with safety-critical requirements that need to operate independently. However, the operation of

RAIM and ARAIM require sufficient measurements to be available and may demand high computational resources due to the complexity of RAIM/ARAIM algorithm.

Alternatively, Satellite Based Augmentation Systems (SBAS) provide integrity to users through an additional monitoring system. Such systems monitor satellite faults in real-time and generate the associated corrections/integrity message to broadcast to the users. Users acquire integrity on their positioning results by applying such additional information, and thus avoiding complicated receiver design and redundant satellite measurements. SBASs shift the responsibility of achieving integrity from the users to the service provider, with minimal computational load on the users, and can achieve better performance over covered regions than RAIM. Currently operating SBAS include systems such as the Wide Area Augmentation System (WAAS) operated in the U.S. [19], the European Geostationary Navigation Overlay Service (EGNOS) [20] in Europe, the Multi-functional Satellite Augmentation System (MSAS) in Japan [21], the GPS Aided GEO Augmented Navigation (GAGAN) in India [22], and the Korea Augmentation Satellite System (KASS) in Korea [23]. While these ground-based integrity monitor systems provide an integrity guarantee for the users following the aviation standard, they currently only support the GPS-L1 signal using snapshot solutions. Although these systems satisfy the global standard from the International Civil Aviation Organization (ICAO) [24] and RTCA DO-229, Minimum Operational Performance Standard (MOPS) [25], their coverage and performance are regionally designed to their own operational area.

This thesis focuses on the design and establishment of a global integrity monitor system. The system is developed to provide integrity assurance to users worldwide with GPS signals. Furthermore, the proposed integrity assurance framework targets Precise Point Positioning (PPP) services and protects users against arbitrary numbers and types of satellite clock and ephemeris faults.

First, this chapter provides the background knowledge to construct the global integrity monitor for precise positioning service.

1.1 Precise Point Positioning

PPP [26] is a positioning technique in GNSS that can achieve decimeter level accuracy for kinematic users and centimeter level for static users with only the code phase and carrier phase signals. The extremely accurate positioning result is attained using precise corrections that are externally provided in addition to the nominally broadcast information like ephemeris signals internally available. As of 2026, several PPP services are in operational/trial status. For the government-based services, Japan Aerospace Exploration Agency (JAXA) proposed the Multi-GNSS Advanced Demonstration tool for Orbit and Clock Analysis (MADOCA) [27] that aims to realize centimeter-level positioning accuracy worldwide. It supports GNSS signals from the U.S. global positioning system (GPS), the Russian global navigation satellite system (GLONASS), European Galileo, Chinese BeiDou, and Japanese Quasi-Zenith satellite system (QZSS). On April 1, 2024, the MADOCA-PPP service entered into operational status and the correction information is broadcast on the L6E signal via QZSS and is also available on JAXA's FTP server [ftp://mgmds01.tksc.jaxa.jp/]. Galileo declared initial service of its High Accuracy Service (HAS) [28] on January 24, 2023. HAS broadcasts precise correction information through Galileo E6B signals and a terrestrial network (Internet). The correction information from HAS supports both GPS and Galileo. Non-profit organizations like the International GNSS Service (IGS) [29] provide precise corrections through the internet [https://igs.org/products-access/]. The above mentioned services are all free-of-charge. On the commercial side, companies such as Trimble® CenterPoint RTX [30] and TerraStar by Hexagon [31] provide PPP services through the internet and their private satellite signals. Of all the correction services mentioned above, Trimble is the only service provider to claim to have integrity in the correction information they provide [32].

1.1.1 Measurement Model

As a standard approach, PPP utilizes the Ionospheric-Free combination from dual-frequency measurements to eliminate ∼99.9% of ionospheric effects. This choice is due to the fact that ionospheric effects are dispersive and thus GPS signals have different degrees of ionospheric effects given the carrier frequencies. The Iono-Free combinations of the pseudorange measurements are as follow:

$$\rho_{\mathrm{IF}} = \frac{{f_1}^2 \cdot \rho_{f_1} - {f_2}^2 \cdot \rho_{f_2}}{{f_1}^2 + {f_2}^2} \tag{1.1}$$
$$\phi_{\mathrm{IF}} = \frac{{f_1}^2 \cdot \phi_{f_1} - {f_2}^2 \cdot \phi_{f_2}}{{f_1}^2 + {f_2}^2}$$

where f_1 and f_2 are frequencies of both carrier waves, and ρ and ϕ are the code phase and carrier phase pseudorange measurements, respectively, in meters. ρ_{IF} and ϕ_{IF} are the combined code and carrier phase measurements that eliminate the first order ionospheric effect.

The Iono-Free measurement models for the pseudorange are defined as follow:

$$\rho_{\mathrm{IF}}^{(i)} = \| \overrightarrow{x}_{sat}^{(j)} - \overrightarrow{x}_{rx} \| + \left(b - b_s^{(i)}\right) + m^{(i)} \nabla \hat{T} + R_m + \epsilon_\rho^{(i)} \tag{1.2}$$
$$\phi_{\mathrm{IF}}^{(i)} = \| \overrightarrow{x}_{sat}^{(j)} - \overrightarrow{x}_{rx} \| + \left(b - b_s^{(i)}\right) + m^{(i)} \nabla \hat{T} + A^{(i)} + R_m + \epsilon_\phi^{(i)} \tag{1.3}$$

where $\rho_{\mathrm{IF}}^{(i)}$ is the Iono-Free code phase measurement, i represents ith measurement, $\overrightarrow{x}_{sat}^{(j)}$ is the satellite position from ephemeris data, j represents jth satellite, $\overrightarrow{x}_{rx}$ is the receiver position, b is the receiver clock bias, $b_{sat}^{(i)}$ is the satellite clock bias, $m^{(i)}$ is the mapping function of tropospheric delay error, $\nabla \hat{T}$ is the tropospheric wet delay, and ϵ_ρ and ϵ_ϕ are the un-modeled effects for code and carrier, respectively. $\phi_{\mathrm{IF}}^{(i)}$ is the carrier phase ionospheric-free measurement. $A^{(i)}$ is the floating ambiguity. R_m consists of all the modeled effects for both code and carrier phase measurements, which include satellite antenna phase center offset, relativistic effects, solid earth tide modeling, tropospheric delay, phase wind up for carrier phase measurements, and differential code bias. Notice the ionospheric-free measurements are computed from dual-frequency measurements, and no single frequency measurements are used in the

estimation, so no ionospheric model is required, and all terms are converted to meters. Multipath errors are accounted for by the un-modeled effects, $\epsilon_{\rho}^{(i)}$ and $\epsilon_{\phi}^{(i)}$.

1.1.2 Observation Modeling

The accuracy of the positioning results directly relies on how accurately the pseudorange errors are modeled, as well as the prior uncertainties of each of the error terms used during the operation. The priors of the errors are the most important part of the error modeling. They represent the uncertainties of the error models applied to the measurements and are the source of confidence for the estimation results. If the priors do not properly represent the actual error distribution, the estimation results cannot be trusted, and therefore integrity cannot be assured.

For estimators such as the Extended Kalman Filter (EKF), the estimation minimizes the variance of errors. The applied priors are such that the actual distributions are bounded by assumed Gaussian distributions. Essentially, the covariance of the error models must bound the actual distribution of the errors. In terms of the EKF, this is shown as determining sufficient initial noise, process noise, and measurement noise during the process of the estimation.

The rest of the section introduces the error modeling and priors on each of the terms in Eq. 1.2 and Eq. 1.3 that are used in the estimation results shown throughout this dissertation.

Satellite Clock Bias and Ephemeris

In the standard positioning process, satellite clock bias, $b_s^{(i)}$, and ephemeris, $\overrightarrow{x}_s^{(j)}$, are assumed to be known to the user.

The satellite clock bias and ephemeris can be obtained in the navigation message along with the broadcast GPS signals through a series of equation conversions given the limitation of bit rate from GPS signals. This is the standard procedure to perform trilateration with peudorange and is well documented in official Interface Specifications, IS-GPS-200N [33]. The broadcast satellite clock bias and ephemeris can achieve 3∼5 meters of position accuracy in users' snapshot solutions.

Alternatively, the PPP service provides more precise satellite clock bias and ephemeris estimates to the users through secondary communications such as internet, telecommunication, or dedicated satellite signals. This implies higher bandwidth and therefore enables the direct transmission of bias and ephemeris values over a period of time. For instance, the IGS ultra-rapid product provides precise orbits and satellite clock biases 4 times a day (every 6 hours), and each product contains 48 hours of data with 15 minute intervals, where the first 24 hours are past observation and the last 24 hours are prediction. The product is in .SP3 format [34]. PPP users simply perform numerical interpolation to obtain precise satellite clock bias and ephemeris for the pseudorange model.

For the satellite clock bias product, a simple linear interpolation is sufficient for centimeter accuracy. As for the ephemeris product, the 15-minute window requires a more sophisticated interpolation method. This dissertation applies an 11th degree Lagrange interpolating polynomial, with 6 data points earlier and another 6 data points later than the desired time. A typical PPP user can achieve centimeter position accuracy in nominal conditions. The residual clock and ephemeris errors are accounted for in the un-modelled effects in $\epsilon_\rho^{(i)}$ and $\epsilon_\phi^{(i)}$.

Receiver Clock Bias

The receiver clock bias, b, is part of the user states to be estimated from the pseudorange measurements. It represents the inaccuracy of the local clock used by the receiver and acts as a common-mode bias that affects all satellite LOS measurements equally. Since the quality of the local clock of the receivers cannot be guaranteed, the receiver clock bias is treated as an ambiguous floating value with infinite uncertainty and can only be determined from the measurements at each time step. That is, the bias may take on any value at any time in theory.

In practice, the propagation of receiver clock bias is often modeled as an affine function because modern GPS receivers continuously monitor and maintain their local clocks. However, the process noise is often set to a very large value to account for the

uncertain nature of the local clock.

$$b_{t+1} = b_t + \nabla b_t \cdot \Delta t \tag{1.4}$$

where ∇b_t is the clock drift, i.e., first order time derivative of the clock bias b_t at time step t, and Δt is the time interval.

Although the clock behavior is heavily dependent on how the clock signal is sourced, the type of receiver make, version of firmware, etc., there are two common tracking patterns in IGS stations.

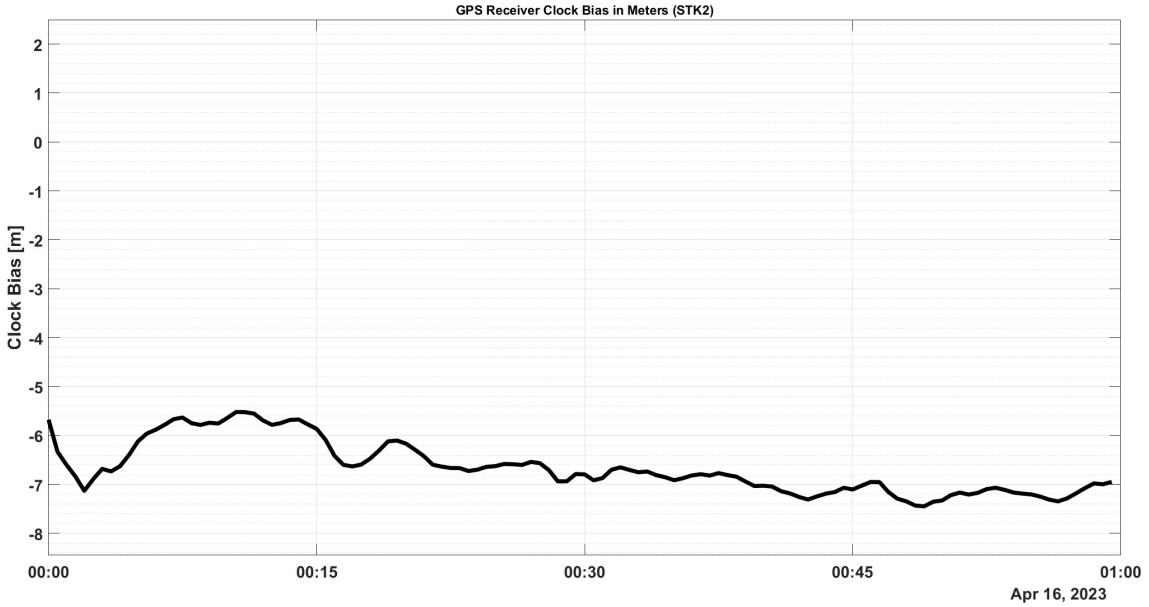


Figure 1.1: Receiver clock bias (in meters) of STK2 station from IGS network.

The first pattern of the receiver clock bias is shown in Fig. 1.1. It is the estimate of the receiver clock bias from STK2, an IGS station at Shintotsukawa, Japan. The estimation period is 1 hour, 00:00:00∼01:00:00 on April 16th, 2023. The receiver clock bias is well tracked at $-6 \sim -7$ m in range domain. The clock drift is optional for this receiver due to the relatively mild variations in the clock bias.

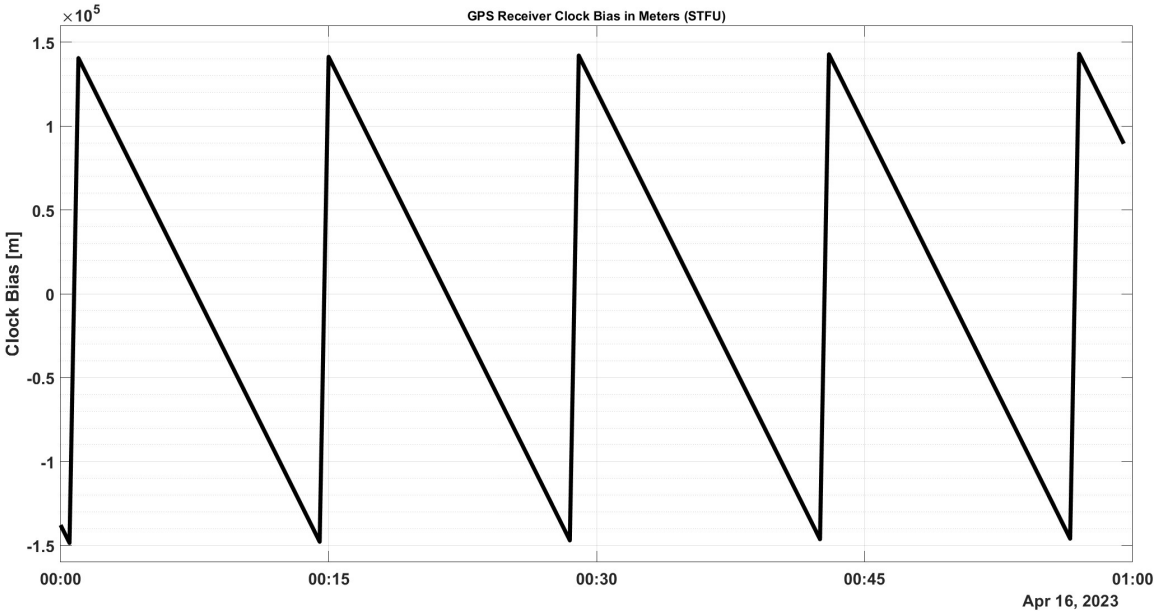


Figure 1.2: Receiver clock bias (in meters) of STFU station from IGS network.

Fig. 1.2 shows the second pattern of receiver clock in IGS stations. This is the STFU station located at Stanford University, USA. The estimation period is the same as STK2, 00:00:00∼01:00:00 on April 16th, 2023. It can be seen that the clock bias of this receiver is consistently drifting. Eq. 1.4 is well suited for this type of clock behavior. Additionally, a typical behavior of such receiver is that the bias is not unbounded. It is limited to be within a range of 0.5 milliseconds, which corresponds to around 15 km of difference in range domain. This zigzag behavior may cause numerical inaccuracy in the floating-point computer when the clock bias suddenly jumps from $\sim +150$ to ~ -150 km, so a logical adjustment may be applied to the deterministic propagation model of the clock bias to avoid numerical issue.

$$b_{t+1} = \begin{cases} b_t + \nabla b_t \cdot \Delta t, & \text{if } \|b_t + \nabla b_t \cdot \Delta t\| < c \cdot 10^{-3}. \\ b_t + \nabla b_t \cdot \Delta t + SGN(\nabla b_t) \cdot c \cdot 10^{-3}, & \text{otherwise.} \end{cases} \tag{1.5}$$

where c is the speed of light and SNG is the sign function. Eq. 1.5 essentially forces the clock bias to be limited within 1 millisecond during the propagation to allow continuous estimates.

Tropospheric Effect

The troposphere is the lowest layer of atmosphere above the earth. It is the region where all weather events occur, such as rain and hurricanes. One important aspect of the troposphere is that the medium is non-dispersive, meaning the induced time delays of radio waves (i.e., the GPS signal) when passing through the troposphere are equally affected across the frequency bandwidth. Therefore, utilizing dual frequency signals cannot mitigate the delay in troposphere. Mathematical models are established that allow one to estimate and reduce the effect of tropospheric delay on GPS signals.

The empirical model adopted in this dissertation is the UNB3 model [35], where the hydrostatic delay, which is induced by the tropospheric components (78% N_2, 21% O_2...), and the wet delay, which is induced by water vapor. The resulting total delay in the zenith direction is projected through a simple thin-shell model to account for the additional slant delay.

$$T_0 = m \cdot \text{UNB3}(doy, height, latitude) \tag{1.6}$$

where T_0 is the total slant delay introduced when the signal traveling through the troposphere, and m is the mapping function. The UNB3 model is a function of day of year (doy), height above sea level and latitude. UNB3 has adopted the mapping of Niell [36] as the mapping function.

The tropospheric correction is part of the modeled effect:

$$R_m = T_0 + \text{other modeled effects specified in Eqs. 1.2 and 1.3} \tag{1.7}$$

In addition to the deterministic correction from empirical models such as UNB3, the measurement model includes an additional stochastic state representing residual wet tropospheric delay due to local weather variations, i.e., $m^{(i)}\nabla\hat{T}$ in Eq. 1.2 and Eq. 1.3.

The deterministic propagation model is employed:

$$\nabla\hat{T}_{t+1} = \nabla\hat{T}_t \tag{1.8}$$

Ionospheric Effect

The charged particles in the upper part of the atmosphere, i.e., ionosphere, interact with the GNSS signals and introduces delay to the code phase, and advances to the carrier phase measurements. Since the ionosphere is a dispersive medium, the ionospheric effects are dependent on the carrier frequencies, and can be almost entirely eliminated with the Iono-Free combination (Eq. 1.1). This is also the pseudorange measurement used in this dissertation.

The higher order ionospheric effects are shown to only affect less than 0.2% of the effects [37]. When the total ionospheric effects are very large (> 100 m), the higher order terms could potentially accounts for tens of centimeters of range errors in low latitude and elevations [38]. Nevertheless, more than 99.8% of the ionospheric effects are removed, and the additional ionospheric effects are accounted for within the measurement noise model and are not modeled explicitly.

Relativistic Effect

Due to the relative velocity of the onboard satellite clock, variations in the gravitational field as the satellite traverses its orbit, and the change in reference frame as signals propagate to users on Earth, relativistic effects must be accounted for. Otherwise, the resulting clock bias error could accumulate to several meters per day [39]. There are four major relativistic effects. The coordinate frame is assumed to be Earth Center Earth Fixed (ECEF).

The first correction is derived based on a Keplerian satellite orbit. The fixed orbit allows the clock rate difference on the satellite and on the ground to be approximated as a constant offset. This effect is accounted for before launching the satellite by adjusting the satellite clock to be slower. For instance, for GPS L1 carrier frequency of 10.23 MHz, the onboard satellite clock frequency is adjusted lower:

$$\left(1 - 4.4647 \times 10^{-10}\right) \times 10.23 \text{ MHz} = 10.22999999543 \text{ MHz} \tag{1.9}$$

This is the most prominent relativistic effect and since it is already adjusted in the hardware so there is no need to correct this term in measurements.

The second relativistic effect is the additional correction introduced in the non-circular orbit and is described by the eccentricity of the Keplerian orbit.

$$\Delta t_e = +\frac{2 \cdot \overrightarrow{x}_s \cdot \overrightarrow{v}_s}{c^2}[\text{sec}] \tag{1.10}$$

where Δt_e is the resulting correction in seconds, $\overrightarrow{x}_s$ is the satellite position, $\overrightarrow{v}_s$ is the satellite velocity, and c is the speed-of-light.

The third relativistic effect is due to the additional path the signal traverse when propagating to a rotating reference frame on Earth. This is also called the Sagnac effect.

$$\Delta t_{\text{sag}} = +\frac{\omega_E \cdot (x_s \cdot y_{rx} - y_s \cdot x_{rx})}{c^2}[\text{sec}] \tag{1.11}$$

where $\omega_E = 7.2921151467 \times 10^{-5}\ [rad \cdot s^{-1}]$ is the Earth rotational rate, x_s and y_s are the satellite position in x and y directions, i.e., the first two entries of $\overrightarrow{x}_s = \begin{bmatrix} x_s & y_s & z_s \end{bmatrix}^T$, and similarly x_{rx} and y_{rx} are the receiver position in x and y directions.

The fourth relativistic effect is the Shapiro delay and is introduced when the signals traverse through the Earth gravitational field from the space to the ground.

$$\Delta t_{\text{sh}} = +\frac{\Phi_0}{c^2} \cdot \frac{l}{c} + \frac{2 \cdot \text{GM}_E}{c^3} \ln\left(\frac{\|\overrightarrow{x}_s\| + \|\overrightarrow{x}_{rx}\| + l}{\|\overrightarrow{x}_s\| + \|\overrightarrow{x}_{rx}\| - l}\right)[\text{sec}] \tag{1.12}$$
$$l = \|\overrightarrow{x}_s - \overrightarrow{x}_{rx}\|$$

where $\frac{\Phi_0}{c^2} = -6.96927 \times 10^{-10}$ is the Earth gravitational potential evaluated at the equator, and $\text{GM}_E = 3.986004415 \times 10^{14}$ is the Earth gravitational constant.

For GPS signals, the resulting relativistic corrections for PPP users with centimeter accuracy is:

$$R_m = c \cdot (\Delta t_e + \Delta t_{\text{sag}} + \Delta t_{\text{sh}}) + \text{other modeled effects specified in Eqs. 1.2 and 1.3} \tag{1.13}$$

Multipath Effect

Multipath is the additional range errors introduced when the antenna picks up signals that have one or more reflections from the surrounding environment. The reflected signals contain extra delays that are not fit to the pseudorange model of satellites that are LOS to the receivers.

For an antenna in an open sky area, the ground is the primary reflector so the intensity of the multipath effect is directly related to the elevation angle. This thesis adopts a simple inverse sinusoidal function to model the degradation of pseudorange measurements due to low elevation angle.

$$\begin{aligned} R_{\rho_{\mathrm{IF}}^{(i)}} &= \sigma_{\rho_0}^2 \cdot \frac{1}{\sin\theta_i{}^2} \\ R_{\phi_{\mathrm{IF}}^{(i)}} &= \sigma_{\phi_0}^2 \cdot \frac{1}{\sin\theta_i{}^2} \end{aligned} \tag{1.14}$$

where σ_{ρ_0} and σ_{ϕ_0} are the zenith measurement noise for code and carrier phase, respectively, θ_i is the elevation angle of the ith satellite, and $R_{\rho_{\mathrm{IF}}^{(i)}}$ and $R_{\phi_{\mathrm{IF}}^{(i)}}$ are the resulting measurement variances to be used in the estimator. Eq. 1.14 is adapted from [40], where the stochastic model accounts for the slant depth of the signals from satellites with low elevation angle.

For antennas operating in complex environments such as urban or mountainous areas, signals from satellites at high elevation angles may still be reflected by surrounding buildings or terrain, resulting in severe multipath effects that cannot be adequately characterized by the elevation angle alone. A simple indicator for reflectivity is the Carrier-to-Noise Density Ratio ($\mathrm{C/N_0}$) since the signal power is significantly weakened after a few reflections. Additionally, $\mathrm{C/N_0}$ can also be decreased for other reasons such as obstruction, foliage, scintillation, equipment failure, and etc. A simple weighting adapted from [41] is adopted:

$$\begin{aligned} \tilde{R}_{\rho_{\mathrm{IF}}^{(i)}} &= R_{\rho_{\mathrm{IF}}^{(i)}} \cdot 10^{-0.1\cdot\left(C/N_{0_i} - C/N_{0_{\max}}\right)} \\ \tilde{R}_{\phi_{\mathrm{IF}}^{(i)}} &= R_{\phi_{\mathrm{IF}}^{(i)}} \cdot 10^{-0.1\cdot\left(C/N_{0_i} - C/N_{0_{\max}}\right)} \end{aligned} \tag{1.15}$$

where C/N_{0_i} is the Carrier-to-Noise Density Ratio in [db-HZ] and $C/N_{0_{\max}}$ is a constant maximum noise ratio to be referenced. This is essentially penalizing the measurements with low power level.

Eq. 1.14 is applied to all the measurements in the example data evaluated in this thesis. Since IGS antennas are generally placed in areas with clear open sky views, the sinusoidal weighting function is sufficient. However, some WAAS receivers are placed in complex environments surrounded by different equipment, or near buildings, so measurements from WAAS receivers are further deweighted by Eq. 1.15 to account for severe multipath effect.

Earth Solid Tide Adjustment

The solid tide adjustment originates from the periodic pull of gravitational effects from large celestial bodies, i.e., the moon and the sun, that induce crust deformation on Earth causing centimeter level variations on static user position. This is on the order of PPP accuracy and must be corrected. These variations are modeled as a function of moon and sun positions, time of a year, and user position [42].

Carrier Phase Wind-up

Effects are unique to carrier phase measurements. These are due to the continuous attitude adjustments required on the satellites to point the transmitter antenna toward the center of the Earth affecting the right-hand circularly polarized carrier. The polarized electromagnetic wave would be offset by 1 wavelength if the satellite rotated 360° relative to a fixed user. This effect and correction model are well documented in [43].

Satellite Antenna Phase Center Offset and Variation

Antenna Phase Center Offset (PCO) arises because the satellite positions provided from the precise ephemeris product in IGS at the center of the mass, rather than the antenna position, so there are 1$\sim$3 meters of physical translational offset for the precise ephemeris products. This is a fixed translation in the satellite reference

frame, and must be applied since it is significant for PPP with centimeter accuracy requirement. PCO varies depending on the type of satellite and can be accessed from IGS as an ANTEX (`.atx`) file.

Antenna Phase Center Variation (PCV) refers to the fact that the actual transmitted point of the electromagnetic signal is not at the physical location of the antenna. The transmitting point is several millimeters above the antenna. PCV affects the carrier phase and is also calibrated and characterized in the ANTEX (`.atx`) file.

Receiver Antenna Phase Center Offset and Variation

Antenna PCO also exists for receivers since the received location of the signals is at the phase center. It requires additional translation after obtaining the position estimates to a physical location. The reference point is typically the Antenna Reference Point (APR) defined by the receiver manufacture, or is up to the user.

PCV also affects the carrier phase measurements on the receiver side. Depending on the type of receiver antenna, elevation and azimuth angle, the variations can be up to several centimeters. The PCVs of common survey-grade antennas are also recorded in ANTEX (`.atx`) files.

Code PCV can can reach up to 20∼30 cm for high end antennas [44]. Although PCV in code phase exists, it is typically not corrected due to the noise floor level of code phase measurements being several meters.

Ambiguity

The carrier phase is an additional pseudorange measurement along with the code phase. It is the number of wavelengths of the carrier signals counted by the receiver. While being extremely accurate, the carrier phase is ambiguous at an integer number of wavelengths, i.e., ambiguity $A^{(i)}$, and must be estimated along with code phase measurements.

Since ambiguity is a fixed number, there is no propagation model for this state:

$$A_{t+1}^{(i)} = A_t^{(i)} \tag{1.16}$$

Ambiguity may randomly jump to another fixed value due to loss of lock on the signals at the receiver. This event is called cycle slip and must be detected. The whole estimation of ambiguities must be reset to prevent corrupting the estimation result, since ambiguity states after convergence typically have relatively small variance.

Cycle Slip Detection

The detection of cycle slip relies on detecting discontinuities in the carrier phase measurements. The Melbourne-Wübbena (MW) combination applies [45, 46] to dual-frequency code and carrier phase. Code phases from both frequencies are combined via Narrow-Lane combination, and carrier phases are combined via Wide-Lane combination.

$$\rho_{NL} = \frac{f_1 \cdot \rho_{f_1} + f_2 \cdot \rho_{f_2}}{f_1 + f_2} \tag{1.17}$$

$$\phi_{WL} = \frac{f_1 \cdot \phi_{f_1} - f_2 \cdot \phi_{f_2}}{f_1 - f_2} \tag{1.18}$$

The Narrow-Lane combination on the code phase reduces the effective wavelength and measurement noise. The Wide-Lane combination creates longer wavelength and thus introduces more noise to the combined measurements, but since it is carrier phase, the noise level remains small. This combination preserves the same geometric range as the Narrow-Lane combination, and allows first-order ionospheric effects to cancel out when differencing these two combinations.

$$B_{MW} = \rho_{NL} - \phi_{WL} = A_{MW} + b_{MW} \tag{1.19}$$

where B_{MW} is the MW combination, A_{MW} is the resulting combined ambiguities, and b_{MW} is the bias from the remaining terms. The MW combination is geometry-free and ionosphere-free, making it a useful metric for cycle-slip detection.

A typical algorithm for using MW combination to detect cycle slip is tracking a certain amount of B_{MW} from the past measurements, and if the latest B_{MW} is an outlier compared to the B_{MW} from the past, a cycle slip event is declared and the corresponding ambiguity state is reset.

Another cycle slip detection method is Phase Ionospheric Residual (PIR) [47]. This method only uses the carrier phase to detect cycle slip. The combination is simply the difference of the carrier phase from different frequencies.

$$B_{PIR} = \phi_{f_1} - \phi_{f_2} = -I_{f_1} + I_{f2} + A_{PIR} + b_{PIR} \tag{1.20}$$

where B_{PIR} is the PIR combination, ϕ_{f_1} and ϕ_{f_2} are the ionospheric effects, respectively, A_{PIR} is the combined ambiguity, and b_{PIR} is the remaining bias. PIR combination removes the geometric range but leaves the ionospheric effect. Therefore, the detection of cycle slip events may be subject to ionospheric activity. A strong ionospheric storm that changes instantly between epochs may result in a false alert on cycle slip detection. Nevertheless, PIR combination works well in detecting additional cycle slip combinations that might be missed from other methods.

The new ambiguity values after cycle slip events are random and cannot be predicted. Thus, it is possible when the new cycle slip numbers are similar to the ratio of the measurement combinations, that cycle slip events go undetected.

For instance, the ambiguity from MW combination is constructed as:

$$A_{MW} = \frac{f_1 \cdot A_{f_1} - f_2 \cdot A_{f_2}}{f_1 - f_2} \tag{1.21}$$

For GPS L1 and L2 signals, the wavelengths are 0.1905 m and 0.2445 m respectively, the corresponding ambiguities are:

$$A_{MW} = \frac{f_1 \cdot (0.1903) \cdot Z_{f_1} - f_2 \cdot (0.2442) \cdot Z_{f_2}}{f_1 - f_2} \tag{1.22}$$

$$\approx (0.8619) \cdot Z_{f_1} - (0.8619) \cdot Z_{f_2} \tag{1.23}$$

which means if the new cycle number after cycle slip has a ratio close to 1 : 1, for

instance, if both f_1 and f_2 have 3 additional cycles, the MW combination cannot detect such a cycle slip since A_{MW} does not change. In a more extreme example, if both f_1 and f_2 have 100 additional cycles after a cycle slip event, which corresponds to 19.03 m and 24.42 m of differences in pseudorange, the MW still cannot detect the cycle slip.

Similarly, the undetectable ratio for the PIR combination is exactly 0.2442 : 0.1903, that is, approximately 4 : 3.

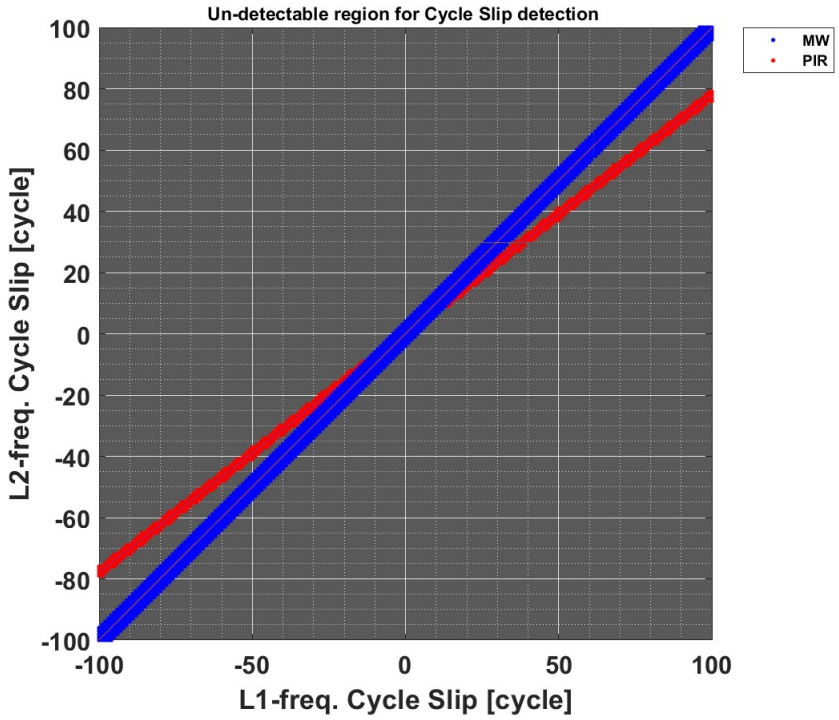


Figure 1.3: Region where MW and PIR cycle slip detectors cannot detect.

Fig. 1.3 illustrates the undetectable regions for both MW and PIR combinations for GPS L1 and L2 signals. If the new cycle slips offset from L1 and L2 signals reside in the blue region, the MW combination will fail to detect the cycle slip event, and similarly for the PIR combination. The MW combination is usually more noisy, thereby using larger margins than the PIR combination to declare cycle slip events. This is the reason for blue region to be thicker than the red region.

The importance of using at least two different cycle slip detection methods can

be observed from the above figure. If only one type of cycle slip detector is adopted, it could potentially miss large cycle slip events and result in inconsistent estimation results. Using two cycle slip detectors with different undetectable regions prevents large cycle slips from going undetected.

This thesis adopts both the MW and PIR cycle slip detectors in the estimation process.

1.1.3 Estimator

The Extended Kalman Filter (EKF) is the de facto standard estimator for PPP. The EKF assumes linearity of the system around the linearized point with the a priori being normally distributed. Therefore, for the estimation results to represent the actual state, the error models are required to bound the error distribution. That is, under the Gaussian assumption, the initial noise, process noise, and measurement noise must bound the actual distribution of errors.

Standard EKF includes Time-Update and Measurement-Update.

Time-Update:

$$\overrightarrow{\mu}_{t|t-1} = A_{t-1}\overrightarrow{\mu}_{t-1|t-1} + B_{t-1}\overrightarrow{u}_{t-1} \tag{1.24}$$

$$\Sigma_{t|t-1} = A_{t-1}\Sigma_{t-1|t-1}A_{t-1}^T + Q_{t-1} \tag{1.25}$$

Measurement-Update:

$$\overrightarrow{\mu}_{t|t} = \overrightarrow{\mu}_{t|t-1} + K_t\left(\overrightarrow{y}_t - C_t\overrightarrow{\mu}_{t|t-1}\right) \tag{1.26}$$

$$\Sigma_{t|t} = \Sigma_{t|t-1} - K_tC_t\Sigma_{t|t-1} \tag{1.27}$$

$$K_t = \Sigma_{t|t-1}C_t^T\left(C_t\Sigma_{t|t-1}C_t^T + R_t\right)^{-1} \tag{1.28}$$

where $\overrightarrow{\mu}$ is the mean of the Gaussian estimate of the system states, and Σ is the covariance of the estimate. K_t is the Kalman Gain, which represents the mapping of the measuring and modeling of the system, Q is the covariance of the process noise, and R is the covariance of the measurement noise. The above updates the system estimates from one time step, $t-1$, to the next time step, t.

As a summary, the standard propagation models for the states of PPP users in this thesis are shown below.

$\overrightarrow{x}_{rx,t}$	Receiver Position	$\overrightarrow{x}_{rx,t}=\overrightarrow{x}_{rx,t-1}+\overrightarrow{v}_{rx,t-1}\cdot\Delta t$
$\overrightarrow{v}_{rx,t}$	Receiver Velocity	$\overrightarrow{v}_{rx,t}=\overrightarrow{v}_{rx,t-1}$
b_t	Receiver Clock Bias	$b_t = b_{t-1}+\nabla b_{t-1}\cdot\Delta t$
∇b_t	Receiver Clock Drift	$\nabla b_t = \nabla b_{t-1}$
$\nabla\hat{T}_t$	Tropospheric Wet Delay	$\nabla\hat{T}_t = \nabla\hat{T}_{t-1}$
A_t	Ambiguity	$A_t = A_{t-1}$

Table 1.1: Deterministic Propagation Models

Table 1.1 provides the deterministic propagation models for the range related states in Eq. 1.2 and Eq. 1.3. Note that Δt is the time interval between data epochs. These states are essentially stationary, and only the receiver clock bias is modeled with a first-order model. Therefore, most states do not change in the time update, and only the uncertainties are inflated from the process noise. Eq. 1.5 may be used as the propagation model for the receiver clock bias state if the receiver clock jumps with 1 ms.

The receiver position state, $\overrightarrow{x}_{rx}$, and the associated higher order time derivatives are dependent on the type of user. This thesis only used the most fundamental user kinematic models.

$\sigma_{\vec{x}_{rx,u,0}}$	Receiver Position in u direction	10 [1]	$[m]$
$\sigma_{\vec{v}_{rx,u,0}}$	Receiver Velocity in u direction	1[2]	$[m/s]$
σ_{b_0}	Receiver Clock Bias	1000	$[m]$
$\sigma_{\nabla b_0}$	Receiver Clock Drift	1000	$[m/s]$
$\sigma_{\nabla \hat{T}_0}$	Tropospheric Wet Delay	0.05	$[m]$
σ_{A_0}	Ambiguity	10	$[m]$

Table 1.2: Initial Noise

Table 1.2 shows the initial noise of the states and is referenced from [48]. The selected values for the receiver position and clock bias states are justified by the use of a snapshot solution for their initialization, which provides relatively accurate initial estimates of these states.

$\sigma_{\vec{x}_{rx,u,Q}}$	Receiver Position in u direction	$10 \cdot \sqrt{\Delta t}$ [3] [4]	$[m]$
$\sigma_{\vec{v}_{rx,u,Q}}$	Receiver Velocity in u direction	$1 \cdot \sqrt{\Delta t}$[5]	$[m/s]$
σ_{b_Q}	Receiver Clock Bias	$1000 \cdot \sqrt{\Delta t}$	$[m]$
$\sigma_{\nabla b_Q}$	Receiver Clock Drift	$10 \cdot \sqrt{\Delta t}$	$[m/s]$
$\sigma_{\nabla \hat{T}_Q}$	Tropospheric Wet Delay	$\frac{0.002}{60} \cdot \sqrt{\Delta t}$	$[m]$
σ_{A_Q}	Ambiguity	0	$[m]$

Table 1.3: Process Noise

[1] Position process noise is set to be zero for **Filter 3** (defined in Sec.1.5).
[2] Velocity process noise is set to be zero for **Filter 3** (defined in Sec.1.5).

Table 1.3 provides the process noise of the state, and is referenced from [48] with the value factored by the time interval. The PPP users are assumed to be kinematic. For static users, such as receivers operating as ground reference stations, the process noise associated with the position and velocity states can be set to 0, or alternatively, the position and velocity states can be removed from the state vector if the receiver position is a known parameter.

$\sigma_{\rho_{\text{IF}}}$	Zenith Iono-Free Code Phase	3	$[m]$
$\sigma_{\phi_{\text{IF}}}$	Zenith Iono-Free Carrier Phase	0.04	$[m]$

Table 1.4: Measurement Noise

Table 1.4 is the measurement noise of code and carrier phase in the zenith direction, and are referenced from [48]. The measurement noise is weighed using Eq. 1.14 based on the elevation angle to suppress multipath effect. If an individual receiver has more severe multipath effects for satellites at higher elevation angles, Eq. 1.15 may be used to further de-weight measurements that have low $C/N_{0_{\text{max}}}$.

No correlations are assumed in the priors, so the corresponding covariance matrices are the diagonal matrices. For example, one can construct the covariance matrix of

[3]Position initial noise is set to be zero for **Filter 3** (defined in Sec.1.5).

[4]The user filter (**Filter 1**, defined in Sec.1.5) initializes position state with snapshot solutions, that has tens of meters uncertainty. If the initial point is farther away, it may be require higher process noise and longer convergence time.

[5]Velocity initial noise is set to be zero for **Filter 3** (defined in Sec.1.5).

process noise as:

$$Q = \begin{bmatrix} \sigma^2_{\overrightarrow{x}_{rx,u,Q}} & 0 & 0 & 0 & 0 & 0 & 0 & \dots \\ 0 & \sigma^2_{\overrightarrow{x}_{rx,u,Q}} & 0 & 0 & 0 & 0 & 0 & \dots \\ 0 & 0 & \sigma^2_{\overrightarrow{x}_{rx,u,Q}} & 0 & 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & \sigma^2_{b_Q} & 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & 0 & \sigma^2_{\nabla b_Q} & 0 & 0 & \dots \\ 0 & 0 & 0 & 0 & 0 & \sigma^2_{\nabla \hat{T}_0} & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \end{bmatrix} \tag{1.29}$$

if velocity states are not considered.

1.1.4 PPP User

$$\overrightarrow{x} = \begin{bmatrix} x \\ y \\ z \\ v_x \\ v_y \\ v_z \\ b \\ \nabla b \\ \nabla \hat{T} \\ A^{(1)} \\ \vdots \\ A^{(k)} \end{bmatrix} \tag{1.30}$$

The propagation model and time-update for the state are straightforward for the PPP user; they are identical except for the receiver clock bias where the clock drift

takes place.

$$\overrightarrow{x}_{t+1} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 1 & \Delta t & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & 1 & \Delta t & 0 & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & \Delta t & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & \Delta t & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & m & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \cdot \overrightarrow{x}_t \tag{1.31}$$

Eq. 1.31 is the time-update of the PPP user state defined in Eq. 1.30. The process noise and measurement noise are constructed in a similar manner in the diagonal entries of the covariance matrices with the priors described in Tables 1.3 and 1.4.

The performance of this example PPP user with 30-second interval GPS L1-L2 measurements from the TWTF station on Jan.16th, 2026, Taoyuan, Taiwan is shown in Fig. 1.4.

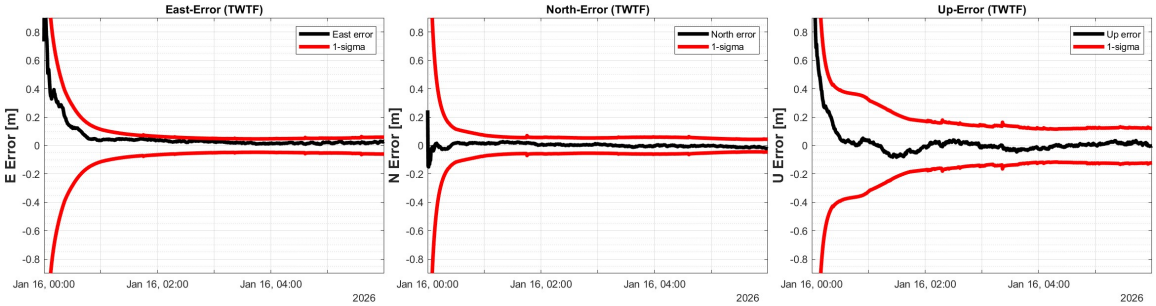


Figure 1.4: Position error estimates in East-North-Up direction.

The black lines in Fig. 1.4 are the position differences between the filter estimates and the reference solution from the Center for Orbit Determination in Europe (CODE) Solution Independent Exchange (SINEX) file. The red lines are the 1-sigma uncertainty estimates. Under nominal conditions, the PPP user filter can achieve centimeter accuracy as shown above.

1.2 Integrity Requirement

For users with safety-critical requirements, the integrity of the users' positioning results can be assured if the correction information broadcast by the correction service is constantly monitored and the users receive and apply the corresponding integrity information. Systems such as the Wide Area Augmentation System (WAAS) [19] have guaranteed meter-level integrity for the broadcast GPS corrections for snapshot solutions over North America for the last 20 years. WAAS utilizes the User Differential Range Error (UDRE) to provide the confidence bound of the satellite clock and ephemeris errors. This parameter is to ensure that broadcast GPS corrections are within the integrity and continuity risk requirements to protect the vertical guidance of aircraft during the landing phase. WAAS has demonstrated its success and trustworthiness as a standalone augmentation system for the integrity of carrier-smoothed code-based GPS users in the field of civil aviation [49].

The idea of monitoring the correction information in real-time on the server side has been explored for precise positioning services such as PPP. The quality of the correction information is shown to be assessed with test statistics from the phase residual [50] to generate health indicators for users to exclude faulty satellites. [51] adopts the isolation forest method to detect and exclude faulted satellites from quality indicators based on phase residuals from the monitoring receivers. [52] specifically applies carrier phase residuals from a network of receivers as test statistics to ensure the requirements of integrity risk and continuity risk are simultaneously satisfied when satellite faults are declared. These methods perform quality monitoring of the correction information in the sense that the phase residuals from the monitoring receivers indicate the anomaly of the satellites' clock and ephemeris, and the

consistency of the statistics validates the quality. Alternatively, [53] monitors the quality of the correction information via the estimates of the user range errors (UREs) at the location of the receivers. The UREs form the Quality Indicator to detect faulted satellites.

In most of these systems, the provision of integrity is generally based on by-products of the estimation filter (the residuals). This makes it difficult to evaluate the observability of potential corrections or satellite faults. This problem is exacerbated by the fact that PPP correction filters usually make strong assumptions about the behavior of the signals and the clocks.

The condition for satisfying the integrity requirement can be generalized as requiring the detection statistics to exhibit a failure rate bounded by the Probability of Integrity Risk (PIR).

The primary test statistic for GNSS is Solution Separation (SS). The test is defined as:

$$|\hat{x}_0 - \hat{x}_i| < T, \ \forall i \in \mathbb{I}_{\text{subset}} \tag{1.32}$$

where $\hat{x}_0$ is the baseline estimates, typically referred to as the All-In-View (AIV) set, $\hat{x}_i$ is the subset estimates that have their corresponding exclusion of fault modes, $\mathbb{I}_{\text{subset}}$ are the subset, and T is the threshold of detection.

The upper bound of the PIR for the SS test can be expressed as:

$$P\left(|\hat{x}_0 - x| > L, |\hat{x}_0 - \hat{x}_i| < T \,|\, \forall i \in \mathbb{I}_{\text{subset}}\right) \leq PIR \tag{1.33}$$

where x is the true state in the system that can be a scalar or vector depending on the system to be monitored. In ARAIM, x is typically defined as the user position, since this is the parameter of interest requiring integrity protection. L is the upper bound that indicates the level of protection that the detection can assure given the possibility of the fault modes. Eq. 1.33 simply states the probability that the estimate exceeds the upper bound, but the SS statistics fail to detect it, given all the threshold tests passed. This is the condition where the system is in danger. For example, for a GNSS users where the tests are in the position domain, this condition states that the

position estimates exceed the upper bound, but the SS tests fail to detect such events. This can endanger applications with safety-critical requirements, so Eq. 1.33 regulates that the probability of such instance must be bounded by a pre-defined probability, that ensures the estimates meet integrity requirements and can be trusted.

An upper bound, L, can be derived. In particular, it can be shown that $P\left(\|\hat{x}_i - x\| > L - T\right)$ is an upper bound of Eq. 1.33.

Let:

$$\begin{aligned} P\left(\|\hat{x}_i - x\| > L - T\right) &\geq P\left(\|\hat{x}_i - x\| > L - T, \|x_i - x_0\| < T\right) \\ &\geq P\left(\|\hat{x}_i - x\| > L - \|x_i - x_0\|, \|x_i - x_0\| < T\right) \\ &= P\left(\|\hat{x}_i - x\| + \|x_i - x_0\| > L, \|x_i - x_0\| < T\right) \end{aligned} \tag{1.34}$$

By triangle inequality:

$$P\left(\|\hat{x}_i - x\| + \|x_i - x_0\| > L, \|x_i - x_0\| < T\right) \geq P\left(\|\hat{x}_0 - x\| > L, \|x_i - x_0\| < T\right) \tag{1.35}$$

As a result:

$$P\left(\|\hat{x}_0 - x\| > L, \|x_i - x_0\| < T\right) \leq P\left(\|\hat{x}_i - x\| > L - T\right) \tag{1.36}$$

and let

$$P\left(\|\hat{x}_i - x\| > L - T\right) \leq PIR \tag{1.37}$$

Finding the upper bound of the approximation in Eq. 1.37 is easier than finding the one in Eq. 1.33.

By the law of total probability, and given that the probability of mis-detection is considered in the fault rate of the events, Eq. 1.37 is re-written to be conditioned on the monitored fault modes.

$$P\left(\|\hat{x}_i - x\| > L - T\right) = \sum_{i=0}^{N-1} P\left(\|\hat{x}_i - x\| > L - T \mid H_i\right) P\left(H_i\right)$$
$$= \sum_{i=0}^{N-1} 2 \cdot Q\left(\frac{L-T}{\sigma_i}\right) \cdot P\left(H_i\right) \tag{1.38}$$

where H_i is the occurrence of the monitored fault mode. Given the Gaussian assumptions on the estimates of URE.

Substitute Eq. 1.38 into Eq. 1.37

$$\sum_{i=0}^{N-1} 2 \cdot Q\left(\frac{L-T}{\sigma_i}\right) \cdot P\left(H_i\right) \leq PIR \tag{1.39}$$

Assume the receiver fault rate is constant, $P(H_i) = P_{fault}$:

$$\sum_{i=0}^{N-1} Q\left(\frac{L-T}{\sigma_i}\right) \leq \frac{PIR}{2 \cdot P_{fault}} \tag{1.40}$$

To simplify computation of the upper bound, PIR is equally allocated to each of the subsets.

$$Q\left(\frac{L-T}{\sigma_i}\right) \leq \frac{PIR}{2 \cdot N \cdot P_{fault}} \tag{1.41}$$

and the upper bound L is simply the maximum of each of the subsets:

$$L \leq \max_{\forall\, i \in \mathbb{I}} \left(T + Q^{-1}\left(\frac{PIR}{2 \cdot N_{\mathbb{I}} \cdot P_{fault}}\right) \cdot \sigma_i\right) \tag{1.42}$$

where Q is defined as the tail probability of the normal distribution and Q^{-1} is the inverse, $N_{\mathbb{I}}$ is the total number of subsets in $\mathbb{I}$, or can also be considered as the number of fault modes being monitored.

In Eq. 1.42, L is bounded by the estimates of the detection subset with their

maximal sigma. L provides the guarantee to the AIV solution with the consideration of the fault mode, and ensures the level of trust for the estimation result given the priors and fault modes that are monitored.

1.3 Continuity Requirement

The continuity requirement is a limit on the probability that the operation is interrupted. Of the contributors to this probability is the probability of alert, which many be due to nominal noise, a true fault, or an unsuccessful exclusion. From the view point of fault detection, the most secure way of detecting fault modes is to set $T = 0$. This would ensure that the PIR is bounded since all estimates would trigger the detection. This is not practical in actual operation where the majority of the time the system is in nominal operation. Also, if a system constantly alerts fault modes, the users simply cannot use the system due to the interruption. Thus, continuity is an operational aspect of consideration in addition to the integrity. The threshold is determined such that the disruption of the system is acceptable during the operation.

For SS statistics, the continuity must satisfy:

$$P\left(|\hat{x}_0 - \hat{x}_i| > T \mid \forall i \in \mathbb{I}_{\text{test}}\right) \leq P_{fa} \tag{1.43}$$

where $\hat{x}_0$ is the baseline estimate, i.e., the AIV estimate that is compared with the subset estimate, $\hat{x}_i$. $\mathbb{I}_{\text{test}}$ is the set of SS tests to be performed. P_{fa} is the probability of false alert. Eq. 1.43 states the condition when the SS test fails, i.e., interruption, and such probability should be bounded by the false alert rate, or can also be referred to as the continuity requirement.

Assume each threshold test is I.I.D. Gaussian:

$$P\left(|\hat{x}_0 - \hat{x}_i| > T \mid \forall i \in \mathbb{I}_{\text{test}}\right) = \sum_{i=1}^{N_{\text{test}}} P\left(|\hat{x}_0 - \hat{x}_i| > T\right) \tag{1.44}$$

$$= \sum_{i=1}^{N_{\text{test}}} 2 \cdot Q\left(\frac{T}{\sigma_{ss,i}}\right) \cdot \leq P_{fa} \tag{1.45}$$

where N_{test} is the total number of threshold tests within the given period, and $\sigma_{ss,i}$ is the sigma of the ith SS statistics.

Here, assume the continuity budget is equally distributed to each test:

$$Q\left(\frac{T}{\sigma_{ss,i}}\right) \leq \frac{P_{fa}}{2 \cdot N_{\text{test}}} \tag{1.46}$$

So the threshold, T, can be determined with normal quantile function, Q^{-1}:

$$T \geq Q^{-1}\left(\frac{P_{fa}}{2 \cdot N_{\text{test}}\cdot}\right) \cdot \sigma_{ss,i} \tag{1.47}$$

Thus, the minimal number of sigma for the detection threshold that satisfies the continuity requirement is:

$$K = Q^{-1}\left(\frac{P_{fa}}{2 \cdot N_{\text{test}}}\right) \tag{1.48}$$

and

$$T = K \cdot \sigma_{ss,i} \tag{1.49}$$

The absolute estimate of $\sigma_{ss,i}$ is difficult to compute in practice, thus, a conservative threshold is applied via the following relation:

$$\sigma_{ss,i} \leq \sigma_i \tag{1.50}$$

where $\sigma_{ss,i}$ is bounded by the sigma from the subset solutions. This relation is derived and the condition is specified in Sec.4.1.2 as well.

Therefore, the threshold for the SS tests can be determined as:

$$\hat{T} = K \cdot \sigma_i \tag{1.51}$$

where $\hat{T}$ is a conservative threshold for the subset i in actual comparison. It ensures that the probability of false alert is bounded, and thereby guarantees the continuous

operation of the system.

1.4 Contributions

This thesis makes several contributions to the provision of integrity for precise GNSS services and their users.

1.4.1 Integrity Monitor

Developed a high-integrity monitor for PPP corrections and users that provides an assured level of protection against satellite clock and ephemeris faults. Users can achieve 10∼20 cm of position accuracy, and ∼ 1 meter of error bound in the horizontal, and ∼2 meter in the vertical direction.

- Developed a fault-tolerant filter that can estimate all GNSS satellite clock and ephemeris errors with minimal prior assumptions on these satellite errors using measurements from multiple static receivers at known receiver locations.
- Achieved decimeter level accuracy in user range error estimates, and better than 2 meters assured error bounds in the range domain, enabling decimeter position accuracy for users.
- Attained decimeter level accuracy in the range estimates by identifying and calibrating the satellite differential code biases among each reference station.
- Developed a method to isolate and remove receiver failure from the estimation filter making the monitor robust against satellite, correction, and receiver faults.

1.4.2 Network Design

Developed an efficient simulation of network performance that can rapidly evaluate the lower error bounds for any receiver geometry. It further can be used in optimally designing, selecting, and placing reference stations when planning a monitor network.

It addresses questions about how many stations and where to place those stations to satisfy the performance requirements of a ground-based integrity monitor.

1.4.3 Tropospheric Error Analysis

Determined appropriate error distribution bounds for the international GPS aviation standards and identified a necessary update to the bounding equation to properly account for error correlation. The work is done by evaluating 10 years (2012∼2021) of global tropospheric delay data from over 400 GPS stations.

1.4.4 Publications

This dissertation is based on several works published in the publications listed below, of which I am a primary or secondary author.

Conference Papers

- Y.-F. Lai, J. Blanch, and T. Walter, "Troposphere delay model error analysis with application to vertical protection level calculation", in Proceedings of the 2023 International Technical Meeting of The Institute of Navigation, 2023, pp. 903–921.

- J. Blanch, T. Walter, and Y.-F. Lai, "Evaluation of tropospheric and mixed single-dual frequency error models in Advanced ARAIM", in Proceedings of the 2023 International Technical Meeting of The Institute of Navigation, 2023, pp. 892–902.

- Y.-F. Lai, J. Blanch, T. Walter, E. Kahr, E. Leahy, P. Silva, and C. Ellum, "Prototyping integrity monitors for PPP fault detections", in Proceedings of the 36th international technical meeting of the satellite division of the institute of navigation(ION GNSS+ 2023), 2023, pp. 2592–2605.

- J. Blanch, Y.-F. Lai, T. Walter, and P. Silva, "Models for the effect of integrity monitors on GNSS user error range distributions", in Proceedings of the 37th

International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS+ 2024), 2024, pp. 58–68.

- Y.-F. Lai, J. Blanch, and T. Walter, " State determination in ekf using singular value decomposition for PPP integrity monitor system", in Proceedings of he 37th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS+ 2024), 2024,pp. 1651–1663.

- Y.-F. Lai, J. Blanch, and T. Walter, "PPP range error integrity using a network of ground monitors", in Proceedings of the 2024 International Technical Meeting of The Institute of Navigation, 2024, pp. 211–225.

- R. Wang, Y.-F. Lai, J. Blanch, D. Akos, E. Phelts, and T. Walter, "Characterization of recent GPS satellite clock jumps", in Proceedings of the 2024 International Technical Meeting of The Institute of Navigation, 2024, pp. 226–240.

- Y.-F. Lai, J. Blanch, and T. Walter, "Ground-based integrity monitors for PPP correction services", in Proceedings of the ION 2024 pacific PNT meeting, 2024, pp. 514–526.

- J. Blanch, Y.-F. Lai, and T. Walter, "Integrity-based estimators for precise point positioning and other applications", in Proceedings of the ION 2024 Pacific PNT Meeting, 2024, pp. 477–490.

- Y.-F. Lai, J. Blanch, and T. Walter, " Protecting against receiver faults to a PPP ground-based integrity monitor system", in Proceedings of the 2025 International Technical Meeting of The Institute of Navigation, 2025, pp. 343–358.

- Y.-F. Lai, J. Blanch, and T. Walter, "User range error corrections with latency from PPP ground-based integrity monitor system", in Proceedings of the 38th International Technical Meeting of the Satellite Division of The Institute of Navigation(ION GNSS+ 2025), 2025, pp. 178–191.

- Y.-F. Lai, J. Blanch, and T. Walter, "Performance evaluation for ground-based integrity monitor network", in Proceedings of the ION 2026 pacific PNT meeting, 2026, pp. 272–285.

Journal Paper

- Y.-F. Lai, J. Blanch, and T. Walter, "Integrity Monitoring for Precise GNSS" NAVIGATION: Journal of the Institute of Navigation (Under Review)

1.5 Outline

There are several variants of Kalman filters used throughout this thesis, and they are listed below for clarity.

1. A user filter applies precise clocks and ephemerides and trusts the correction (standard PPP).
2. A user filter applies precise clocks and ephemerides, but does not trust the correction and performs self-monitoring.
3. A network filter uses groups of reference receivers, and does not trust the correction (self monitoring). It generates integrity information.
   (a) Apply precise clocks and ephemerides as reference.
   (b) Apply broadcast clocks and ephemerides as reference.
   (c) A receiver fault-tolerant implementation.
4. A user filter applies precise clocks and ephemerides in order to evaluate its integrity information from **Filter 3**.

Filter 1 is introduced in Sec.1.1.4, and is used in Figs. 1.4, 2.2 and 2.3, 3.4.
Filter 2 is introduced in Sec.2.1.1, and is used in Figs. 2.4, 2.5, 2.6, 2.7, and 2.8.
Filter 3.a is introduced in Sec.2.1.2, and is used in Figs. 2.10, 2.11, 2.12, 2.13, 2.14, 2.16, 2.18, and 4.2.

Filter 3.b is introduced in Sec.2.1.2, and is used in Figs. 2.15, 2.19, 2.20, 2.21, 3.1, 3.2, and 3.3.

Filter 3.c is introduced and used in Chap. 4.

Filter 4 is introduced in Chap. 3, and is used in Figs. 3.5, A.1, A.2, A.3, and A.4.

This thesis presents a ground based integrity monitor system design that can provide integrity to precise satellite navigation service, and the tropospheric error analysis that ensures the error model for the tropospheric model can be trusted. The thesis is organized as follows. Following the introduction in Chapter 1, Chapter 2 presents one of the most important tool in this thesis, the fault-tolerant filter, that enables the direct estimates of the satellite clock and ephemeris faults in real-time operation. In this thesis, the fault-tolerant simply refers to the robustness against the satellite faults, because they are being estimated. Chapter 3 verifies the usage of this fault-tolerant filter, and demonstrates the performance of PPP users with the integrity guarantee provided by the integrity monitor. Chapter 4 establishes the integrity of the monitor itself against potential receiver failure within the network. Chapter 5 defines and evaluates the performance of the entire network from the perspective of the users. Chapter 6 develops the methodologies to optimally design the geometric structures of the ground-based monitor network. In particular, the location and number of receivers is assessed. Chapter 7 analyzes the spatial correlations of the tropospheric delay error models, and provides an update to the equation for the SBAS vertical protection levels that bound the correlations. Chapter 8 concludes the thesis with the contributions made.

Chapter 2

Ground-Based Integrity Network

2.1 Fault-Tolerant Filter

The fault-tolerant filter is the most essential component in the operation of the integrity monitor system. It includes the satellite fault modes (i.e., satellite clock or ephemeris faults) as part of the estimation state so that the other filter states are unaffected by those fault modes being present in the measurements.

The filtering method is a standard EKF that requires measurement models and priors to correctly represent and bound the potential errors.

2.1.1 Single Receiver Self Monitoring

The simplest integrity monitor consists of only a single GNSS receiver, and the integrity is self-monitored. Given the geometry of the receiver with respect to the satellites that have LOS to it, the receiver can only monitor a combination of the clock and ephemeris errors.

To enable monitoring of a satellite orbit and clock fault, one additional state, a satellite clock error state $\zeta^{(k)}$, must be added to the standard PPP user filter. Although it is labeled a clock error state here, because only one receiver is included in this filter, the state can actually represent the combination of the projected orbit error together with the clock error

$$\overrightarrow{x}_s = \begin{bmatrix} x \\ y \\ z \\ v_x \\ v_y \\ v_z \\ b \\ \nabla b \\ \nabla \hat{T} \\ A^{(1)} \\ \vdots \\ A^{(k)} \end{bmatrix} \tag{1.30}$$

With the measurement models:

$$\rho_{\text{IF}}^{(k)} = \| \overrightarrow{x}_{sat}^{(k)} - \overrightarrow{x}_{pos} \| + \left(b - b_{sat}^{(k)} \right) + m^{(k)} \nabla \hat{T} + R_{\rho}^{(k)} + \zeta^{(k)} + \epsilon_{\rho}^{(k)} \tag{2.1}$$

$$\phi_{\text{IF}}^{(k)} = \| \overrightarrow{x}_{sat}^{(k)} - \overrightarrow{x}_{pos} \| + \left(b - b_{sat}^{(k)} \right) + m^{(k)} \nabla \hat{T} + A^{(k)} + R_{\phi}^{(k)} + \zeta^{(k)} + \epsilon_{\phi}^{(k)} \tag{2.2}$$

where $\zeta^{(k)}$ is the satellite clock and orbit error from the kth satellite.

$\zeta^{(k)}$ is unique in that there is no dynamic model so the estimates are reset to zero in every time-update, and the process noise is as large as the receiver clock bias. The non-dynamic and large process noise of $\zeta^{(k)}$ represent no prior being assumed in the modeling of the satellite clock error, meaning the filter is theoretically capable of dealing with any type and any magnitude of satellite clock errors.

$\zeta^{(k)}$ effectively reduces the available satellites. Given the linear measurement model of EKF:

$$\overrightarrow{y} = \mathbf{C} \cdot \overrightarrow{x} \tag{2.3}$$

which is equivalent to

$$\overrightarrow{y} = \begin{bmatrix} -\mathbf{1}^{(1)} & 0 & 0 & 0 & 1 & 0 & m^{(1)} & 1 & 0 & 0 & 0 & \dots \\ -\mathbf{1}^{(2)} & 0 & 0 & 0 & 1 & 0 & m^{(2)} & 0 & 0 & 0 & 0 & \dots \\ -\mathbf{1}^{(3)} & 0 & 0 & 0 & 1 & 0 & m^{(3)} & 0 & 0 & 0 & 0 & \dots \\ -\mathbf{1}^{(1)} & 0 & 0 & 0 & 1 & 0 & m^{(1)} & 1 & 1 & 0 & 0 & \dots \\ -\mathbf{1}^{(2)} & 0 & 0 & 0 & 1 & 0 & m^{(2)} & 0 & 0 & 1 & 0 & \dots \\ -\mathbf{1}^{(3)} & 0 & 0 & 0 & 1 & 0 & m^{(3)} & 0 & 0 & 0 & 1 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ z \\ v_x \\ v_y \\ v_z \\ b \\ \nabla b \\ \nabla \hat{T} \\ \zeta^{(s=1)} \\ A^{(1)} \\ \vdots \\ A^{(k)} \end{bmatrix} \tag{2.4}$$

where $\mathbf{1}$ is the LOS directional vector pointed from the estimated receiver to the corresponding satellites. These vectors vary from satellite to satellite but are unit length. $\mathbf{C}$ is the Jacobian matrix of measurement models Eq.(2.1) and Eq.(2.2). $\zeta^{(k)}$ is always presented on the measurements from PRN-1 which form a one-vector in $\mathbf{C}$ for measurements from PRN-1.

Since the error state is always set to zero in the time-update, the value of the error state at measurement update is always zero.

$$\overrightarrow{y} = \begin{bmatrix} -\mathbf{1}^{(1)} & 0 & 0 & 0 & 1 & 0 & m^{(1)} & 1 & 0 & 0 & 0 & \dots \\ -\mathbf{1}^{(2)} & 0 & 0 & 0 & 1 & 0 & m^{(2)} & 0 & 0 & 0 & 0 & \dots \\ -\mathbf{1}^{(3)} & 0 & 0 & 0 & 1 & 0 & m^{(3)} & 0 & 0 & 0 & 0 & \dots \\ -\mathbf{1}^{(1)} & 0 & 0 & 0 & 1 & 0 & m^{(1)} & 1 & 1 & 0 & 0 & \dots \\ -\mathbf{1}^{(2)} & 0 & 0 & 0 & 1 & 0 & m^{(2)} & 0 & 0 & 1 & 0 & \dots \\ -\mathbf{1}^{(3)} & 0 & 0 & 0 & 1 & 0 & m^{(3)} & 0 & 0 & 0 & 1 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ z \\ v_x \\ v_y \\ v_z \\ b \\ \nabla b \\ \nabla \hat{T} \\ \zeta^{(s=1)} \\ A^{(1)} \\ \vdots \\ A^{(k)} \end{bmatrix} \tag{2.5}$$

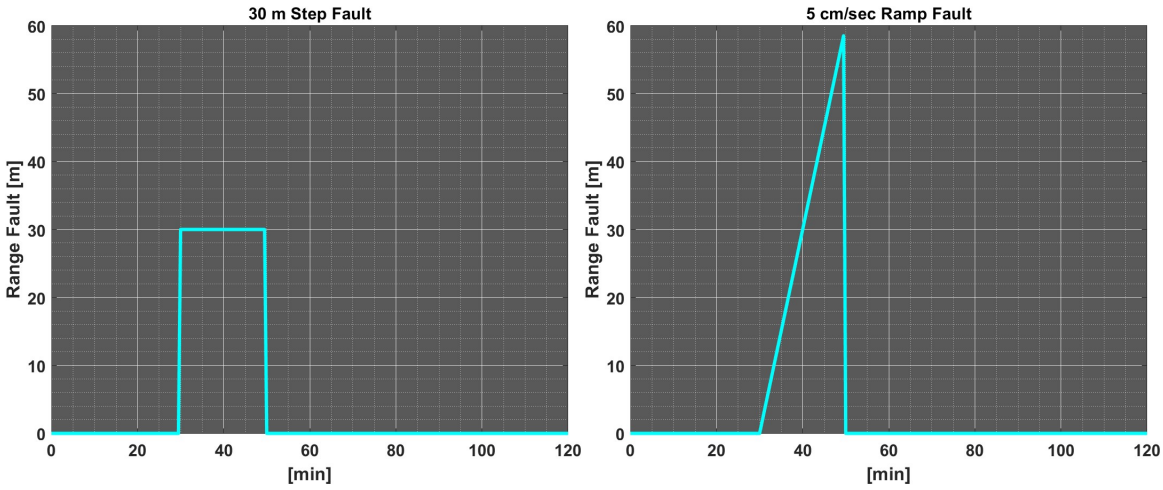
As a result, the column-wise manipulation of the error state makes no difference to the expected measurement, and the resulting measurement model becomes:

$$\overrightarrow{y} = \begin{bmatrix} \mathbf{0}^{(1)} & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & \dots \\ -\mathbf{1}^{(2)} & 0 & 0 & 0 & 1 & 0 & m^{(2)} & 0 & 0 & 0 & 0 & \dots \\ -\mathbf{1}^{(3)} & 0 & 0 & 0 & 1 & 0 & m^{(3)} & 0 & 0 & 0 & 0 & \dots \\ \mathbf{0}^{(1)} & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & \dots \\ -\mathbf{1}^{(2)} & 0 & 0 & 0 & 1 & 0 & m^{(2)} & 0 & 0 & 1 & 0 & \dots \\ -\mathbf{1}^{(3)} & 0 & 0 & 0 & 1 & 0 & m^{(3)} & 0 & 0 & 0 & 1 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ z \\ v_x \\ v_y \\ v_z \\ b \\ \nabla b \\ \nabla \hat{T} \\ \zeta^{(s=1)} = 0 \\ A^{(1)} \\ \vdots \\ A^{(k)} \end{bmatrix} \tag{2.6}$$

where the LOS, clock bias, and tropospheric wet delay of PRN-1 is zero. The above model shows that the addition of $\zeta^{(k)}$ to PRN-1 is equivalent to excluding PRN-1 from the estimation of the filter non-error states: position, clock bias, and tropospheric wet delay. In other words, the estimation of $\zeta^{(k)}$ is based on other satellites that are considered to be functioning normally while the satellites with clock faults are not contributing to the non-error states. As a result, the more satellite clock faults to be monitored from the same receiver, the fewer measurements are available to the non-error states.

Artificial Satellite Clock Fault

In the following examples, two types of artificial faults, step fault and ramp fault, are injected into the IGS observation data to simulate the influence of satellite faults on a PPP user.



(a) 30 meters Step Fault (b) 5 cm/sec Ramp Fault

Figure 2.1: Fault injection types.

The artificial step and ramp faults into be injected to the pseudorange measurements are shown in Fig. 2.1a and Fig. 2.1b. The faults begin at 30 min and end at 50 min. The step fault has a magnitude of 30 m and the ramp fault has a fault rate of 5 cm/sec.

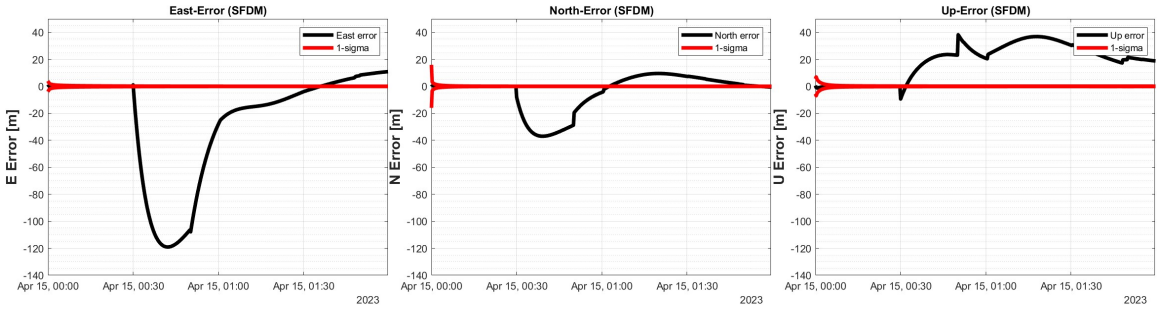


Figure 2.2: Position error in East-North-Up direction with 30 m step fault.

Fig. 2.2 shows the position error of the SFDM station, located near Los Angeles, on April 16th, 2023 when the step fault from Fig. 2.1a is injected into PRN-2. The SFDM station operates as a kinematic PPP user. In other words, for the SFDM station, there is a 30 m step satellite clock error that occurs on PRN-2 from 30 to 50 minutes during the position estimation. It can be seen that the 30 m satellite clock fault causes a significant impact in the position domain, leading to as much as 130 meters of error in East, 30 meters in North and almost 50 meters in the up directions. On the other hand, the uncertainties of the position errors from the estimator remain unchanged, since they are only dependent on the geometric matrix. The actual errors in the range domain have a minimal affect to the estimation of the uncertainties of the position errors. Therefore, although the satellite fault causes hundreds of meters of positioning errors, the user would still be unaware of the errors if they only relied on the estimator itself, even if the measurement residuals are likely to be large. In addition, the impact of fault does not disappear after the end of fault. The temporal correlation of the EKF estimator results in continuing effects of the satellite fault once it enters the system. Depending on the priors of each state, it takes more time than the presence of the fault to recover. In this case, restarting the filter is a better choice.

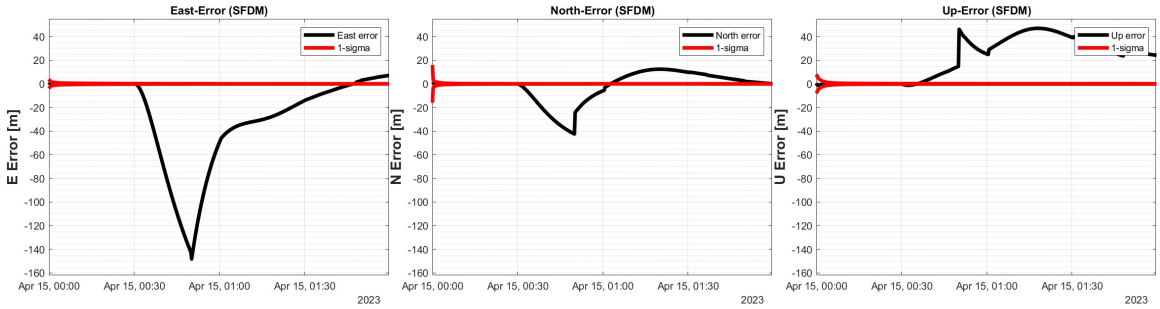


Figure 2.3: Position error in East-North-Up direction with 5 cm/sec ramp fault.

Fig. 2.3 demonstrates the position error of the SFDM station during the same period as shown in Fig. 2.2. The ramp fault causes a different evolution of position errors, with changes of errors slightly lagging behind those caused by the step faults. However, the user still experiences tens and hundreds of meters of errors as the ramp fault builds up.

Figs. 2.2 and 2.3 show the weakness of a standalone PPP user without an integrity guarantee. In the case of satellite fault, the positioning errors would also be far greater than the magnitude of the fault in the range domain, depending on the geometry and a posteriori of the states related to the faulted satellite.

By adding $\zeta^{(k)}$ as an extra error state to the measurement models of PRN-2, such a satellite clock fault can be self-monitored therefore protecting the filter itself.

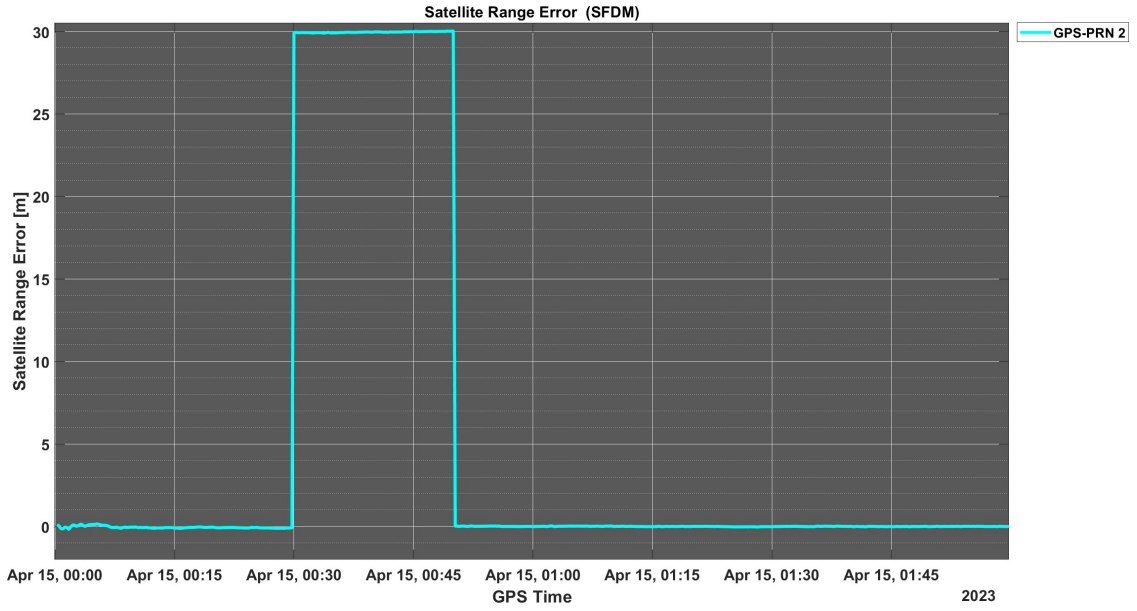


Figure 2.4: Satellite clock error with 30 m step fault injection

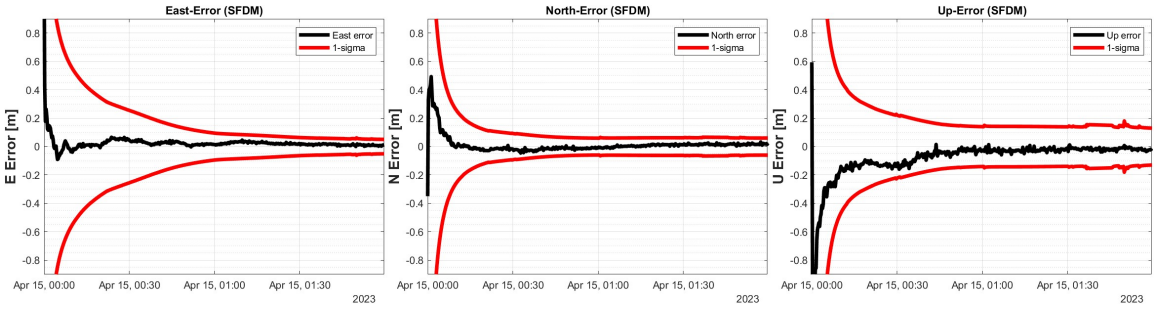


Figure 2.5: Position error in East-North-Up direction with satellite clock error state and 30 m step fault.

Fig. 2.4 shows the value of the satellite error state, $\zeta^{(k)}$. When the 30 m step clock fault occurs on PRN-2 from 30 to 50 min of time, it can be seen that the error state captures the onset and termination of the step fault. Fig. 2.5 shows the position error of the filter; it is unaffected by the step fault. As shown in Eq. 2.6, the addition of $\zeta^{(k)}$ to PRN-2 effectively excludes the measurements of PRN-2 from the computation in the estimator, so all states remain normal throughout the course of the 30 m step fault. Essentially, the estimator is using the measurements, excluding the ones from

PRN-2, to estimate the state of the entire estimator, then applying the resulting receiver position and clock bias as reference to compare with the pseudorange from PRN-2 to obtain the amount of range error presented in the measurement of PRN-2.

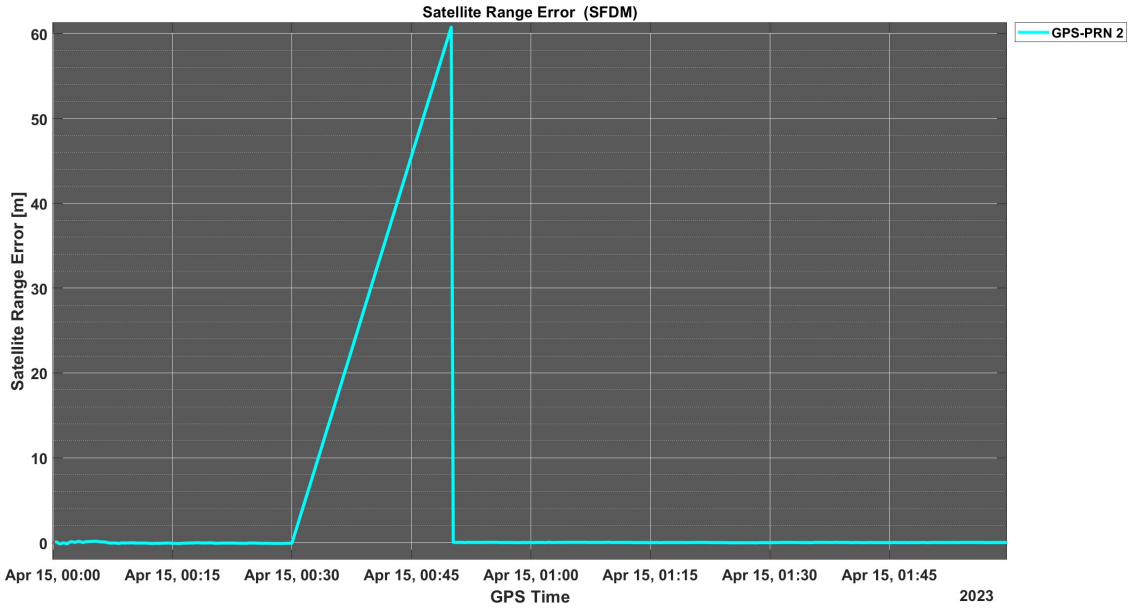


Figure 2.6: Satellite clock error state with 5 cm/sec ramp fault injection

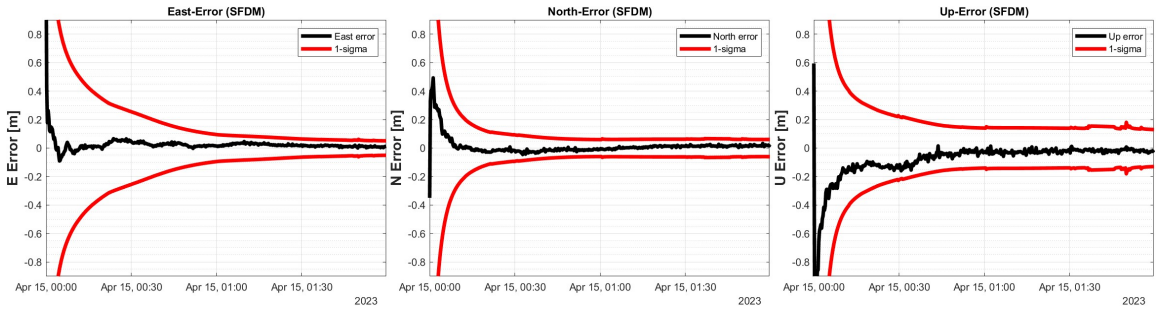


Figure 2.7: Position error in East-North-Up direction with satellite clock error state and 5 cm/sec fault.

Similarly, the 5 cm ramp fault is also successfully estimated. Fig. 2.6 shows the value of the satellite clock error state monitoring for PRN-2. Fig. 2.7 is the position error under the ramp fault, and it is identical to Fig. 2.5.

Actual Satellite Clock Fault on January 25th, 2023, Single Receiver

GPS-PRN-1 (SVN-63) experienced several clock faults and anomalies in 2023. Particularly, there was a step fault that occurred on GPS-PRN-1 on January 25th, an anomaly on January 28th, and a step fault on July 10th. Due to SVN-63 demonstrating unpredictable performance and currently operating on its last clock, it was retired on Aug. 10th and SVN-44 was reactivated and operates under PRN-22 as the replacement to PRN-1.

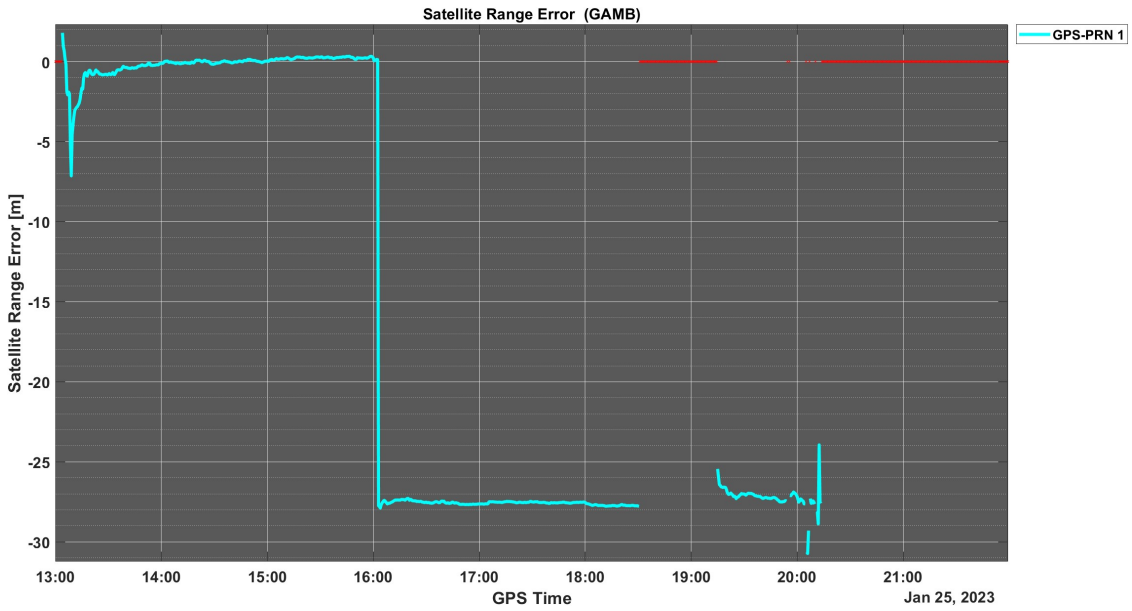


Figure 2.8: Satellite clock error state value on January 25th.

Fig. 2.8 is the value of the satellite clock error state. The fault represented by the error state is consistent with the fault analyzed in [10]. The only difference is it begins a few minutes earlier, at 16:02:30, when the satellite error state drops to around -27 m, then persists until 18:30:30. From 18:31:00 to 19:14:30 the measurements from PRN-1 are missing due to the satellite being switched to non-standard code, and then the signal is received again after 19:15:00 and labeled unhealthy. The users were exposed to a faulted GPS PRN-1 signal for up to 3.5 hours after 16:00:00 on Jan 25th, which threatened the integrity of the user if the measurements from PRN-1 had been used in the position estimation.

The result shown in Fig. 2.8 applied IGS precise ephemeris and broadcast satellite clock bias. Since the IGS precise satellite clock bias is post-processed by IGS, it removed the occurrence of the satellite clock fault. In the rest of the dissertation, unless otherwise specified, the actual satellite fault is using IGS precise ephemeris and broadcast satellite clock bias.

The use of the satellite clock error state, $\zeta^{(k)}$, allows a PPP filter to be tolerant towards both the satellite clock and ephemeris faults, since for the single receiver, the pseudorange error induced by the ephemeris error is absorbed into the satellite clock bias error over the instantaneous estimates of the satellite clock error. It also provides the real-time estimation of any abnormal behavior of the corresponding satellite clock or corrections from the filter in the pseudorange domain. Although the addition of $\zeta^{(k)}$ effectively reduces the observable measurements to the filter and can only monitor the dedicated satellite clock fault, it can provide a light-weight estimation and monitoring of the satellite clock faults.

2.1.2 Multi Receiver Monitor Network

The limitation of monitoring the satellite faults through a single receiver is obvious: it can only monitor the satellite faults that have LOS to the receiver, and cannot separate the effect of satellite clock and ephemeris faults, so the errors are only applicable within close proximity. The size of the region is determined by an upper bound on the magnitude of the ephemeris error. In addition, the error state has to be assigned correctly to the faulted satellite when the satellite fault occurs. If a fault occurs on a satellite without the error state, the estimator would still fail. Because of this, although the estimator is causal and can operate in real-time, the single receiver self-monitoring is more suitable for quick analysis of post-processing where the faulted satellite has been identified. Alternatively, SS on the available satellites forming multiple filters, with each of them having corresponding fault modes, is also possible for operating regionally, which is how standard ARAIM operates [18].

To enable the consistent and complete monitoring of both the satellite clock and ephemeris faults for the entire constellation, measurements from multiple receivers are combined together to form a new fault-tolerant filter against the potential satellite clock and/or ephemeris faults.

From this point onward, all satellites are assumed to be monitored, and the corresponding satellite error states are included in the measurement models for all satellites. This effectively assumes the satellite faults could happen to any satellite within the constellation. The benefit of monitoring the whole constellation is apparent. There is no need to do solution separation on different satellite measurements or assign error states to potential faulted satellites. The satellite fault detection and exclusion mechanics can therefore be avoided on satellite faults. The fault-tolerant filters are directly estimating the satellite faults of the entire constellation from the filter since the satellite clock and ephemeris faults are part of the model.

In this context, the state-vector of each receiver within the network is defined as:

$$\overrightarrow{x_r} = \begin{bmatrix} b_r \\ \nabla b_r \\ \nabla \hat{T}_r \\ A_r^{(1)} \\ \vdots \\ A_r^{(k)} \end{bmatrix} \tag{2.7}$$

where $\overrightarrow{x_r}$ is the state-vector of the rth receiver in the network. The receivers are assumed to be static and the positions are known parameters so position and velocity states are removed.

Then, the state-vector of the fault-tolerant filter that monitors the entire

constellation with multiple receivers is defined as:

$$\overrightarrow{x}_f = \begin{bmatrix} \overrightarrow{x_1} \\ \overrightarrow{x_2} \\ \vdots \\ \overrightarrow{x_r} \\ \vdots \\ \overrightarrow{x_n} \\ \overrightarrow{D_1} \\ \overrightarrow{D_2} \\ \vdots \\ \overrightarrow{D_k} \\ \vdots \\ \overrightarrow{D_m} \end{bmatrix} \tag{2.8}$$

where there are n receivers and m satellites within the integrity monitor, and $\overrightarrow{D_k}$ is defined as the satellite clock and ephemeris errors of the kth satellite:

$$\overrightarrow{D_k} = \begin{bmatrix} \zeta^{(k)} \\ \Delta\overrightarrow{\gamma}^{(k)} \end{bmatrix} \tag{2.9}$$

$$\Delta\overrightarrow{\gamma}^{(k)} = \begin{bmatrix} \gamma_x^{(k)} \\ \gamma_y^{(k)} \\ \gamma_z^{(k)} \end{bmatrix} \tag{2.10}$$

with the measurement models:

$$\rho_{\text{IF},r}^{(k)} = \| \overrightarrow{x}_{sat}^{(k)} - \overrightarrow{x}_{pos,r} + \Delta\overrightarrow{\gamma}^{(k)} \| + \left(b_r - b_{sat}^{(k)} \right) + m_r^{(k)} \nabla\hat{T}_r + R_{\rho,r}^{(k)} + \zeta^{(k)} + \epsilon_{\rho,r}^{(k)} \tag{2.11}$$

$$\phi_{\text{IF},r}^{(k)} = \| \overrightarrow{x}_{sat}^{(k)} - \overrightarrow{x}_{pos,r} + \Delta\overrightarrow{\gamma}^{(k)} \| + \left(b_r - b_{sat}^{(k)} \right) + m_r^{(k)} \nabla\hat{T}_r + A_r^{(k)} + R_{\phi,r}^{(k)} + \zeta^{(k)} + \epsilon_{\phi,r}^{(k)} \tag{2.12}$$

Eq. 2.11 and Eq. 2.12 are the measurement models of the fault-tolerant filter.

They are similar to Eq. 2.1 and Eq. 2.2 except for the additional error states, $\zeta^{(k)}$ and $\Delta\overrightarrow{\gamma}^{(k)}$, that account for the satellite clock errors and ephemeris errors, respectively. $\gamma_x^{(k)}$, $\gamma_y^{(k)}$, and $\gamma_z^{(k)}$ are the satellite position errors in the x, y and z directions. Similarly, $\zeta^{(k)}$ and $\Delta\overrightarrow{\gamma}^{(k)}$ are assumed without any prior knowledge and reset to zero at each time-update with their process noise as large as the receiver clock bias states. This reflects the fact that there are no prior knowledge about when and what type of satellite clock and ephemeris fault might occur at any given time.

The time-update model for the satellite error states is:

$$\overrightarrow{D}_{k,t+1} = \overrightarrow{0} \tag{2.13}$$

where $\overrightarrow{0}$ is simply the zero vector to ensure no priors are applied to the satellite error states. The filter is dynamically estimating the zenith tropospheric errors and the floating ambiguities, while the rest of the states are essentially snapshots.

Satellite Clock Error and Ephemeris Error

$\zeta^{(k)}$ and $\Delta\overrightarrow{\gamma}^{(k)}$ are what make the filter tolerant to the satellite clock and ephemeris faults. They correspond to the potential satellite clock error and ephemeris error that might occur in the assumed values. All satellites have these two additional error states that allow the filter to not only be tolerant to the satellite clock and ephemeris errors, so the filter states are not corrupted by the faults, but also allow it to consistently estimate the errors induced in the satellite clock and ephemeris, even when multiple satellite faults occur simultaneously. Only one fault-tolerant filter is required to detect satellite clock and ephemeris faults, and can estimate the errors in the pseudorange domain.

For the fault-tolerant filter to react to any type of fault, including step jump or slow/fast drift in both satellite clock and ephemeris, an important assumption on $\zeta^{(k)}$ and $\Delta\overrightarrow{\gamma}^{(k)}$ is that there are no constraining priors on satellite clock and ephemeris errors. The satellite clock and ephemeris faults may occur at any instance of time, and can have any type of dynamics. To represent this assumption in EKF, $\zeta^{(k)}$ and $\Delta\overrightarrow{\gamma}^{(k)}$ are set to have process noise as large as receiver clock bias, which are infinite

in theory, and have no dynamic model. This is represented by a time update with zero vectors in Eq. 2.13. It means that in each time update, the satellite clock and ephemeris errors are independently estimated by the filter. In other words, the filter is tolerant to satellite clock and ephemeris faults because it assumes all satellites are faulted all the time, and the filter is estimating the associated errors for the entire constellation.

The system of the fault-tolerant filter becomes seemingly unobservable when receiver clocks, satellite positions, and clocks are unknown, and there are no dynamics at all for the errors. However, the system is still observable. The observability of the system is discussed in the next subsection.

The process noise for satellite error states is shown in Table 2.1, and no correlations are assumed. The numerical values are selected in align with the process noise of the receiver clock bias states.

These values are significantly greater than the real faults occurred in the operational history of GPS [11], and are functionally comparable to having non-constraining priors to the satellite faults. Larger values are possible but may lead to numerical inaccuracy due to formulation of ill-conditioned matrices in the computer.

$\sigma_{\zeta^{(k)}}$	Satellite Clock Bias Error	$1000 \cdot \sqrt{\Delta t}$	$[m]$
$\sigma_{\gamma_x^{(k)}}$	Satellite Position Error in x direction	$1000 \cdot \sqrt{\Delta t}$	$[m]$
$\sigma_{\gamma_y^{(k)}}$	Satellite Position Error in y direction	$1000 \cdot \sqrt{\Delta t}$	$[m]$
$\sigma_{\gamma_z^{(k)}}$	Satellite Position Error in z direction	$1000 \cdot \sqrt{\Delta t}$	$[m]$

Table 2.1: Process Noise of Satellite Errors

Functional Observability Analysis

The functional observability of the system can be analyzed by forming the observability matrix. To simplify the construct, here only clock bias is considered, b_r, for the receiver, and satellite clock bias error, $\zeta^{(k)}$, for the satellite in the fault-tolerant filter. First, assume the fault-tolerant filter only consists of one receiver and

one satellite. The associated state vector is $\begin{bmatrix} b_1 & \zeta^{(1)} \end{bmatrix}^T$, and the time-update model A_1, the measurement model C_1, and the observability matrix $\mathcal{O}_1$, are:

$$A_1 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix},\ C_1 = \begin{bmatrix} 1 & 1 \end{bmatrix},\ \mathcal{O}_1 = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$$

where the rank number is the same as the number of states. So, the system is mathematically observable. This means one satellite clock bias error only requires one reference receiver to observe. For such a single receiver monitor, the effect of b_1 and $\zeta^{(1)}$ are equivalent to the pseudorange error, and the filter simply distributes the amount of biases to receiver and satellite bias states based on the priors, i.e., the initial noise, process noise, and covariance estimate. For instance, assume the absolute clock bias for receiver and satellite are $b_{1,abs} = 10[m]$ and $\zeta^{(1)}_{abs} = 20[m]$. When both states have the same priors and conditions, the filter would equally distribute the resulting biases as $b_{1,est} = 15[m]$ and $\zeta^{(1)}_{est} = 15[m]$. The filter can estimate the joint effects of both biases, but cannot estimate the absolute bias and it distributes them according to the priors on these two states.

In the case of one receiver and two satellites, the system is still observable since it can always be decomposed into one on one receiver-satellite relation, but the clock biases from both satellites are distributed differently with the receiver. The receiver can only observe and be affected by the difference of the satellite clock biases, since the receiver clock bias state removes the common mode bias between the two satellites. For example, let the absolute biases be $b_{1,abs} = 0[m]$, $\zeta^{(1)}_{abs} = 5[m]$, and $\zeta^{(2)}_{abs} = 10[m]$. They have the equivalent effect to the receiver as $b_{1,est} = -5[m]$, $\zeta^{(1)}_{est} = 0[m]$, and $\zeta^{(2)}_{est} = 5[m]$, or any other values that preserve the relative differences. So, the fault-tolerant filter is estimating these relative differences on the receiver and satellite clock biases. The concept is more clear in the case of one receiver and three satellites. Let $b_{1,abs} = 0[m]$, $\zeta^{(1)}_{abs} = 2[m]$, $\zeta^{(2)}_{abs} = 6[m]$, and $\zeta^{(3)}_{abs} = 16[m]$, the filter is actually measuring the differences of the clock biases from the pseudorange measurements.

The satellite clock differences are:

$$\begin{aligned}\zeta_{est}^{(1)} - \zeta_{est}^{(2)} &= -4 \\ \zeta_{est}^{(1)} - \zeta_{est}^{(3)} &= -14 \\ \zeta_{est}^{(2)} - \zeta_{est}^{(3)} &= -10\end{aligned} \tag{2.14}$$

where there are three equations and three unknowns (clock bias can be arbitrary so it is not counted here), but the equations are always underdetermined by one degree of freedom since it is constructed from the difference of the biases, so there are only two effective equations and three unknowns. The filter estimates these differences of the biases, and represents them as absolute biases based on the priors. Although the system is observable, the resulting estimates may be arbitrary since the filter is representing an relative difference with absolute values without a fixed reference. Therefore, if an additional relation for the biases is applied, the filter can have a consistent way of representing the differential relations. For example, assume the mean of the satellite clock biases is zero, $\zeta_{est}^{(1)} + \zeta_{est}^{(2)} + \zeta_{est}^{(3)} = 0$, Eq. 2.14 can then be determined as $\zeta_{est}^{(1)} = -6[m]$, $\zeta_{est}^{(2)} = -2[m]$, and $\zeta_{est}^{(3)} = 8[m]$, and the receiver clock bias state absorbs the common mode, $b_{1,est} = 8[m]$. The absolute clock biases for receiver and satellites cannot be determined if the errors states exist, but the relative differences are observable, and by adding an additional relation, such as a zero-mean constraint, on the satellite clock bias states, the filter can have a consistent reference to represent such relative estimates. In fact, the fault-tolerant filter is essentially estimating the single difference on the satellites. For GPS constellations with 31 satellites, there are a total of 31 choose 2 combinations of differences, $PRN1 - PRN2$, $PRN1 - PRN3$, ..., $PRN31 - PRN32$, but instead of forming 465 error states, the fault-tolerant filter only forms 31 satellite clock error states. For users, their range errors only require the relative difference of the satellite clock bias errors since the common mode bias is compensated by their receiver clock bias state, so absolute corrections are not needed. Consequently, adding an additional constraint does not add any information to the system, it merely provides a consistent reference point for the filter to use. Finally, since the difference of the clock errors only lack by one

degree, only one constraint should be added.

When the fault-tolerant filter involves the estimation of both the satellite clock and ephemeris errors, at least four receivers per satellite are required otherwise the system cannot be made functionally observable. For a filter of four receivers and one satellite, the associated state vector is $\begin{bmatrix} b_1 & b_2 & b_3 & b_4 & \zeta^{(1)} & \gamma_x^{(1)} & \gamma_y^{(1)} & \gamma_z^{(1)} \end{bmatrix}^T$, and the time-update model A_4, the measurement model C_4, and the observability matrix $\mathcal{O}_4$ after row reduction are:

$$A_4 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{2.15}$$

$$C_4 = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & u_{x_1}^{(1)} & u_{y_1}^{(1)} & u_{z_1}^{(1)} \\ 0 & 1 & 0 & 0 & 1 & u_{x_2}^{(1)} & u_{y_2}^{(1)} & u_{z_2}^{(1)} \\ 0 & 0 & 1 & 0 & 1 & u_{x_3}^{(1)} & u_{y_3}^{(1)} & u_{z_3}^{(1)} \\ 0 & 0 & 0 & 1 & 1 & u_{x_4}^{(1)} & u_{y_4}^{(1)} & u_{z_4}^{(1)} \end{bmatrix} \tag{2.16}$$

$$\mathcal{O}_4 = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & u_{x_1}^{(1)} & u_{y_1}^{(1)} & u_{z_1}^{(1)} \\ 0 & 1 & 0 & 0 & 1 & u_{x_2}^{(1)} & u_{y_2}^{(1)} & u_{z_2}^{(1)} \\ 0 & 0 & 1 & 0 & 1 & u_{x_3}^{(1)} & u_{y_3}^{(1)} & u_{z_3}^{(1)} \\ 0 & 0 & 0 & 1 & 1 & u_{x_4}^{(1)} & u_{y_4}^{(1)} & u_{z_4}^{(1)} \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{2.17}$$

where $u_{x_r}^{(1)}$ is the LOS unit-vector component in the x direction from receiver r to satellite 1, and same for the y and z directions. $\mathcal{O}_4$ is full rank only when the LOS

vectors do not coincide with each other. This corresponds to requiring four receivers to have different LOS directions with respect to the satellite in order to have the system be observable. Similarly, the fault-tolerant filter is essentially estimating the single difference of the LOS satellites from the range domain. Therefore, while the system is observable, an additional constraint can be added to have consistent estimation from the filter.

In an actual integrity monitor network where there are multiple receivers and different LOS satellites, the formulation of the observability matrix can be complicated, but it can always be decomposed into one satellite with respect to multiple receivers. Thus, the relative pseudorange errors of the entire system is observable as long as each satellite has LOS to at least four receivers that have a sufficient overlap of common satellites with other neighboring receivers as is the case for the type of receiver networks considered in this thesis. However, because the distribution of the error estimates depends on various factors, including prior assumptions and receiver clock bias estimates, the satellite error estimates may be arbitrarily assigned in an absolute sense, while their relative relationships remain meaningful. In fact, the errors cannot be absolutely estimated. The addition of a constraint is to make absolute estimates from the filter consistent and stable; it does not add any information to the system. The choice of the constraint can be arbitrary and is discussed in Sec. 2.1.2. Finally, including tropospheric state for the receiver does not change the required number of receivers per satellite due to the presence of the mapping function, and the prior on the wet delay parameter.

Projection and User Range Error

Although the satellite clock and ephemeris errors are directly estimated by the filter, they are not obvious indicators to describe the influence of those errors on the user's position accuracy. As a result, an important transformation is conducted; i.e., the projection, to interpret the impact of $\zeta^{(k)}$ and $\Delta\overrightarrow{\gamma}^{(k)}$ in the range domain. The projected result from $\zeta^{(k)}$ and $\Delta\overrightarrow{\gamma}^{(k)}$ in the range domain is then the User Range Error (URE) at the projected location induced from the satellite clock and ephemeris

errors. The projection can be expressed as:

$$\beta_u^{(k)} = \begin{bmatrix} 1 \\ \overrightarrow{e}_{u,k} \end{bmatrix}^T \cdot \begin{bmatrix} \zeta^{(k)} \\ \Delta\overrightarrow{\gamma}^{(k)} \end{bmatrix} \tag{2.18}$$

where $\beta_u^{(k)}$ is the URE at a potential user location, u, from the kth satellite, and $\overrightarrow{e}_{u,k} \in \mathbb{R}^{3\times 1}$ is the LOS unit-vector from u to k^{th} satellite. This allows us to obtain the estimate of the URE at the designated location and so to correct the projected error at that location since we are evaluating the same satellite clock and ephemeris.

Similarly, the uncertainty of the estimated URE is calculated with the linear projection:

$$\sigma_u^{(k)^2} = \begin{bmatrix} 1 \\ \overrightarrow{e}_{u,k} \end{bmatrix}^T \cdot COV\left(\begin{bmatrix} \zeta^{(k)} \\ \Delta\overrightarrow{\gamma}^{(k)} \end{bmatrix}\right) \cdot \begin{bmatrix} 1 \\ \overrightarrow{e}_{u,k} \end{bmatrix} \tag{2.19}$$

where $\sigma_u^{(k)}$ is the sigma corresponding to $\beta_u^{(k)}$.

$\beta_u^{(k)}$ and $\sigma_u^{(k)}$ are the interpretation of $\zeta^{(k)}$ and $\Delta\overrightarrow{\gamma}^{(k)}$ in the range domain for any potential user at the location u. They are inherently the integrity information with which users protect themselves from satellite clock and ephemeris faults. The monitoring of the satellite fault modes and the generation of integrity information both stem from these estimates. In the rest of the thesis, they will be used to provide integrity to the users of the PPP correction service.

Recall from the single receiver self-monitoring (Sec. 2.1.1), it implicitly projected the satellite clock and ephemeris errors onto its receiver location estimated by other measurements without the error states. The result is local. On the other hand, the use of multiple receivers can separate satellite clock and ephemeris errors and allows global projection of the satellite range errors.

Static Constraint

Since ground-based GNSS receivers are static in IGS, and there is no reason to monitor the satellite errors with kinematic receivers, the position of the satellites can be

surveyed beforehand and become part of the constant in the measurement models.

Zero-Mean Constraint

The choice of constraint is fundamentally arbitrary because it does not add any information into the system, but as discussed in Sec. 2.1.2, only one constraint can be applied. Additionally, the constraint also affects the linearization point for the system given that the measurement models are not linear.

Similarly, for a filter that monitors both satellite clock and ephemeris errors, a constraint should be applied to provide a consistent reference. However, due to the spatial variations of ephemeris error to the receivers in the range domain, another choice of design emerges for the ephemeris errors. In the range domain, satellite clock errors have equal effects regardless of the location of projection, but the range error contributed by the ephemeris errors may vary. Given that only one effective constraint should be added to the system, the projection location should be near the service area for the users. Since the target users of the integrity monitor considered in this thesis are global users, the range errors are projected onto the entire surface of the Earth. Assuming the ellipsoid shape of the Earth, the equal projection of range error to the Earth is equivalent to projecting to the geometric center, that is the center of the Earth. Thus, the Zero-Mean Constraint for the fault-tolerant filter can be expressed as:

$$\sum_{k=1}^{n_{\text{sat}}} \beta_{u_0}^{(k)} = 0 \tag{2.20}$$

where n_{sat} is the number of satellites in the constellation, and u_0 is the location of the center of the Earth.

Alternatively, the reference point should be dependent on the user regions to which the range errors are projected. Since the target user regions are related to the locations of the receivers within the monitor network, the Zero-Mean Constraint can be chosen to be enforced at the average location of the receivers.

$$\beta_r^{(k_r)} = \begin{bmatrix} 1 \\ \overrightarrow{e}_{r,k_r} \end{bmatrix}^T \cdot \begin{bmatrix} \zeta^{(k_r)} \\ \Delta\overrightarrow{\gamma}^{(k_r)} \end{bmatrix} \tag{2.21}$$

where $\beta_r^{(k_r)}$ is the range error projected from the LOS satellites , k_r, of the receiver, r.

The constraint to be enforced can then be formed:

$$\sum_{r=1}^{n_{\text{rec}}} \sum_{k_r=1}^{n_{\text{sat,r}}} \beta_r^{(k_r)} = 0 \tag{2.22}$$

where n_{rec} is the number of receivers in the network and $n_{\text{sat,r}}$ is the number of LOS satellites for receiver r.

Eq. 2.22 is the general Zero-Mean-Constraint that can be used for arbitrary combinations of the receiver network. Since the reference point is the instantaneous average of the LOS satellites for each receiver, it is aligned with the potential user area for the projection of the satellite errors.

Linearity and Precise Product in Fault-tolerant Filter

If the system were truly linear, the estimation results would be independent of the initial clocks and orbits. Since there are no dynamics and only very loose priors on the satellite errors states, the filter largely acts as a comparator between the satellite states and the measurements, that measuring the range errors between the pseudorange measurements and the precise satellite clock and ephemeris. When precise products are used for the initial clock and ephemeris states, the URE estimates would be expected to be zero. For the remainder of this paper we will be utilizing precise products as the linearization point for the filter estimates of UREs. This serves two purposes: as a reference for the filter estimate, and as truth to verify the system. In other words, the UREs induced from measurement errors are effectively compared against the precise product.

Due to the fact that the actual pseudorange measurement is not linear, it requires a

linearization point to construct a locally linear system to be used within the framework of Kalman Filter. Therefore, the fault-tolerant filter is not completely independent to the input product, since it uses the information from that product as the linearization point. Nevertheless, the impact of the linearization point is negligible in the scale of the fault that is being monitored. Specifically, the filter uses the input satellite clock bias to synchronize measurements in time, and input ephemeris as linearized model for satellite position, i.e., the LOS unit-vector. For instance, 300 km of satellite clock error in range domain corresponds to ~ 1 millisecond in time domain, and for GPS satellite with $\sim$20200 km altitude, 200 km of translational error in satellite position corresponds to $\sim 1\%$ of variation in LOS unit-vector. Even for hundreds of kilometers of satellite clock and ephemeris errors, the linearized model from the precise product can still be well representative for the pseudorange measurements. Such a condition would require range errors on the order of thousands of kilometers from the satellite clock, ephemeris, or input products, causing the linearization assumptions of the fault-tolerant filter to break down. In other words, as long as the satellite clock and ephemeris errors are within this linearized region for the linear model constructed based on the input product, any satellite clock and ephemeris faults with magnitude within tens or few hundreds of kilometers can still be accurately captured by the fault-tolerant filter.

Demonstration of the Multi-Receiver Network

For the purpose of demonstrating integrity monitor performance with multiple receivers, 56 IGS stations form a ground-based monitor network. The station codes are: AC23, ADIS, AGGO, AREG, AREQ, BRMG, CCJ2, CHTI, CRO1, CZTG, DGAR, DRAO, DUND, FAA1, FALK, FLIN, FRDN, HAL1, HARB, HKWS, HOLP, IISC, IQAL, JOG2, KMNM, KRGG, LAMA, LMMF, MAL2, MANA, MCIL, MIKL, MIZU, MKEA, MQZG, NAUS, P051, PIMO, POL2, POVE, RESO, REYK, RIO2, SCTB, SFDM, STHL, SUTH, TASH, TRO1, TSK2, UNBD, VILL, WARK, WSRT, YELL, and ZAMB.

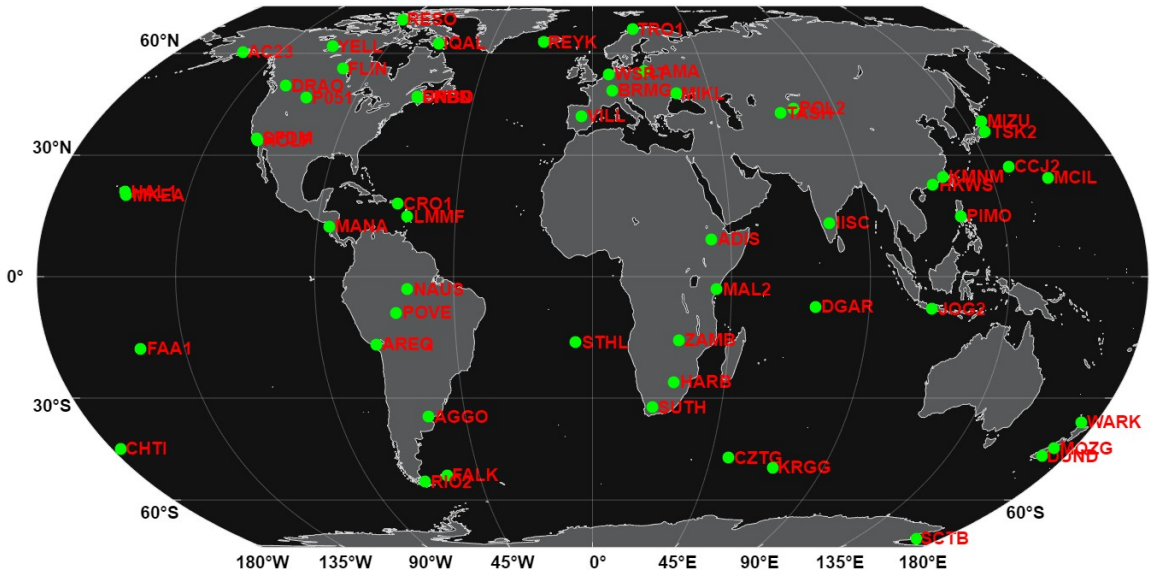


Figure 2.9: 56 selected IGS stations.

The locations of these 56 stations are shown in Fig. 2.9. The stations are selected such that each satellites should be LOS to at least 8 receivers at any time in a day in normal conditions without interruptions such as cycle slips or loss of tracks.

Unless otherwise specified, the 56-station network will be used as the global receiver network in the rest of the thesis.

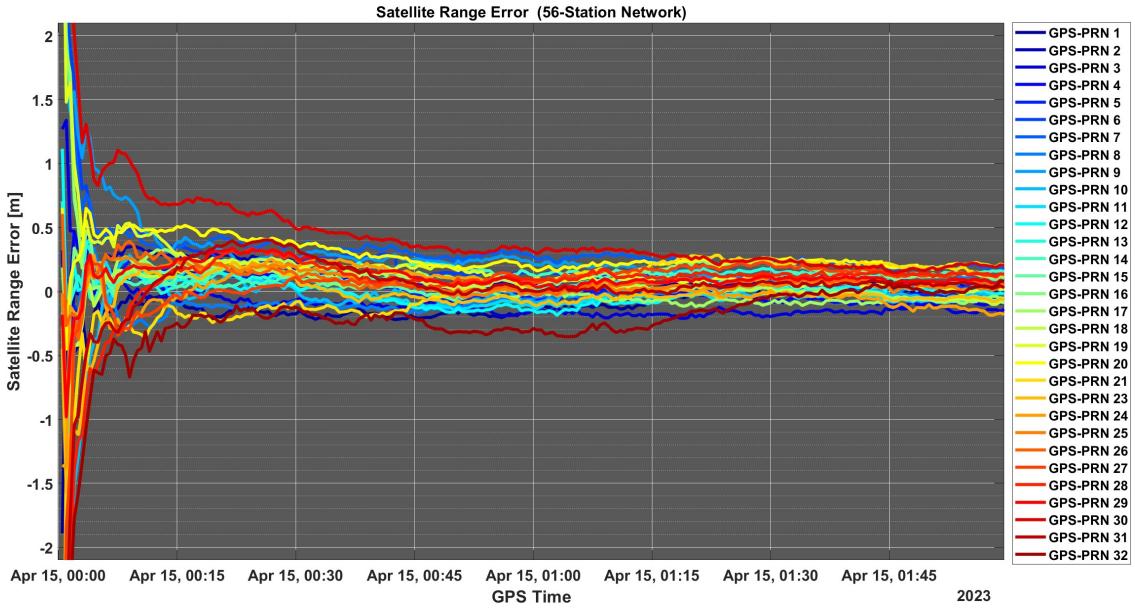


Figure 2.10: UREs of 56-station network on a normal day at JPLM.

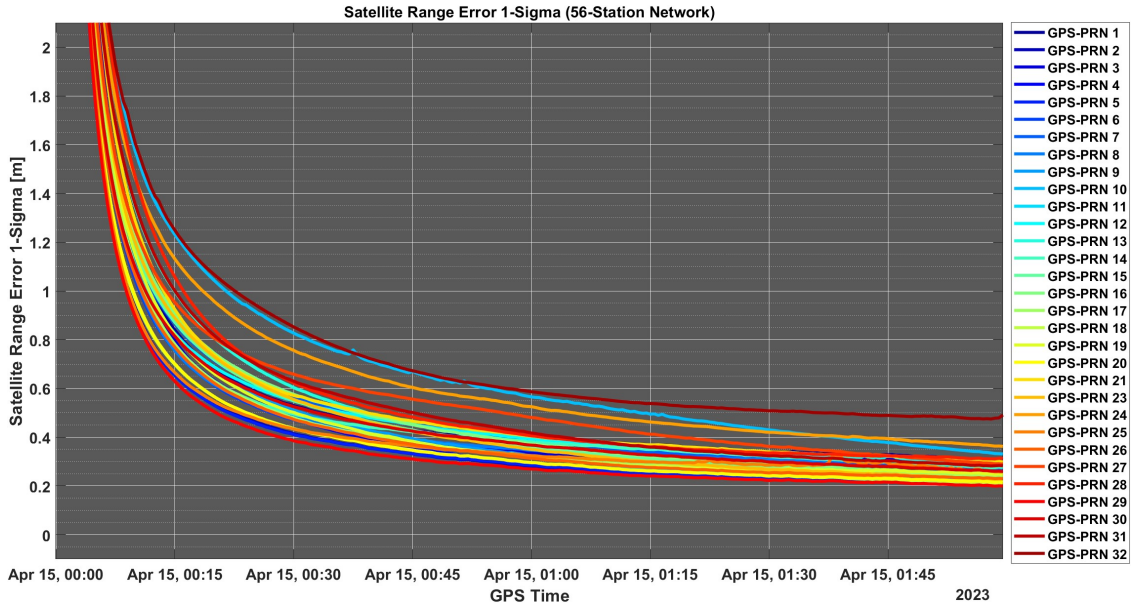


Figure 2.11: URAs of 56-station network on a normal day at JPLM.

Fig. 2.10 is the direct projection of the satellite clock and ephemeris error estimates from the estimator to an IGS station, JPLM, which is independent from

the 56-station network and is located at the Jet Propulsion Laboratory. The JPLM station will be used as the example PPP user in this thesis unless otherwise specified. The UREs shown above are the nominal performance of the GPS constellation on April 16th, a normal day, at JPLM, after applying precise product. Since the precise product has millimeter accuracy, the UREs shown here represent the measurement errors effect on the filter estimates.

Fig. 2.11 shows the 1-sigma uncertainty corresponding to the URE estimates in Fig. 2.10, i.e., the User Range Accuracy (URA). The 56-station network can achieve 30$\sim$50 cm of standard deviation in less than 2 hours of convergence.

The above results only represent a small period of operation for the integrity monitor. This does not represent the robustness of the integrity monitor, rather, this is to build understanding of the structure and mechanism of the integrity monitor. The analysis of a longer period of real data using the integrity monitor will be conducted in Chap. 4.

Determination of Faulted Satellites

In the case of satellite faults, the URE estimates of different satellites would correlate due to the Zero-Mean Constraint. For instance:

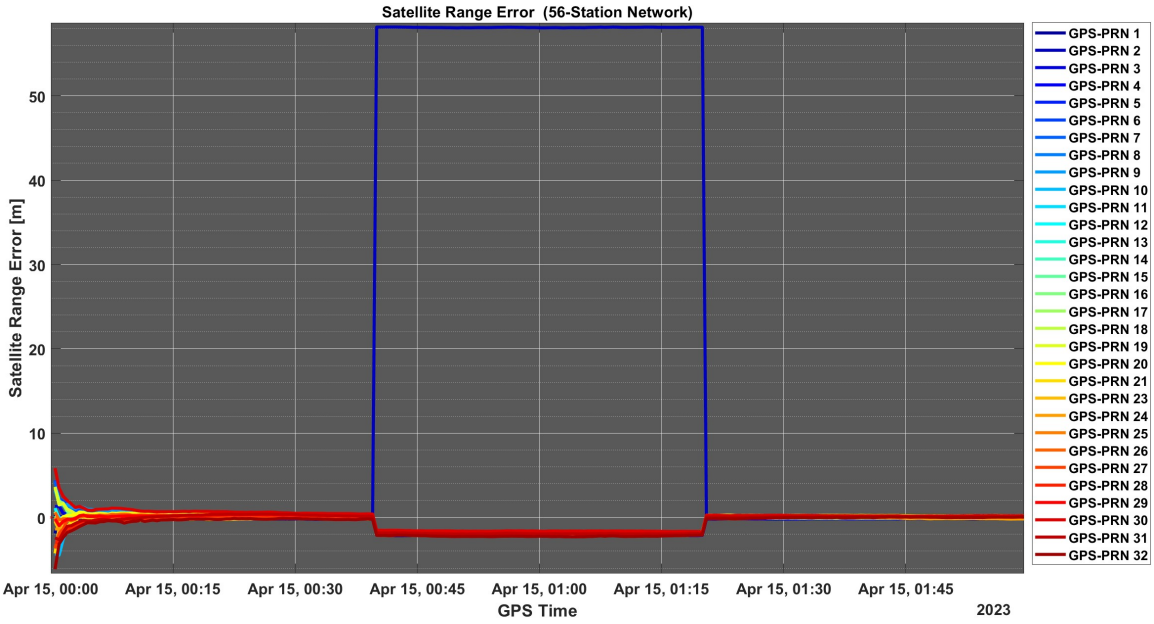


Figure 2.12: UREs of 56-station network with 60 m artificial step fault on GPS-PRN-1 at JPLM.

Fig. 2.12 uses the same URE estimates as Fig. 2.10 but with an artificial 60 m step fault injected to PRN-1 from 00:40:00∼01:20:00 during the 2-hour data period. It can be seen that the URE of PRN-1, the dark blue line, instantly reacts to the 60 m step fault. However, the normal satellites also change. This is due to the Zero-Mean Constraint that constrains the sum of all the UREs after projection is approximated to zero. As a result, the UREs of normal satellites react accordingly to the faulted satellite to compensate for the overall changes caused by the 60 m step fault from PRN-1.

The Zero-Mean Constraint specifies the reference value for the UREs among all the satellites. Similar to the role of receiver clock bias, it ensures all the satellites share a common mode bias in their URE estimates, and the actual informative estimates are the differences between the UREs. As shown in Fig. 2.12, the range difference of PRN-1 and all the other normal satellites are exactly 60 m.

For a user such as JPLM during this step fault event, all it needs is to compensate a 60 m difference to PRN-1, or all the other satellites except PRN-1, to make the measurements consistent. A more extreme example would be adding 180 m offset to

the pseudorange of PRN-1, and 120 m to all the other satellites. The absolute values do not matter since the receiver clock bias of the user would adjust accordingly.

For the purpose of consistency, the absolute value of the normal satellites are zeroed out, so a new reference point is chosen at each time instance of estimates to which the "faulted" satellites are determined based on distance away from the congregation of satellites.

Assuming the UREs of each satellite are independent, a simple outlier exclusion from a uni-variate Gaussian distribution can be formed:

$$X = \sum_{k=1}^{N_{sat}} \frac{X_k}{N_{sat}} \tag{2.23}$$

$$X \sim \mathcal{N}\left(\bar{\beta}, \bar{\sigma}^2\right) \tag{2.24}$$

$$X_k \sim \mathcal{N}\left(\beta_k, \sigma_k^2\right) \tag{2.25}$$

$$\bar{\beta} = \sum_{k}^{N_{sat}} \frac{\beta_k}{N_{sat}}$$

$$\bar{\sigma}^2 = \sum_{k}^{N_{sat}} \frac{\sigma_k^2}{N_{sat}^2}$$

where X_k is the random variables of the UREs from the kth satellite.

The faulted satellite can be excluded with, for instance, a 5% exclusion threshold, given the distribution of X. After identifying the faulted satellites, the normal satellites can be offset such that the mean of them is zero.

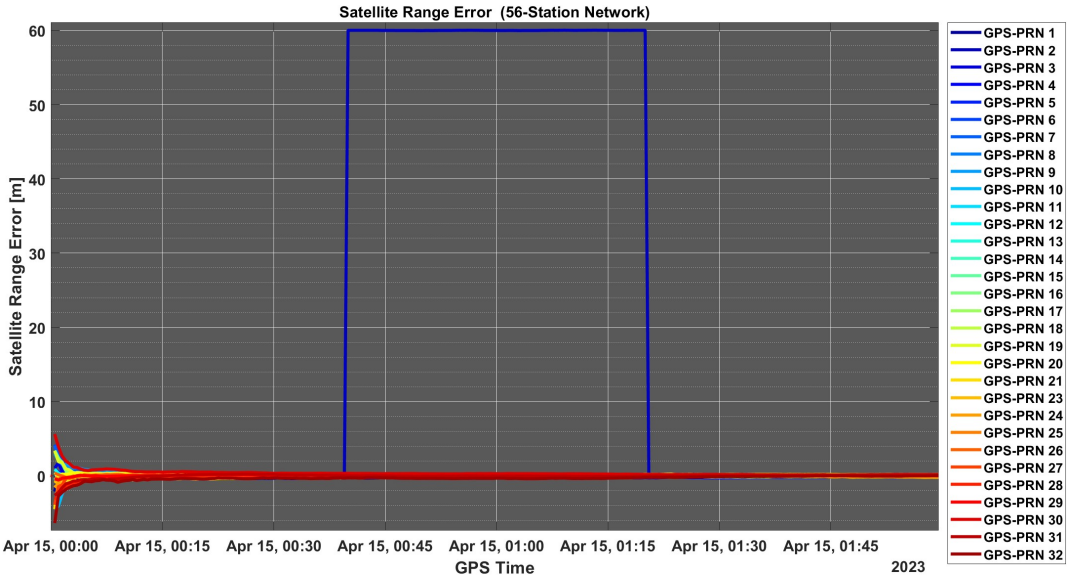


Figure 2.13: Centered UREs of 56-station network with 60 m artificial step fault on GPS-PRN-1 at JPLM.

Fig. 2.13 shows the UREs of the GPS constellation with 60 m artificial fault injected to PRN-1. The only difference from Fig. 2.12 is that the normal satellites are centered to zero mean after excluding the faulted satellites with the distribution formed in Eq. 2.23. It can be seen that the URE of PRN-1 is exactly 60 m during the injection of the step faults. The process of zeroing out the mean of the UREs of normal satellites does not alter the relative differences of the UREs, so no new information is added to the system and the covariance remains unchanged. The faulted satellite is inferred from the assumption that, the majority of GPS satellites are normal at a given time. For a zero latency system, any reference point is acceptable. Whenever fewer than half of the satellites are faulted, this approach can identify the faulted satellites.

Artificial Multi Satellite Faults

As mentioned in the previous section, the filter is tolerant against multiple satellite faults and does not require solution separation.

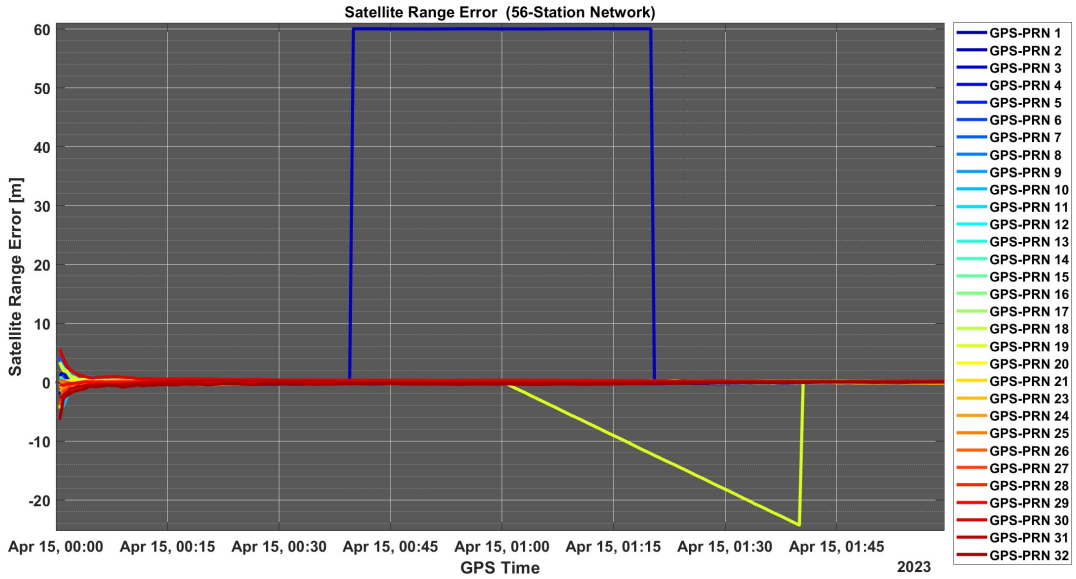


Figure 2.14: UREs of 56-station network with 60 m artificial step fault on PRN-1 and -1 cm/sec ramp fault on PRN-19 at JPLM.

Fig. 2.14 shows the UREs of the GPS with two artificial satellite clock faults. The first fault is the 60 m step fault as used in Fig. 2.13, the second fault is a -1 cm ramp fault injected from 01:00:00∼01:40:00. Although the two faults started and ended at different time, the fault-tolerant filter can accurately estimate the values of the satellite faults.

Actual Satellite Clock Fault on January 25th, 2023, Multi Receivers

Contrary to the single receiver monitor, where the period of monitoring is limited to the LOS of the faulted satellites, using multiple receivers enables un-interrupted monitoring of the entire constellation.

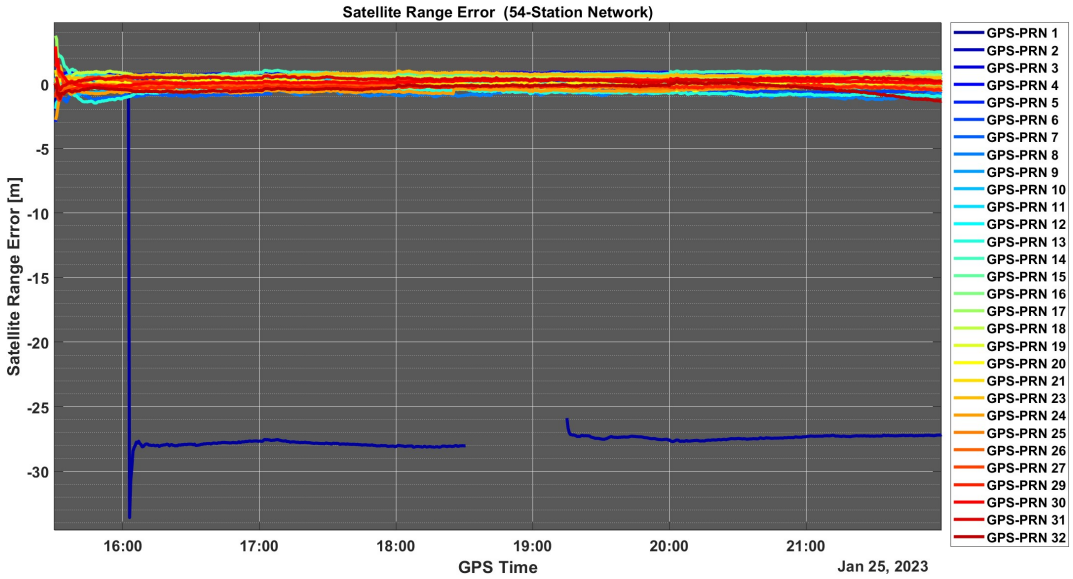


Figure 2.15: UREs on January 25th from multi-receiver network.

Fig. 2.15 shows the UREs of the GPS constellation during the satellite clock fault that occurred on January 25th. During this period, two of the IGS stations, HAL1 and NASU, lack data. This is possibly due to power outage or hardware maintenance, both common among the data from IGS, so only 54 stations are used. It can be seen from the figure that the URE of PRN-1 drops to around 27 m at 16:03:00 and stays for over 2.5 hours before switching to non-standard code at 18:31:00. The signal from PRN-1 is broadcast again starting at 19:15:00 and is labeled unhealthy. With the measurements from multiple receivers across the world, the presence of the clock fault can be monitored consistently without being interrupted by the LOS of a particular receiver. The receiver network is essentially monitoring the entire GPS constellation.

2.1.3 Calibration of Receiver Differential Code Biases

Since the accuracy of the projected UREs from the fault-tolerant filter affected by the Receiver Differential Code Biases (RDCBs), they need to be calibrated to reduce the estimation errors. In other words, the closer the UREs for normal satellites, the more accurate the URE estimates for abnormal satellites given the deviations would be more sensitive for identification by the combined Gaussian distribution. RDCBs

change slowly so they can be calibrated with prior measurements.

In the nominal condition, the satellite clock and ephemeris error states, $\zeta^{(k)}$ and $\Delta\overrightarrow{\gamma}^{(k)}$, in the estimator are essentially estimating the common mode bias in the code phase measurement between the satellites. Therefore, if there is inconsistency due to an individual receiver, the measurement of the same satellite from a difference receiver would not agree and would induce nominal errors in the URE estimates. The receiver clock bias can reduce such biases within the receiver itself, but cannot remove the RDCB caused by hardware delays for different satellites and frequencies.

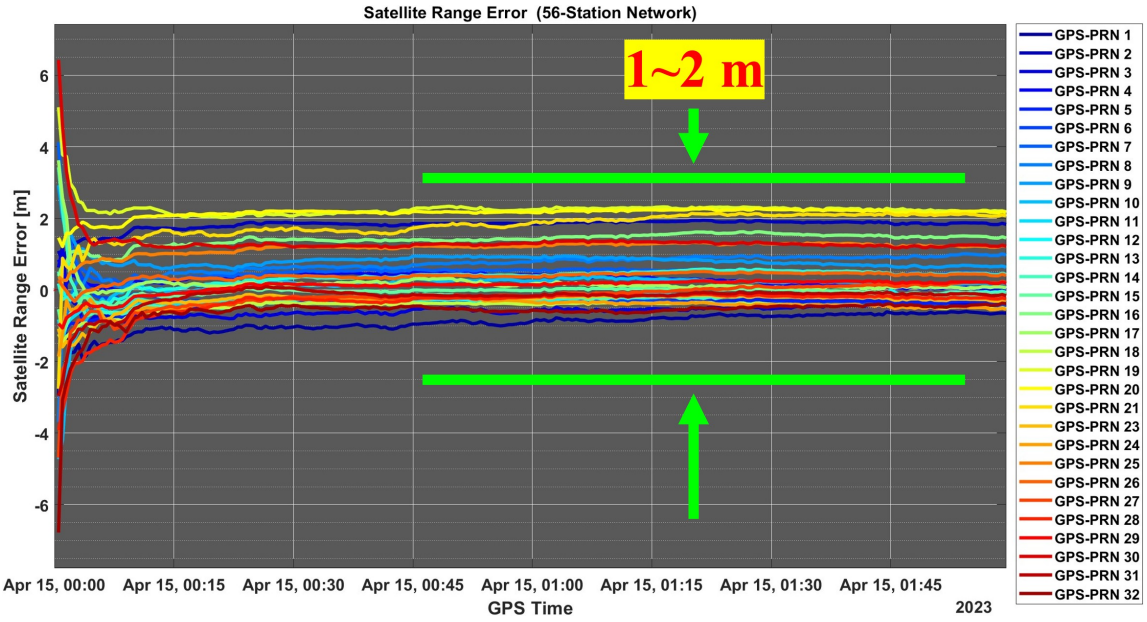


Figure 2.16: Nominal UREs without difference code bias corrections.

Fig. 2.16 shows the nominal UREs of the GPS constellation at JPLM without applying RDCB. Due to inconsistencies of code phase measurements in each of the receivers in the network, the spread of the UREs reaches 1∼2 m even after convergence. This means the URE of a faulted satellite would be un-noticed until its magnitude exceeded 1∼2 m. For this reason, to ensure the accuracy of the URE estimates from the fault-tolerant filter, it is necessary to calibrate the biases that are specific to the receivers.

The RDCB can be estimated independently using the code phase measurements of

individual receivers, and it represents the additional biases between the satellites that the receiver received. They can be estimated from the difference of the pseudorange between the satellites per receiver:

$$\begin{aligned}\rho^{(i)}_{\mathrm{IF},r} - \rho^{(j)}_{\mathrm{IF},r} = &\| \overrightarrow{x}^{(i)}_{sat} - \overrightarrow{x}_{pos,r} \| - \| \overrightarrow{x}^{(j)}_{sat} - \overrightarrow{x}_{pos,r} \| \\ &+ \left(b^{(j)}_{sat} - b^{(i)}_{sat} \right) \\ &+ \left(m^{(i)}_{r} \nabla \hat{T}_r - m^{(j)}_{r} \nabla \hat{T}_r \right) \\ &+ \left(R^{(i)}_{\rho,r} - R^{(j)}_{\rho,r} \right) + b^{i,j}_{r} + \sqrt{2} \cdot \epsilon^{(k)}_{r} \end{aligned} \tag{2.26}$$

where $b^{i,j}_r$ is the RDCB between satellites i and j at receiver r. Eq. 2.26 is simply performing single difference on the receiver to estimate the receiver-specific biases.

For the GPS constellation of 31 satellites, there are a total of $\binom{31}{2} = 465$ combinations, but since each receiver would only be LOS to 8∼12 satellites at a time, the number of combinations for static receivers calibrated for a day typically contains 350∼400 per receiver.

The estimation results of $b^{i,j}_r$ are fitted with second order polynomial model as a function of the elevation angles of the satellites, i and j.

$$\hat{b}^{i,j}_r = \alpha_{r,0} + \alpha_{r,1} \cdot sin(\theta^i_r) + \alpha_{r,2} \cdot sin(\theta^j_r) \tag{2.27}$$

where Eq. 2.27 is the model to be fitted by the estimation result from $b^{i,j}_r$. $\alpha_{r,0}$ is the constant bias, $\alpha_{r,1}$ is the first-order multiplier for satellite i, θ^i_r is the elevation angle from satellite i, and similarly for satellite j.

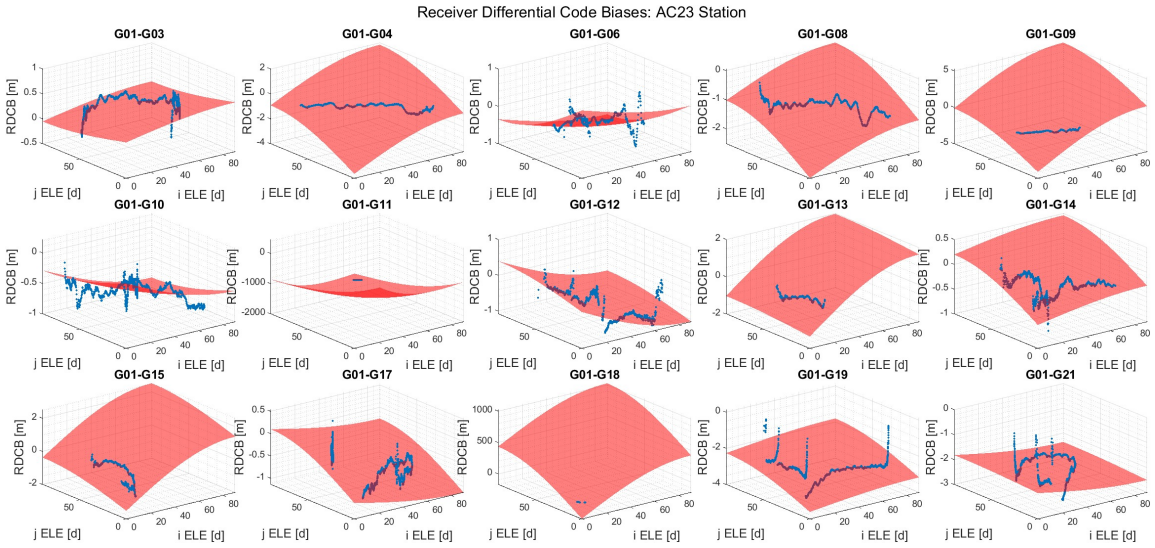


Figure 2.17: Fitting result of the first 15 RDCBs of AC23 station.

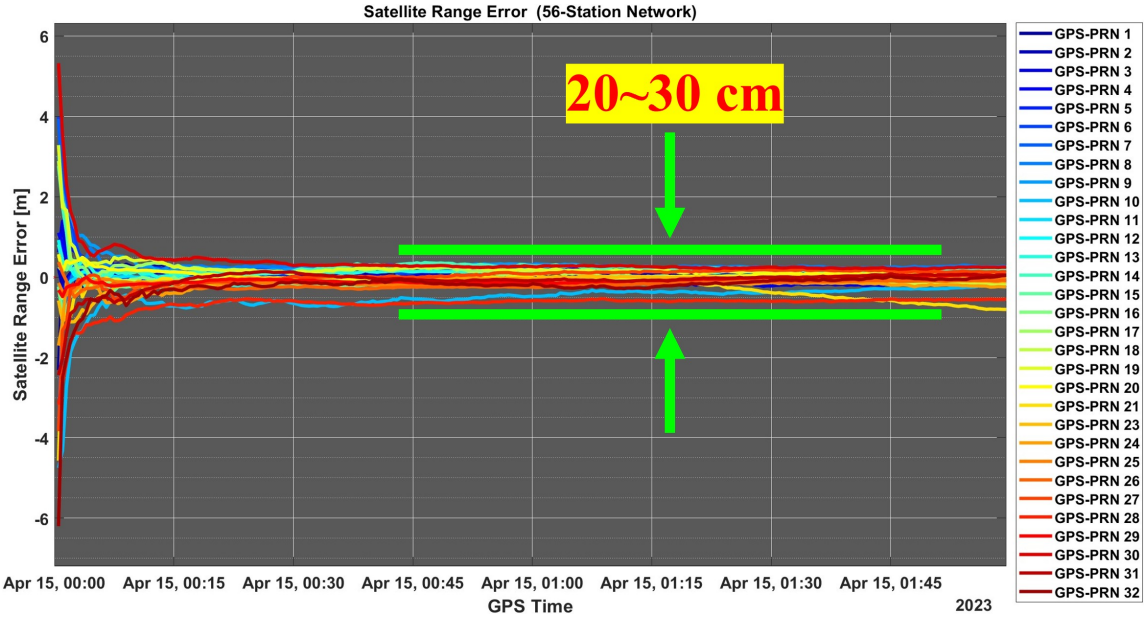


Figure 2.18: Nominal UREs with difference code bias corrections.

Fig. 2.17 is the first 15 RDCB estimations and fitting results from the AC23 station with data from January 24th, 2023. The blue lines represent the RDCBs

between two satellites, and the red surface indicates the fitted model from the first-order polynomial, Eq. 2.27, where the RDCBs and the associated 1-sigma from the estimator are the fitting data. In practice, each station has a calibration process where 1∼3 days of nominal code phase measurements are used to estimate RDCBs, and fit the model. Unless any of the hardware changes, such as new satellite service, new antenna, new receiver, or firmware update, the fitted models can last for several months.

Fig. 2.18 shows the nominal GPS UREs after applying the RDCB models, where the precise product was used as truth. The calibration greatly reduces the spread of the UREs from 1∼2 m to 20∼30 cm. The models are fitted with the RDCBs generated from 24 hours of observation data on January 24th, and are still valid for the data on April 15th as shown in the figure. This enables decimeter-level accuracy for monitoring any type and number of satellite faults using this integrity monitor framework.

2.2 Global Integrity Monitor

With the calibration of the RDCBs, the 56-station network can provide a direct analysis on the effect of satellite faults on users in the range domain. In this section, several actual GPS faults will be evaluated using the fault-tolerant filter. Since all of the GPS faults are satellite clock faults that have the same pseudorange error on users at any location, the UREs are all projected to JPLM. As the precise ephemeris and clock products would remove the satellite faults, for this section we use the broadcast orbit and clock information instead

2.2.1 GPS Clock Fault at PRN-22 on October 2nd, 2022

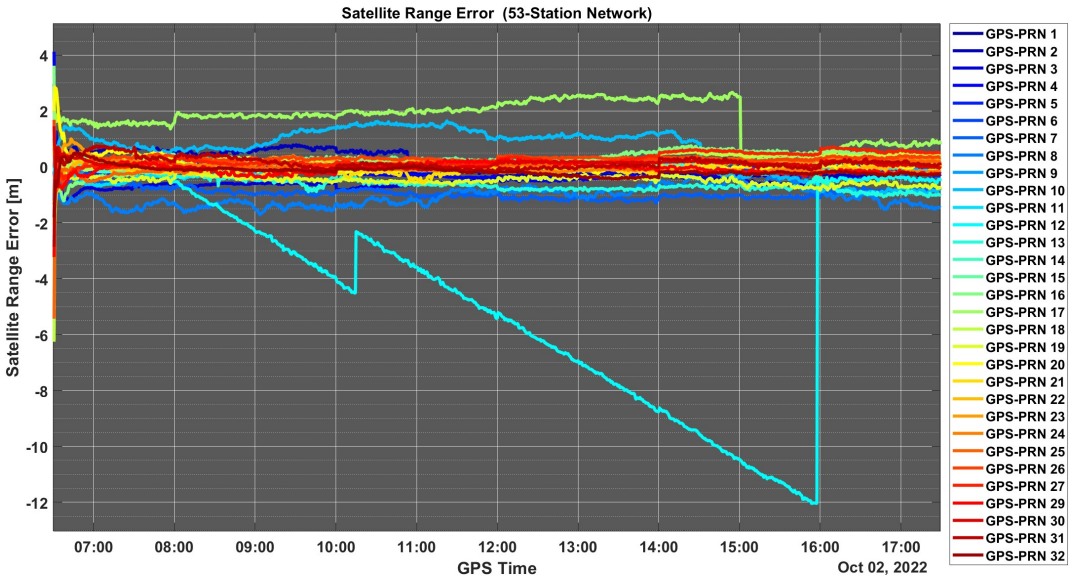


Figure 2.19: UREs on October 2nd, 2022 from multi-receiver network.

Fig. 2.19 shows the UREs of GPS during the clock fault at PRN-22 on October 2nd, 2022. This is a slow ramp started at 07:55:00 that reached -12 m at 16:00:00 before it was manually corrected by the ground segment. The small jumps of the UREs along the estimations are due to alternating with different Issue of Data Clock (IODC) from the broadcast message. Since the fault is examined with the broadcast satellite clock biases, the magnitude of the URE estimates are larger, but they still reflect the clear onset and offset of the ramp fault in the range domain.

2.2.2 GPS Clock Fault at PRN-1 on January 25th, 2023

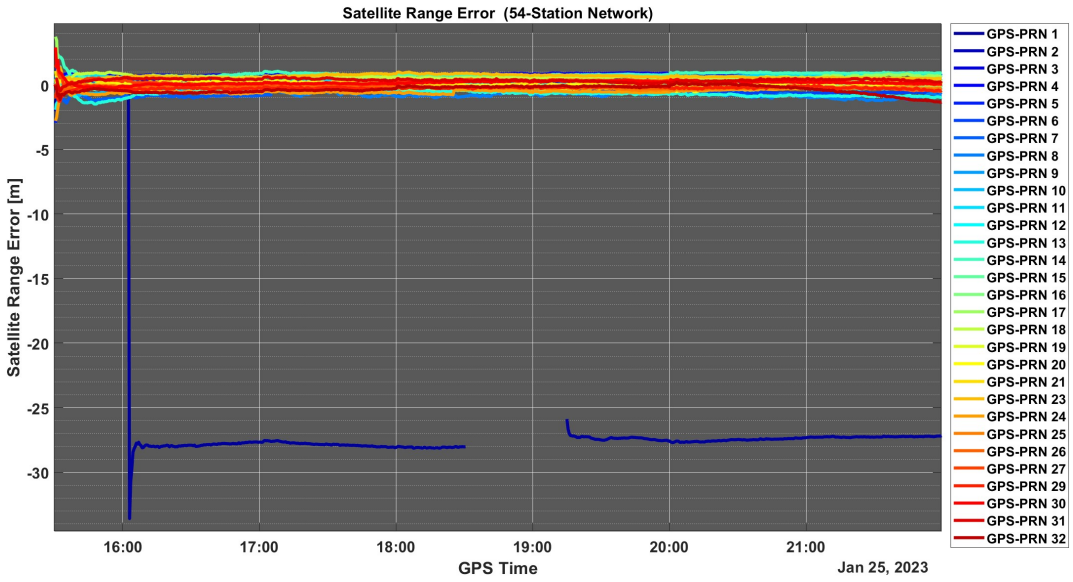


Figure 2.15: UREs on January 25th, 2023 from multi-receiver network.

Fig. 2.15 is the GPS clock fault at PRN-1 on January 25th that was shown and analyzed earlier in Sec. 2.1.2 and is shown again here for the sake of completeness. The URE estimates are more stable and accurate compared to only using a single receiver as shown in Fig. 2.8.

2.2.3 GPS Clock Fault at PRN-1 on July 10th, 2023

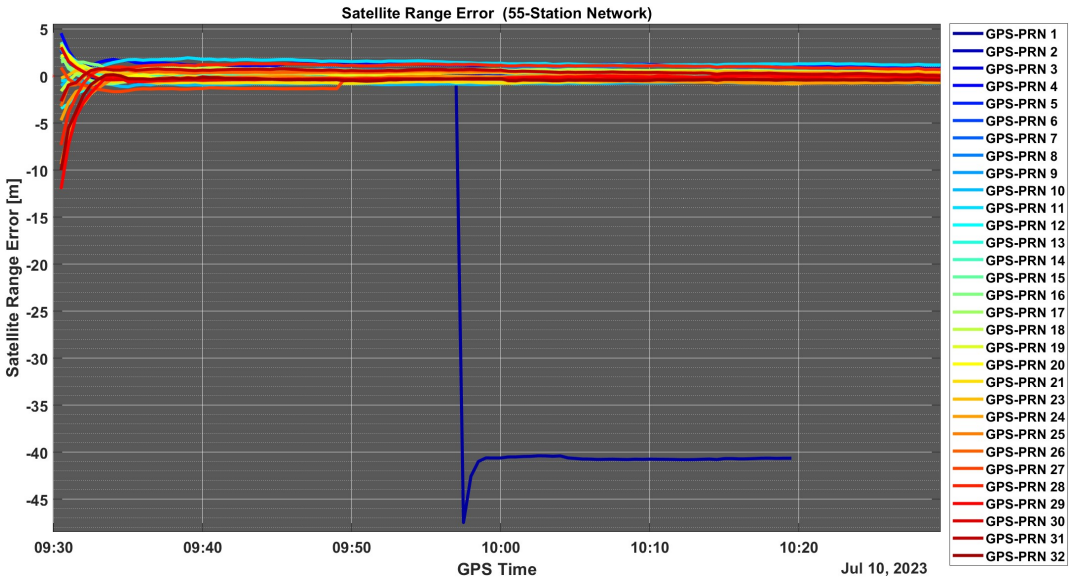


Figure 2.20: UREs on July 10th, 2023 from multi-receiver network.

Fig. 2.20 shows the UREs of GPS on July 10th, 2023. The URE of PRN-1 drops to -40 m at 09:57:30 and stays at this level until 10:19:30 before PRN-1 switched to non-standard code and was set unhealthy.

2.2.4 GPS Clock Fault at PNR-7 on May 25th, 2024

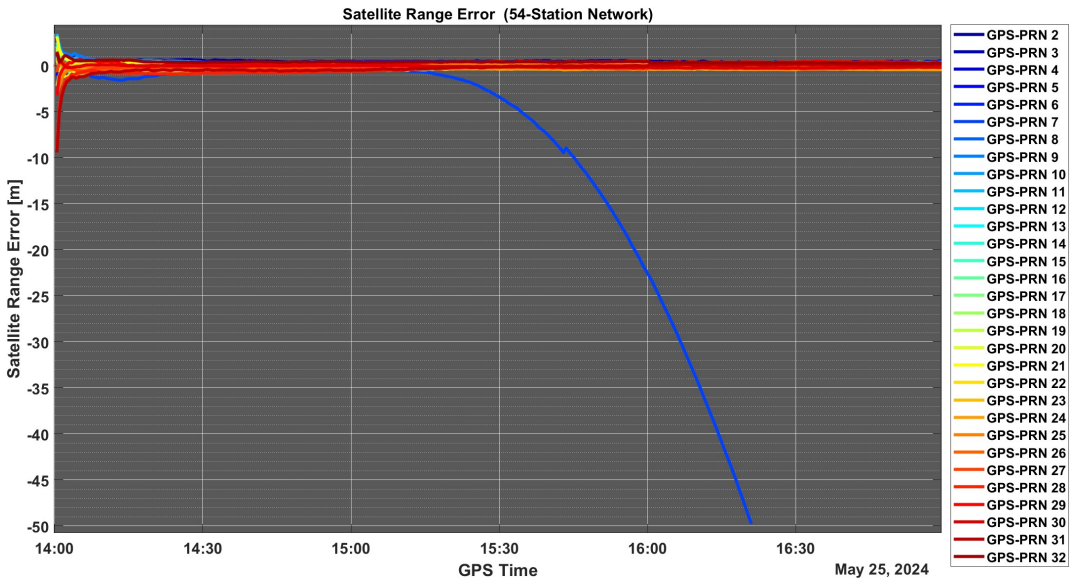


Figure 2.21: UREs on May 25th, 2023 from multi-receiver network.

Fig. 2.21 shows the UREs of GPS on May 25th, 2024. The fault is a quick ramp at PRN-7 started at 15:10:00 and reaching -50 m at 16:20:00 before switching to non-standard code and being set unhealthy.

2.3 Regional Integrity Monitor

The coverage area of the ground-based integrity monitor depends on the placement and number of the receivers within the network. The differences between global and regional networks are only the locations and total number of receivers. For the fault-tolerant filter, the operation remains the same with either a global or a regional network. However, the performance and continuity of the URE estimates differs during the LOS restrictions of the regional network.

There are 38 WAAS stations located in North America as shown in Fig. 2.22. This form an SBAS regional integrity monitor network for GPS L1 snapshot users.

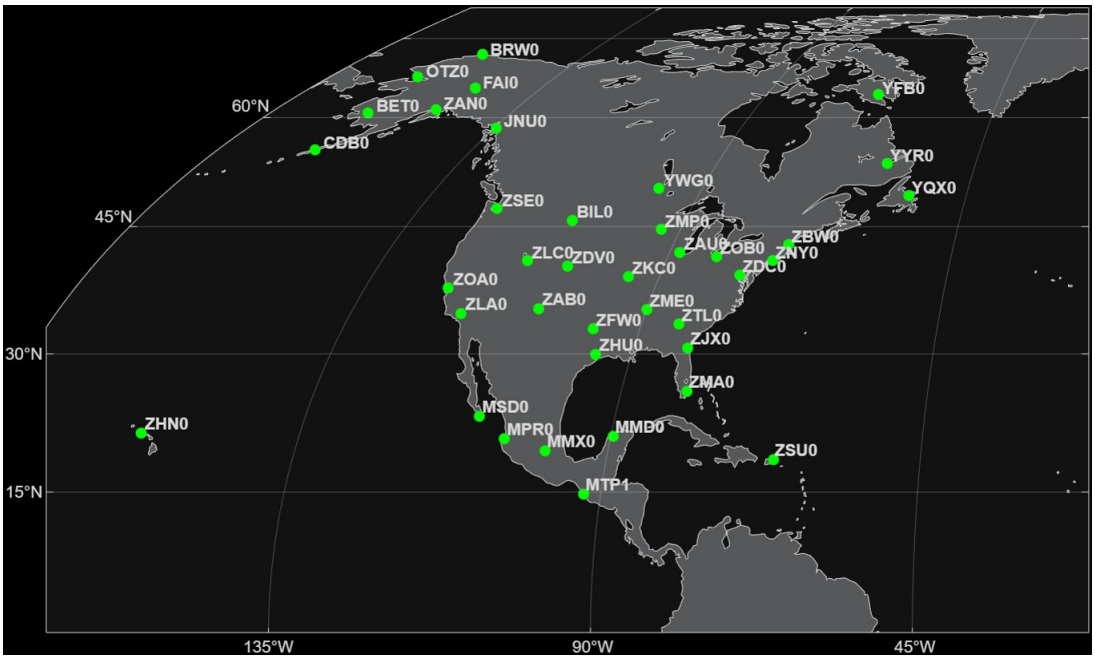


Figure 2.22: 38 WAAS stations.

2.3.1 Limitations of a Regional Network

The primary limitations of a regional network are the the number of LOS receivers and geometries to the satellites. Since the fault-tolerant filter has no prior on the satellite clock and ephemeris errors, the accuracy of the URE estimates relies on multiple receivers having LOS to the satellites. The ephemeris errors are especially reliant on having LOS from different directions of the receiver. A regional network, such as WAAS, might have insufficient geometries to achieve decimeter-level accuracy when the estimator assumes no priors on the ephemeris errors as compared with the global network.

Due to regional placement of the receivers, the network could not monitor the satellites without interruption by LOS of the satellites. Since the estimator relies on the carrier phase measurement to achieve decimeter-level accuracy and uncertainty to the UREs, it takes tens of minutes to converge the uncertainty of the floating ambiguities from the new satellites in the estimator. Multipath effects are also more prominent during the transition between LOS and NLOS, when all the pseudorange measurements are from lower elevation angles.

2.4 Summary

In summary, the fault-tolerant filter introduced in this chapter can achieve decimeter-level accuracy when estimating the UREs without constraining priors on satellite clock and ephemeris faults in a global monitor after convergence.

The fault-tolerant filter is essentially an estimator that has taken into account the potential existence of the satellite clock and ephemeris errors and are as part of the state estimates. Without constraining priors on the estimates of satellite clock and ephemeris errors, the filter can accurately estimate the impact of those errors as URE and URA, thus providing direct analysis on potential satellite faults. There are several variants of fault-tolerant filters introduced in this chapter.

Filter 2 introduced in Sec. 2.1.1 is the prototype of the fault-tolerant filter, which uses single receiver with dedicated error state corresponds to the faulted satellite. It

is basically doing self-monitoring by estimating URE of the dedicated faulty satellite using measurements from other satellites that are considered to be normal. It light-weight and quick to operate, but lacks geometric diversity and the estimated UREs are only valid within close proximity.

Filter 3.a and **3.b** introduced in Sec. 2.1.2 extended the idea of direct estimation of the URE into all the satellites in-view, so the filter can consistently monitor the UREs of the entire constellation. They differ only by the applied clocks and ephemerides, where the former uses precise and the latter uses broadcsat. In combining with a network of reference receivers, the URE estimates became more accurate and comprehensive. The geometric diversity of multiple receivers also allows the estimation of the ephemeris errors. The complete coverage of the network receivers provide un-interrupted estimates of the UREs induced from arbitrary types of satellite clock or ephemeris errors for the entire constellation. Thus, it can give global coverage of the integrity monitoring results with a network of reference receivers. This section also discussed the observability of the fault-tolerant filter.

Sec. 2.2 demonstrated the performance of such fault-tolerant filters with real data from actual GPS fault events using broadcast clocks and ephemerides with a global reference receiver network.

Sec. 2.3 discussed the global v.s. regional network, where they only differ by the placement of the receivers. The regional network may be not consistently estimate the ephemeris errors, so it might be necessary to include error models for the orbit kinematics if the fault-tolerant filter is to be applied to a regional network. The transition between LOS and NLOS conditions also introduces errors due to more pronounced multipath, which must be properly modeled within the regional network framework. Specifically, lower weights should be applied to measurements at lower elevation angle or lower SNR value.

WAAS has been successfully operated as an SBAS in North America since 2003, targeting users that only using carrier smoothed GPS L1 code measurements. However, the same 38 WAAS stations might not be adequate to provide integrity to precise positioning services without adjusting the priors used in the global network. In particular, the noise due to multipath and ephemeris errors with limited geometries

can cause difficulties when compared to a global network.

Chapter 3

User Integrity from the Monitor

Chap. 2 formulated the estimator that can directly estimate the satellite clock and ephemeris errors with non-constraining priors on the error models using a ground-based receiver network. The estimation results are the range errors. This chapter focuses on how the range errors affect the PPP users under the coverage of the global integrity monitor.

Filter 3.a (defined in Sec. 1.5) is used in this chapter for the integrity monitor result shown in Fig. 3.1, 3.2 and 3.3.

Filter 4 (defined in Sec. 1.5) is introduced in this chapter and is shown in Fig. 3.4 and 3.5c.

The PPP user, who is assumed to have a single receiver using the measurement models specified in Eq. 1.2 and 1.3, can attain integrity on its positioning solution with the support from the integrity information generated by the integrity monitor. Faults that the users can protect from depend on what type of faults the integrity monitor is monitoring. The fault-tolerant filter formulated in Chap. 2 can monitor the satellite clock and ephemeris faults without priors on the satellite error models with the global receiver network, so the user can also be protected from the threat of arbitrary satellite clock and ephemeris.

This chapter incorporates the user into the operation of the integrity monitor. The purpose is to demonstrate the integrity performance of a PPP user with the assistance of the integrity monitor developed in Chap. 2.

In practice, due to hardware, computational, and transmission delays, the integrity information inevitably contains latency between real-time satellite errors from the user receiver and the satellite errors estimated by the integrity monitor. Appendix.A describes the temporal modeling of the latency under Gaussian assumptions to be consistent with the estimator.

3.1 User Integrity and Performance

The UREs and their uncertainties (sigmas), $\beta_u^{(k)}$ and $\sigma_u^{(k)}$, obtained from projecting the estimates of the satellite clock and ephemeris errors onto user locations are defined as the integrity information for users to preserve their integrity.

Given that the fault-tolerant filter is invariant to the threat of the satellite clock and ephemeris faults, the users can simply apply the estimate of such satellite faults as part of the corrections in their measurement models. The user range errors induced by satellite faults are removed with the integrity information, $\beta_u^{(k)}$ and $\sigma_u^{(k)}$, thereby protecting the user positioning solution and ensuring its integrity.

The modified pseudorange models with $\beta_u^{(k)}$ as extra corrections are:

$$\tilde{\rho}_{\mathrm{IF},u}^{(k)} = \| \overrightarrow{x}_{sat}^{(k)} - \overrightarrow{x}_{pos,u} \| + \left(b_u - b_{sat}^{(k)} \right) + m_u^{(k)} \nabla \hat{T}_u + R_{m,\rho,u}^{(k)} + \beta_u^{(k)} + \epsilon_{\rho,u}^{(k)} \tag{3.1}$$

$$\tilde{\phi}_{\mathrm{IF},u}^{(k)} = \| \overrightarrow{x}_{sat}^{(k)} - \overrightarrow{x}_{pos,u} \| + \left(b_u - b_{sat}^{(k)} \right) + m_u^{(k)} \nabla \hat{T}_u + A_u^{(k)} + R_{m,\phi,u}^{(k)} + \beta_u^{(k)} + \epsilon_{\phi,u}^{(k)} \tag{3.2}$$

where $\beta_u^{(k)}$ is added to the code and carrier phase pseudorange as extra corrections.

Here, the UREs are assumed to be uncorrelated across satellites, and the measurement noise of the code and carrier phase with the UREs as corrections is:

$$R_{\tilde{\rho}_{\mathrm{IF},u}^{(k)}} = R_{\rho_{\mathrm{IF},u}^{(k)}} + {\sigma_u^{(k)}}^2 \tag{3.3}$$

$$R_{\tilde{\phi}_{\mathrm{IF},u}^{(k)}} = R_{\phi_{\mathrm{IF},u}^{(k)}} + {\sigma_u^{(k)}}^2 \tag{3.4}$$

where the measurement noise is inflated with the variance of the UREs, i.e., the square of the sigmas. This is the **Filter 4** mentioned in Sec. 2.1.

The rest of the computation for the user filter remains unchanged from a standard

PPP user. The only difference with the standard PPP user (**Filter 1** defined in Sec. 2.1) is the addition of the integrity information, $\{\beta_u^{(k)}, \sigma_u^{(k)}\}$, generated from the fault-tolerant filter in each of the LOS satellite measurements. By applying this information, the users acquire integrity on their positioning solution since the threats of the satellite clock and ephemeris faults are constantly monitored and corrected by the integrity monitor.

As a result, the protection levels of the user position solution are simply a constant multiple of the user position sigmas from the user filter:

$$PL_e = K \cdot \sigma_e \tag{3.5}$$

$$K = Q^{-1}\left(1 - \frac{PIR}{2}\right) \tag{3.6}$$

where σ_e is the 1-sigma uncertainty of position estimates from the user filter in the direction of e, e.g., x-y-z or East-North-Up. Q is the CDF of the normal distribution and PHMI stands for Probability of Hazardously Misleading Information (PHMI). PHMI is assumed to be 10^{-7} in the following tests.

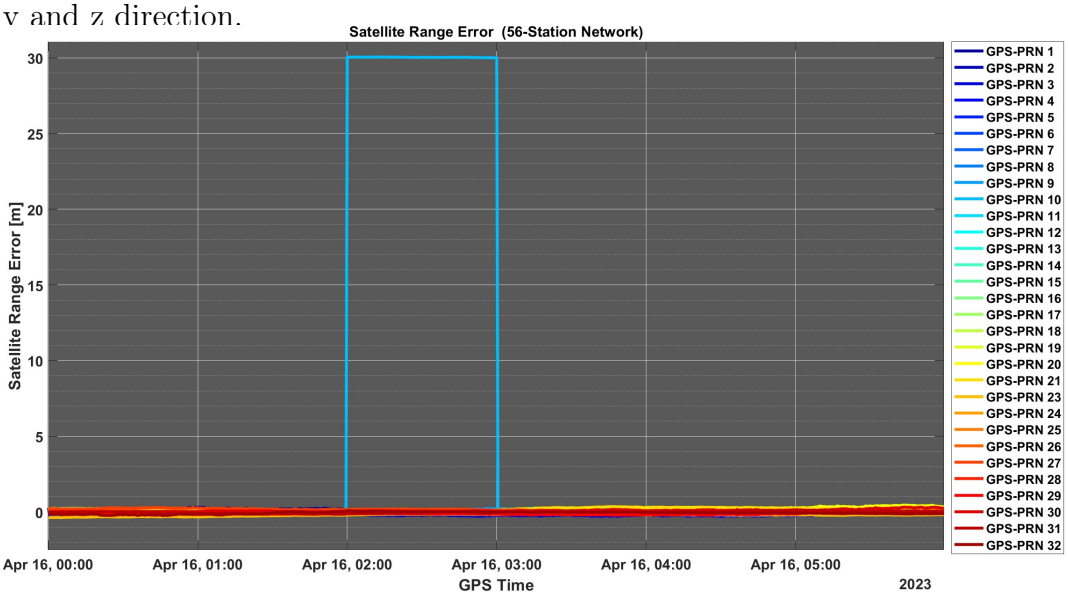
3.1.1 Dataset and Monitor Network

To demonstrate the performance of the PPP users with the support from the integrity information, the same 56-station global network introduced in Chap. 2 (Fig. 2.9) is adopted to generate the integrity information for a user at the JPLM station.

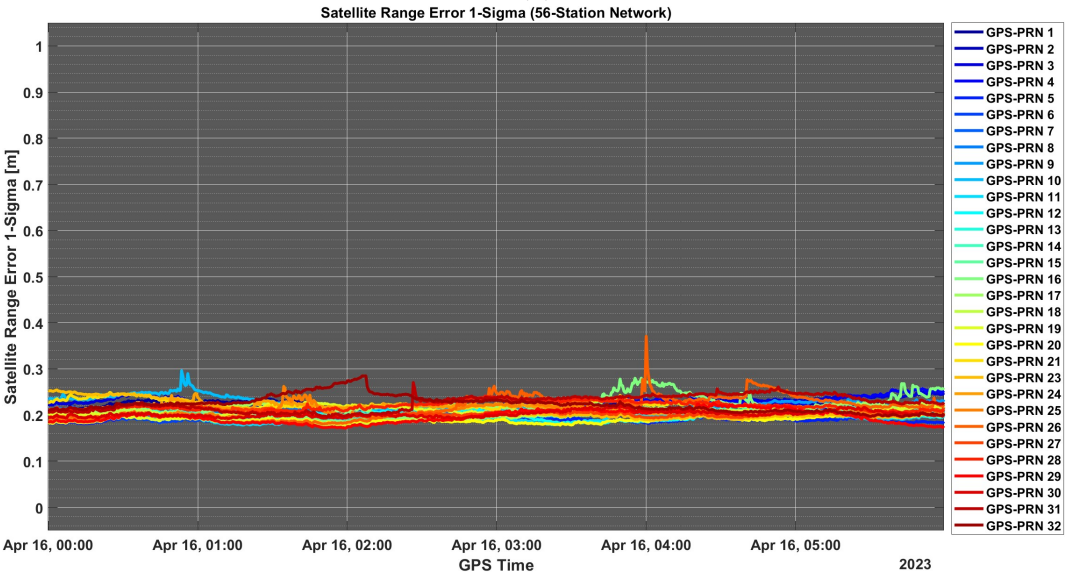
In this demonstration, GPS L1 and L2 code and carrier phase peudorange measurements collected on April 15th and 16th, 2023, from the 56 stations form the global integrity monitor that monitors the satellite clock and ephemeris errors in the range domain. This is the same global network used in Chap. 2. Twenty-four hours of observation data on April 15th is used to converge the fault-tolerant filter, and the experiments are conducted on the April 16th data.

3.1.2 Test Setup

Three types of artificial satellite faults are injected into the observation data of GPS-PRN-10 from 02:00:00 to 03:00:00. The faults are: 1.) 30 m step fault on satellite clock. 2.) $\frac{1}{120}$ m/s ramp fault on satellite clock. 3.) 30 m step offset on ephemeris in y and z direction.



(a) UREs



(b) Sigmas

Figure 3.1: UREs and Sigmas of 30 m Step Fault at JPLM.

Fig. 3.1 shows the UREs and their sigmas of the GPS constellation as an estimation result from the 56-station ground-based integrity monitor at JPLM from 00:00:00 $\sim$ 06:00:00. In Fig. 3.1a, the fault-tolerant filter captures the 30 m step fault injected into the clock of GPS-PRN-10. Fig. 3.1b shows the sigmas of UREs shown in Fig. 3.1a. The sigmas of the UREs are approximately 20 cm after a 24-hour convergence period completed on April 15th.

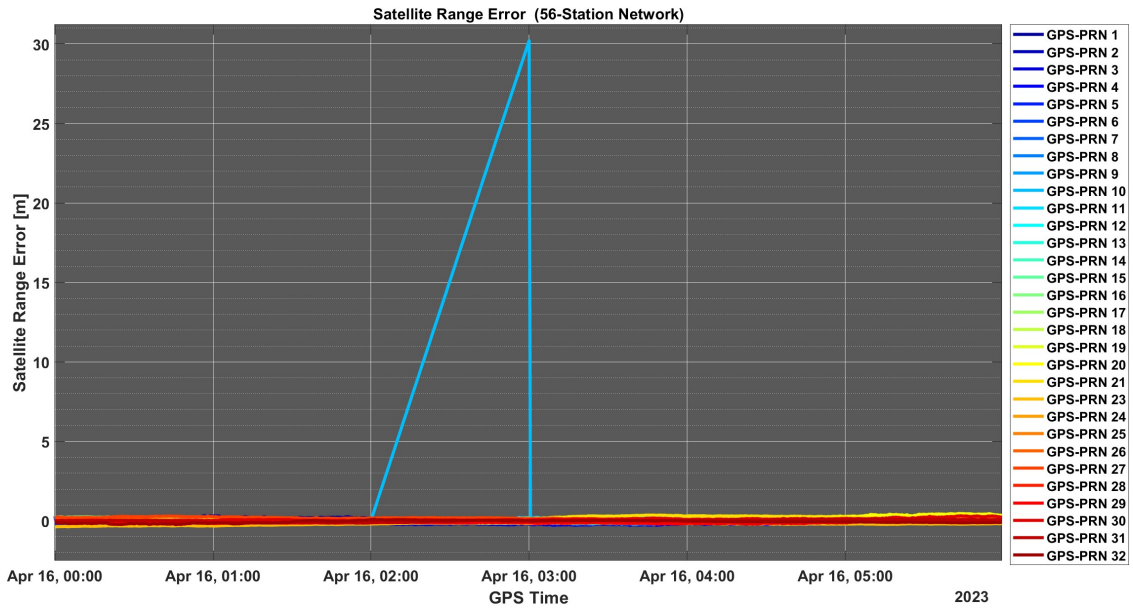


Figure 3.2: UREs of $\frac{1}{120}$ m/s ramp fault.

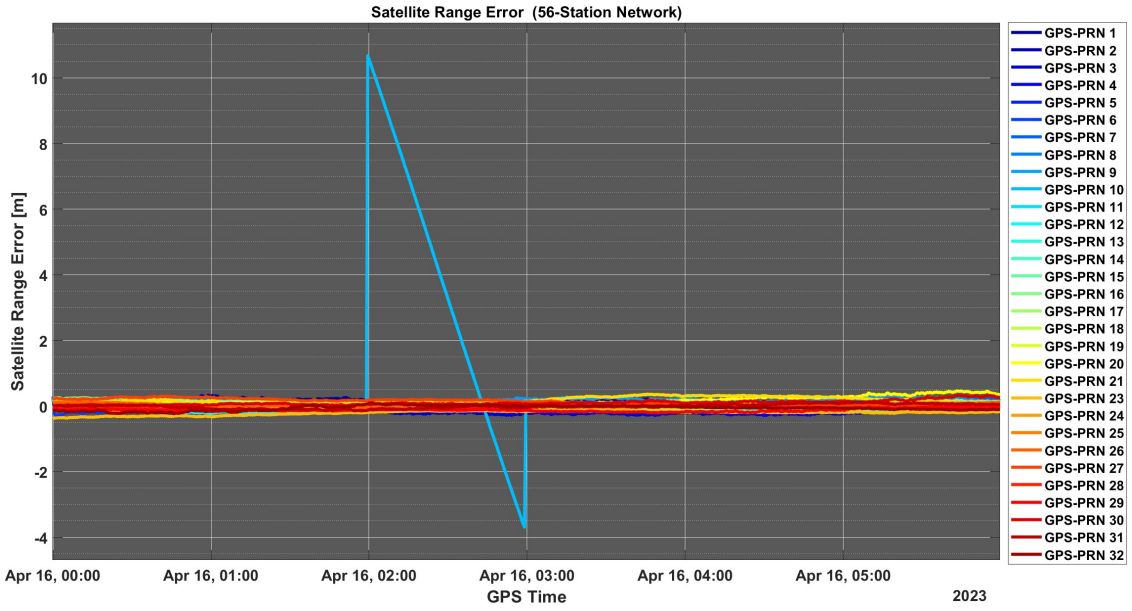
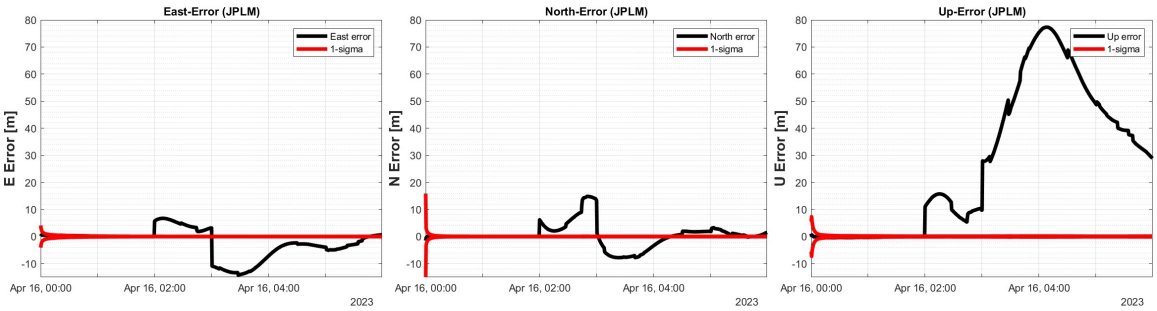


Figure 3.3: UREs of ephemeris faults with 30 m step offset on y and z directions.

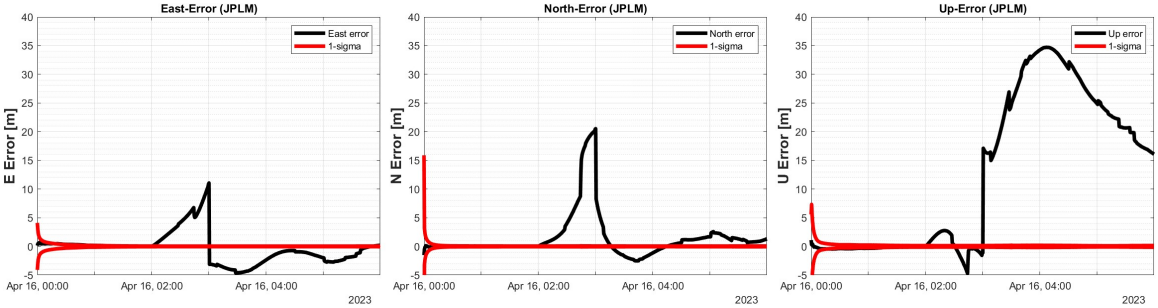
Similarly, Fig. 3.2 provides the UREs with $\frac{1}{120}$ m/s ramp fault injected to the clock of PRN-10, and Fig. 3.3 is the UREs with 30 m step fault injected to the ephemeris of PRN-10 in the y and z directions. The sigmas of the corresponding UREs are identical to those shown in 3.1b since the only difference between them are the artificial faults injected to the observation data, which are nearly independent of the covariance update in EKF.

Since satellite clock faults are spatially independent and users at any location would experience the same amount of pseudorange errors, the UREs of the clock faults are identical to their pseudorange injection as shown in Figs. 3.1a and 3.2. On the other hand, the ephemeris fault is dependent on the location of the user. As shown in Fig. 3.3, the 30 m step fault on the ephemeris of PRN-10 in the y and z directions is represented as a slightly curved ramp fault at the location of JPLM due to the orbital movement. The above UREs and sigmas are the estimation results of the artificial faults from the fault-tolerant filter. Users receive these estimates as integrity information, which they apply to their measurement models as additional corrections.

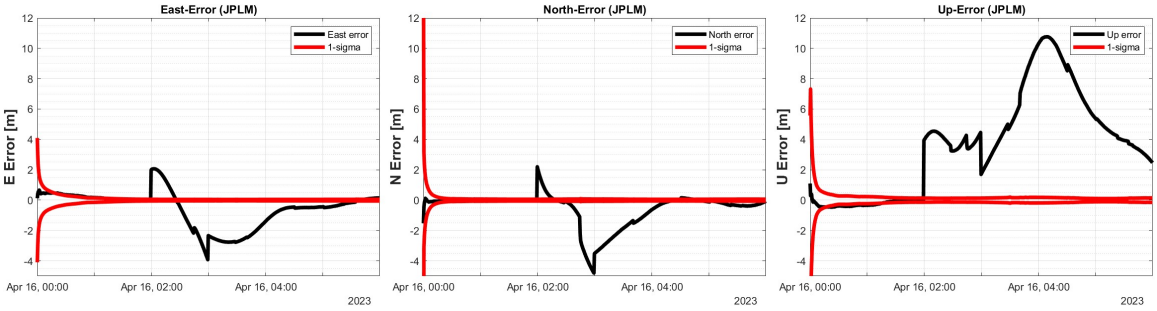
3.1.3 Test Results



(a) Position Errors and 1-sigma at JPLM during 30 m satellite clock step fault without integrity information.



(b) Position Errors and 1-sigma at JPLM during $\frac{1}{120}$ m/s satellite clock ramp fault without integrity information.

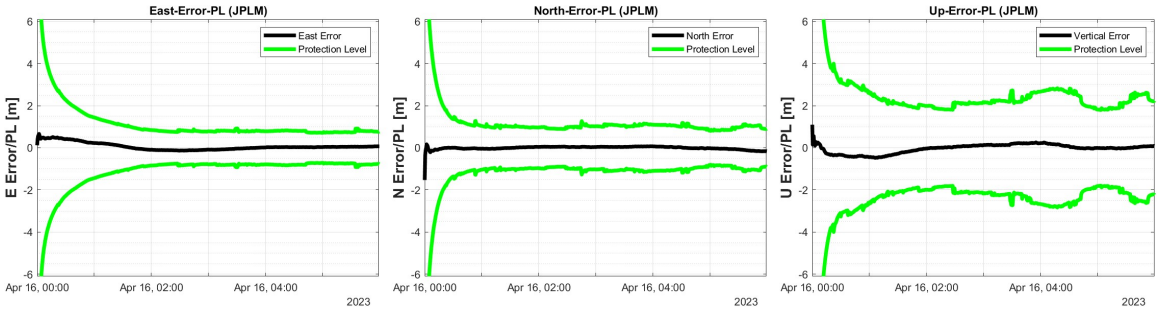


(c) Position Errors and 1-sigma at JPLM during 30 m y and z ephemeris step fault without integrity information.

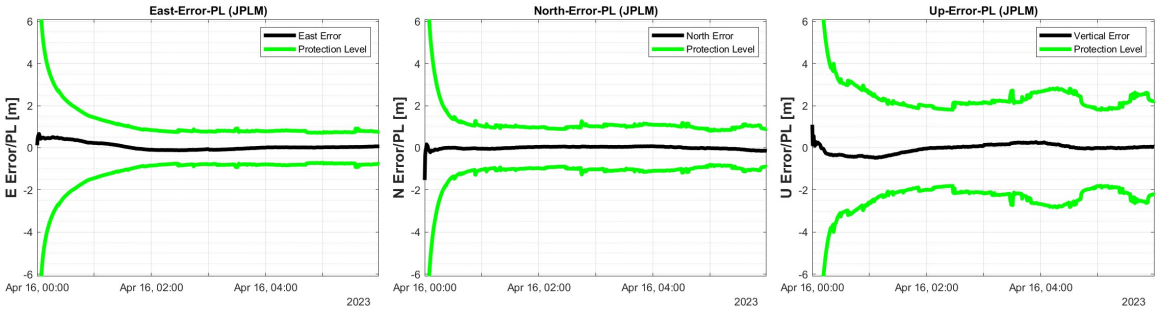
Figure 3.4: Position Errors and Sigmas at JPLM with integrity risk.

Fig. 3.4 shows the position errors of the JPLM station under the clock step, clock ramp, and ephemeris step faults as described previously. The user does not apply the integrity information and so without other consistency checks such as RAIM, the estimates of the user position are completely deteriorated due to the inability to

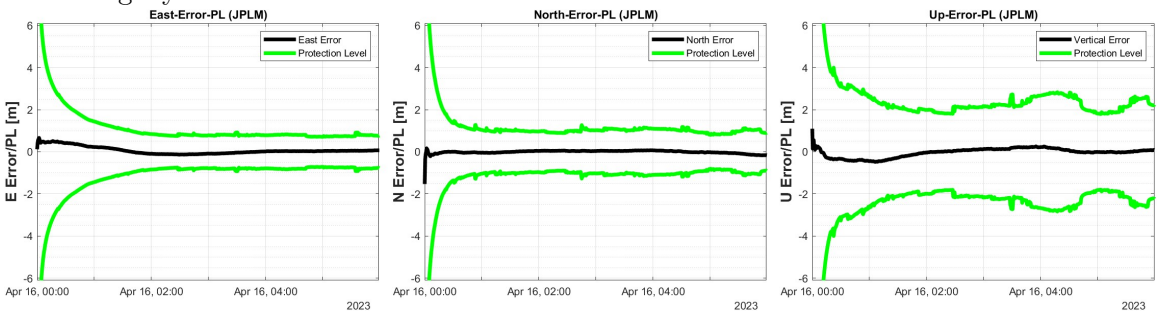
identify and exclude the faulted satellites. Each satellite fault induces position errors ranging from meters to tens of meters in each coordinate direction.



(a) Position Errors and protection levels at JPLM during 30 m satellite clock step fault with integrity information.



(b) Position Errors and protection levels at JPLM during $\frac{1}{120}$ m/s satellite clock ramp fault with integrity information.



(c) Position Errors and protection levels at JPLM during 30 m y and z ephemeris step fault with integrity information.

Figure 3.5: Position Errors and protection levels at JPLM with integrity information and zero latency.

Fig. 3.5 shows the position errors of the JPLM station under the satellite faults but with the integrity information as extra corrections to both the code and carrier

phase measurements (Figs. 3.1, 3.2 and 3.3). The user is applying the integrity information with zero latency, meaning that the user is directly correcting the UREs in the measurement models. As can be seen in all three cases, the user has their horizontal errors under 10 cm and vertical errors under 30 cm during the period of satellite faults. The sigmas of the horizontal errors are inflated to 15∼20 cm and 30∼40 cm for the vertical errors as a result of applying integrity information, which correspond to ∼1 m of horizontal protection level and ∼2 m of vertical protection level. The user exhibits almost identical performance regardless of what type of satellite faults are injected. This indicates that for integrity information with zero latency, the user is completely free from the satellite clock and ephemeris faults since the integrity monitor is actively correcting them with real-time estimation, and the positioning performance of the user is only dependent on the accuracy of the integrity information. This is the nominal performance for the user with integrity guarantee from the monitor.

3.2 Summary

The estimation results shown in the tests demonstrated the performance of the PPP users with the assistance of the integrity monitor in an ideal operational scenarios where no latency is introduced. The users simple takes in integrity information from the integrity monitor as additional corrections. The UREs are the pseudorange corrections and the URAs inflates the nominal measurement noise.

As shown in Appendix. A, when the latency of the integrity information is included, the users might still need to exclude satellites from the solution until a valid correction can be supplied again. This is especially true when forward predicting from past data.

Another way of preserving the integrity of the user position under the influence of latency is only performing estimation after receiving the integrity information, where the temporal modeling of the integrity information can be prevented. Therefore, the positioning solutions from the GPS signals would not be real-time and would instead a lag behind with the latency of the integrity information. Although such positioning

solution only reflects the user position in the past, it is guaranteed to have integrity. In combination with other sensors such as an Inertial Measurement Unit (IMU), the user can still have a positioning solution in real-time with integrity.

The user integrity is protected given the satellite errors are monitored and corrected from the ground-based monitor. However, the user side errors, such as multi-path, are local to the user environment and are not observable from the reference receivers. Thus, for users operating in complex area, such as urban canyon, additional measures need to be implemented locally to provide integrity to the user.

Chapter 4

Network Integrity of the Monitor

The fault-tolerant filter alone is capable of monitoring the satellite clock and ephemeris faults without the need to do solution separation on satellites, given that it already assumes all satellites could have satellite clock and ephemeris errors, $\zeta^{(k)}$ and $\Delta\overrightarrow{\gamma}^{(k)}$. However, **Filter 3.a** (defined in Sec. 1.5) described in Sec. 2.1.2 relies on all the receivers providing correct measurements and functioning properly. If any of the receivers offer faulty measurements without being excluded, regardless of which satellites, the least-square minimizing nature of the EKF would force all the states to assimilate with the faulty measurements. Depending on how inconsistent the measurements are, the entire filter might potentially be corrupted. Receiver failures pose non-negligible threats to the integrity of the network itself and must be considered in the design of any integrity monitor providing integrity to applications with safety-critical requirements.

This chapter develops **Filter 3.c** (defined in Sec. 1.5) that takes into the consideration of the receiver failure within the network.

4.1 Receiver-wise Solution Separation Statistics

The receiver failures are mitigated using Solution Separation (SS) statistics, with receiver failures treated as the underlying fault modes. SS is the standard consistency check technique used in RAIM for individual users. The process involves forming

sets of estimators, each consisting independent combinations of the measurements with corresponding exclusions on the fault modes that SS monitors. The fault mode is detected when the estimation results from each estimator do not agree. For instance, take a RAIM user with measurements from eight satellites, who is protecting against satellite faults. If the fault mode is 1-out, meaning at most one satellite can fail at a time, the user can form a total of $1 + \binom{8}{1} = 9$ solutions from the nine independent estimators. One solution uses measurements from all eight satellites, and eight of other solutions each have different exclusions on the satellites and only use measurements from seven of the satellites. The solution that utilizes all eight satellites is referred to as the All-In-View (AIV) solution, and the solutions that only use part of the satellites are referred to as subset solutions. By comparing the difference between the AIV and subset solutions, the statistics can detect the outliers, i.e., the satellite fault modes, with confidence specified by the desired probability when the errors are bounded by the normal distributions. The estimators for the AIV and subset solutions are independent and can therefore be implemented in parallel.

In the case of the integrity monitor, since the satellite faults have been included as part of the models, they do not pose a threat to the system. As a result, instead of the satellite faults, the fault modes considered here are the receiver failures, and are detected in the range domain. In this section, the receiver fault modes are included into the consideration of the integrity monitor, and a Fault Detection and Exclusion (FDE) criteria is derived in detail under Gaussian assumptions.

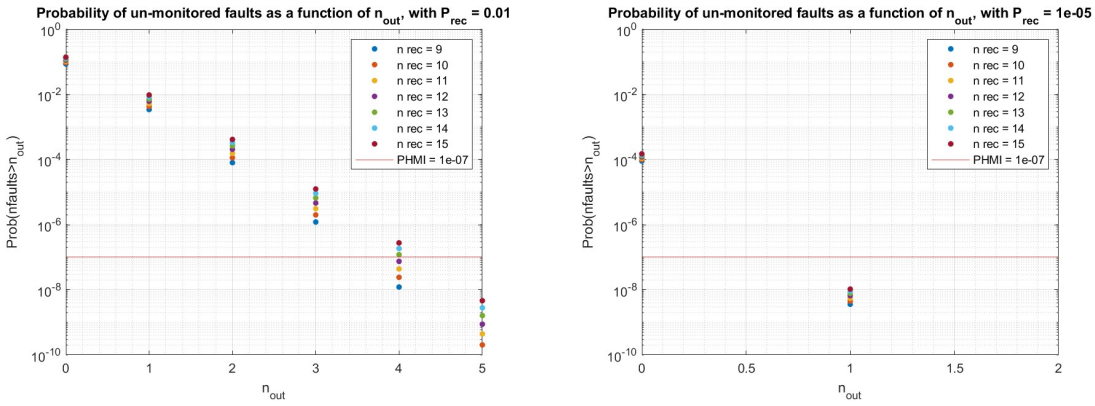
4.1.1 Probability of Receiver Failure

The number of receiver exclusions needed to be carried out in the subset solutions is determined by the performance requirement of the monitoring network. Assume that the receiver failure occurs with the same probability, P_{rec}, and is Independent and Identically Distributed (I.I.D.). The probability that the number of receiver failures exceeds the number of receivers excluded in the subset solution can be modeled using

a binomial distribution.

$$P\left(n_{\text{fail}} > n_{out}\right) = 1 - \sum_{y=0}^{n_{out}} \binom{n_{rec}}{y} p_{rec}^{y} \left(1 - p_{rec}\right)^{n_{rec}-y} \tag{4.1}$$

where n_{out} is the number of receivers to be excluded in the subset solutions, n_{fail} is the number of receiver failures, and n_{rec} is the total number of receivers in the monitoring network. Eq.(4.1) simply states the probability of having more receiver failures than the number of receivers out in the subset, which is the case where the SS would fail to detect the receiver fault. It can determine the set of receiver faults that need to be explicitly monitored (single and double faults, for example) to satisfy the requirement of network integrity.



(a) Probability of Receiver Fault = 0.01 (b) Probability of Receiver Fault = 0.00001

Figure 4.1: Probability of Number of Receiver Faults Greater than Number of Receiver Exclusions.

The above figures show the probability of having more receiver faults than the number of exclusions. From Fig. 4.1a we can see when the receiver fault rate is 10^{-2}, the network needs to perform at least 4-out solution separation for the network size of 9∼12 receivers, and 5-out for the size of 13∼15. Similarly, when the receiver fault rate is 10^{-5}, Fig. 4.1b shows that only 1-out subset is required to have $P\left(n_{\text{fault}} > z\right)$ below the desired PIR. The 10^{-5} prior bound will need to be verified.

4.1.2 Receiver Fault Detection

The SS statistics lead to the formation of subset solutions within the monitoring network. Since the fault modes to be monitored are the measurements at the receiver level, a bank of fault-tolerant filters that are a subset of the AIV filter are formed and operated in parallel. Each subset filter has a different receiver exclusion from the network, such that they are tolerant to the corresponding receiver failure. The formulation is similar to ARAIM in PPP user integrity [54], for which the exclusion is at the satellite level, whereas here the exclusion is at the receiver level for the filter. Thus, at each projected location, a list of estimates in the range domain from the AIV and subset filters are formed.

$$(\beta_{u,i}^{(k)}, \sigma_{u,i}^{(k)}) \mid i \in \left\{ 0, 1, ..., \binom{n_{\text{rec}}}{n_{\text{out}}} \right\} \tag{4.2}$$

where i is the index of all the subset filters. By convention, $i = 0$ represents the AIV set, and $i \in \{1, \binom{n_{\text{rec}}}{n_{\text{out}}}\}$ are the indices of subsets with different receiver exclusions.

The exclusion subset for detection is defined as follows:

$$\mathbb{I} := \left\{ 1, ..., \binom{n_{\text{rec}}}{n_{\text{out}}} \right\} \tag{4.3}$$

where $\mathbb{I}$ contains all the subset combinations that correspond to the fault modes to be detected for the receiver network.

For solution separation statistics, the detection of the fault modes relies on evaluating the differences between the AIV and the subset estimates. The presence of receiver failures is determined with the SS test as introduced in Eq. 1.32:

$$\left| \hat{x}_{u,0}^{(k)} - \hat{x}_{u,i}^{(k)} \right| < T, \ \forall i \in \mathbb{I} \tag{4.4}$$

where $\hat{x}_{u,i}^{(k)}$ represents the estimates of URE, $\beta_{u,i}^{(k)}$, and T is the threshold for the solution separation test. If a satellite fails the test in a subset, it indicates that a receiver fault mode exists within the network and an exclusion must be performed to preserve the integrity of the monitoring system itself.

The choice of the threshold, T, is dependent on the required continuity. From the view point of finding outliers using SS, a consistent way of determining the threshold is through the normalized statistics.

$$\Sigma_{ss} = \Sigma_i - \Sigma_0 \tag{4.5}$$

This is exact representation of Σ_{ss} for the snapshot estimator derived in [55]. Although the variations between Σ_0 and Σ_i exist, it is not uncommon to observe in the estimation results that the statistics are normalized with near zero value of σ_{ss} that is close to numerical singularity. For example, σ_{ss} may approach as small as $\sim 10^{-9}$, where the numerical errors are amplified and disrupt the operation. For this reason, an approximated upper bound of Σ_{ss}, i.e., the subset covariance, Σ_i, is adopted to evaluate the statistics to avoid numerical issues.

$$\Sigma_{ss} \preceq \Sigma_i \tag{4.6}$$

Given that all the covariance matrices are positive definite, Σ_i naturally bounds Σ_{ss} from Eq. 4.5. This ensures that when evaluating the statistics with the selected threshold, the normalized statistics do not overestimate the discrepancies between the AIV and subset. This introduces conservatism into the threshold test but allows continuous operation of the SS check. The conservatism is also represented by the resulting upper bounds in the range domain, that is, $L_u^{(k)}$ from Eq. 1.42 is inflated accordingly.

By definition, any linear projection, $\overrightarrow{a}$, satisfies:

$$\overrightarrow{a}^T \Sigma_{ss} \overrightarrow{a} \leq \overrightarrow{a}^T \Sigma_i \overrightarrow{a} \tag{4.7}$$

let $\overrightarrow{a}$ be the linear projection of the satellite clock and ephemeris errors to range domain, Eq. 4.7 becomes:

$$\overrightarrow{a}^T \Sigma_{ss} \overrightarrow{a} \leq \overrightarrow{a}^T \Sigma_i \overrightarrow{a} \Rightarrow {\sigma_{u,i,ss}^{(k)}}^2 \leq \sigma_{u,i}^{(k)^2} \Rightarrow \sigma_{u,i,ss}^{(k)} \leq \sigma_{u,i}^{(k)} \tag{4.8}$$

where $\sigma_{u,ss}^{(k)}$ is the sigma of the solution separation statistics between the AIV and subset i from the kth satellite at user u, and $\sigma_{u,i}^{(k)}$ is the sigma of the URE estimate from the subset filter.

The threshold can then be determined for the corresponding SS test as:

$$T = \sigma_{u,i}^{(k)} \cdot K \tag{4.9}$$

where K is the number of sigmas that defines the outliers in the statistics.

The condition for $\sigma_{u,i}^{(k)}$ to be used is that the AIV and subset filters must strictly have identical priors, initial noise, process noise, weightings, etc, with the only discrepancy being the measurement exclusions.

4.1.3 Receiver Fault Exclusion

In order to be consistent with the detection result, the exclusion of the fault modes also adopts the solution separation statistics. Thus, an extra number of exclusions is required to identify which fault mode results in the failure of the threshold tests. The extra exclusions can form a sub-subset for the detection set:

$$\mathbb{J}_i := \left\{ 1, ..., \binom{n_{\text{rec,i}}}{1} \right\} \tag{4.10}$$

And the threshold test for the fault exclusions:

$$\left| \hat{x}_{u,i}^{(k)} - \hat{x}_{u,ij}^{(k)} \right| < T, \ \forall i \in \mathbb{I}, \ j \in \mathbb{J}_i \tag{4.11}$$

where $n_{\text{rec,i}}$ is the number of receivers in subset i after exclusion for detection, and $\mathbb{J}_i$ corresponds to the sub-subset of the subset that is used for fault mode detection.

The same solution separation threshold tests are performed for each subset. The subset that passes all the tests and has the lowest statistics corresponds to the most

likely fault mode exclusion. The fault exclusion can be expressed as:

$$\hat{i} := \arg\min_{i} \left\{ \max_{\left|\hat{x}_{u,i}^{(k)} - \hat{x}_{u,ij}^{(k)}\right|} \left\{ \forall j \in \mathbb{J}_i \cap \left| \hat{x}_{u,i}^{(k)} - \hat{x}_{u,ij}^{(k)} \right| < T \right\} \mid \forall i \in \mathbb{I} \right\} \tag{4.12}$$

where $\hat{i}$ is the subset with the most likely exclusion on the receiver fault modes detected by the solution separation statistics from the kth satellite at the position u. It can be seen as the new AIV solution to the integrity monitor, where the faulted receivers are excluded. The subset with the faulted receivers should be handled properly, e.g., reset the filters, until they can re-join the integrity monitor and the computation of the integrity information. If no fault modes are detected, the AIV set remains unchanged, i.e.

$$i_{\text{AIV}} := \begin{cases} 0, & \text{if no fault detected} \\ \hat{i}, & \text{fault detected and excluded} \end{cases} \tag{4.13}$$

and the subset:

$$\mathbb{I}_{\text{AIV}} := \begin{cases} \mathbb{I}, & \text{if no fault detected} \\ \mathbb{J}_{\hat{i}}, & \text{fault detected and excluded} \end{cases} \tag{4.14}$$

4.1.4 Robust Integrity Information

The detection and exclusion scheme described in Sections. 4.1.2 and 4.1.3 relies on the solution separation statistics to successfully identify the inconsistency between the AIV and subset estimates. The failure to detect the potential fault modes results in Integrity Risk (IR) to the entire system. It is required that the Probability of IR (PIR) be bounded such that the detection results are trustworthy.

$L_u^{(k)}$ is the upper bound to the UREs that bounds the probability of the receiver failures, PIR. The computation of such upper bound is derived in Eq.1.42.

$$L \leq \max_{\forall i \in \mathbb{I}} \left(T + Q^{-1} \left(\frac{PIR}{2 \cdot N_{\mathbb{I}} \cdot P_{fault}} \right) \cdot \sigma_i \right) \tag{1.42}$$

where in the context of receiver failures, P_{fault} represents the probability of receiver failures.

These upper bounds can be expressed in terms of the sigma of a normal distribution, facilitating efficient transmission and interpretation by users, under the assumption that the user error is dominated by the corresponding range error. The sigma is approximated as follows.

$$\begin{aligned} \sup_{\tilde{\sigma}_u^{(k)}} . \quad & \tilde{\sigma}_u^{(k)} \\ \text{s.t.} \quad & Q\left(\frac{L_u^{(k)} - \hat{x}_{u,i_{\mathrm{AIV}}}^{(k)}}{\tilde{\sigma}_u^{(k)}} \right) \leq \frac{PIR}{2} \end{aligned} \tag{4.15}$$

where $\tilde{\sigma}_u^{(k)}$ may be treated by the user as a sigma of a normal distribution.

Finally, the integrity information that is robust against receiver failure for range errors is defined as follows:

$$(\beta_{u,i_{\mathrm{AIV}}}^{(k)}, \tilde{\sigma}_u^{(k)}) \tag{4.16}$$

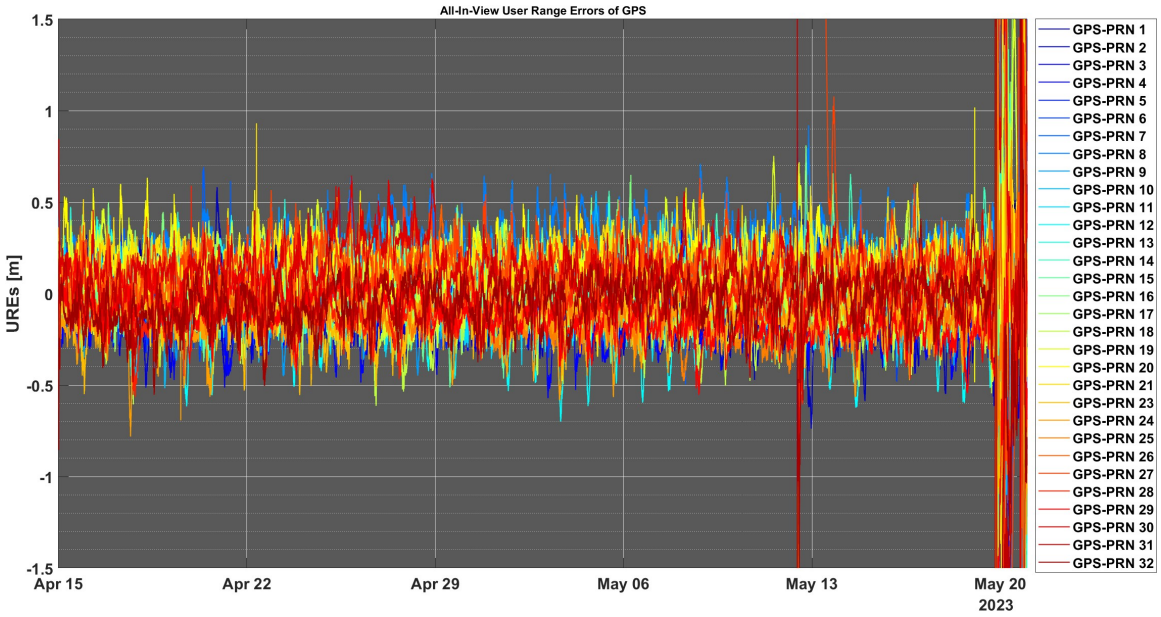
where $\beta_{u,\hat{i}}^{(k)}$ is the URE from the kth satellite at the location u and $\tilde{\sigma}_u^{(k)}$ is the approximated sigma. Both are the results of receiver fault detection and exclusion from solution separation statistics that ensure that the PIR of the receiver fault modes are bounded, given that priors correctly bound the nominal errors.

4.2 Evaluation Using Global Data

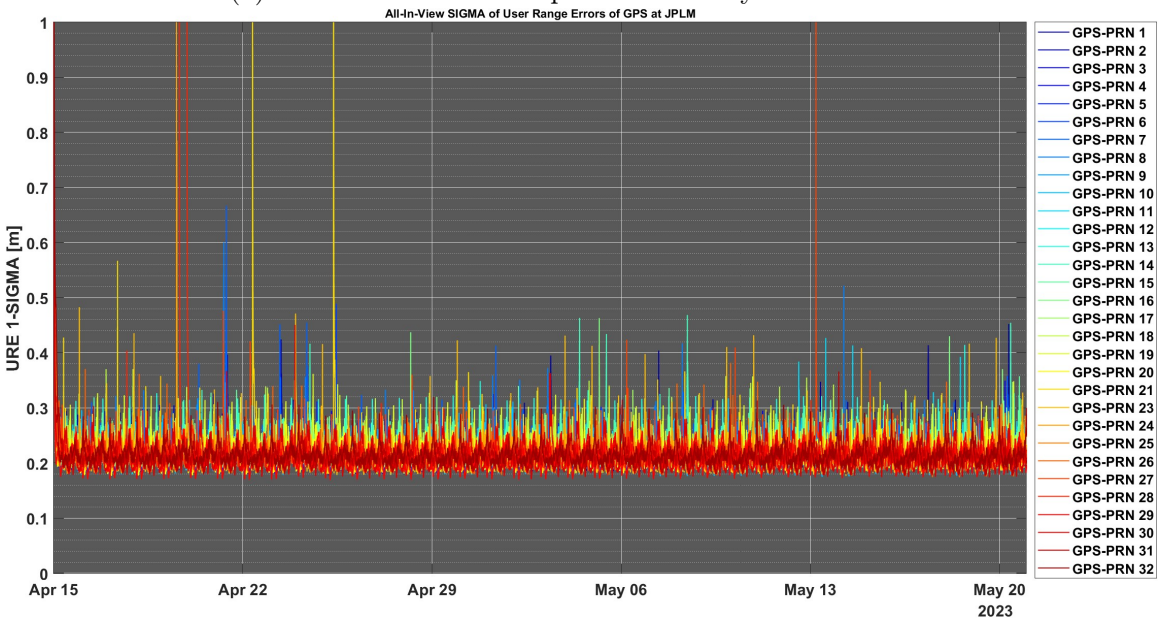
Using the network introduced in Fig. 2.9, the operation of the proposed integrity monitor is demonstrated. A total of 36 days of observations from April 15th to May 20th in 2023 are evaluated. A total of 36 days of observations from April 15th to May

20th in 2023 was evaluated. Within the span of the 36 days, a receiver failure was discovered at the MIKL station in Mykolaiv, Ukraine. This fault posed a potential threat to the integrity monitor system, and was later confirmed to be a spoofing event.

4.2.1 AIV UREs and Sigmas



(a) UREs of GPS from April 15th to May 20th in 2023..



(b) Sigmas of UREs from April 15th to May 20th in 2023..

Figure 4.2: Estimation results of the UREs from the 56-station network with data from April 15th to May 20th in 2023.

Fig. 4.2a shows the URE estimates for the 31 GPS satellites. It can be seen there are large URE jumps on May 12th, 2023 at 10:59:00, May 19th, 2023 at 19:27:00, and May 20th, 2023 at 17:45:00, where the UREs of several satellites jump to as high as hundreds of meters of errors while the sigmas (i.e., the uncertainty) stay as low as a few decimeters. From the perspective of post-processing, it is known that no such GPS failures occurred during those periods of time, so it was receiver failures that caused such drastic changes in the URE estimates. However, for an integrity monitor system in real-time operation, it is also possible for it to declare a constellation fault to GPS, given that the URE estimates of all the GPS satellites are extremely large. This is where the integrity monitor system fails to deliver integrity to the users. In the next section, it is demonstrated how the detection and exclusion process described in Sec. 4.1 would protect the user against such a fault.

For URE estimates on normal days, the overall range is from -0.5 m $\sim$ 0.5 m, with some occasions where certain satellites have higher UREs due to insufficient geometry and/or cycle slips. These occasions are also represented by the estimates of the sigmas (i.e., the uncertainty) of the UREs in Fig. 4.2b. In Fig. 4.2b, the sigmas of the UREs range from 0.2 m $\sim$ 0.25 m when satellites are monitored with enough receivers to provide sufficient geometry. There are a few instances where the sigmas of some GPS satellites rise. This is caused by temporary degradation of ambiguity re-convergence from cycle slips.

Due to the lack of receivers physically present in remote areas such as the South Atlantic and South Pacific, the fault-tolerant filter inevitably having fewer stations monitoring the satellites that travel across those regions. Another cause of the loss of accuracy is the occurrence of cycle slips. If multiple receivers declare cycle slips to the same satellite simultaneously, the filter temporarily loses range accuracy until the convergence of the ambiguity states. This phenomenon can only be mitigated by having enough receivers in the network. These are the main causes of the instances of less accurate results of the fault-tolerant filter. Nevertheless, these occasions are reflected by high sigmas as shown in Fig. 4.2b, so the integrity of the filter itself is not at risk. The fault-tolerant filter is properly reflecting the estimation results of the UREs and sigmas from the available observations.

4.2.2 Detection of Receiver Failure

In this example, receiver failure is assumed to occur independently at a fault rate of 10^{-5}, i.e., $P_{rec} = 10^{-5}$. The PIR is assumed to be 10^{-7} and is sufficient to monitor fault modes of at most one receiver failing at a time. Therefore, the number of receivers out in the detection subset is one, i.e., $n_{out} = 1$, and 2 for the exclusion subset. In the case of the example network, there are $\binom{56}{1} = 56$ detection subsets and $\binom{56}{2} = 1540$ exclusion subsets. Including the AIV set, a total of $1 + 56 + 1540 = 1597$ solutions are required for the full operation of an integrity monitor that is robust against receiver faults.

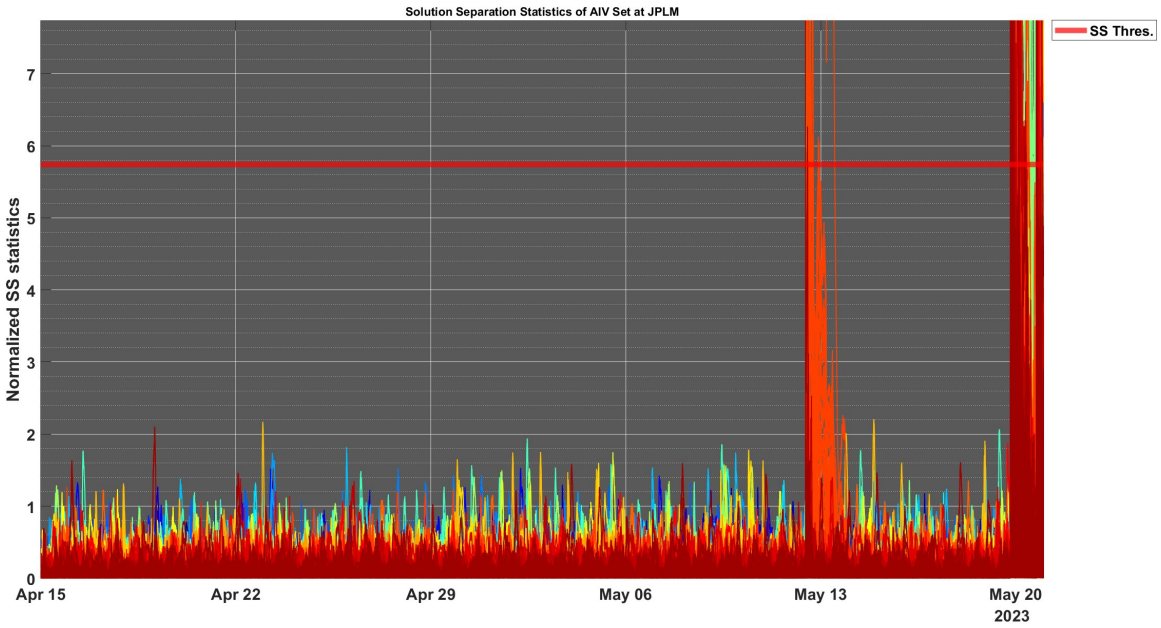
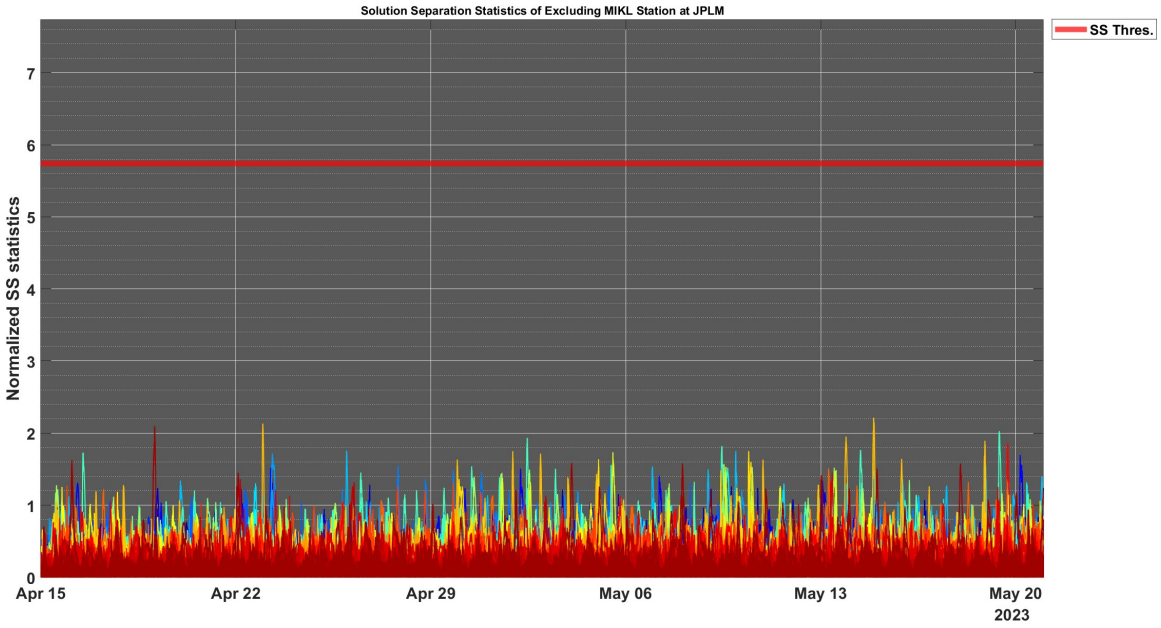


Figure 4.3: SS statistics of AIV set with 5.74 sigma threshold.

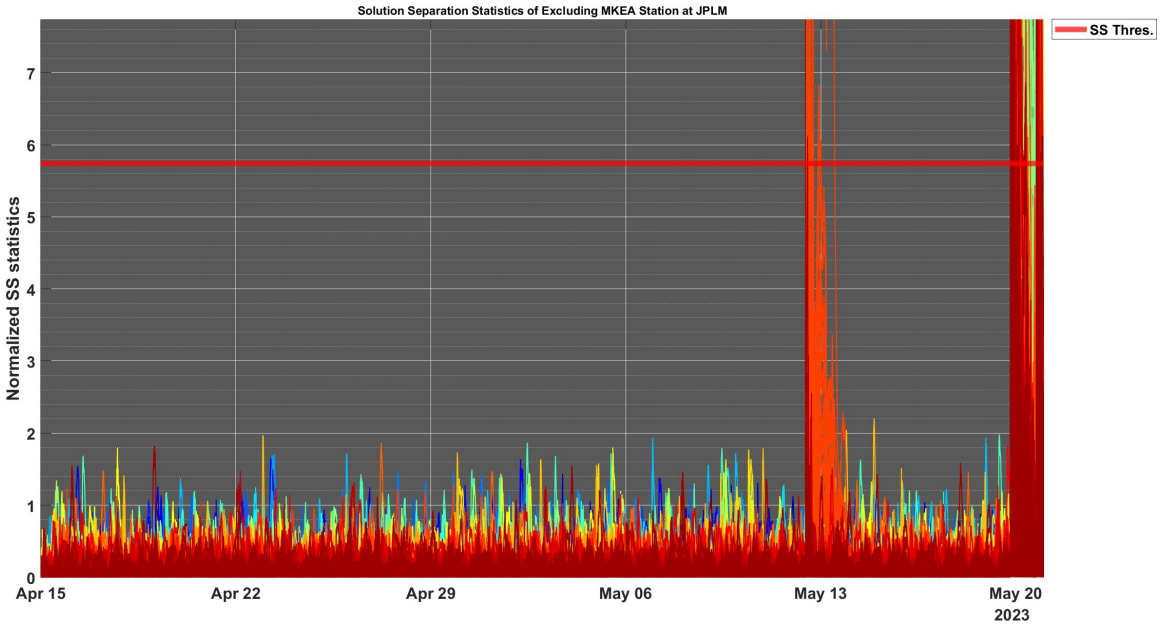
Fig. 4.3 shows the SS statistics of the detection subset (Eq. 4.4). All 56 subsets and 31 GPS satellites are shown in the figure. The threshold, $T = 5.74$, satisfies the continuity requirement specified by Eq. 1.48, with false alert rate $P_{fa} = 8 \times 10^{-6}$ per 15 seconds when only projecting to JPLM, represented as a horizontal red line. Any of the satellites in a subset that fail the SS test indicate the detection of the receiver failure, and exclusion must be attempted. It can be clearly seen that the SS test fails on May 12th, May 19th, and May 20th.

The use of SS statistics allows the integrity monitor to distinguish between satellite and receiver faults. Since the fault-tolerant filter has no constraining priors on the potential satellite clock and ephemeris faults, any type of UREs are possible, including the UREs in Fig. 4.2a. The SS test indicates that a receiver fault event occurs, so the UREs on May 12th, May 19th, and May 20th in Fig. 4.2a do not reflect the actual UREs.

4.2.3 Exclusion of Receiver Failure



(a) SS statistics of excluding MIKL receiver.



(b) SS statistics of excluding MKEA receiver.

Figure 4.4: SS statistics of the correct and incorrect exclusions at JPLM.

Knowing that there is a receiver fault present in the system, the act of excluding such a fault mode again occurs by evaluating the SS statistics on additional exclusions on each detection subset. For the sake of simplicity, only 2 of the 56 detection subsets are shown. One is the subset that exclude the faulted receiver, and another one fails to exclude the faulted receiver, which results in an SS statistics trip.

Fig. 4.4 shows the SS statistics of two of the detection subsets. Fig. 4.4a is the statistics that exclude the MIKL receiver, and Fig. 4.4b is the statistics that exclude the MKEA receiver. The SS statistics of the former subset pass all the tests, whereas the latter one fails on the event of the receiver fault. Since excluding the MIKL receiver from the network results in passing all the SS tests, and all the other SS statistics are similar to Fig. 4.4b, it is concluded that the fault mode is due to the failure of the MIKL receiver. Thus, the detection subset that excludes the MIKL receiver becomes the new AIV set for the rest of the operation until the MIKL receiver is recovered and added back to the system. This is the process of exclusion as described in Sec. 4.1.3.

The MIKL receiver is located in Mykolaiv, Ukraine. The observation data from MIKL during the period of the receiver failure, as shown in Fig. 4.3, indicates the C/N0 value on the L2 signal as low as 18 dB-Hz, while the L1 signal is as high as 50 dB-Hz. This is an unusual phenomenon on L1 and L2 signals with such a discrepancy in their C/N0 values. Moreover, the pseudorange value of the L1 signal jumps to hundreds of kilometers away the moment its C/N0 value suddenly peaks at around 50 dB-Hz. It is suspected that these might be a mixture of GPS spoofing and jamming events. The L1 signal is spoofed, and the L2 signal is jammed.

As an independent verification, RTKLIB [56] is used to estimate the MIKL position as a PPP user. As shown in Fig. 4.5, the position solution jumped to locations near Belarus, Moscow, and the Black Sea within a 24-hour period, indicating that multiple spoofing events likely occurred at the MIKL station.



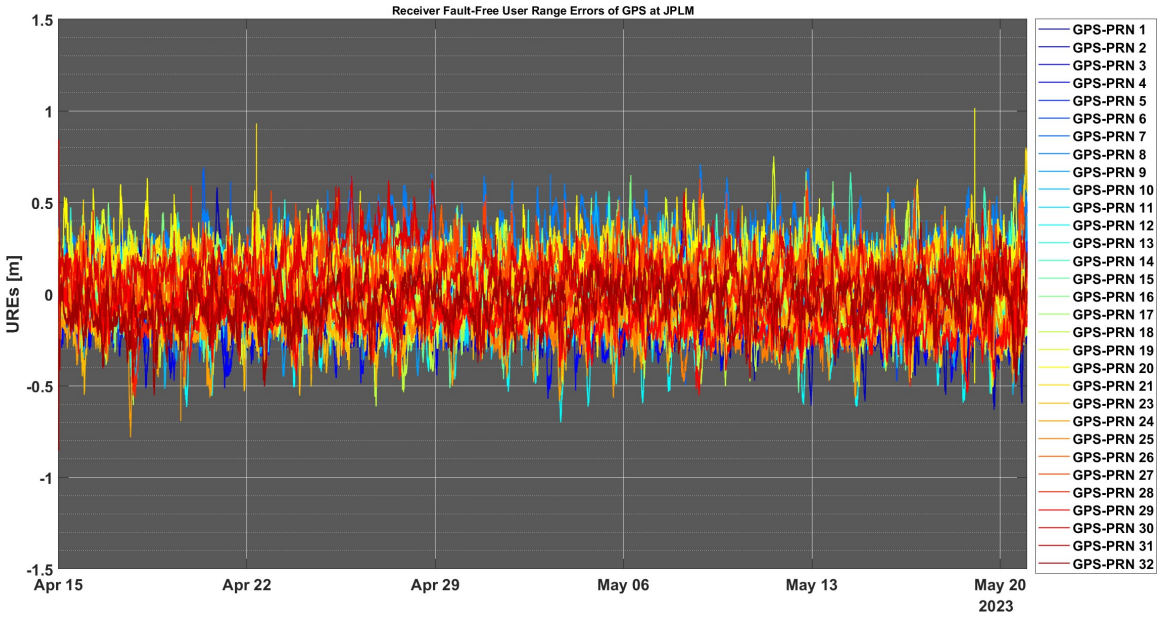
Figure 4.5: Positioning result from MIKL on May 19th, 2023, using RTKLIB as PPP user.

Although the intended function of the integrity monitor system is not about finding spoofing events, it does show the potential vulnerability of the system towards receiver failures. The receiver-wise SS process allows the system to remain consistent in the event of receiver failure. In this case, the receiver failure is probably due to spoofing events at the MIKL receiver. The spoofing events are successfully detected and excluded by the SS statistics, as they only affect the MIKL receiver and are classified as part of the receiver failure. The events demonstrate the effectiveness of the integrity monitor with SS statistics against unexpected receiver failures in the network.

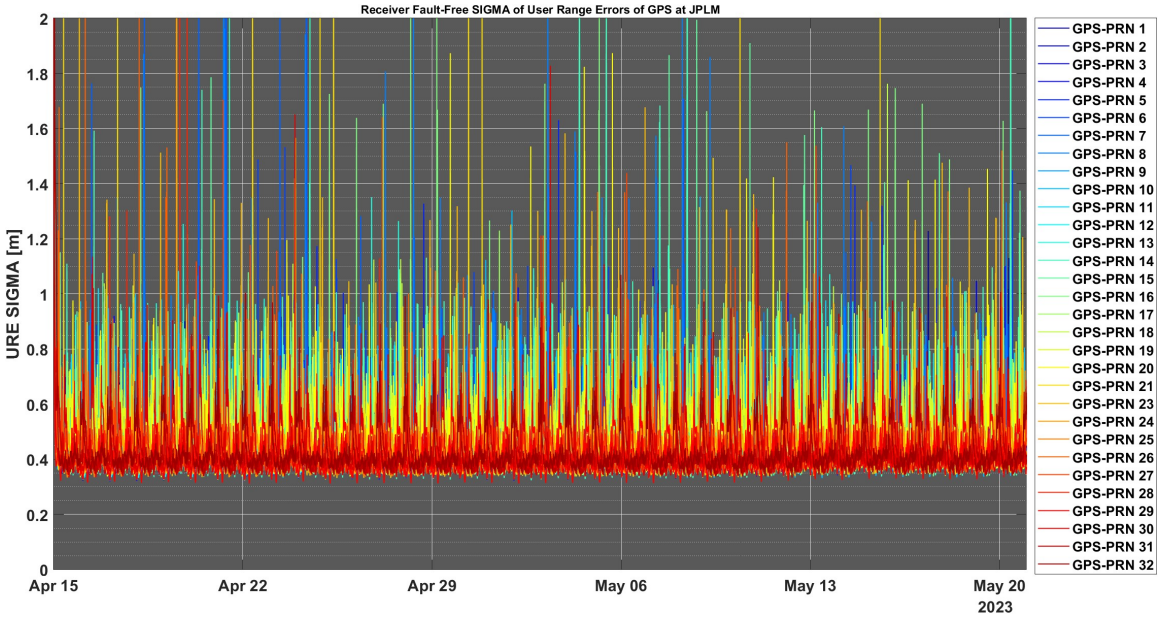
4.2.4 Robust Integrity Information

The integrity information (Eq. 4.16) is computed based on the SS test result. The AIV fault-tolerant filter correctly estimates the UREs of the GPS constellation from April 15th to May 19th before the SS tests fail. After the tests fail, the exclusion test from Fig. 4.4a indicates the failure of the MIKL receiver, so only the detection subset that excludes the MIKL receiver contains the correct estimates of the UREs. The UREs are formed from the combination of the AIV set and subset excluding the MIKL receiver, as shown in Fig. 4.6a. That is, Fig. 4.6a consists of the estimation result of the AIV set until May 12th, 2023 at 10:59:00, where the MIKL station failed. After that, the subset that excludes MIKL station succeeded, forming the UREs for the integrity information.

Finally, the sigma of the associated UREs is shown in Fig. 4.6b. They are computed from the approximation of the upper bound of the UREs, which overbounds PIR with the presence of the receiver fault modes as illustrated in Eqs. 1.42 and 4.15. The resulting sigmas of the overbounds range from 0.35 m $\sim$ 0.45 m after convergence. The sigmas are computed using Eq. 4.15 from the AIV set and its 1-out subset before the SS test failed, and the subset that excludes MIKL station and its 1-out sub-subset after. The sigmas are inflated due to the test threshold and receiver fault modes. The underlying difference between Figs. 4.2b and 4.6b is that the former is vulnerable towards receiver failures, for which the values fail to represent the correct UREs when the receiver fault occurs, whereas the latter includes the receiver failures as part of the fault modes, and proves to be reliable against receiver fault modes with the probability of mis-detection bounded by PIR.



(a) Integrity information: UREs.



(b) Integrity information: Sigma of upper bounds.

Figure 4.6: Integrity Information.

4.2.5 Suspected Ephemeris Anomaly

Since the receiver fault has been identified and excluded from the SS process, it can now identify if there are any faults/anomalies from the satellites or the corrections by directly evaluating the UREs and the sigmas.

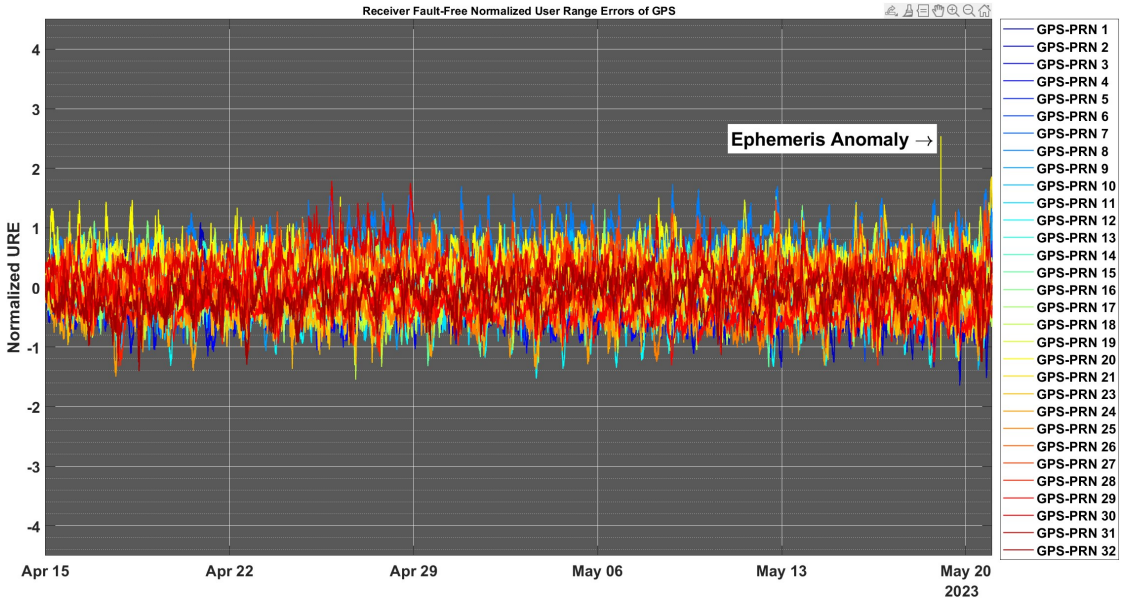
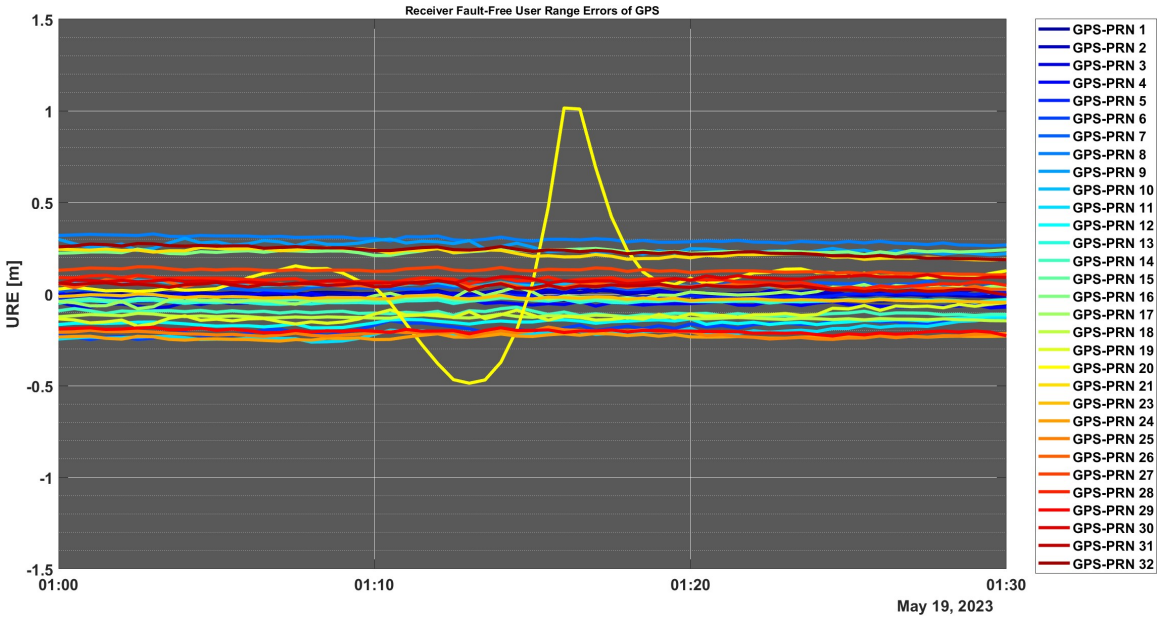
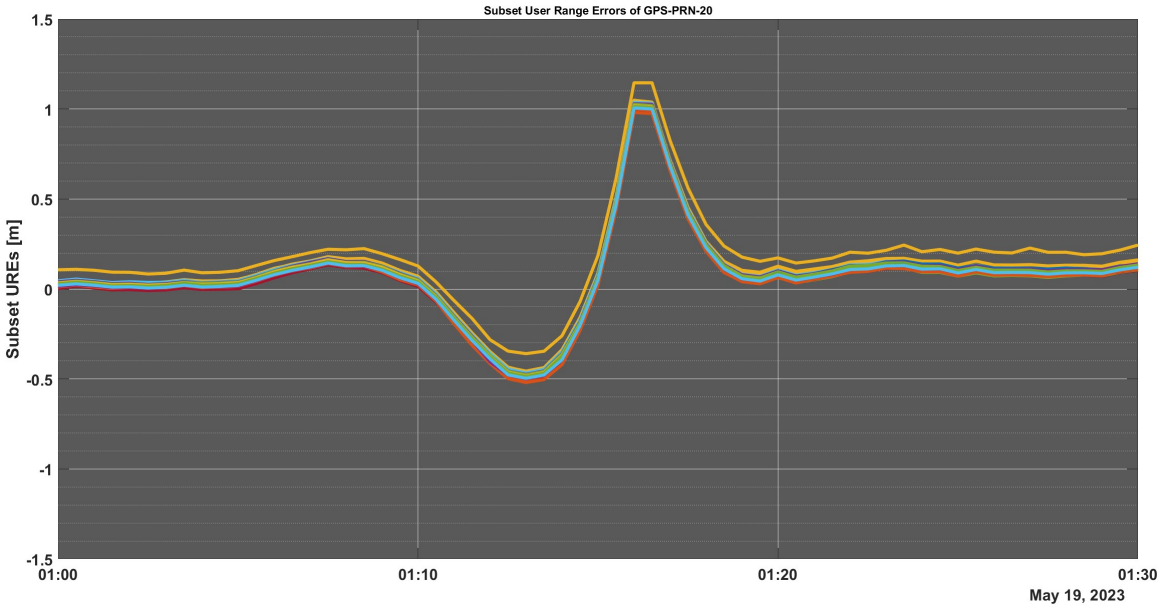


Figure 4.7: Normalized Integrity Information.

Fig. 4.7 shows the normalized integrity information. With the assumption that the URE of a satellite in nominal condition should be overbounded by a normal distribution, any potential clock and ephemeris failure can be identified if the normalized UREs are greater than a certain number of its sigmas. In the example network of the 36 days of data, there is only one instance of this occuring (PRN-20 on May 19th, 01:05:00 $\sim$ 01:20:00), which is pointed out in the figure, where the normalized URE exceeds 2.4 sigmas that is substantially different from the ambient normalized UREs, indicating a potential clock and ephemeris fault or anomaly.



(a) UREs during the anomaly.



(b) Subset UREs during the anomaly.

Figure 4.8: Subset UREs of PRN-20 during the anomaly.

Fig. 4.8a focuses on the anomaly event on the UREs of GPS. It can be seen as the blue line, the URE of GPS-PRN-20, experiences an abnormal shift from 01:05:00

$\sim$ 01:20:00 on May 19th that causes the normalized URE to exceed 2.4 sigmas. As a comparison, Fig. 4.8b shows all the subset UREs of GPS-PRN-20. It is clear that this event is not caused by the receiver failure since the separated solutions all exhibit the same pattern and no SS tests fail as shown in Fig. 4.4a. Therefore, the integrity monitor system strongly suggests that the anomaly is due to a clock and/or ephemeris event.

According to the NOTICE ADVISORY TO NAVSTAR USERS (NANU) announced by the United States Coast Guard (UCSG) [https://www.navcen.uscg.gov/sites/default/files/gps/nanu/2023/2023025.nnu], SVN51 (PRN20) was set to be unusable from May 18th, 22:30:00 until May 19th, 10:30:00, which coincides with the anomaly event we observed from the AIV UREs and its subset. As a result, we conclude that the abnormal URE is due to a satellite maneuvering which caused a short period of disturbance on its orbit and that was not as accurately be captured by the model that generates the precise ephemeris corrections. The impact of such ephemeris anomaly to the users are demonstrated in Sec.4.3. Since the fault-tolerant filter does not assume any dynamics on the clock and ephemeris of the satellites, this kind of short disturbance is successfully reflected in our integrity monitor system. Typically, we do not consider the health status of a satellite when utilizing precise products for evaluation of our filter. For the sake of evaluating the nominal performance of integrity information, this suspected anomaly event is removed from the computation of the overbound in the next subsection.

4.2.6 Global Projection

The projection to the JPLM station reflects the detection and exclusion of the receiver fault modes local to the projected location. To ensure that the receiver fault modes are monitored with global user coverage, the evaluation of the SS statistics in the range domain must also vary globally.

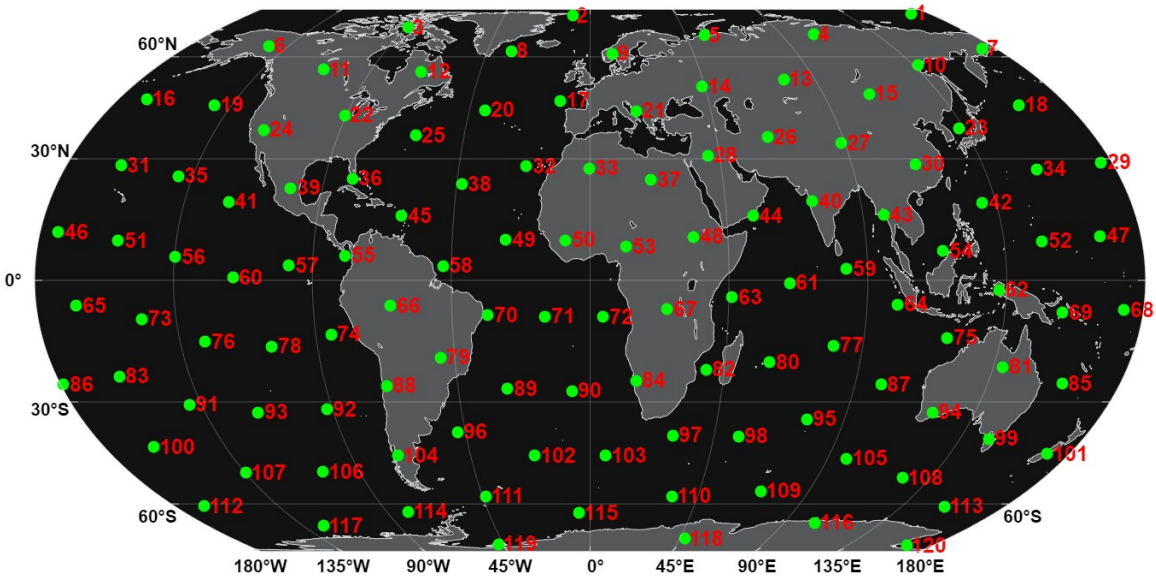


Figure 4.9: 120 uniformly distributed global users.

Fig. 4.9 shows the locations of all 120 global user projections. The users are uniformly distributed across the Earth. Instead of only projecting to JPLM user, the UREs are projected to all 120 users to demonstrate the global applicability.

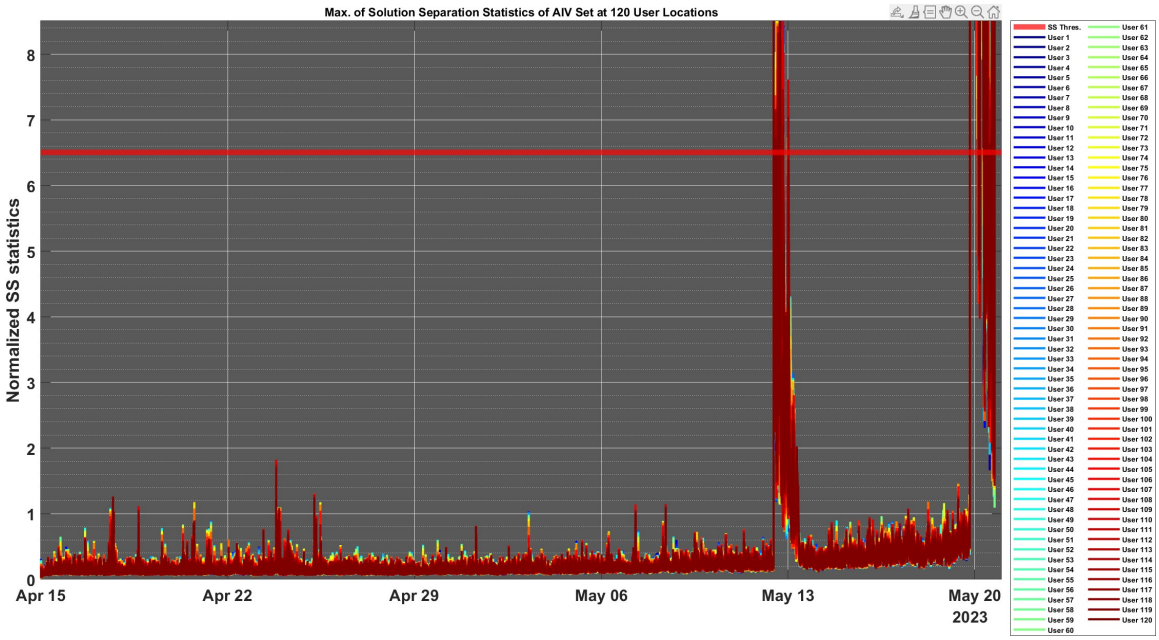
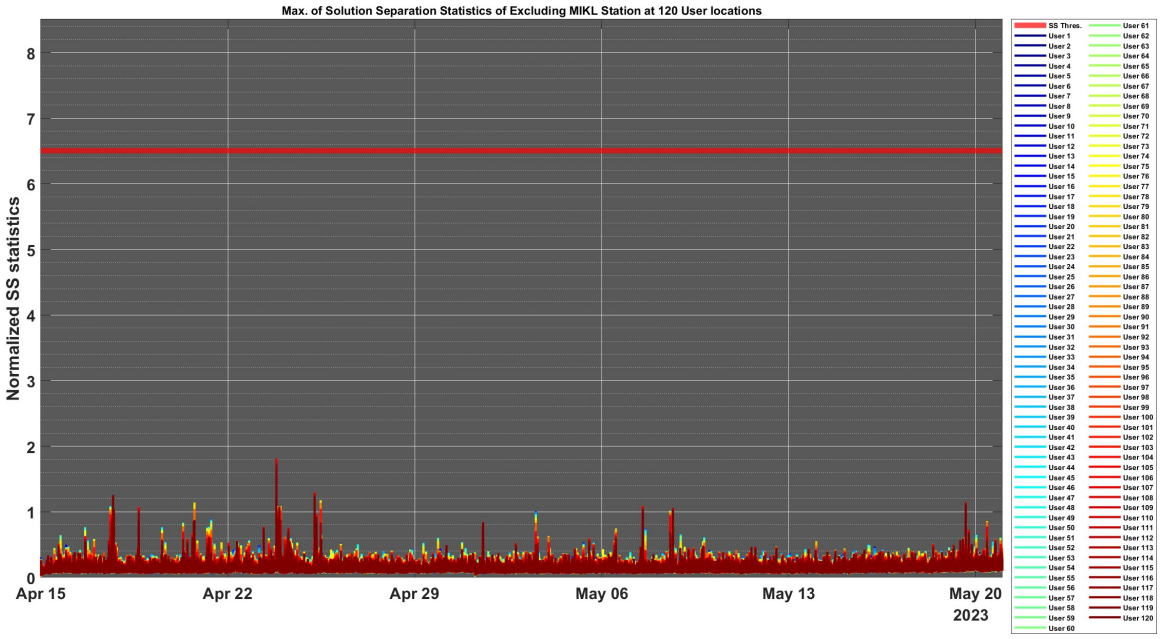
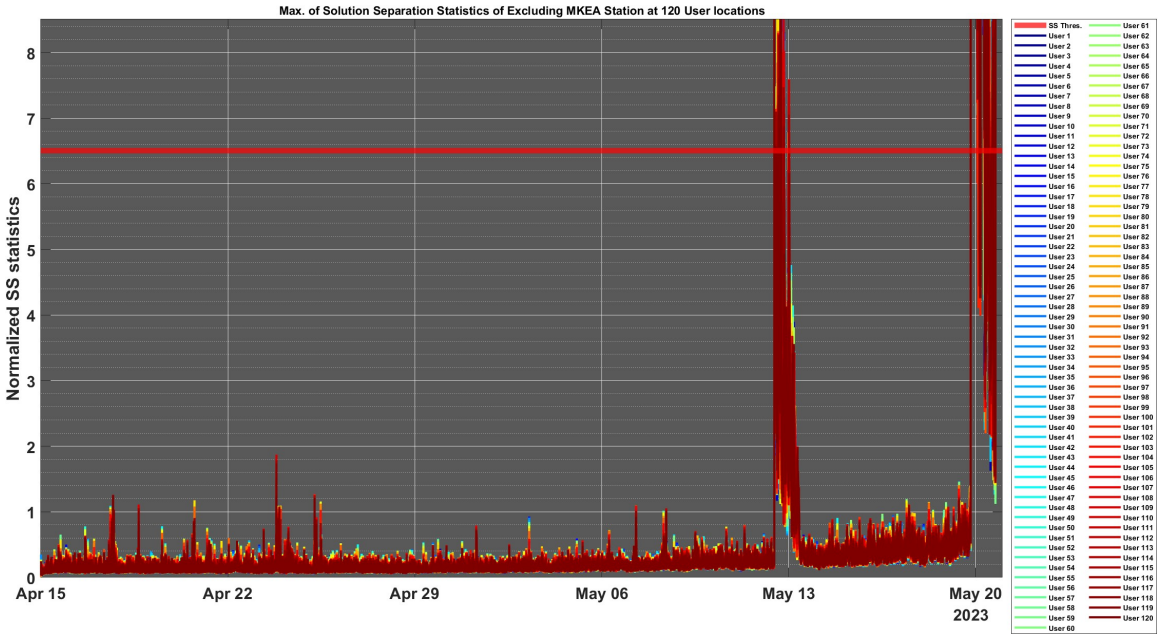


Figure 4.10: Max. SS statistics of AIV set with 6.5 sigma threshold at all 120 locations.

The same projection process is performed at 120 user locations as shown in Fig. 4.9 to demonstrate the sensitivity of the SS statistics at the global scale. The threshold, $T = 6.5$, is inflated due to projecting to 120 locations, thus the test numbers are 120 times more than only projecting to JPLM. Additionally, only the maximum of the test statistics at each user location is shown for simplicity. This does not alter the detection result given that the threshold test is tripped by the largest statistics.



(a) Max. of SS statistics of excluding MIKL receiver over 120 global users.



(b) Max. of SS statistics of excluding MKEA receiver over 120 global users.

Figure 4.11: SS statistics of the correct and incorrect exclusions at all 120 locations.

Similarly, Figs. 4.11a and 4.11b are the exclusion results of the SS statistics of all

120 user locations. The variations due to projecting to different geometric locations of the users are almost unchanged, and the result suggest the MIKL station is the faulted receiver.

4.2.7 Integrity of the Estimator

The underlying design of the integrity monitor is based on assuming that the errors in the URE estimates are overbounded by normal distribution whose sigmas correspond to the associate filter estimates. This section uses the precise products as source of truth, so the evaluation method essentially assumes that the precise products contain negligibly small errors compared to the filter estimates. Aside from the abnormally large error identified in Sec. 4.2.5, this assumption has been observed to be true for the remainder of the precise ephemerides examined. Therefore, it can be verified that the empirical distribution of the normalized URE estimates is consistent with the estimator assumptions.

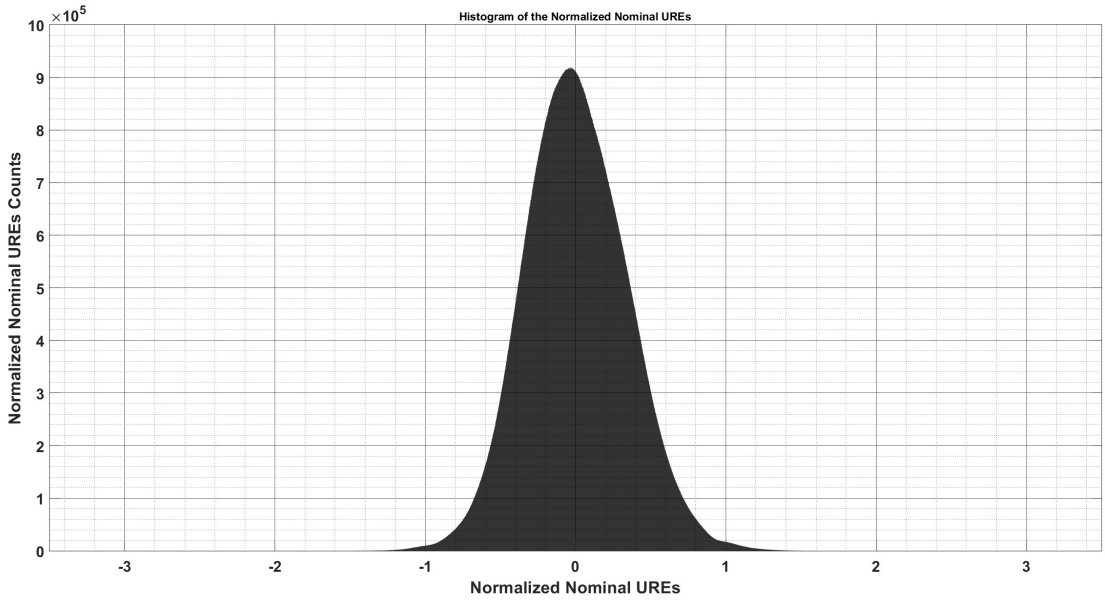


Figure 4.12: Histogram of the normalized integrity information. The ephemeris anomaly event is excluded from the counts.

The UREs obtained from Sec. 4.2.4 are normalized with their associated sigma

values. Fig. 4.12 shows the histogram of the normalized UREs from across 36 days and all 31 GPS satellites over 120 global users.

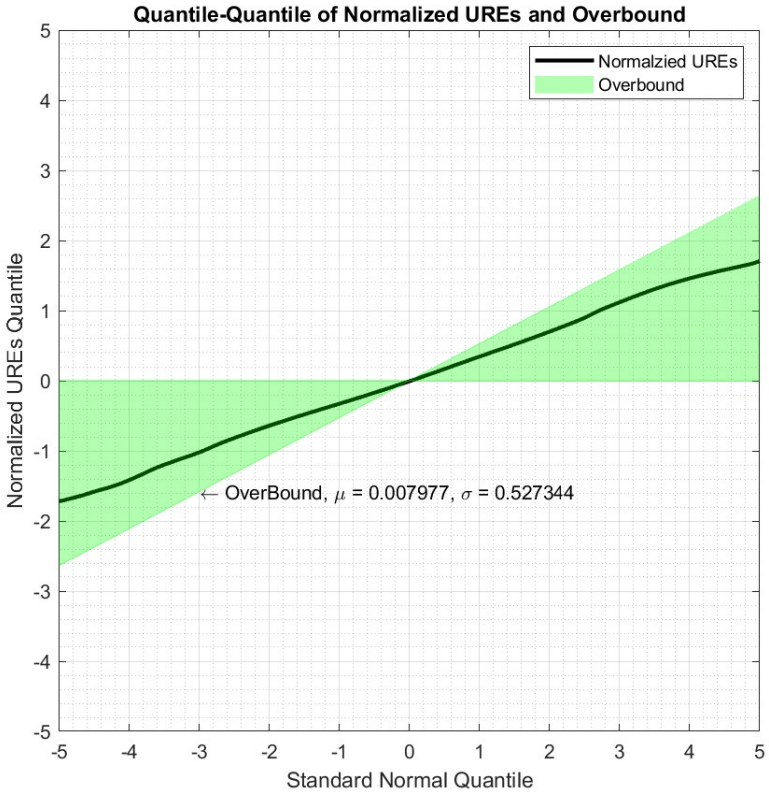


Figure 4.13: QQ-plot of the Empirical CDF of the Normalized UREs and the corresponding overbound.

The Quantile-Quantile (QQ) relation of the empirical distribution of the normalized UREs is shown in Fig. 4.13 as the black line. The green area represents the standard distribution with a minimal mean that can overbound with a zero-mean, unity-variance Gaussian. The overbound has the mean (μ) of 0.008 and the sigma (σ) of 0.5273 indicating significant margin between the estimated sigma values and the minimum overbounding empirical sigmas. It indicates that the estimator of the fault-tolerant filter, including the SS methodology, is sufficiently conservative for the 36 days of data processed.

4.3 User Integrity from the Robust Integrity Information

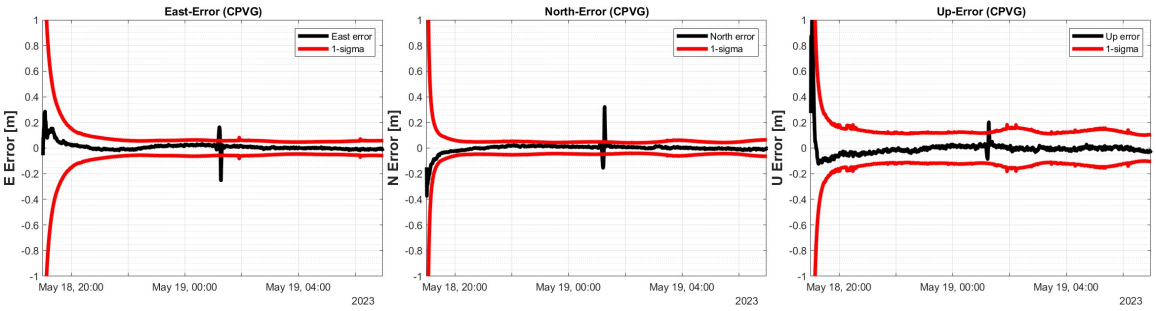


Figure 4.14: Position errors of CPVG in East-North-Up direction under the ephemeris anomaly at PRN-20 on May 19th.

Fig. 4.14 is the position estimates of a PPP user at CPVG, an IGS station independent of the network receivers located at Espargos, Cabo Verde, during the ephemeris anomaly discovered in Sec. 4.2.5. From 01:10:00 to 01:20:00 on May 19th, the position ephemeris anomaly caused noticeable disturbances to the positioning accuracy of the user: approximately 20 cm errors in East, 30 cm in North, and 20 cm errors in the up directions. The position errors are not significant due to the small magnitude of the ephemeris anomaly, and the fact that PRN-20 was flagged unhealthy at that time posed no integrity risk to the users. Nevertheless, considering that the IGS precise products claimed ephemeris accuracy of $\sim$2.5 cm, the resulting errors show the importance of integrity monitoring in PPP services.

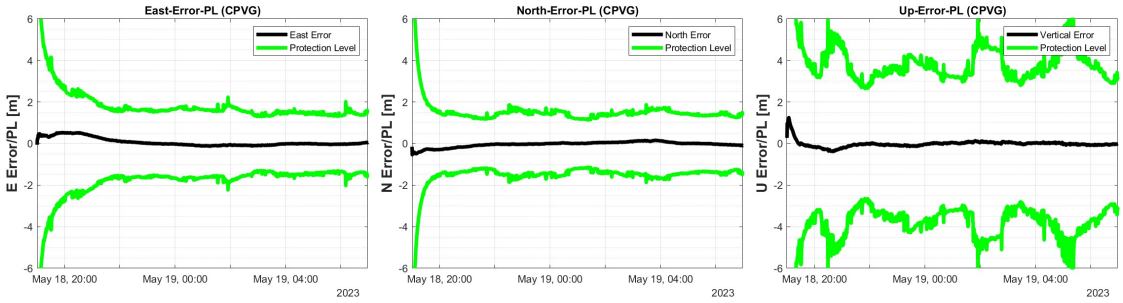


Figure 4.15: Position errors and protection levels of CPVG in East-North-Up direction under the ephemeris anomaly at PRN-20 on May 19th, with robust integrity information.

Fig. 4.15 shows the position errors and protection levels at CPVG with the robust integrity information as extra corrections. The influence of the ephemeris anomaly is removed and errors remain at the decimeter, even centimeter level, in ENU directions. The protection levels generally achieve 1.2∼1.4 m in the horizontal and 2.8∼3.5 m in the vertical direction. The resulting protection levels protect the users from arbitrary satellite clock and ephemeris faults and 1-out receiver fault from the monitoring network.

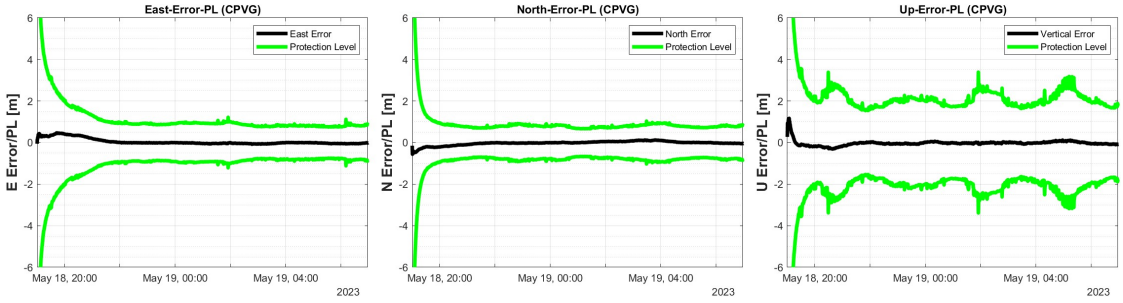


Figure 4.16: Position errors and protection levels of CPVG in East-North-Up direction under the ephemeris anomaly at PRN-20 on May 19th, with AIV integrity information.

As a comparison, Fig. 4.16 gives the position errors and protection levels applying AIV integrity information, without assuming receiver faults. It can be seen that the protection levels are less than the user with robust integrity information as shown in Fig. 4.15, achieving 0.7∼0.9 m in horizontal and 1.8∼2 m in the vertical direction. The direct effect of including receiver fault modes into the system is the inflation of

the protection level. In other words, the protection levels shown in Fig. 4.16 are the lower bound that the example network can provide to the users. Depending on the priors of receiver failure within the system, the protection levels provided to users are inflated accordingly.

In both cases, the integrity and accuracy of the CPVG user is preserved with the assistance of the integrity information from the fault-tolerant filter. It shows that the proposed integrity monitor system can capture the instance where such satellite faults occurs, and provide the corresponding corrections against the faults to the users instantly.

4.4 Summary

A key property of the proposed integrity monitoring system is the priors it requires. The only prior the fault-tolerant filter requires other than a normal PPP user is the static locations of the GNSS receivers within the network, which can be surveyed beforehand and maintained for an extended period of time for the period of the operation of the integrity monitor system. The corrections it needs are available to the monitor given that they are to be broadcast to the user in real-time. The observations are code and carrier phase measurements from any type of GNSS receiver. In principle, the fault-tolerant filter is no different than the standard PPP user with a single GNSS receiver, except that the fault-tolerant filter is tolerant to satellite clock and ephemeris faults by assessing the consistency of the observations and the static locations across multiple receivers within the network. The Solution Separation process on the receiver network requires simply running the fault-tolerant filter with different combinations of the network receiver to detect and exclude potential receiver failure. In fact, if the failure rate of the receiver is less than PIR, then no SS is required, and the estimation results from the AIV filter are the integrity information. The SS threshold could potentially be lowered with tighter SS sigmas, and therefore lowering the robust URE sigmas and the protection levels for the users. The entire framework of the integrity monitor system is therefore theoretically viable to operate live. Thus, the proposed framework can provide integrity information to users in

real-time.

Chapter 5

Network Performance Evaluation

The ground-based integrity monitoring system is fundamentally based on the network of receivers. Under the assumption that measurements from all receivers share identical prior distributions, only two factors can be controlled to affect the monitoring results.

1. Number of receivers

2. Location of receivers

The first factor, the effect of the number of receivers, is straightforward. A larger number of receivers provides more pseudorange measurements, thereby increasing the information available to the system and improving confidence in the estimation results. However, as pointed out in Sec. 4.1, more receivers results in higher probability of receiver failure within the system. The second factor, the location of receivers refers to the spatial distribution of the network. The estimations of the ephemeris errors rely on the instantaneous geometric relations between the satellite and the receivers' LOS. This is especially important for the fault-tolerant filter since no priors are assumed on the kinematics of the satellite orbits.

Following the design of the integrity monitor, the problem can be formulated as determining the two factors: the required number of receivers and their spatial configuration. To start with, this chapter is dedicated to defining and evaluating the

performance of a network based on these two factors. It is necessary to first establish the performance metrics between different designs before one can be chosen to be the desired solution.

Additionally, this chapter transitions from the use of real data, as presented in Chapters 2–4, to simulation-based analyses for evaluating network performance. This transition is a crucial step in the process of designing the network, as the objective is to determine the optimal placement of reference receivers. Evaluating network geometry only after deployment would be impractical; therefore, simulations provide an essential tool for assessing performance prior to having actual receivers in the field.

5.1 Defining Performance of an Integrity Network

From the perspective of PPP users, the most important metric for them to assess is the accuracy of the positioning result. There might be cases where PPP users focus on other states such as ionospheric effect, receiver clock bias for time synchronization, etc., but the majority of PPP users are concerned with position. Since the objective of the integrity monitor for the precise GNSS service is to provide users with precise accuracy while having integrity on the position results, the protection levels are what measure integrity guarantee for the users. In particular, given that the protection levels for PPP users with integrity information are simply the constant multiple of the sigmas of the position estimates when no receiver failures are considered, the criteria to evaluate the user uncertainty can also indicate the performance of the source of the integrity information, i.e., the integrity monitor.

The fault-tolerant filter that contains the integrity monitor provides integrity information to the users. However, the process of projection makes the integrity information correlate with the user locations. As a result, the protection levels from users at different places may vary and must be taken into account when considering the contribution of the integrity monitor to the users as a whole.

In theory, there could be an infinite number of UREs as a result of projecting the satellite clock and ephemeris errors to difference places on Earth. To simplify the evaluation process, the performance of the integrity monitoring system is defined

in terms of the user protection levels at the location of interest, given the integrity information as additional corrections. In other words, the lower the protection levels for a user, the better the performance of the integrity monitor locally from the view point of the user. However, since the protection levels only reflect from a single point, the performance of the integrity monitor should be characterized by protection levels across users from the region of interest.

5.2 Steady-State Kalman and Information Filter

So far, computation of the protection levels has been based on real observation data. This reflects the authentic performance of the 56-station network and the user with integrity guarantee, but it limits the performance evaluation of the integrity monitor to only the places that have access to GNSS receivers. To enable the performance evaluation at places without the observation data, the steady-state covariance of the user filter is adopted as the lower bound protection levels that users can achieve.

In standard EKF, computation of the protection levels for a PPP user involves the process of convergence, typically introduced by the initial uncertainty of each filter state, that is modeling the temporal correlations of the error. However, this implies that the protection levels are not consistent depending on the time sequence of the measurements available to each user of interest, and the beginning of the convergence.

To eliminate the influence of the temporal correlations on the computation of the protection levels, protection levels are computed from the users in the form of the steady-state uncertainty. This allows the performance evaluation of the integrity network to be focused on geometric relations, which are only relevant to the placement and number of network receivers.

5.2.1 Steady-State Covariance from Discrete-Time Algebraic Riccati Equation (DARE)

The spatial correlation of standard EKF solutions emerges from the LOS geometries between users and satellites, and the temporal correlation is from the converging

process of the sequential update. In general, the estimation result of a PPP user is available only after convergence. That is, the performance of an integrity monitor when evaluating at the users can only be comparable if the temporal effects during the convergence are removed. As a result, the performance evaluation focuses on the spatial correlations from the integrity monitor to the user by computing the user's protection level at steady-state.

The equation of the steady-state covariance is derived by substituting Eqs. 1.25 and 1.28 into Eq. 1.27:

$$\Sigma = A\Sigma A^T + Q - \left(A\Sigma A^T C^T + QC^T\right)\left(CA\Sigma A^T C^T + CQC^T + R\right)^{-1}\left(CA\Sigma A^T + CQ\right) \tag{5.1}$$

where Σ is the steady-state covariance, and the subscript t is abbreviated for simplicity. Eq. 5.1 is also known as a form of Discrete-Time Algebraic Riccati Equation (DARE) [57] that can be solved efficiently in Matlab with its numerical solver, *idare*. Σ in Eq. 5.1 represents the uncertainty of the defined state given the linear kinematic model (A), process noise (Q), measurement model (C), and measurement noise (R) at steady-state without the influence of the temporal correlation from the converging process.

The computation of the protection levels for any potential users is straightforward. Given the priors of the user filter and the associated integrity information, one can solve Eq. 5.1 for the position covariance Σ_{xyz}, and thereafter the protection levels at steady-state from Eq. 3.5.

Due to the fact that the process noise of ambiguity states is zero, meaning they will converge to zero at steady-state, the following is true:

$$\Sigma = \begin{bmatrix} \sigma_x^2 & \dots & \sigma_{x,\nabla\hat{T}_r} & \sigma_{x,A_r^{(1)}} & \dots & \sigma_{x,A_r^{(k)}} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{\nabla\hat{T}_r,x} & \dots & \sigma^2_{\nabla\hat{T}_r} & \sigma_{\nabla\hat{T}_r,A_r^{(1)}} & \dots & \sigma_{\nabla\hat{T}_r,,A_r^{(k)}} \\ \sigma_{A_r^{(1)},x} & \dots & \sigma_{A_r^{(1)},\nabla\hat{T}_r} & \sigma^2_{A_r^{(1)}} & \dots & \sigma_{A_r^{(1)},A_r^{(k)}} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{A_r^{(k)},x} & \dots & \sigma_{A_r^{(k)},\nabla\hat{T}_r} & \dots & \dots & \sigma^2_{A_r^{(k)}} \end{bmatrix} \rightarrow \begin{bmatrix} \sigma_x^2 & \dots & \sigma_{x,\nabla\hat{T}_r} & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{\nabla\hat{T}_r,x} & \dots & \sigma^2_{\nabla\hat{T}_r} & 0 & \dots & 0 \\ 0 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & \dots & \dots & 0 \end{bmatrix} \tag{5.2}$$

Therefore, it is mathematically equivalent and numerically beneficial to ignore the ambiguity states for the steady-state covariance computation. The dimension of the state for a PPP user filter is then reduced to nine and is consistent regardless of the number of satellites LOS.

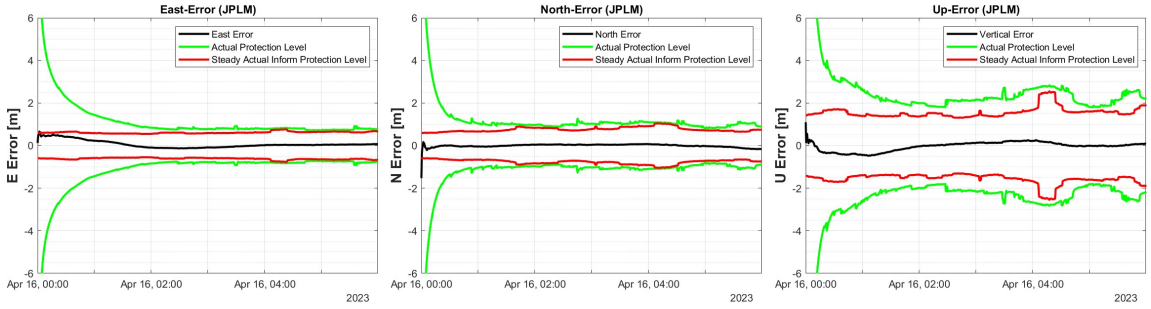


Figure 5.1: Actual and predicted steady-state protection levels.

Fig. 5.1 shows the position estimates at the JPLM station for a PPP user. The green lines are the protection levels computed from the assistance of the 56-station network, shown in Fig. 2.9. The figure represents the actual protection levels the user can achieve. The red lines are the predicted protection levels with the position sigma from Eq. 5.1 at steady-state. These are effectively the lower bound that the user can achieve given the geometry of the receiver network since it neglects the uncertainty from the temporal convergence of the filter, and only depends on the geometric relations. Although real data can be used to obtain actual protection levels, the resulting evaluation reflects the performance of the integrity monitor only

at a specific point.

5.2.2 Evaluation of the Protection Levels

To demonstrate the performance of the 56-station network across the surface of the Earth, users are simulated via the precise ephemeris product who have LOS to the corresponding satellites, with priors identical to the real PPP user filter, and measurement noise inflated with the sigmas of the integrity information, as shown in Eq. 3.3, and Eq. 3.4. Finally, the covariance is computed from Eq. 5.1. The resulting position covariance is transformed into the East-North-Up direction, and then multiplied by K to obtain the protection levels. Since no pseudorange is required when computing the covariance, the whole process can be performed at any place where protection levels are of interest.

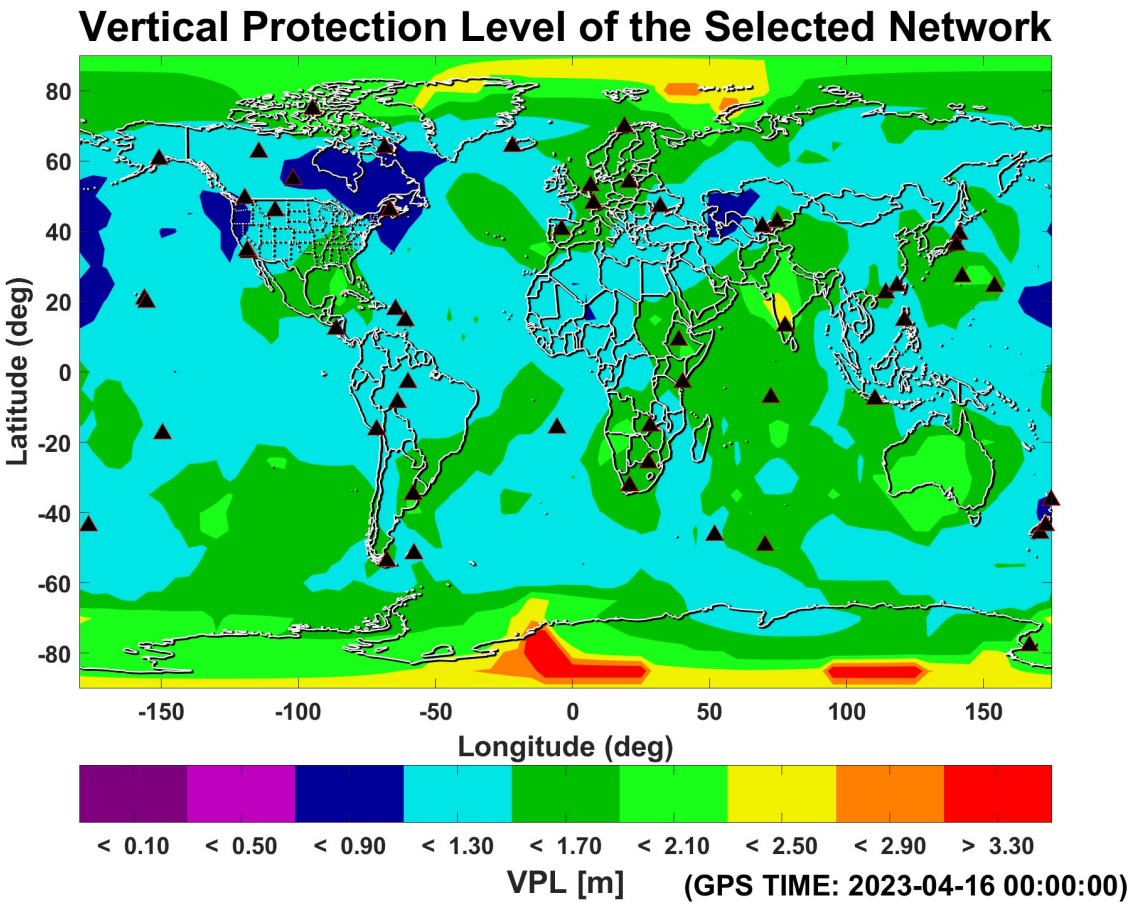


Figure 5.2: Steady-State snapshot VPL from 56-station network using real integrity information.

Fig. 5.2 shows the vertical protection levels (VPLs) at steady-state across the world on April 16th, 2023 at the instance of 00:00:00. The black triangles represent the locations of the network receivers used to generate the integrity information, i.e., the 56-station network. The integrity information is computed from the actual IGS code/carrier phase data and precise product from the network receivers, and the global VPL map consists of $5^\circ \times 5^\circ$ granularity grids. Each grid point acts as a PPP user applying the integrity information to its pseudoragne model with the geometric matrices from from the precise ephemeris used by the integrity monitor at steady-state. The only constraint on LOS and NLOS satellites for grid users is the cut-off elevation angle. Different color regions indicate different magnitudes of the protection

levels at steady-state at different places.

The global VPL map is straightforward in terms of evaluating the performance of the integrity monitor at one instance. It can be seen that steady-state VPL achieved 1.3 $\sim$ 1.7 m in most parts of the world, and as low as $\sim$1 m in some parts of North America. In comparison, the network performs relatively worse in regions such as the polar regions, South America, India, and Australian continent. It is useful to identify the regions within the network where the integrity monitor performs relatively better or worse, thereby providing feedback when designing network geometries, since the steady-state protection levels can only be modified through the number and spatial distribution of the network receivers, assuming fixed prior distributions.

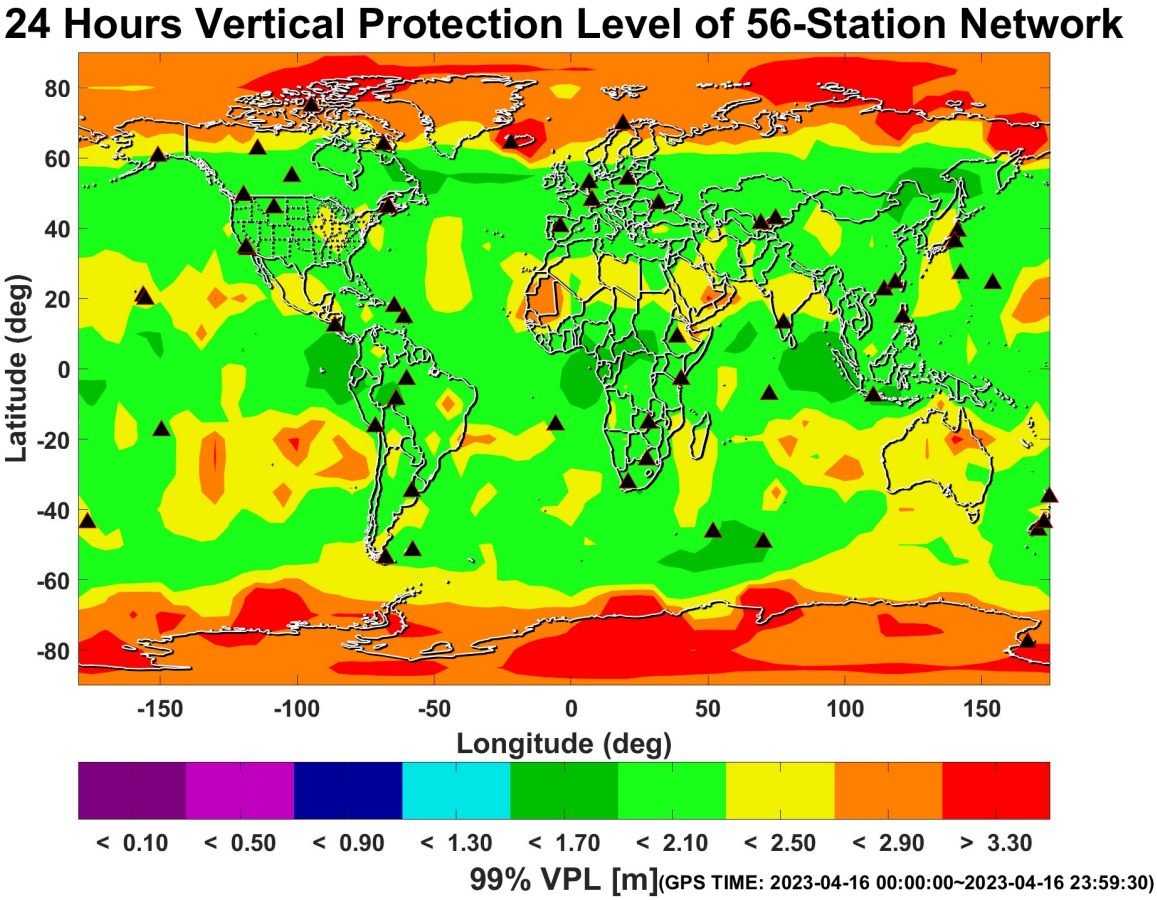


Figure 5.3: 99% Availability Steady-State VPL from real network.

Fig. 5.3 shows the steady-state VPL of the 56-station network with 99%

availability over the course of 24 hours. That is, 99% of the time over a 24-hour period, users can achieve a VPL less than or equal to the value shown at the specified location.

The global steady-state VPL map provides a straightforward view of the performance of the network at different parts of the world over an extensive period time. It can be seen that the VPL is generally worse in the polar regions with VPL often greater than 3 m. The 56 station network also does not perform comparatively well in regions such as Australia, Japan, India, and the Arab peninsula. This information cannot be obtained by computing the VPL using only real user data, since the observations may be limited to specific times and locations.

The Horizontal Protection Levels (HPLs) are generally much smaller than the VPLs and are therefore provided in Appendix.B for reference. The VPLs are chosen as the fundamental metric to quantitatively evaluate the performance of the integrity monitor.

5.3 Steady-State Information Approximation of Satellite Errors

Both PPP users and the fault-tolerant filter can solve for steady-state covariance using Eq. 5.1. The user can easily compute the steady-state covariance with the simplification of zero ambiguity states, where the dimensionality is limited to nine for each user. However, even after ignoring ambiguities at steady-state, the dimensionality of the fault-tolerant filter can still be hundreds or up to thousands depending on the total number of receivers within the network as expressed in Eq. 2.8. To reduce the computation of the covariance of the satellite clock and ephemeris errors, Eq. 5.1 can be transformed into information form and utilize the fact that no dynamic model is adopted for the satellite errors.

The states correspond to the receiver and satellite error and can be expressed as

follows:

$$\Sigma = \begin{bmatrix} \Sigma_{rcv} & \Sigma_{1,2} \\ \Sigma_{2,1} & \Sigma_{sat} \end{bmatrix} \tag{5.3}$$

$$A = \begin{bmatrix} A_{rcv} & 0 \\ 0 & A_{sat} = 0 \end{bmatrix} \tag{5.4}$$

$$C = \begin{bmatrix} C_{rcv} & C_{sat} \end{bmatrix} \tag{5.5}$$

$$Q = \begin{bmatrix} Q_{rcv} & 0 \\ 0 & Q_{sat} \end{bmatrix} \tag{5.6}$$

where the subscript rcv corresponds to the receiver states, sat corresponds to the satellite clock and ephemeris errors, $\Sigma_{1,2}$ are the cross-covariances between the receiver and satellite error states, and A_{sat}, is zero since satellite error states have no dynamic model.

Substituting the above expressions into Eq. 5.1, the covariance of the satellite errors, Σ_{sat} becomes:

$$\Sigma_{sat} = Q_{sat} - Q_{sat}C_{sat}^T \left(C_{rcv}A_{rcv}\Sigma_{rcv}A_{rcv}^T C_{rcv}^T + C_{rcv}Q_{rcv}C_{rcv}^T + C_{sat}Q_{sat}C_{sat}^T + R\right)^{-1} C_{sat}Q_{sat} \tag{5.7}$$

Approximate the satellite error covariance by simplifying the effect of receiver covariance at steady-state, i.e., $C_{rcv}A_{rcv}\Sigma_{rcv}A_{rcv}^T C_{rcv}^T \approx 0$:

$$\left(C_{rcv}A_{rcv}\Sigma_{rcv}A_{rcv}^T C_{rcv}^T + C_{rcv}Q_{rcv}C_{rcv}^T + C_{sat}Q_{sat}C_{sat}^T + R\right)^{-1} \approx \left(C_{rcv}Q_{rcv}C_{rcv}^T + C_{sat}Q_{sat}C_{sat}^T + R\right)^{-1} \tag{5.8}$$

The reason that the receiver covariance can be simplified is that without the ambiguity state, the prior of the receiver states are dominated by the receiver clock bias (the tropospheric state is negligible), which theoretically has infinite process noise, meaning the information from the prior of the filter is near zero. Thus, the information matrix of the prior is approximately equivalent without considering the

prior information of the filter. This results in the covariance of the filter at steady-state is effectively snapshot.

Thus,

$$\Sigma_{sat} \approx Q_{sat} - Q_{sat}C_{sat}^T \left(C_{rcv}Q_{rcv}C_{rcv}^T + C_{sat}Q_{sat}C_{sat}^T + R\right)^{-1} C_{sat}Q_{sat} \tag{5.9}$$

In practice, this simplification makes less than 1 mm of difference on the sigma of UREs.

Eq. 5.9 expresses Σ_{sat} entirely with the measurement models for each of the receivers, and decouples the correlations between the receiver states that allow efficient computation in information form.

Given the Gaussian assumption of the error distribution, the information matrix is the inverse of the covariance.

$$Y_{sat} = \Sigma_{sat}^{-1} \tag{5.10}$$

where Y_{sat} is the information matrix of the satellite error.

Invoking the Woodbury Matrix Identity, Y_{sat} becomes:

$$Y_{sat} = \Sigma_{sat}^{-1} = \left[Q_{sat} - Q_{sat}C_{sat}^T \left(C_{rcv}Q_{rcv}C_{rcv}^T + C_{sat}Q_{sat}C_{sat}^T + R\right)^{-1} C_{sat}Q_{sat}\right]^{-1} \tag{5.11}$$

$$= Q_{sat}^{-1} + C_{sat}^T \left(C_{rcv}Q_{rcv}C_{rcv}^T + R\right)^{-1} C_{sat} \tag{5.12}$$

Since each receiver will only affect the LOS satellites and the measurement noises are IID, the matrix inversion is equivalent to the sum of each receiver to its LOS satellite errors.

$$Y_{sat} = Q_{sat}^{-1} + \sum_{r=1}^{n} C_{sat,r}^T \left(C_{rcv,r}Q_{rcv,r}C_{rcv,r}^T + R_r\right)^{-1} C_{sat,r} + C_{sat,const}^T R_{const}^{-1} C_{sat,const} \tag{5.13}$$

where $C_{sat,r}$, $C_{rcv,r}$, R_r correspond to the measurement models and noise of each

receiver respectively and independently. $C_{sat,const}$ and R^{-1}_{const} are the zero-mean constraints described in Sec. 2.1.2.

Eq. 5.13 computes the information matrix of the satellite error by summing over the contributions of all receivers with their LOS satellites, and finally using a matrix inversion to convert the information matrix into a covariance that enables the projection to the range domain. Each receiver would be LOS to approximately tens of satellites at a time, meaning it is relatively lightweight in terms of matrix computation, and the final matrix inversion will have a fix dimensionality of $m \cdot 4$ that only needs to be computed one time. This allows for extremely efficient evaluation of the steady-state satellite error covariance regardless of the number of total receivers within the network.

5.3.1 Heterogeneous Receiver Priors

The examples shown in this chapter adopts identical priors for each receivers. In particular, the measurement noise, R_r in Eq. 5.13, is weighted equally across the receivers.

Alternatively, if priors of the network receivers are not identical, i.e., heterogeneous, the associated measurement noise is weighted accordingly to reflect different priors among the receivers.

5.4 Full Simulation Via Almanac

While the VPL at any potential user locations can be computed efficiently with the integrity information and steady-state covariance (Eq. 5.1), it still relies on the ephemeris product and LOS from the observation data of each receiver in the network.

To enable rapid evaluation of the performance of an arbitrary network, and since the computation of the VPL only requires the covariance of the satellite clock and ephemeris errors before projecting onto the user locations, the almanac is sufficient to simulate the geometry between the satellites and receivers.

The GPS almanac is used to generate the ephemeris for the purpose of simulation.

In particular, the ephemeris from the almanac data allows for quick evaluation of the receiver to satellite geometry and the covariance computation when combined with Eq. 5.1. The simulation of the WAAS performance evaluation is established similarly in [58].

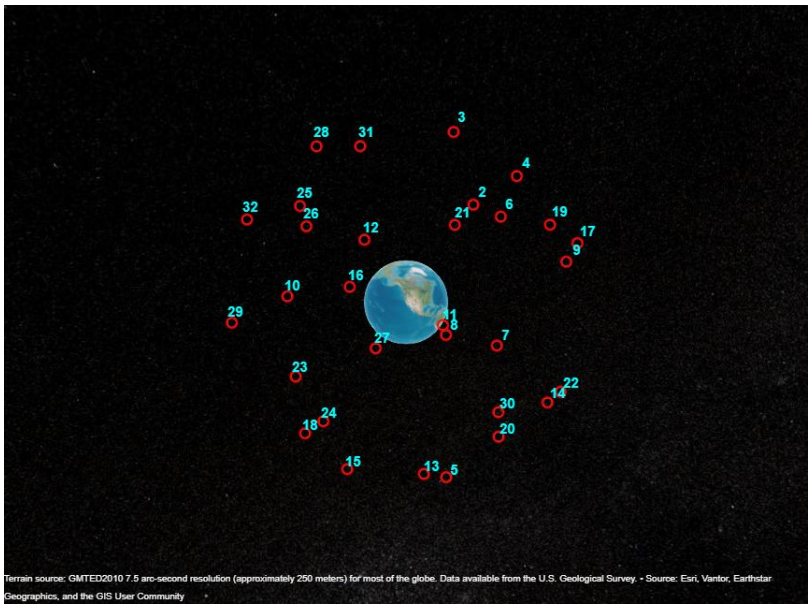


Figure 5.4: Ephemeris simulation from almanac.

Fig. 5.4 shows a snapshot of the ephemeris simulation of the GPS constellation from the almanac. The red circles with satellite numbers represent the corresponding PRN. The almanac provides the ephemeris in ECEF coordinates using a coarse Keplerian orbit model. The accuracy of such orbits are at the scale of $\sim$1 km compared with the precise product. This is not accurate enough for state estimation given the kilometer level of pseudorange errors, but is good enough for the approximations of LOS vector and elevation angle, for which the modeling error of $\sim$1 km is normalized by the GPS orbital altitude of $\sim$20200 km. As a result, the almanac itself is suitable for the simulation of the covariance estimates as well as the computation of the protection levels.

Given the receiver locations from the integrity monitor, the LOS vectors can then be simulated from the almanac using the elevation angle as the cut-off criterion. To some degree, the protection levels computed from such a simulation represent the

lower bound of the protection levels that the geometry of the integrity monitor can achieve. In terms of LOS vector availability, the almanac provides more stable satellite visibility than actual receiver observations, which may include events such as cycle slip, lost of track, or multi-path, etc., that would degrade the available satellites for the receiver. Even though the LOS vectors from the precise product and almanac are almost identical, the covariance computed from the actual receivers will be greater than or at least equal to the one computed from the almanac where the only cut-off criteria for the almanac is elevation angle.

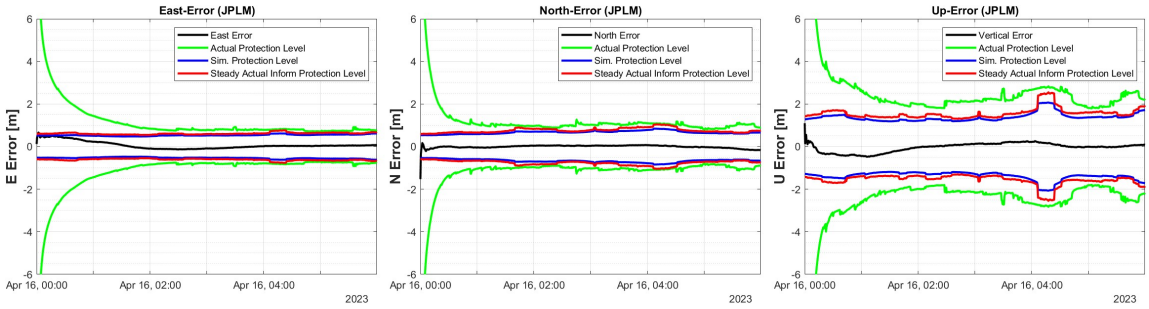
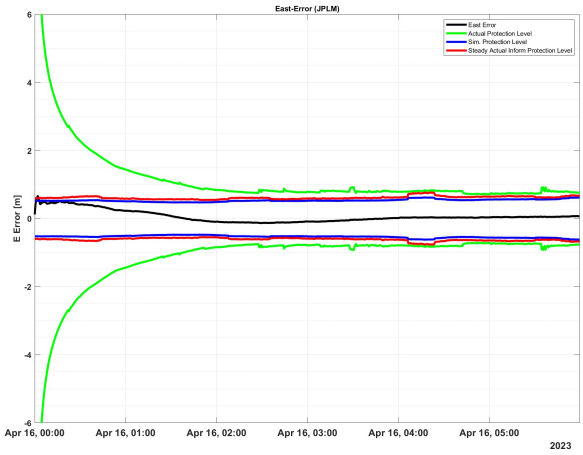


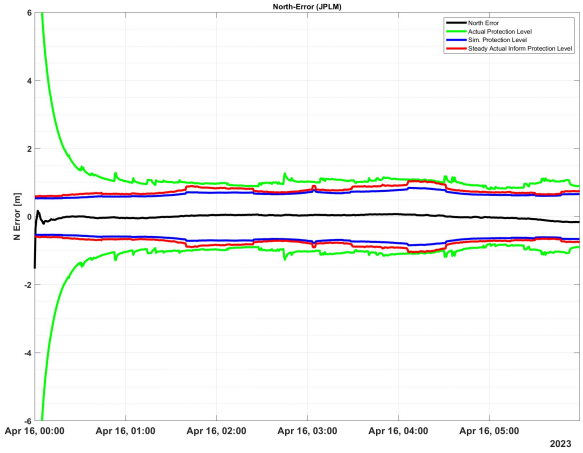
Figure 5.5: True and steady-state protection levels.

Similar to Fig. 5.1, Fig. 5.5 shows the protection levels at JPLM, with additional protection levels computed completely from the simulation. Such protection levels are shown as blue lines, where the ephemeris were generated from almanac, and the integrity information is projected from the covariance of satellite errors approximated by Eq. 5.13. This results in the most optimal estimates of protection levels that the user can achieve. It can be seen that the fully simulated protection levels are the tightest bounds. It is the theoretical lower bound that the user can achieve given the geometries of the satellites and receiver network.

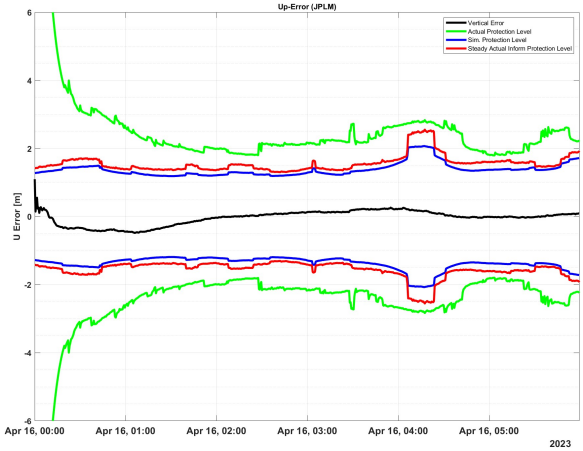
Fig. 5.6 is the larger version of Fig. 5.5 for reference.



(a) East direction.



(b) North direction.



(c) Up direction.

Figure 5.6: Larger true and steady-state protection levels.

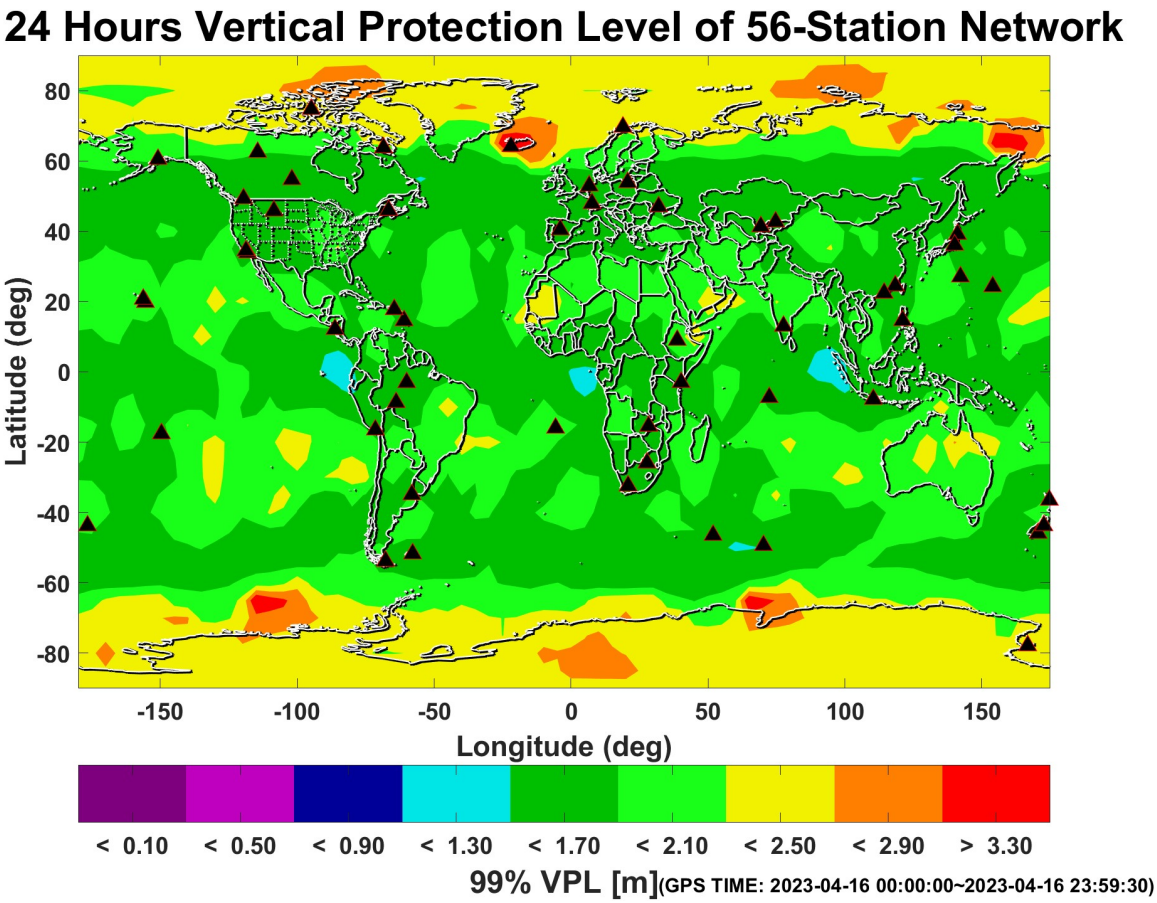


Figure 5.7: 99% Availability Steady-State VPL from Almanac.

Fig. 5.7 shows the VPL map of 99% availability on April 16th with the same network and time as shown in Fig. 5.3. The main differences between these two VPL maps are the sources of the ephemeris product and integrity information. The integrity information is computed from the ephemeris generated by the almanac and the steady-state approximation of the information as described in Eq. 5.13. It can be seen that the VPLs are lower everywhere than those shown in Fig. 5.3. For the network receivers and grid users, the only criterion to limit the available satellites is the elevation angle, whereas in the real data, there are events such as cycle slips, loss of track, multi-path, and temporal correlations that would inevitably deteriorate the performance of the fault-tolerant filter. As a result, the VPLs shown in Fig. 5.7 are inherently the lower bounds to the actual VPLs the user could have achieved with

the real ephemeris.

5.5 Summary

Despite the optimistic nature of the estimation results from simulating the user performance from steady-state solutions, they nevertheless provide insight into the relative performance at different locations within the simulation. For instance, the VPLs in polar regions, Australia, Japan, India, etc., are sill relatively worse than the other regions, which corresponds with the VPLs from the actual data. The simulation provides the optimal estimates of the performance of the integrity network when only focusing on the geometric relations with the satellites. Since real data is not required, it enables efficient evaluation of the performance of arbitrary network combinations anywhere on Earth.

Chapter 6

Network Geometry Design

This chapter is dedicated to the geometric design of the network. The problems to be resolved are simple.

1. Where to place the receivers?

2. How many receivers are needed?

The objective is to design a network geometry such that the performance of the network is optimal, that is, the protection levels of the users of interest are minimized. This problem is formulated as sensor selection under the framework of convex optimization [59].

It is impractical to place receivers in the field to collect GNSS data at the stage of designing the network to evaluate the performance of each design. Therefore, in this chapter the integrity information and ephemeris data are generated from Eq. 5.13 and almanac simulation.

6.1 Introduction

In the scale of the Earth, the placement of the receivers is not sensitive to its exact location. Displacements on the order of tens to a few hundreds of kilometers do not appear to introduce significant differences in the performance of the global integrity

monitor. Therefore, the placement of the receivers for the integrity monitor can be approximated with the selection from a pool of potential available receivers.

Mathematically, the problem can be formulated as the selection of K receivers from N candidate receivers.

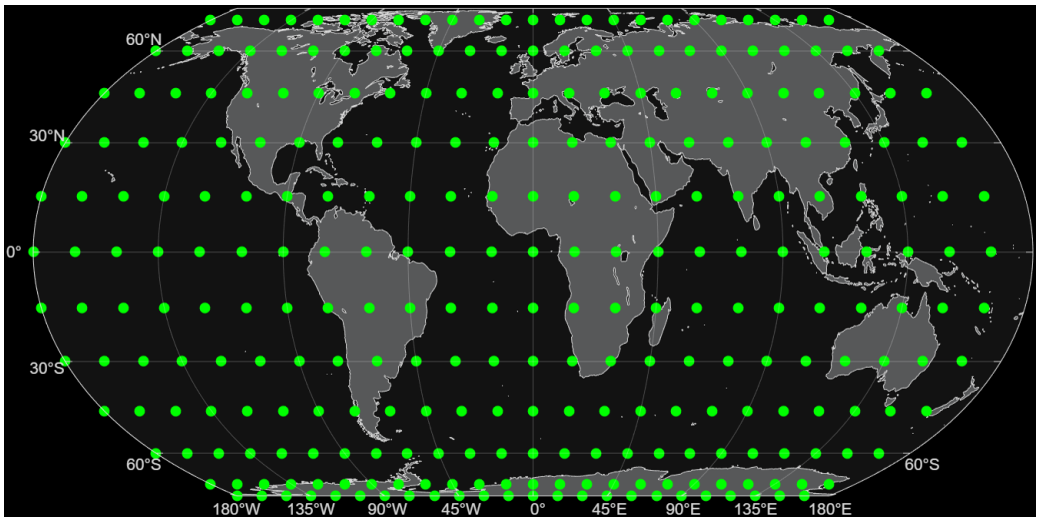


Figure 6.1: 5°×5° Global Grids.

For instance, for a global network size of 60 receivers, one can determine where to put the receivers by selecting 60 out of all the possible receiver locations shown in Fig. 6.1, consisting of grid points every 5 degrees in latitude and longitude. For now, ignore the fact that some parts of the grid might be infeasible for receiver placement.

The naive way of placing the 60 receivers is to iterate through all the combinations possible with the candidate receivers, then compare the performance of each of the combinations as illustrated in Chap. 4 until the resulting protection levels satisfy the requirement. This would involve evaluating $\binom{N}{k}$ combinations. In this example the actual number would be $\binom{312}{60}$. Unfortunately, this is computationally infeasible with today's computing capabilities.

In terms of the number of receivers, one principle is that using more receivers in a network, results in better performance of that network. Ignoring the possibility of faulted receivers, the overall information will only increase as the number of receivers increases. In the most extreme scenario, if the performance of the network does

not satisfy the requirement after iterating through all $\binom{312}{60}$ combinations, i.e., the protection level of certain locations is not low enough, the number of the candidate receivers can be increased by one, followed by re-evaluation of the whole $\binom{312}{61}$ until the performance is satisfied. Similarly, the number of receivers can also be decreased to reduce the load of the network while the desired network performance is met. Since the number of receivers is always an integer, it could be determined incrementally or decrementally by a fixed value.

Although the naive solution of placing the receivers is not feasible, instead of trying to place the receivers on the map, the more systematic way of doing it is first identifying all the possible locations of the receivers for the network, then choosing a certain number from them. This is the main methodology behind the design of the receiver geometry for the ground-based integrity monitor in this thesis. One can always add more candidate receivers to increase the available combinations, but as mentioned earlier, at the scale of the GNSS constellation, the LOS differences are rather insensitive in terms of absolute distances between the receivers.

In the rest of the chapter, the idea of choosing a desired number of receivers will be used as the way of designing the ground-based integrity network.

6.2 Heuristic Receiver Selection Via Monte-Carlo

In addition to choosing the combination of the receivers sequentially, one shortcut is choosing the network at random, evaluating the performance of the combination, and determining if it is to be kept for the next iteration.

This process is called Monte-Carlo [60] and there is no guarantee of producing optimal or even sub-optimal solutions from the receiver combinations that are randomly generated. However, generally after a few million trials the result is decently usable as a network.

The 56 stations introduced in Fig. 2.9 are chosen randomly from 100 million trials from the available IGS stations with some manual adjustments. For instance, if the observation data is unavailable from the chosen station, a nearby one is selected to be part of the network.

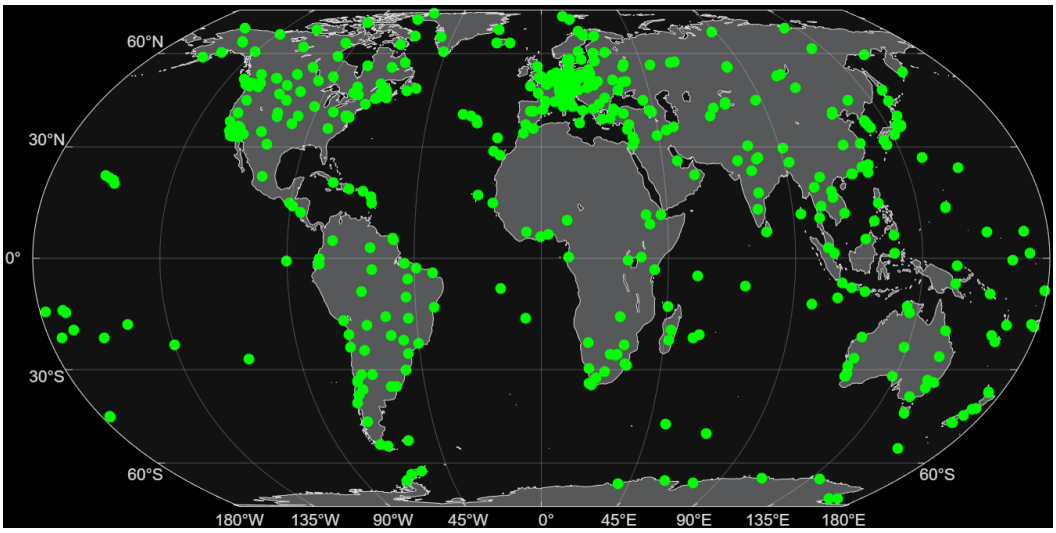


Figure 6.2: 519 IGS stations.

As of 2026, there were 519 stations available (Fig. 6.2), forming a pool of candidate receivers for the integrity monitor design and constraining receiver placement to locations that are physically feasible.

As a reminder, Fig. 2.9 is repeated here showing the 56-station network used throughout the thesis.

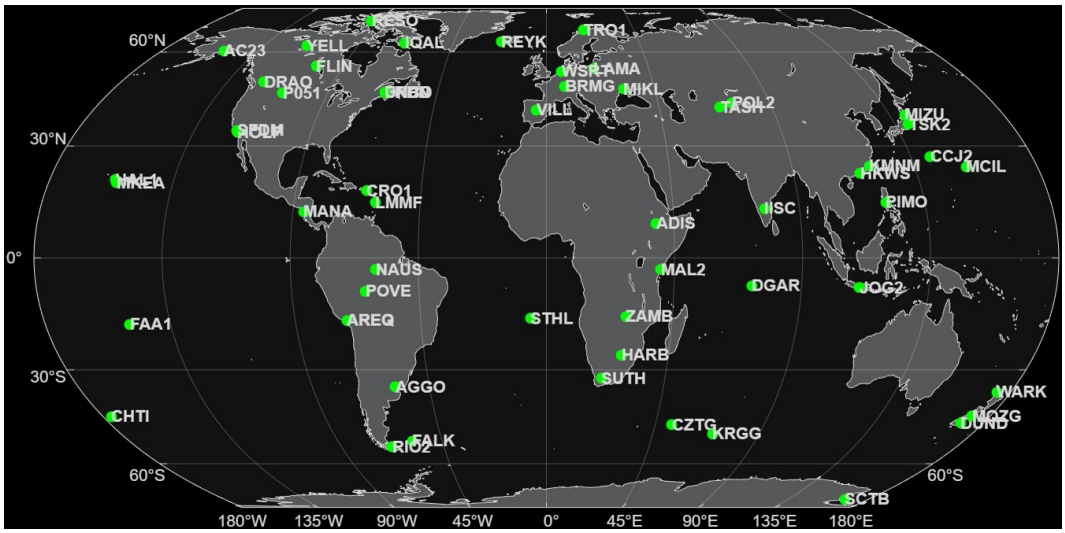


Figure 2.9: 56 selected IGS stations.

Simplified Criteria

Instead of evaluating the VPLs every time a new receiver combination is available, a simplified criteria is adopted, only counting the number of available stations per satellite with the cut-off elevation angle as the LOS. A new combination will replace the old one if the number of available stations are strictly greater for all the satellites over the period of interest. The simplified criteria can reduce the computations and greatly speed up the process of random trials, but will lose the geometric properties.

6.3 Receiver Location Selection Via Convex Optimization

The disadvantage of selecting receivers via the Monte-Carlo method is the optimality. There could be hundreds of millions of random trials that still end up with unsatisfactory performance of the receiver network. There is no guarantee on the optimality of the resulting performance of the receiver network. Since the receiver placement problem can be formulated as a selection problem, convex optimization with a relaxed integer constraint can be employed to determine the optimal receiver configuration [59].

6.3.1 Problem Formulation

The optimization is focused on the geometric relations between the satellites and the receivers, so no temporal relations are considered.

Recalling the steady-state information form of the satellite errors from Eq. 5.13,

an initial form of receiver selection via convex optimization can be written as follows:

$$
\begin{aligned}
&\text{maximize} \quad \textbf{obj.} \\
&\text{s.t.} \quad Y_{sat} = Q_{sat}^{-1} \\
&\qquad + \sum_{r=1}^{n} \mathbf{Z}_r \cdot C_{sat,r}^{T} \left(C_{rcv,r} Q_{rcv,r} C_{rcv,r}^{T} + R_r \right)^{-1} C_{sat,r} \\
&\qquad + C_{sat,const}^{T} R_{const}^{-1} C_{sat,const} \\
&\qquad , 0 \leq \mathbf{z}_r \leq 1 \; \forall \; r \in \{0, 1, \ldots, n-1\} \\
&\qquad , \mathbf{1}^{T} \cdot \mathbf{Z}_r = k
\end{aligned}
\tag{6.1}
$$

where **obj.** is the objective of the optimization problem and is not specified for now, $\mathbf{Z}_r = [\mathbf{z}_0, \mathbf{z}_1, \ldots, \mathbf{z}_{n-1}]^T \in R^{n \times 1}$ is the floating point relaxation of the indicator for the receiver selection ranging from 0 to 1. $\mathbf{Z}_r$ comprises also the design variables for the problem. It controls the contribution of a receiver towards the information matrix of the satellite errors at steady-state from the receiver network, where there are a total of k receivers to be chosen from n receivers. Y_{sat} is the information matrix of the satellite errors with fractional information contribution from each of the receivers.

Eq. 6.1 describes the relation of the satellite errors to all the potential receivers to be selected; this is universal regardless of the choice of design for the problem. However, the objective function is to some extent arbitrary depending on the requirements. Since the overall goal of the integrity network is to have as small a protection level as possible, a straightforward choice of the objective function for the problem would simply be minimizing the protection levels at the target user locations. Unfortunately, the position covariance used to compute the protection levels involves solving Discrete-time Algebraic Riccati Equation (DARE), which includes non-convex operations such as matrix inversion. Although the DARE can be relaxed with the Schur-complement [61], there is no guarantee for this relation to be met given the freedom of inequalities from the Schur-complement. Moreover, the formulation of the Schur-complement makes the problem cumbersome and heavily deteriorates the numerical stability of the solver. Only a small scale set of receiver selection can be solved reliably with the numerical solver. In general, any attempt to perform

matrix inversion with the convex relaxation of the Schur-complement is numerically demanding. The dimensionality and the number of variables must be kept small for the problem to be solved numerically. As a result, instead of directly optimizing the protection levels, the goal is to minimize the measurement noise associated with the users.

Since the sigmas of the UREs from the integrity information are direct additions to the measurement noise of the user filter as described in Chapter 3:

$$R_{\tilde{\rho}^{(k)}_{\mathrm{IF},r}} = R_{\rho^{(k)}_{\mathrm{IF},r}} + {\sigma^{(k)}_u}^2 \tag{3.3}$$

$$R_{\tilde{\phi}^{(k)}_{\mathrm{IF},r}} = R_{\phi^{(k)}_{\mathrm{IF},r}} + {\sigma^{(k)}_u}^2 \tag{3.4}$$

The minimal protection levels the users can achieve occur when the sigmas approach zero:

$$\sigma^{(k)}_u \to 0$$

Therefore, minimizing $\sigma^{(k)}_u$ is proportional to minimizing the protection levels.

To avoid the numerical instability from the Schur-complement, the goal here is to minimize the sigmas of the UREs from the information matrix.

Given the vector $B_{u,k} = [1 \ \overrightarrow{e}^T_{u,k}]^T$ that projects the satellite clock and ephemeris errors from the kth satellite to the user u, the same projection vectors on the covariance and information matrices are:

$${\sigma^{(k)}_u}^2 = B^T_{u,k} \cdot \Sigma_{sat} \cdot B_{u,k} \tag{6.2}$$

$${\alpha^{(k)}_u}^2 = B^T_{u,k} \cdot Y_{sat} \cdot B_{u,k} \tag{6.3}$$

where $\sigma^{(k)}_u$ is the URE sigmas and $\alpha^{(k)}_u$ is the linear protection onto the information matrix that is loosely related to $\sigma^{(k)}_u$:

$$4 \cdot \frac{1}{\kappa} \leq {\sigma^{(k)}_u}^2 \cdot {\alpha^{(k)}_u}^2 \leq 4 \cdot \kappa \tag{6.4}$$

Eq.6.4 can be shown as follows.

κ is the condition number of the covariance matrix Σ_{sat}.

$$\kappa = \frac{\lambda_{max}\left(\Sigma_{sat}\right)}{\lambda_{min}\left(\Sigma_{sat}\right)} \tag{6.5}$$

Since

$$\sigma_u^{(k)^2} \cdot \alpha_u^{(k)^2} = B_{u,k}^T \cdot \Sigma_{sat} \cdot B_{u,k} \cdot B_{u,k}^T \cdot Y_{sat} \cdot B_{u,k} \tag{6.6}$$

and

$$\Sigma_{sat} = Y_{sat}^{-1} \tag{6.7}$$

Given

$$\begin{aligned} \lambda_{min}\left(B_{u,k}^T \cdot \Sigma_{sat} \cdot B_{u,k}\right) \leq B_{u,k}^T \cdot \Sigma_{sat} \cdot B_{u,k} \leq \lambda_{max}\left(B_{u,k}^T \cdot \Sigma_{sat} \cdot B_{u,k}\right) \\ \lambda_{min}\left(B_{u,k}^T \cdot\ Y_{sat} \cdot B_{u,k}\right) \leq B_{u,k}^T \cdot\ Y_{sat} \cdot B_{u,k} \leq \lambda_{max}\left(B_{u,k}^T \cdot\ Y_{sat} \cdot B_{u,k}\right) \end{aligned} \tag{6.8}$$

where $B_{u,k}$ is the linear projection of the clock and ephemeris error to the range domain, and would at most contributes to the multiple of a constant 2 due to the projection of the satellite clock bias and ephemeris error by LOS vector. Eq.6.8 becomes:

$$\begin{aligned} 2 \cdot \lambda_{min}\left(\Sigma_{sat}\right) \leq B_{u,k}^T \cdot \Sigma_{sat} \cdot B_{u,k} \leq 2 \cdot \lambda_{max}\left(\Sigma_{sat}\right) \\ 2 \cdot \lambda_{min}\left(Y_{sat}\right) \leq B_{u,k}^T \cdot\ Y_{sat} \cdot B_{u,k} \leq 2 \cdot \lambda_{max}\left(Y_{sat}\right) \end{aligned} \tag{6.9}$$

Since $\lambda_{min}\left(\Sigma_{sat}\right) = \frac{1}{\lambda_{max}(Y_{sat})}$ and $\lambda_{max}\left(\Sigma_{sat}\right) = \frac{1}{\lambda_{min}(Y_{sat})}$, Eq. 6.9 becomes:

$$\begin{aligned} 2 \cdot \lambda_{min}\left(\Sigma_{sat}\right) \leq B_{u,k}^T \cdot \Sigma_{sat} \cdot B_{u,k} \leq 2 \cdot \lambda_{max}\left(\Sigma_{sat}\right) \\ 2 \cdot \frac{1}{\lambda_{max}\left(\Sigma_{sat}\right)} \leq B_{u,k}^T \cdot\ Y_{sat} \cdot B_{u,k} \leq 2 \cdot \frac{1}{\lambda_{min}\left(\Sigma_{sat}\right)} \end{aligned} \tag{6.10}$$

Therefore, the product in Eq. 6.6 is bounded by the condition number of $\Sigma_u^{(k)}$:

$$4 \cdot \frac{1}{\kappa} \leq C_{u,k} \leq 4 \cdot \kappa \tag{6.4}$$

where

$$\sigma_u^{(k)^2} \cdot \alpha_u^{(k)^2} = C_{u,k} \tag{6.11}$$

Thus, $\alpha_u^{(k)^2}$ is the reciprocal of the URE variance with a multiplier that is bounded by the condition number of $\Sigma_u^{(k)}$:

$$\alpha_u^{(k)^2} = \frac{C_{u,k}}{\sigma_u^{(k)^2}}$$

That is, maximizing $\alpha_u^{(k)^2}$ is equivalent to minimizing the URE variance, $\sigma_u^{(k)^2}$, weighted by the positive $C_{u,k}$ that is bounded by the condition number of $\Sigma_u^{(k)}$, and is therefore equivalent to minimizing the protection levels for the users.

Since the condition number of the outer product of the projection vector, is composed of the LOS vector given the geometry between the satellites and the users, the satellites with better geometry with the users will generally have LOS more well-conditioned. As a result, $C_{u,k}$ weights more on the satellites that have poor geometry with the users, which are exactly the satellites that would contribute to poorer protection levels for the users.

The process of receiver selection can thereby be formulated as a complete convex optimization problem.

$$
\begin{aligned}
\text{maximize} \quad & \min\{{\alpha_{u,t}^{(k)}}^2\} \\
\text{s.t.} \quad & Y_{sat,t} = Q_{sat,t}^{-1} \\
& \quad + \sum_{r=1}^{n} \mathbf{Z}_r \cdot C_{sat,r}^T \left(C_{rcv,r,t} Q_{rcv,r,t} C_{rcv,r,t}^T + R_r\right)^{-1} C_{sat,r,t} \\
& \quad + C_{sat,const,t}^T R_{const}^{-1} C_{sat,const,t} \\
& \quad , 0 \leq \mathbf{z}_r \leq 1 \; \forall \; r \in \{0, 1, \ldots, n-1\} \\
& \quad , \mathbf{1}^T \cdot \mathbf{Z}_r = k \\
& \quad , {\alpha_{u,t}^{(k)}}^2 = B_{u,k,t}^T \cdot Y_{sat,t} \cdot B_{u,k,t}
\end{aligned}
\tag{6.12}
$$

where the objective to be maximized is the minimal of the weighted reciprocal of the projected URE variance at the location of the target user u at time t. The maximization process connects the projection of the URE variance with different time steps over a period of time such that the receivers selected include the temporal relations of the satellite ephemeris. Finally, the receivers with the top k values of $\mathbf{z}_r$ are selected to be the network receivers.

6.3.2 Selection Result

This section demonstrates the results of selecting 56 stations out of the candidate receivers for the integrity network by maximizing the objective and constraints specified in Eq. 6.12. The constraint optimization problem is solved in Matlab using CVX, a package for specifying and solving convex programs [62] [63].

Evenly Spread Candidate Receivers

The pool of candidate receivers are subject to the design of the network. They represent the locations where reference stations are feasible to be placed and operated for the integrity monitor.

As shown in Fig. 6.1, while the 5°×5° global grids provide full coverage to the surface of the Earth for the candidate receivers, the spacing between the receivers are not uniform due to the curvature of the earth. The candidate receivers are concentrated in higher latitudes, creating an inefficient distribution that does not fully capture the spatial geometries of the potential network during the optimization process.

Therefore, in this example, 56 stations are selected from a pool of 200 candidate receivers that are evenly spread-out across the surface of the Earth. The spatially uniform candidate receivers are shown in Fig. 6.3 and their locations are generated using uniform sampling techniques [64], where the random seed is set to one to ensure the distribution is deterministic throughout the optimization. Whether the receivers are physically able to be placed at the candidate locations is ignored for now. The objective of this example is to demonstrate the optimal design of a global network by efficiently selecting from the $\binom{200}{56}$ possible combinations via convex optimization.

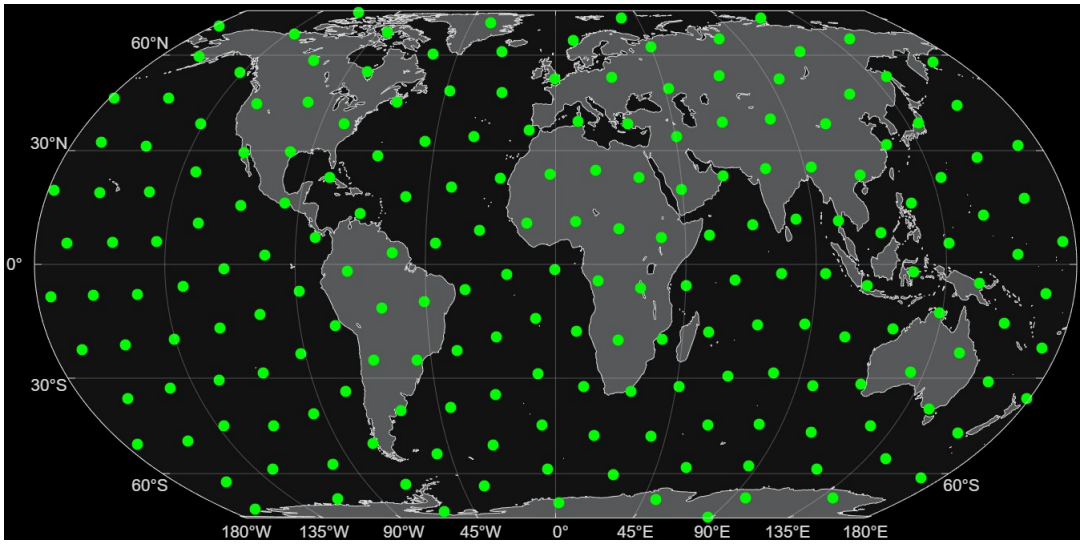


Figure 6.3: 200 evenly spread-out candidate receivers.

In addition to determining the candidate receivers, the target user locations must also be specified in the actual optimization problem. The user locations are the points to which the satellite errors are projected. Essentially, Eq. 6.12 is only minimizing the protection levels of the user locations that are specified by $B_{u,k,t}$, where

a predetermined user location is required to do the projections. Therefore, depending on the target area of service, for instance, similar evenly distributed locations can be used as target users when the goal is to equally minimizie the protection levels around the world. Alternatively, the target users can be specified based on the service requirements that could be locally concentrated to specific regions. In this example, a global service is desired so 80 evenly spread-out target users are adopted.

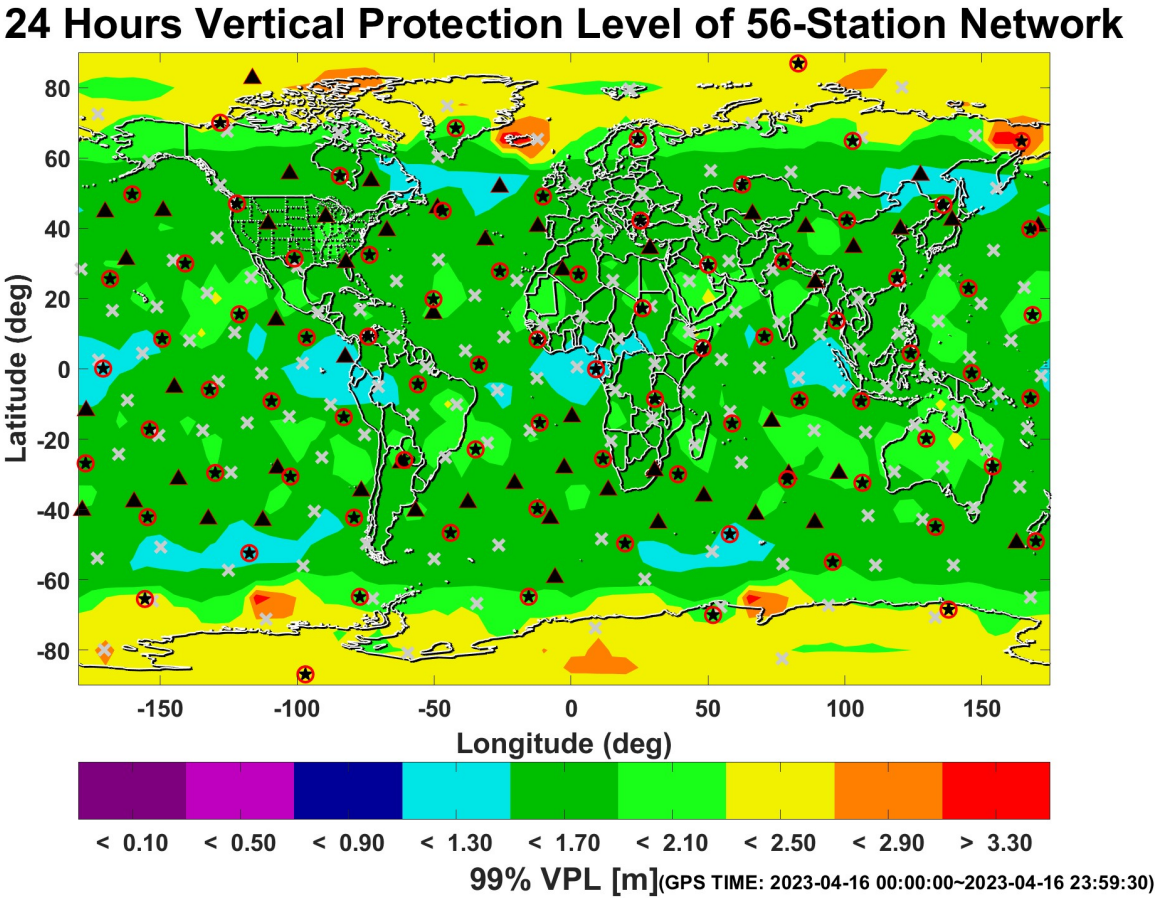


Figure 6.4: Evenly spread-out Receiver Selections

Fig. 6.4 shows the VPLs of 99% availability for the network size of 56 receivers chosen from solving Eq. 6.12. There are 200 uniformly distributed candidate receivers and 80 uniformly distributed target users in the problem. The black asterisks within the red circles represent the target users, the black triangles are the 56 receivers selected, and the gray crosses are the other 144 receivers that are not selected by

the optimization process. The network is obtained by optimizing over a period of 24 hours with 10 minute intervals using ephemeris data on April 16th.

In comparison with Fig. 5.7, the VPLs are generally smaller than the original 56-station network as shown in Fig. 6.4, generated using the Monte-Carlo method. Without changing the number of receivers in the network, the selection results give the new geometries of the receiver network that can provide smaller protection levels than the old design when ignoring the temporal effects.

IGS Candidate Receivers

In this example, the candidate receivers are selected from the IGS network rather than from uniformly distributed locations, therefore ensuring that the receiver sites are physically realizable for station deployment.

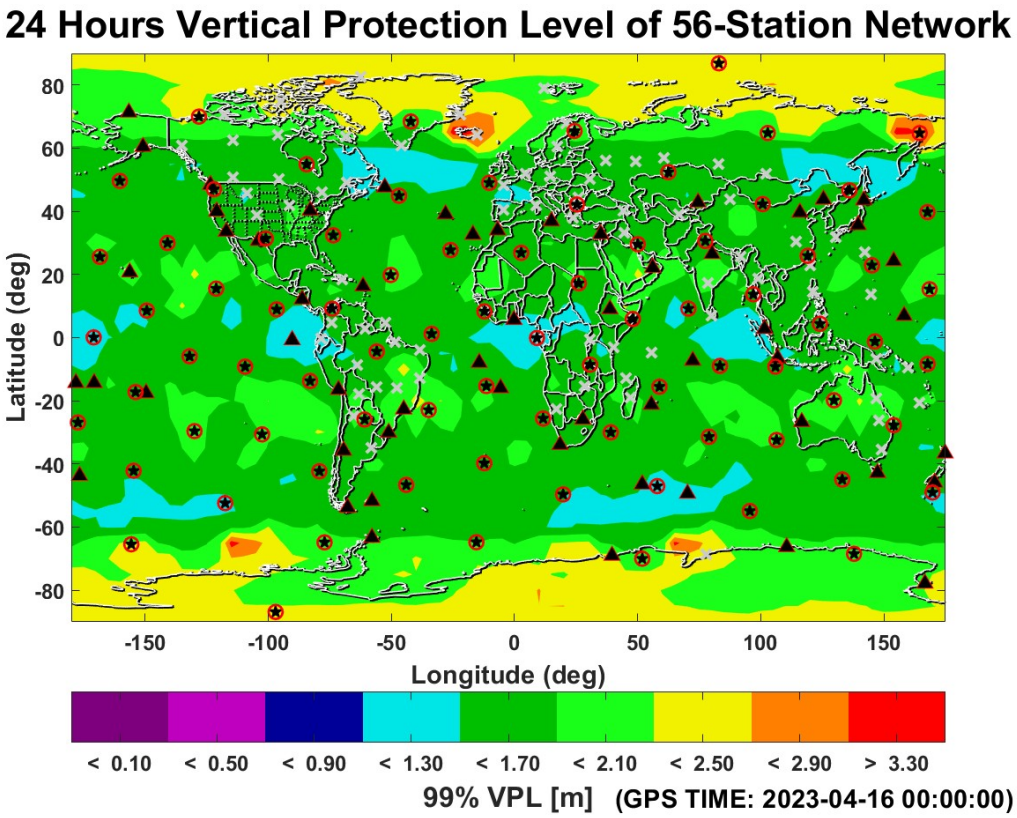


Figure 6.5: IGS Receiver Selections

Fig. 6.5 shows the VPL of 99% availability of the 56 receiver network chosen from the IGS network. To avoid falling into clustering receivers when selecting the receivers, excessive IGS stations are removed to make sure each of the candidate receivers are at least 700 km apart in distance. This reduces the total number of candidate receivers to 133, resulting in a total of $\binom{133}{56}$ possible combinations. The VPLs are also generally smaller than Fig. 5.7, and are comparable with Fig. 6.4.

6.3.3 Analysis of Coverage

While the availability map reflects the spatial distribution of the protection levels, it is difficult to quantitatively compare the coverage of the service. Depending on the requirement, the service provider of the integrity monitor might prefer smaller

protection levels at specific service areas and ignore other places.

In the examples, the object is to provide an integrity monitor with global coverage that treats users equally all over the world. Thus, a simple criteria to compare the overall performance of different geometric designs of the network is the percentage of protection level coverage.

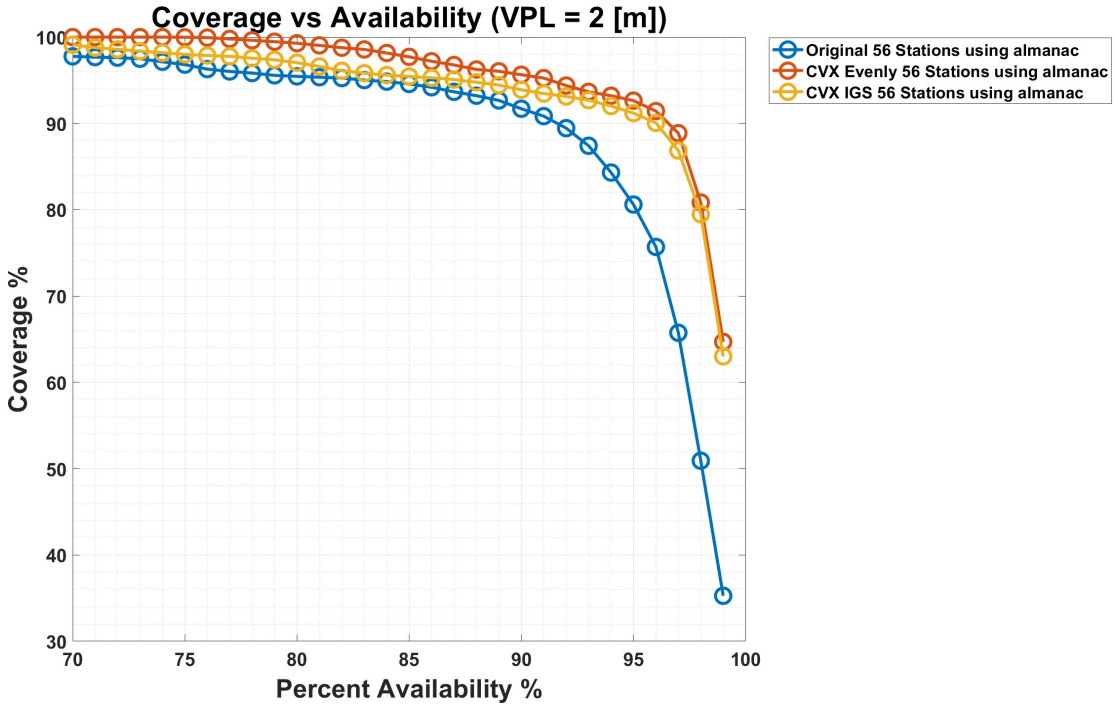


Figure 6.6: Coverage vs. Availability for 3 different networks with 56 receivers.

Fig. 6.6 is the coverage vs. availability of three different networks, where a 2 m threshold is chosen as an example. The three networks correspond to the one selected by the Monte-Carlo method, Fig. 2.9, by convex optimization with uniformly distributed candidate receivers, Fig. 6.4, and by convex optimization with candidate receivers from the IGS network, Fig. 6.5. In Fig. 6.6, the percentage of the coverage is defined as the percentage of the surface area on earth that has VPL less than 2 m given the percentage of the availability. For instance, at 99% availability, the blue line indicates that Fig. 2.9 has 35.3% of the surface under 2 m VPL, the orange line indicates Fig. 6.4 has 64.7% of the surface under 2 m VPL, and the yellow line

indicates Fig. 6.5 has 63.0% of the surface under 2 m VPL. It can be seen that the network selected from uniformly distributed candidate receivers performs slightly better than the one selected from the IGS network, and both of them outperform the original 56-station network selected from the Monte-Carlo method. This plot provides a direct performance comparison of different network designs in terms of global coverage at various availability levels.

6.4 Receiver Number Selection

The determination of the number of receivers becomes straightforward once the process for determining receiver locations is streamlined. The more receivers included in the network, the more information is collected, resulting in better performance.

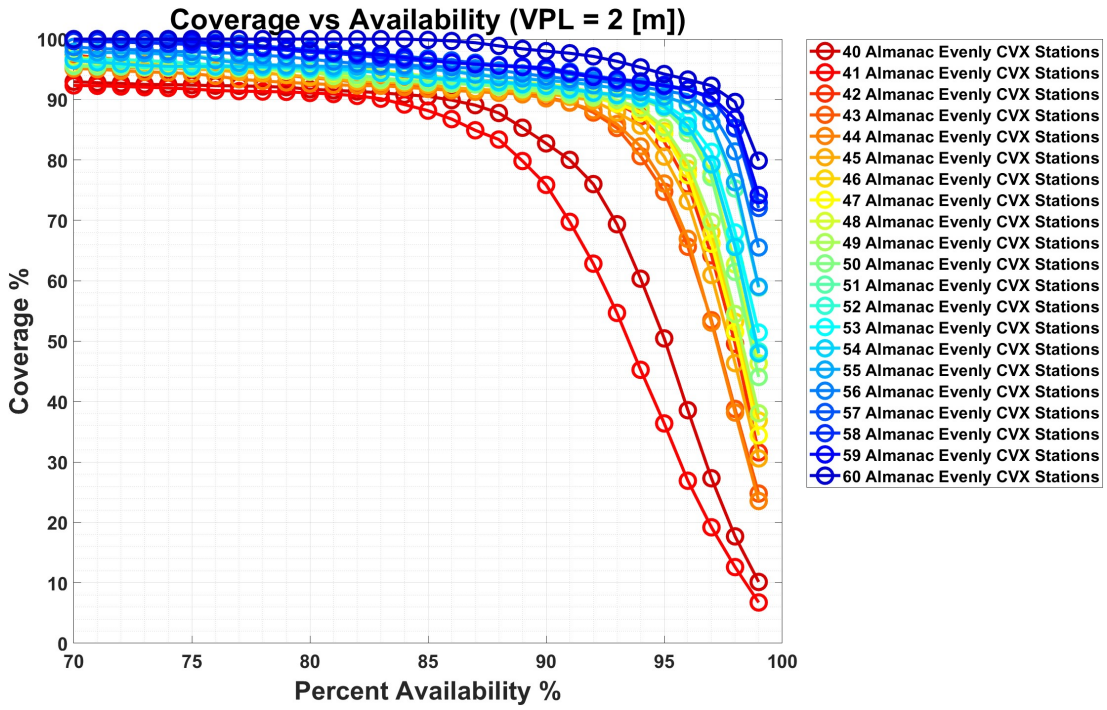


Figure 6.7: Coverage vs. Availability for evenly spread-out network with receiver number ranges $40 \sim 60$.

Fig. 6.7 illustrates this principal, where the evenly spread-out 200 candidate receivers and 80 target users are re-used as a demonstration. The only variable here

is the number of desired receivers, k, in Eq. 6.12. Each line is the solution from solving the same convex optimization problem, but with different k values ranging from 40 to 60, representing the size of the desired network. It can be seen that coverage of the network is better with more receivers.

Thus, the process of determining the receiver number can be accomplished by simply enumerating the solutions from a range of acceptable numbers, and choosing the one that satisfies the requirement. Since the cost and probability of receiver failure (Eq. 4.1) increases along with the number of receivers, this becomes a trade-off between the performance (more receivers are better) and operational limit (fewer are better), both of which depend on the design requirement.

The solutions of Eq. 6.12 with different receiver number, k, may not strictly monotonically increase due to the relaxation of the integer constraints, but as shown in Fig. 6.7, the performance of the network is positively related to the number of receivers.

One can always add additional receivers to an existing network, which guarantees monotonically increasing performance. However, this heuristic approach may lead to suboptimal performance as more receivers are added due to the incremental extension of a pre-existing network.

6.5 Summary

The design of the receiver network for the ground-based integrity monitor is simplified to solving the issues of where to place receivers and how many receivers are needed.

For the first problem, two design methods are introduced: Monte-Carlo and Convex Optimization. The major distinction between the designs of Monte-Carlo and convex optimization is that the former is stochastic and has no guarantee on the optimality of the selection results, and the latter is deterministic where the solutions based on integer relaxation and convex optimization outperforms the Monte-Carlo approach. While the globally optimal solution might be attainable by the Monte-Carlo method, the convex optimization provides a consistent optimal selection of the network receivers given the geometries and approximation of the integrity information

formulated in Eq. 6.12, where it is indirectly minimizing the protection levels of the target users. Furthermore, convex optimization provides the flexibility for the designers to design the network subject to different requirements using different formulations of the constraints, such as assigning target users at different locations or with various weightings. Therefore, convex optimization is preferred for designing the receiver network of the ground-based integrity monitor during the initial phases of the overall network design process.

Once the process of determining the receiver locations is accomplished, determining the required number of receivers becomes relatively straightforward. It can be obtained by selecting the design with the smallest number of receivers that satisfies the performance requirement from the range of designs produced by solving Eq. 6.12.

Chapter 7

Tropsphere Error Analysis

7.1 Introduction

The SBAS MOPS [49] regulate the operation and application of the SBAS correction signals. The definitions of the protection level equations is essential to the integrity of the SBAS users. Currently, the L1 SBAS MOPS considers tropospheric delay errors as uncorrelated across satellites when calculating the protection level for the position solution. It was acknowledged that the tropospheric errors are almost always correlated since tropospheric delay on the user side primarily only differs by the line-of-sight elevation angle to the satellites. Although MOPS implemented the protection level as if the tropospheric error is always uncorrelated, this simplification has served well for the standard requirement when the Alert Limit (AL) is above 35 m. However, this assumption might not always be sufficient when for Vertical Alert Limit (VAL) below 10 m.

Additionally, the correlations of the tropospheric errors are local to the proximity of the users, which cannot be measured nor corrected by the reference stations of the integrity monitor. Thus, to ensure the integrity that the ground-based monitor provides to the users are not threatened by the tropospheric errors, the magnitudes of the tropospheric errors are needed to be analyzed. This is to provide guarantee to the users of the integrity monitor from the threat of the tropospheric errors.

This chapter focuses on the VPL equations for the L1 single frequency, which is

already well established in the L1 SBAS MOPS, [65] using the tropospheric error analysis from dual frequency SBAS and Advanced Receiver Autonomous Integrity Monitoring (ARAIM).

Three standard tropospheric models were evaluated for accuracy, UNB3 [35], GPT2W [66], and GPT3 [67], against 10 years of tropospheric delay data collected and processed by IGS. The data were used to determine the paired-bounds statistics that bound the model error distribution over the past 10 years. Without changing the nominal tropospheric correction model, an alternative form of the VPL equation was proposed and standardized in the dual frequency SBAS MOPS. This new equation incorporated the effects of correlation and bounded the modeling error associated with the adopted UNB3 tropospheric model in MOPS. [68] identified that the simplification of neglecting the tropospheric correlation among the satellites could lead to error in VPL bounding, and [69] previously identified potential improvements to tropospheric modeling and bounding for improved performance.

7.2 Troposphere Delay Models: UNB3, GPT2W, and GPT3

Two main components in the troposphere cause delay to the GPS signal, the hydrostatic delay and the wet delay. The first part, hydrostatic delay, or dry delay, is the delay induced by the dry gases in the troposphere (78% N_2, 21% O_2...). The dry part of the troposphere typically causes 2.3 $\sim$ 2.6 m of delay to the signal in the zenith direction, and up to 20 m in lower elevation angles. The second part, wet delay, is the delay caused by water vapor and condensed water. This part of the delay varies from 0 $\sim$ 40 cm in the zenith direction, and 0 $\sim$ 2 m in lower elevation angles. The dry component is considered to be stable due to the fact that dry gases vary predictably through years and locations. On the other hand, the wet component is highly stochastic in the sense that water vapor and condensed water can change from 0 to 100% humidity in a very short period of time, making it difficult to model. Thus, the wet delay behavior is relatively unpredictable compared to the dry delay and is

the main source of randomness when estimating tropospheric delay. So, even though the dry delay is several times greater than the wet delay in most cases, the wet delay is the dominant contributor to the uncertainties of tropospheric delay estimation when applying tropospheric models.

The Mapping function is another important part of tropospheric models. It maps dry and wet delay components from the zenith direction to the LOS direction. Since GPS signals travel farther in the troposphere if the satellite is at low elevation angle, the signals stay longer inside the troposphere and more delay is incurred. As a result, the lower the elevation angle, the greater the mapping function in order to account for the additional delay. Although satellites are all at different locations and thereby different elevation angles, the tropospheric delay is all concentrated to a small region around the receiver, resulting in strong correlation. This is the relation simplified in the current VPL equation.

The tropospheric delay models examined in this chapter (UNB3, GPT2W, and GPT3) have similar structure. First, the models predict the dry and wet delay components in the zenith direction, then compute mapping functions for both components in terms of the satellite locations, then apply mapping functions to the zenith values, eventually combining them to obtain the slant total delay (STD). The UNB3 [35] model is adopted by SBAS MOPS as their tropospheric delay model for correction broadcast. It relies on a simple 15° latitude-wise lookup table for estimating parameters such as barometric pressure, temperature, and water vapour pressure. The resulting parameters are then applied to an improved Sasstamoinen model [70] to obtain dry and wet delays in the zenith direction, and the Black and Eisner Mapping function [71] is applied for the mapping function values. GPT2W [66] and GPT3 [67] were developed for the purpose of geodesy application, and were designed for accuracy. They use a $1^\circ \times 1^\circ$ grid map built from 10 years of empirical data to estimate necessary parameters such as pressure, temperature, and mapping factors, and more complicated mapping functions. GPT3 is the successor of GPT2W. The GPT series are typically more precise than UNB3 but require more computations and larger storage space for the table.

7.3 The Dataset and Residuals

In 2022, the International GNSS Service [1] had around 400 GNSS receiver stations world-wide and provided free access to various GNSS products (Fig. 7.1). The number of stations available may vary from day to day due to new stations introduced, old stations retired, or station maintenance. In the experiment, every station available from year 2012 to 2021 is included to verify the effectiveness of the Gaussian paired-bound established. The tropospheric Zenith Path Delay (ZPD), or Zenith Total Delay (ZTD), is the GNSS product examined in the experiment. ZTD is the summation result of both dry and wet delay in the zenith direction from the corresponding stations, and is treated as the truth in the analysis of tropospheric models.

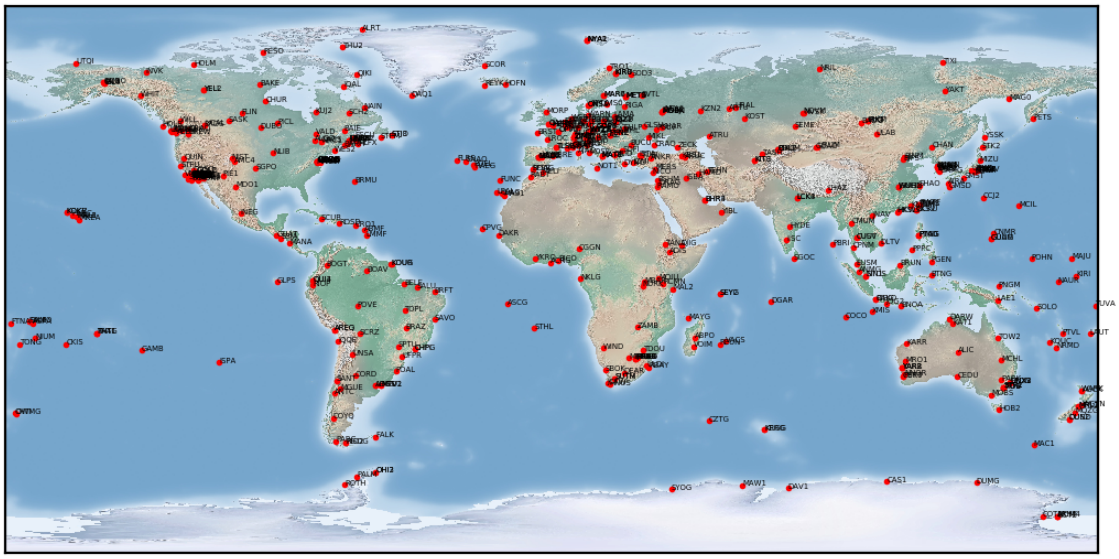


Figure 7.1: IGS Map and stations in 2022. [1]

Tropospheric zenith residuals are defined as the difference between ZTD from IGS and ZTD predicted by models:

$$R_{model} = ZTD_{Model} - ZTD_{IGS}$$

where R_{model} is the residual of the model, ZTD_{Model} is the ZTD estimated from the model, and ZTD_{IGS} is the ZTD data from IGS.

In some infrequent cases, the magnitude of the residuals exceed 28 cm, the maximum deviation of the UNB3 model used in MOPS. This is sometimes due to a reset of the IGS filter, which may take time to converge to the actual value from an initial guess, or sometimes due to losing track of the signals, unusual weather conditions such as a hurricane, extremely dry or wet day, monsoon season, or lack of maintenance to the receiver itself. Weather reports are reviewed on days with abnormal residual values, in particular, the days with residual magnitudes greater than 28 cm since the ZTD value is expected to be within a certain reasonable range. For example, even in the most extreme cases, where the temperature is 60° Celsius, pressure is at the extreme limit, and humidity is 0% or 100% saturated, the possible tropospheric delay value that can be observed on earth is bounded. Without loss of integrity to the tropospheric delay value, weather reports on those anomalous days are used as reference, and measurements are excluded if the weather reports show nothing abnormal.

In addition to IGS data, the Jet Propulsion Laboratory (JPL) produces a similar tropospheric delay product using GipsyX [72], in their Making Earth System Data Records for Use in Research Environments (MEaSUREs) program [73]. While the number of stations (around 1800 stations) JPL processes is significantly greater than IGS, the majority of the stations are located in the North American region. So, JPL data are a better choice for evaluating the North American region but the global analysis performs better using IGS data. Given global station distribution across the two data sets, the JPL data serve as supplemental verification to IGS data, thus limiting the scope of cross-validation on IGS stations.

7.4 Residuals Analysis

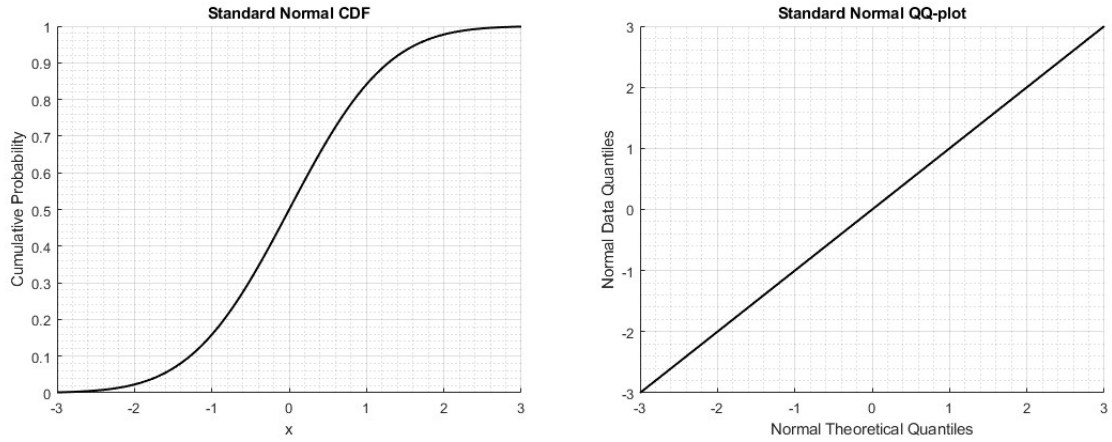
7.4.1 The Grouping Criteria

IGS stations are grouped every five stations and every 5° latitude, meaning each group has at most 5 stations and lies across at most 5 degrees in latitude. The purpose is to look for a dependency on latitude. The reason for grouping stations in this way is

to retain a minimum number of stations per group and location without having an excessive number of stations concentrated in one single group. Note that IGS stations are located in $30°N \sim 50°N$, thus necessitating this approach.

7.4.2 Quantile-Quantile Plot

A Quantile-Quantile Plot (QQ-plot) is a way of representing data distribution in terms of the quantile of data and quantile of a reference distribution. If the data present on the QQ-plot is a straight line with slope of one, the data is distributed the same as the reference distribution. A common choice for reference distribution is standard normal distribution.



(a) Normal Empirical CDF (b) QQ-plot

Figure 7.2: Normal CDF and QQ-plot

Fig. 7.2a is a conventional representation of CDF, having the often seen S-curve. Fig. 7.2b is an example of a QQ-plot. For standard normal quantiles, the slope of the line represents standard deviation, and displacement in the vertical direction at zero quantile represents mean. The QQ-plot provides a more convenient way of determining the paired-bound condition for empirical data. The Gaussian paired-bound is a pair of straight lines in the QQ-plot and the bounded region becomes a simple interval between them. In the rest of the chapter, the QQ-plot is treated as the default representation of the tropospheric residual distribution.

7.4.3 Residual Distribution

The residual distribution varies little throughout the years, so in this section the data in 2021 is examined.

UNB3

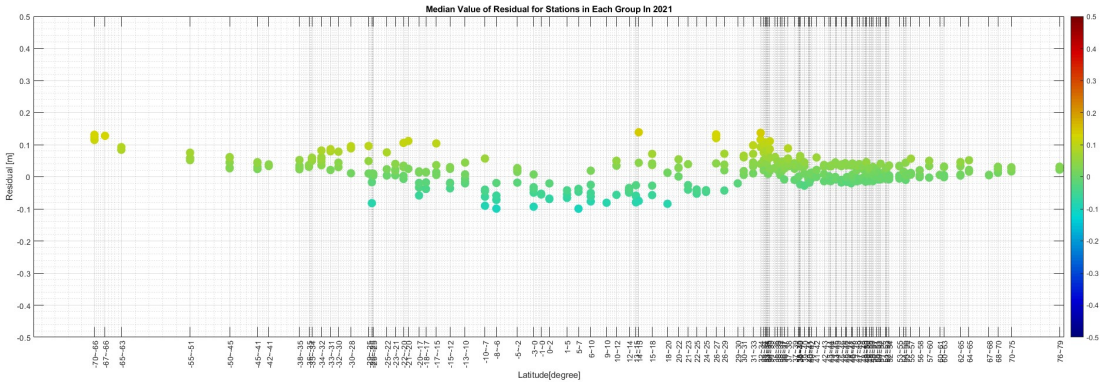


Figure 7.3: Median residuals of UNB3 model in 2021.

Fig. 7.3 shows the median (i.e., residual with half cumulative probability) of UNB3 at all IGS stations in 2021. The medians are grouped in latitude and are displayed in scale with the actual latitude. The medians are densely compact in latitudes $30° \sim 50°$ because the majority of IGS stations are located in these latitudes. This region is also where most of the data used to create the UNB3 model was taken from, so it can be seen that UNB3 is relatively stable in this region and above, with the medians generally smaller than 10 cm. There are fewer stations below 30° south, and the medians are less stable and less accurate as in the southern hemisphere, with many stations having medians between 10 cm and -10 cm. There are only 25 stations located from $-40° \sim -70°$. The medians rise as the latitude becomes lower, and are all greater than 10 cm for stations below $-65°$ (stations in Antarctica). The reason for this is UNB3 appears to have simplified the model by assuming symmetry to weather environments in the northern and southern hemispheres, in the sense that the temperature and pressure conditions are similar in the same latitude region for

each hemisphere. This is a reasonable approximation as the medians are in general distributed symmetrically with respect to 0°, the equator. However, the symmetry does not hold for latitudes below −65° where the Antarctica continent is located. Despite the symmetric approximation, all stations have medians less than 17 cm and greater than -10 cm. The median values of stations are an essential factor to the paired-bound of the residual distributions as they limit the minimal value of the bounding bias.

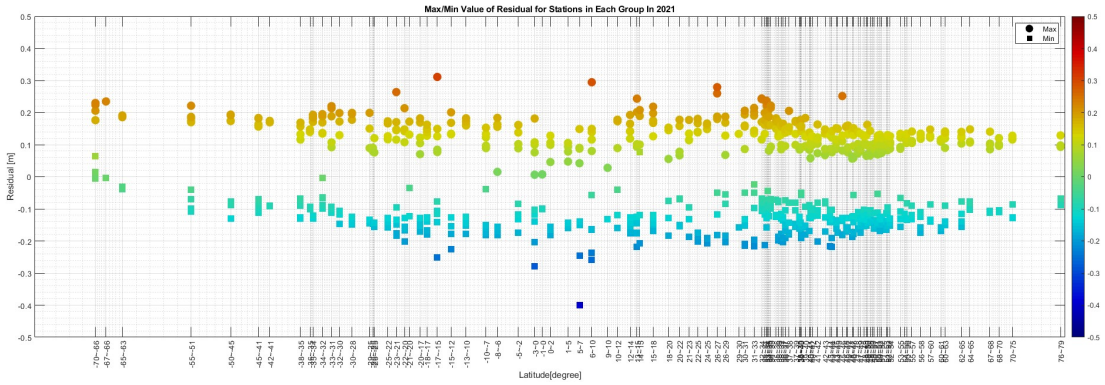


Figure 7.4: Maximal and Minimal residuals of UNB3 model in 2021.

Fig. 7.4 shows the maximal and minimal residuals for all stations in 2021, where the circle represents maximum and the square represents minimum. The extremum indicate the worst estimation of UNB3 at the stations, and the widely-spread distribution.

Here it can be observed that there exist multiple stations having maximal residual as high as 30 cm, and minimal residual down to -40 cm, which are unusually large residual error magnitudes. The definitive parameters to tropospheric delay include pressure, temperature, humidity, and altitude. For one location to have -40 cm residual, it demands extreme weather conditions such as low pressure, low humidity, and high altitude, and there are limitations on how low a tropospheric delay can be. For instance, the humidity cannot go below complete unsaturation or above complete saturation, so if a station reports unusually low or high residuals, there must be an extreme weather event happening nearby the receiver. Consequently, the weather

records is evaluated on the days the abnormal residual is computed. Data points are excluded from the computation of the residual distribution if the weather record reports a nominal result. The station with -40 cm residual in 2021 is ACRG, a station located in Accra, Gahna, where the receiver is 83 m above sea level. The weather record shows fairly moderate weather conditions on that day, April 20th, 2021. The temperature ranged from 79°F $\sim$ 93°F, humidity from 20% $\sim$ 75%, pressure from 29.58 inch $\sim$ 31.53 inch, and no precipitation. From this point of view, the day with the -40 cm residual is no different from the days before and after, when the residual magnitudes were well within 28 cm. Thus, it is concluded that the -40 cm residual does not reflect the true residual and should be removed from the data set. The cause of this failure has not been identified but may be due to filter reset, loss of signal, or device glitch.

Without loss of integrity, weather records are evaluated for the days on which station residual magnitudes exceeded 28 cm. Data points are then excluded if nominal temperature, humidity, and pressure were present. For some stations where JPL data are available, the JPL data are used to cross-validate the IGS data along with the weather record. 28 cm is the threshold selected and is not the basis for data rejection. The basis for rejection is when the weather conditions were not extreme, the filter behavior is unusual, and it does not match a cross-comparison when available. Other numbers such as 25 or 22 cm can be easily adopted for subsequent analysis.

In short, if the residual magnitude is greater than 28 cm, the data will be evaluated and excluded if the corresponding weather record is normal and the residual magnitude is not consistent with the day before and after, otherwise they will not be excluded.

There are three days of data containing residual magnitudes greater than 28 cm in 2021 for UNB3: STHL on January 26th, ACRG on April 20th, and SGOC on April 22th. The weather record suggests all three of these should be excluded from the computation of residual distribution.

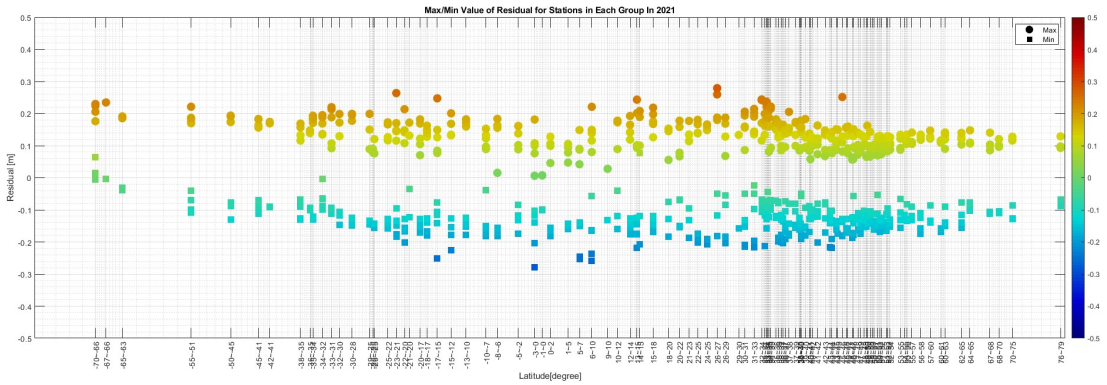


Figure 7.5: Maximal and Minimal residuals of UNB3 model in 2021 after removal of data.

Fig. 7.5 shows the maximal and minimal residuals of all IGS stations in 2021 after data exclusion. As shown in Fig. 7.3, the extreme residuals are symmetric about the equator with the stations in Antarctica biased by about 10 cm compared to stations in the north. The stations in the northern hemisphere are concentrated and relatively stable, with only a few stations having residual magnitudes greater than 20 cm. The extrema affect the bounding in head and tail regions, which may amplify the bound since it is necessary to enclose the residual distribution on the entire residual space. A few extreme cases could enlarge the bias or variance of the bound so as to completely include the residual.

GPT2W and GPT3

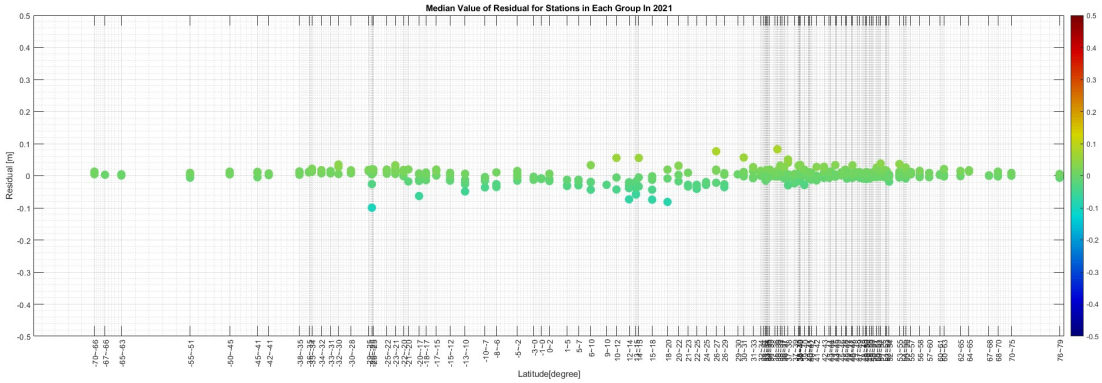


Figure 7.6: Median residuals of GPT2W model in 2021.

Fig. 7.6 shows the median residual of GPT2W in 2021. The medians are more consistent and closer to zero compared to UNB3 in Fig. 7.3. Since GPT2W relies on $1^\circ \times 1^\circ$ empirical grid point data for parameters such as temperature and pressure, there is no symmetric error issue for stations in the Antarctica region, and the median magnitudes are all within 10 cm. This property implies that the minimal bias for bounding GPT2W residuals is consistent throughout different latitudes, and the resulting overall bound is less conservative for use in different locations.

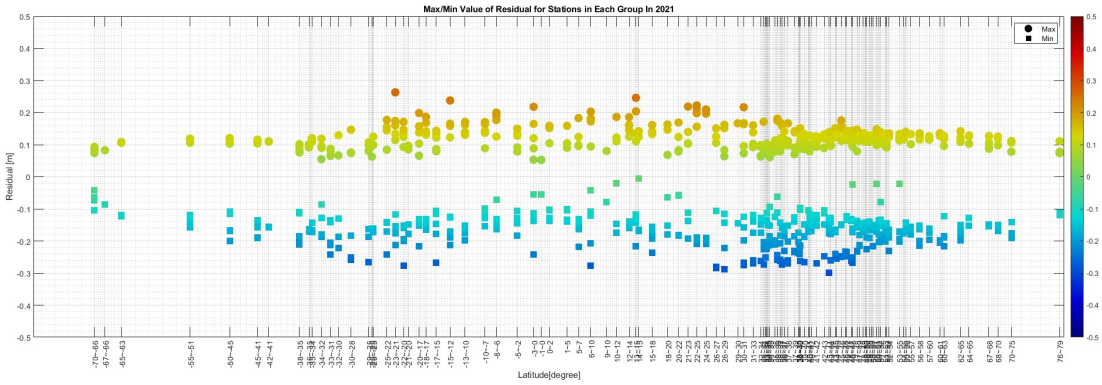


Figure 7.7: Maximal and Minimal residuals of GPT2W model in 2021.

Fig. 7.7 shows the maximal and minimal residuals for GPT2W in 2021 after

excluding outlier data points. Note there are some stations with minimal residual less than -28 cm. The weather records have been evaluated and the data evolved continuously before and after, so it is concluded that the tropospheric delay estimation from the IGS station on those days was reflecting the real situation and should not be excluded from the computation of the distribution. The maximal residuals from GPT2W are more stable and concentrated than the minimal residual. The latter is more scattered and distributed in a way similar to UNB3, which could potentially result in greater paired-bound variance in order to bound the tail distribution. As for the Antarctica region, the distribution of the extrema are more concentrated.

The GPT3 model is the successor of GPT2W with an improved mapping function and empirical grid points for interpolation. However, the resulting residual distributions is nearly identical to GPT2W so the median and extrema residual distributions are omitted here for simplicity.

For data from 2012 $\sim$ 2020, the residual is not shown here for the same reason. The complete 10 years of residual distribution is shown in the paired-bound in the following section.

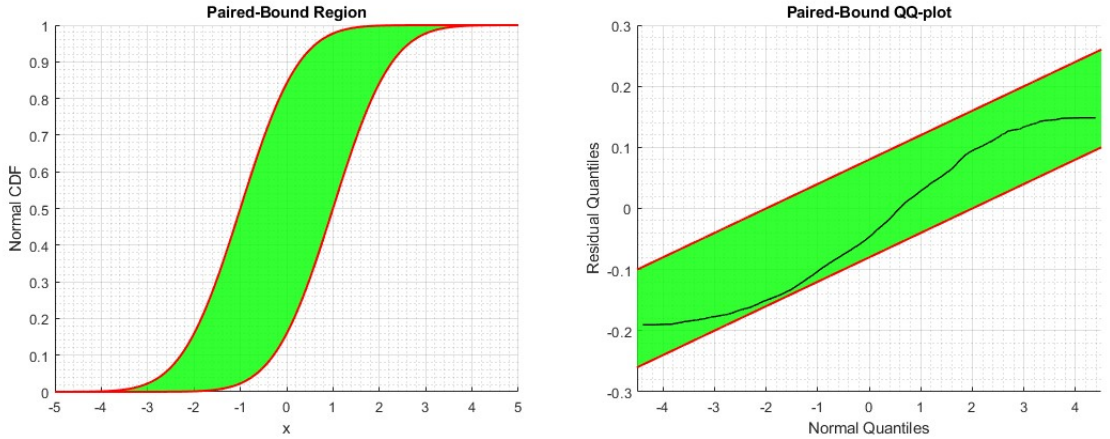
From the perspective of median and extrema residuals, GPT models possess less biased estimation of tropospheric delay, with all of them under 10 cm of residual magnitude, and have no symmetric bias issue in the southern hemisphere. The maximal residual is more consistent and aligns with IGS data, but the minimal residual resembles the distribution of UNB3. In contrast, the median and extrema of UNB3 are more biased and less concentrated. However, UNB3 can compute tropospheric delay much faster and needs less memory storage whilst GPT models require much larger computational resources.

7.5 Paired-Gaussian Bounding

The bounding of a distribution is achieved such that the probability of any interval of a data CDF is lower than the distribution of an upper bound and greater than a lower bound. In the case of the tropospheric residual distribution, the upper and lower bounds imply the worst case scenario for the uncertainties of the tropospheric model

prediction, hence the bounding to the residuals provides a conservative correction for VPL.

The paired-bounding theorem relies on Gaussian distribution as the distribution of the upper and lower bound. The issue with the single Gaussian bound is if the distribution has different median, the bound would inevitably fail at the median. As a result, an additional parameter, bias, is added to ensure the upper and lower bounds are everywhere bounded to the distribution within the bias region.



(a) Gaussian Bounding with Bias (b) Paired-Bound QQ-Plot

Figure 7.8: Gaussian Paired-Bound

Fig. 7.8a shows the paired-Gaussian bound. The paired-bound theorem defines a wider bounding region envelope. Any distribution that falls inside the envelop is considered bounded and the paired-bound can be described with two parameters, σ and β, where σ is the standard deviation of the Gaussian distribution, and β is the bias of the bound, i.e., the displacement of upper and lower bound from zero quantile. Note that while having different σ for the the paired-bounds is possible, here both sides of the bound share the same σ, so the bound must be symmetric. The advantage of defining the bound in this way is that the entire pair-bound can be determined with two parameters, σ and β. Fig. 7.8b shows the Quantile Quantile (QQ) plot representation of the paired-bound. It enables simpler logic for developing the ad-hoc algorithm for searching for the paired-bound since the boundedness is now defined by

two straight lines.

One potential issue with the paired-bound occurs when bounding multiple distributions as for tropospheric residuals. From the perspective of protection level, a paired-bound with larger bias and standard deviation is guaranteed to bound a paired-bound with smaller parameters. However, from the paired-bound perspective, a distribution might not be bounded with larger parameters. For example, if the standard deviation of the paired-bound in Fig. 7.8b is increased, the bounding for the distribution will fail, but the protection level is still valid.

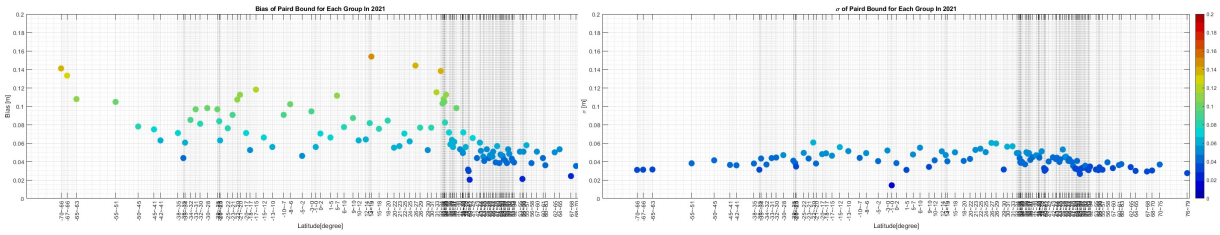
To ensure that the nominal error model conservatively bounds the residual distribution, the ad-hoc algorithm calculates suitable pairs of β and σ on each residual distribution, and retains the maximal β and σ for groups of residuals. The detailed procedures of the ad-hoc algorithm is described in Appendix.C.

7.6 Bounding the Residuals

Two categories of paired-bound are computed in this section. Group paired-bound continues the latitude-wise grouping criteria established in Section.7.4.1. The paired-bounds bound the residual distribution group by group, and as a result, could apply less stringent paired-bound depending on specific locations. Overall, paired-bound essentially combines all stations into a single group for it to be bounded. The overall paired-bound could provide general boundedness to all stations but is a more conservative bound to the tropospheric residual uncertainties. In this section, the residuals of UNB3 are evaluated.

7.6.1 Group Paired-Bound

Minimal Bias



(a) Bias of Group Paired-Bound (b) Std of Group Paired-Bound

Figure 7.9: Group Minimal Bias Gaussian Paired-Bound in 2021.

Fig. 7.9 represents the Gaussian paired-bound for each latitude-wise group. Fig. 7.9a is the minimal bias of the paired-bound computed from the ad-hoc algorithm. The minimal bias is dependent on median value but may also be affected by the tail or head of the residual distribution. Overall, the bias is distributed similarly to Fig. 7.3, with lower biases in latitudes $30° \sim 50°$ and has the same increasing pattern when lower than $-40°$. The median essentially determines how wide the paired-bound must be in order for it to be bounded. If the residual has Gaussian distribution, the median value of the residual is the minimal bias, and any paired-bound with bias lower than that would fail to bound the distribution. Fig. 7.9b presents the standard deviation of the paired-bound. It follows the symmetric pattern of Max/Min residuals of UNB3, with most of them below 6 cm.

Constant Bias and Constant Standard Deviation

Constant bias selects a fixed bias value, then searches for the minimal standard deviation that bounds the residual distribution. This is similar to the minimal bias process except the bias value is specified instead of determined from the search process. Constant standard deviation is the opposite of constant bias; it chooses standard deviation first, then minimizes the corresponding bias value. Minimizing bias and standard deviation brings up the reciprocal relation of these two parameters.

The higher the bias, the smaller the standard deviation, and vice versa. The extreme case in our application is when bias is greater than 28 cm, the bias itself can bound the entire residual distribution such that the corresponding standard deviation can be zero, which is equivalent to assuming all tropospheric delay predictions have 28 cm of error. On the other hand, enlarging standard deviation would not reduce bias to lower than median owing to the fact that the Gaussian paired-bound is a straight line on the QQ-plot. This is the trade-off between bias and standard deviation for paired-bound, and the relative importance of bias and uncertainties depend on the purpose of the application.

7.6.2 Overall Paired-Bound

The issue with the group paired-bound is the validity of bounds depends on the latitude, the bounding parameter biases, and standard deviation needing to change according to the user location. While it offers tighter bounds to the user, it requires extra complexity and storage from the SBAS. The overall paired-bound simplifies the latitude-wise sorting criteria by finding one bound that covers the entire residual distribution for all IGS stations regardless of their location. This overall paired-bound inevitably introduces conservatism to the bounding parameters compared to the group paired-bound, but allows simpler bounding logic for the user.

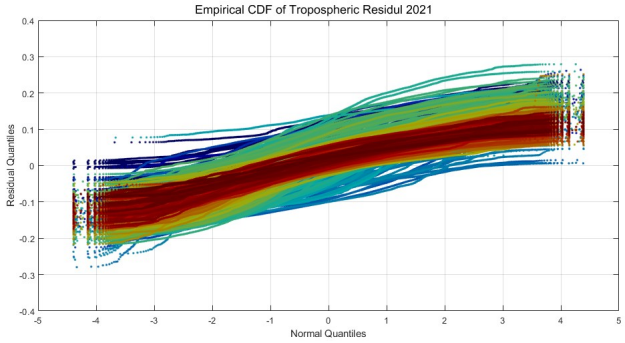
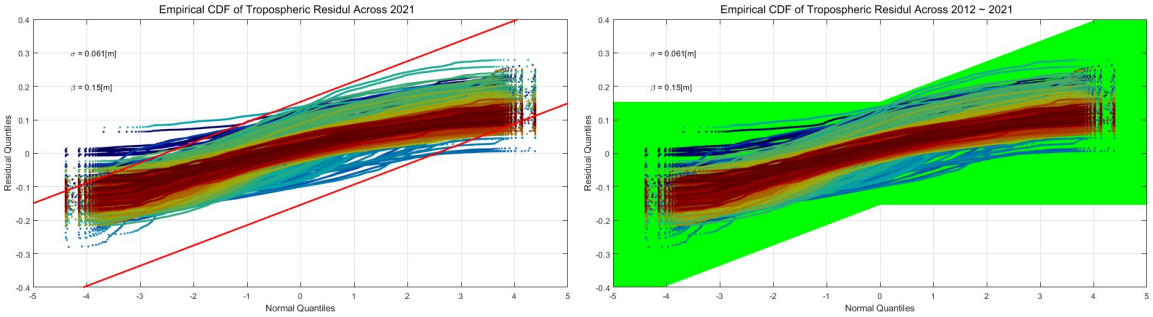


Figure 7.10: Overall residual paired-bound in 2021.



(a) Overall Paired-Bound (b) Overall Paired-Bound - Valid Region

Figure 7.11: Overall Paired-Bound in 2021.

Fig. 7.10 shows the residual distribution in 2021. Fig. 7.11a shows the overall minimal bias paired-bound for the whole residual distribution as two red lines. The bound is produced from the minimal bias algorithm, and the resulting bias is 0.1538 m with standard deviation of 0.0606 m. The red straight lines construct the paired-bound region. It can be seen that some part of the residuals are outside of the bounding region, but since the larger bias and larger standard deviation are chosen for each residual distribution, the seemingly unbounded residuals are in fact properly bounded by the Gaussian paired-bound in VPL. The green region in Fig. 7.11b is the valid region for the paired-bound parameters and all residuals are inside.

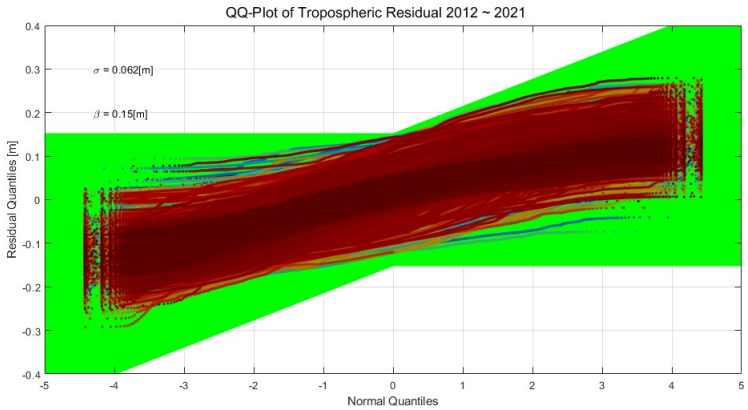


Figure 7.12: Overall residual in 2012∼2021.

Fig. 7.12 provides the entire residual distribution of all IGS stations in 2012 ∼ 2021 and its paired-bound region. The bias is 0.1526 m and the corresponding

standard deviation is 0.0624 m. The bias is the same as the overall bound in 2021 implying that the stations in 2021 have the largest minimal bias among all other stations, and the standard deviation is slightly higher. These two parameters bound the residuals for all IGS stations from 2012 to 2021, and could be the choice for the alternate SBAS VPL equation introduced in Sec. 7.7.

7.7 Alternate SBAS VPL

7.7.1 Current SBAS VPL

The current SBAS VPL adopts the scheme of real-time protection level, which contains:

Variance on satellite i:

$$\sigma_i^2 = \sigma_{i,flt}^2 + \sigma_{i,air}^2 + \sigma_{i,tropo}^2 + \sigma_{i,UIRE}^2 \tag{7.1}$$

where $\sigma_{i,flt}^2$ is fast and long-term correction error, $\sigma_{i,air}^2$ is the airborne error including multipath error, $\sigma_{i,tropo}^2$ is tropospheric error, and $\sigma_{i,UIRE}^2$ is User Ionospheric Range Error. In particular, the tropospheric errors are treated as independent from satellite to satellite.

Geometric Matrix $\boldsymbol{G}$:

$$\overrightarrow{y} = \boldsymbol{G} \cdot \overrightarrow{x} \tag{7.2}$$

where $\overrightarrow{y}$ is a vector of pseudorange measurement from the line-of-sight satellites in view to the user. $\overrightarrow{x}$ is the position and clock bias estimation of the user. $\boldsymbol{G}$ is the geometric matrix that maps position into the line-of-sight domain, and is composed of the normalized line-of-sight vector.

Weighting Matrix $\boldsymbol{W}$:

$$\boldsymbol{W} = \begin{bmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_n^2 \end{bmatrix}^{-1} \tag{7.3}$$

Position Covariance Matrix $\boldsymbol{d}$:

$$\boldsymbol{d} = \left(\boldsymbol{G}^T \cdot \boldsymbol{W} \cdot \boldsymbol{G}\right)^{-1} \tag{7.4}$$

Projection Matrix $\boldsymbol{S}$:

$$\boldsymbol{S} = \left(\boldsymbol{G}^T \cdot \boldsymbol{W} \cdot \boldsymbol{G}\right)^{-1} \cdot \boldsymbol{G}^T \cdot \boldsymbol{W} \tag{7.5}$$

where the projection matrix, $\boldsymbol{S}$, projects the pseudorange measurement into the position domain. Note that it is also the optimal solution to the weighted least-square problem.

VPL:

$$VPL = K_V \cdot d_U \tag{7.6}$$

where K_V is a constant with value of 5.33, corresponding to an inverse standard normal distribution with integrity budget allocated to the vertical direction, and is standard for VAL below 10 m. d_U is the position variance in the vertical direction, or the Up in the East-North-Up (ENU) coordinate system. The current VPL equation depends on variance of the vertical position, and therefore depends on the nominal error model used in the calculation of σ_i^2.

MOPS mapping function:

$$m(EL[i]) = \frac{1.001}{\sqrt{0.002001 + sin^2(EL[i])}} \tag{7.7}$$

$$\overrightarrow{m} = \begin{bmatrix} m(EL[1]) \\ m(EL[2]) \\ \vdots \\ m(EL[n]) \end{bmatrix} \tag{7.8}$$

where $EL[i]$ is the elevation angle of the ith satellite, and n is the total number of satellites in view for the receiver.

To ensure the VPL is robust and fault tolerant in any valid condition for the uncertainties of tropospheric delay in the current nominal error model, an LOS unit-vector for the geometric matrix is randomly generated, $\boldsymbol{G}$. Next, randomly select elevation angles for line-of-sight measurement and tropospheric delay, then projection matrix, $\boldsymbol{S}$, projects the maximal nominal tropospheric delay model uncertainties, 28 cm, along with mapping function, into the position domain. The resulting vertical position error is compared with the corresponding VPL.

Uniform Random Simulation - Current SBAS VPL

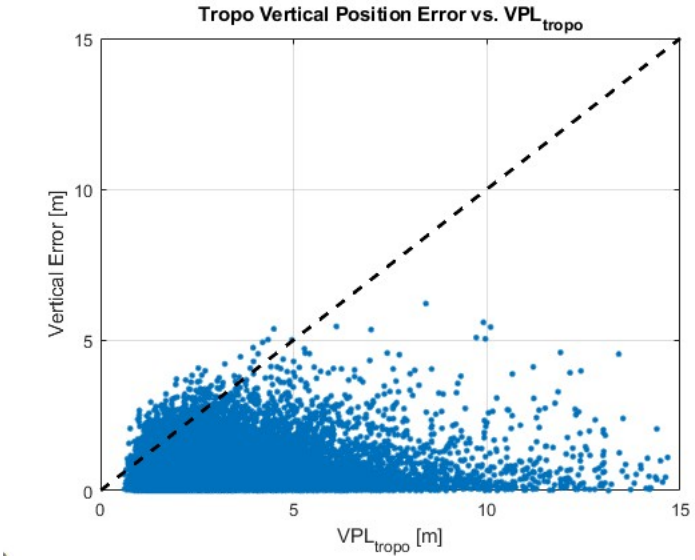


Figure 7.13: SBAS Tropospheric Error vs. VPL

Fig. 7.13 plots the possible positioning error versus the VPL. The positioning error is computed from Eq.(7.5) with 28cm zenith total delay error in measurement, and VPL is from Eq.(7.6). This simulation assumes 12 satellites in-view for one constellation and randomly generates 50000 geometric matrices. A diagonal dash-line is drawn to facilitate the determination of whether the vertical error exceeds the VPL. Note that there is a small portion of vertical errors above the diagonal line, indicating the possibility that vertical errors might be underestimated due to the simplification that tropospheric errors are uncorrelated across satellites.

This does not mean the tropospheric correlation error has resulted in an unbounded position error, it merely shows that there exist combinations of geometric matrices and elevation angles that might result in higher vertical error than the VPL. The simulation takes the most extreme case of tropospheric delay error as the sole factor in the estimation of the VPL, ignoring contributions of dual-frequency correction error, air error, etc. In SBAS, since tropospheric delay error is not the only error considered in position estimation, there is no immediate issue to the integrity of the position solution.

7.7.2 Alternate SBAS VPL

Acknowledging that the current VPL might underestimate the possible effect of nominal troposphere on position error, a modified VPL equation is suggested that includes the correlation from one satellite to another and the uncertainties of the tropospheric residual distribution in the VPL calculation.

To support correlation and biased tropospheric error, the covariance contributions are separated into non-tropospheric and tropospheric.

Non-tropospheric variance:

$$\sigma^2_{i,nt} = \sigma^2_{i,flt} + \sigma^2_{i,air} + \sigma^2_{i,UIRE} \tag{7.9}$$

Non-tropospheric position variance:

$$d^2_{nt,U} = \sum_i^n s^2_{3,i} \cdot \sigma^2_{i,nt} \tag{7.10}$$

Scaling for tropospheric to position error:

$$f_{t,U} = \| \sum_i^n s_{3,i} \cdot m(EL[i]) \| \tag{7.11}$$

New VPL:

$$VPL_{new} = K_V \cdot \sqrt{d^2_{nt,U} + \sigma^2_{TVE} \cdot f^2_{t,U}} + \beta_{TVE} \cdot f_{t,U} \tag{7.12}$$

where $d^2_{nt,U}$ is the position variance without tropospheric effect, $f_{t,U}$ is the mapping from tropospheric error to position error domain. σ^2_{TVE} and β_{TVE} are the variance and bias of the zenith tropospheric residual.

Uniform Random Simulation - Corrected SBAS VPL

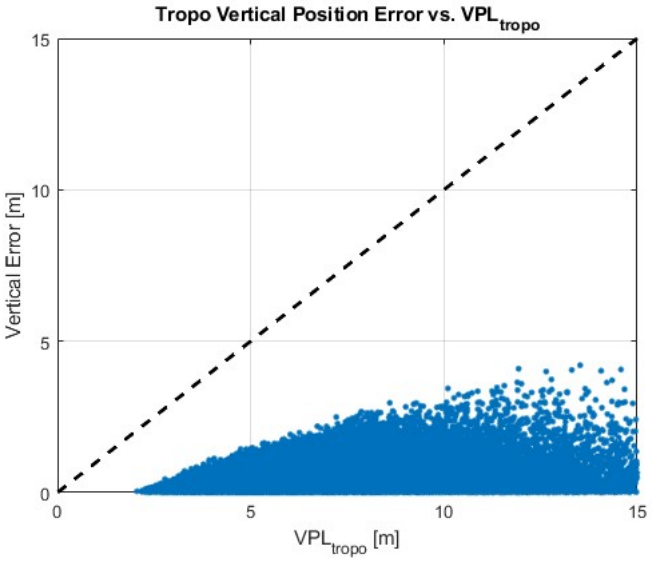


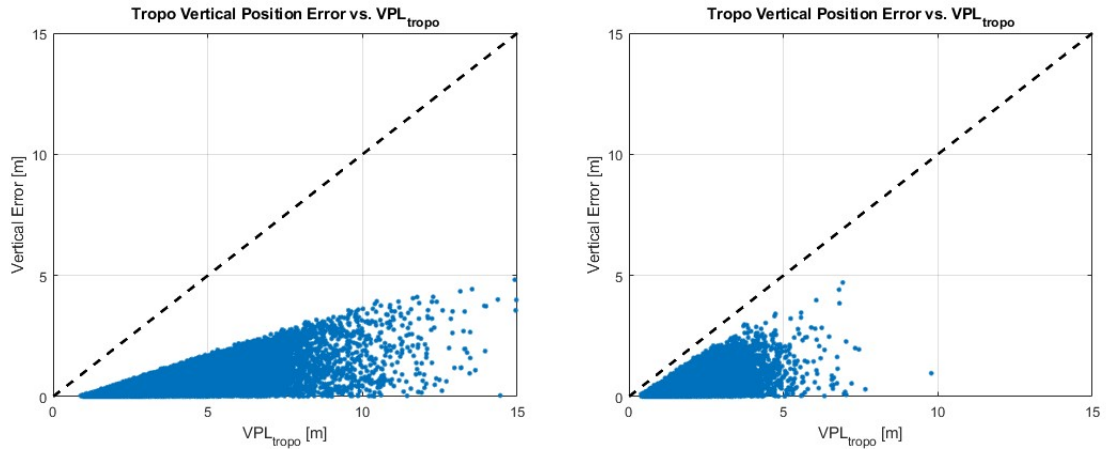
Figure 7.14: SBAS Tropospheric Error vs. Corrected VPL

Fig. 7.14 plots the vertical position error versus the corrected VPL. The only difference between Fig. 7.13 and Fig. 7.14 is the VPL equation. The new VPL equation bounds all the possible combinations of geometric matrices and elevation angle by incorporating the satellite correlations and Gaussian tropospheric residual bound, where the paired-bound parameters are the overall paired-bound of UNB3. σ_{TVE} and β_{TVE} used in the simulation are the overall bound obtained from ten years of data in Sec. 7.6. With the new VPL equation, all possible position errors are below the corresponding VPL.

The modification to the VPL equation does not require anything new to the SBAS except σ_{TVE} and β_{TVE}, which can be two constants added to the system similar to what has been done in the simulation, or a table of values based on locations.

7.7.3 Sigma-Bias Value for Alternate VPL Equation

Other sigma-bias values are also possible for use in the alternate VPL equation.



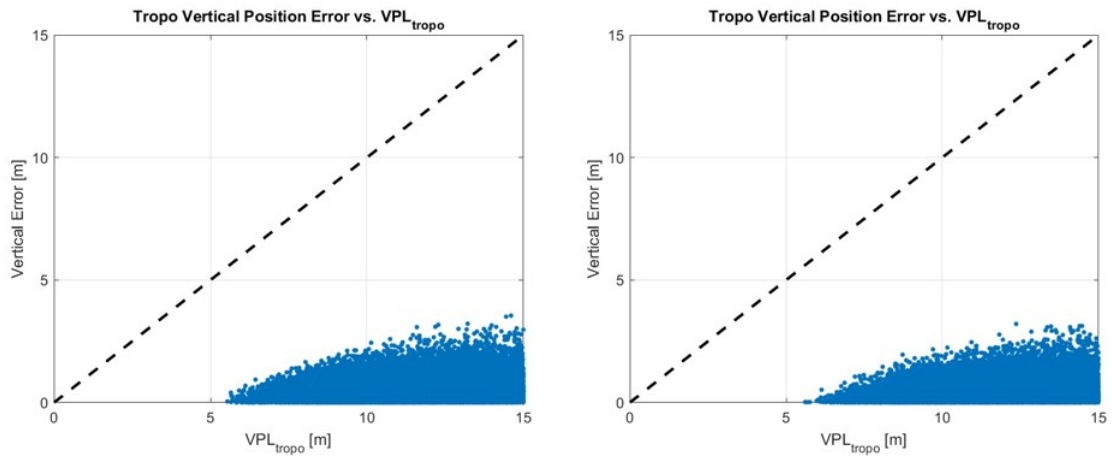
(a) SBAS Tropospheric Error vs. Corrected VPL. $\sigma = 0.12$ m and $\beta = 0$ m.

(b) SBAS Tropospheric Error vs. Corrected VPL. $\sigma = 0$ m and $\beta = 0.28$ m.

Figure 7.15: SBAS Tropospheric Error vs. Corrected VPL without Non-tropo errors

Fig. 7.15 shows how different sigma-bias values can have different vertical error versus alternate VPL when non-tropospheric errors are not included. Fig. 7.15a has $\sigma = 0.12$ m and $\beta = 0$ m. Fig. 7.15b has $\sigma = 0$ m and $\beta = 0.28$ m. Both sigma-bias

values are capable of bounding the tropospheric error, but Fig. 7.15b has VPL closer to vertical error, so the VPLs are smaller and still can bound the error. As a result, $\sigma = 0$ m and $\beta = 0.28$ m is a better sigma-bias value when tropospheric error is the sole factor in the calculation of alternate VPL.



(a) SBAS Tropospheric Error vs. Corrected VPL

(b) SBAS Tropospheric Error vs. Corrected VPL

Figure 7.16: SBAS Tropospheric Error vs. Corrected VPL with Non-tropo errors

Fig. 7.16 shows the vertical error versus VPL when non-tropospheric errors are included in the calculation of VPL, so $d_{nt,U} \neq 0$. Fig. 7.16a has $\sigma = 0.12$ m and $\beta = 0$ m. Fig. 7.16b has $\sigma = 0$ m and $\beta = 0.28$ m. It can be seen with the non-tropospheric error included into the calculation of the alternate VPL, Fig. 7.16a has smaller VPL than Fig. 7.16b. This is contrary to the result without non-tropospheric error. The reason for this is the non-tropospheric error, $d_{nt,U}$, is typically larger than tropospheric error, σ_{TVE}, so after the root-mean-square process in Eq. 7.12, the contribution of tropospheric error to VPL is smaller. For $\sigma = 0$ m and $\beta = 0.28$, the contribution of tropospheric error to VPL is a linear addition so the resulting VPL is larger. Therefore, with non-tropospheric error included in the calculation of alternate VPL, the larger sigma value provides a better result.

7.8 Summary

10 years of tropospheric delay data of all IGS stations from 2012 to 2021 are evaluated. The tropospheric prediction from three models, UNB3, GPT2W, and GPT3, are evaluated and compared with data from IGS stations to form the residual distribution at each station in every year. Some IGS data were excluded from the computation

of residual due to anomalous values after cross-validating against the weather record and JPL data. An ad-hoc algorithm was developed to compute the Gaussian paired-bound to bound the distribution of the residual. There are three types of bounds: minimal bias, constant bias, and constant standard deviation. Minimal bias obtained the lowest bias for paired-bound and allowed for automation of the whole process. Constant bias and constant standard deviation are relaxations to the minimal bias and require manually assigning one of the paired-bound parameters. Residuals from 2012 to 2021 of UNB3, GPT2W, and GPT3 are all successfully bounded with the paired-bound generated by ad-hoc algorithm.

Simulation results show that the current VPL equation may significantly underestimate the nominal tropospheric effect on position error when VAL is below 10 m. Therefore, a modification to the current SBAS VPL equation was proposed such that the tropospheric correlations of satellites can be accounted for in the calculation of VPL. The paired-bound parameters computed are also integrated into the VPL equation. Additional simulation results show the alternate VPL equation with the paired-bound parameters computed from IGS data yielding more conservative VPL for VAL below 10 m. Finally, different sigma-bias values are possible for the alternate VPL equation depending on the non-tropospheric error sources. The new VPL equation was easily adapted with the paired-bound parameters suggested in this thesis.

Chapter 8

Conclusions

The fundamental objective of this thesis is to provide integrity to precise navigation services such that users will have integrity bounds guaranteed on their position solution. Integrity for the users is achieved by ensuring the corrections are free from potential satellite faults. That is, as long as the corrections the users apply to their measurements reflect the actual errors, including the errors induced by satellite faults, the positioning results of the users will be trustworthy no matter how many faults are present or where the users are on earth. This approach achieves the integrity guarantee for PPP users by constructing a ground-based integrity monitor from the perspective of the service provider such that no significant changes are introduced to the algorithm of the PPP users.

8.1 Integrity Monitor

Chapter 2 develops the very basic tools to monitor the instantaneous behaviors of two of the most prominent and important faults, satellite clock faults and ephemeris faults. These two satellite fault modes are incorporated into the models of expected behavior, and enable the direct estimates of the value of the UREs induced by these two fault modes at any user location with decimeter level of accuracy in real-time with the use of multiple ground-based receivers. The only significant priors required for this monitor are the static locations of the receivers, with non-constraining magnitudes

on the faults. The priors on the faults are set to be 1000 m in the experiments for numerical consideration.

Chapter 3 verifies that the performance the PPP users can achieve when the satellites clock and ephemeris faults are completely removed by the integrity monitor in the form of integrity information, which are simply extra corrections to the measurement models of the users. It is demonstrated that the users can achieve 10∼20 cm of position accuracy, ∼ 1 meter of error bound in the horizontal, and ∼2 meters in the vertical direction, while being completely protected from any number and type of satellite clock and ephemeris faults.

Chapter 4 shows that from the perspective of the integrity monitor, the integrity information it generated is bounded with the prior error models, proving that the formulation of the fault-tolerant filter is sufficient even considering the possibility of a receiver failure within the network.

The example network demonstrated and verified in this thesis is a global network that has uninterrupted LOS to the GPS constellation. The users are treated equally around the world. It has been shown that such a global network can achieve decimeter-level user positioning accuracy without priors on satellite errors. However, the same priors on the error distributions from the global network might not be adequate for the regional network. Specifically, two parts of the priors will require further study. The first is the ephemeris errors. Depending on the receiver geometries, the regional network would require priors on the orbit kinematics, thus the user is not completely free from arbitrary ephemeris faults. The second prior that needs to be included for the regional network is temporal correlations. Specifically, the transient response when satellites transition from NLOS to LOS in the network. The converging process of the uncertainty of the satellite errors should be strictly bounded to provide valid integrity information to the users. The multipath effects must also be accounted for given the environments of the receivers used in the network.

8.2 Network Design

Chapter 5 defines and develops an efficient way of evaluating the performance of an integrity monitor network with an arbitrary number and geometry of reference receivers via simulation using almanac data. The horizontal and vertical protection levels of PPP users with the integrity information can be rapidly computed anywhere on earth. These serve as the lower error bound to the actual protection levels that the user can achieve using Riccati equation. The evaluation offers a straightforward view of user performance given different geometries between satellites and reference stations.

Chapter 6 utilizes the performance definition of the integrity monitor defined in Chapter 5, and generalizes the issue of receiver placement by treating the receiver selections as a single convex optimization problem via integer relaxation. It also allows the design of the receivers' location to be determined by the target users' locations. It is shown that the new network selected, even under different criteria such as the number of available receivers or target user locations, easily outperformed the network selected from the Monte-Carlo heuristic method. The process is straightforward, and intuitive for reaching the goal of the integrity network, which is to have the lowest protection level possible for the users of interest.

8.3 Tropospheric Error Analysis

Chapter 7 evaluated ten prior years of global tropospheric delay data from over 400 IGS stations. It is shown that even for the worst case scenarios the modeling errors of the tropospheric zenith delay errors would not exceed 30 cm. It confirms that the fault modes of significant impact can be focused and attributed to the satellite clock and ephemeris fault types. Additionally, the result helps to identify a necessary update to the vertical error bounding with consideration of the error correlations.

Appendix A

Temporal Modeling of User Latency

From the perspective of PPP users, they receive the integrity information, $\beta_u^{(k)}$ and $\sigma_u^{(k)}$, generated using the measurements a period of time later than the current time of operation. This latency is inevitable due to the delays in the delivery of the message. The user receivers and the integrity monitor receive the same satellite signals at nearly the same time, so the integrity information can only reflect the satellite clock and ephemeris faults after it is processed, broadcast, and received by the users. A temporal prediction model can extend the accuracy of the integrity information by acknowledging the fact that the integrity information is not up-to-date by the time the users receive it.

A.1 Gaussian Processes

For a PPP user, u, receiving the integrity information discretely from the service provider of the kth satellite at time t_0, the UREs and sigmas, $\beta_{u,t_p}^{(k)}$ and $\sigma_{u,t_p}^{(k)}$, span across the timestamp p the user received:

$$\{\beta_{u,t_{-N}}^{(k)}, ..., \beta_{u,t_{-2}}^{(k)}, \beta_{u,t_{-1}}^{(k)}\} \tag{A.1}$$

$$\{\sigma_{u,t_{-N}}^{(k)}, ..., \sigma_{u,t_{-2}}^{(k)}, \sigma_{u,t_{-1}}^{(k)}\} \tag{A.2}$$

where N is the total number of integrity information the users received. The difference between t_{-1} and t_0 is the latency of the latest integrity information and the real-time of the users.

One can adopt a model to predict, or extrapolate, the value of $\beta_{u,t_0}^{(k)}$ based on the evolution of $\beta_{u,t_p}^{(k)}$ over time.

Assume $\beta_{u,t_p}^{(k)}$ is a function of time:

$$\beta_{u,t_p}^{(k)} = f\left(t_p\right) \tag{A.3}$$

The model of the URE is then:

$$\hat{\beta}_{u,t_p}^{(k)} = \hat{f}\left(t_p \mid f\left(t_{-N}\right), ..., f\left(t_{-2}\right), f\left(t_{-1}\right)\right) \tag{A.4}$$

For the sake of generality, what model is used is not specified at this point.

In this formulation, the users only consider the UREs in real-time. The other chapters only focus on evaluating the current time of the users at t_0.

$$\hat{\beta}_{u,t_0}^{(k)} = \hat{f}\left(t_0 \mid f\left(t_{-N}\right), ..., f\left(t_{-2}\right), f\left(t_{-1}\right)\right) \tag{A.5}$$

In addition to evaluating the prediction of the URE at t_0, it is also necessary to quantify the uncertainty of $\hat{\beta}_{u,t_0}^{(k)} = \hat{f}\left(t_p \mid \cdot\right)$. Given the assumption that the integrity information is bounded by a Gaussian distribution, Gaussian Processes can be applied to obtain the uncertainty of $\hat{\beta}_{u,t_0}^{(k)}$, which is essentially the conditional distribution of the model over the past integrity information over the same priors as the estimator.

$$\begin{bmatrix} \hat{y} \\ y \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \hat{f}(t_0 \mid \cdot) \\ \hat{f}(t_{-1} \mid \cdot) \\ \vdots \\ \hat{f}(t_{-N} \mid \cdot) \end{bmatrix}, \Sigma\right) \tag{A.6}$$

$$\Sigma = \begin{bmatrix} \hat{\sigma}_{t_0}^2 + \nu_{t_0} & \hat{\sigma}_{t_0}\sigma_{u,t_{-1}}^{(k)}\rho(t_0,t_{-1}) & \cdots & \hat{\sigma}_{t_0}\sigma_{u,t_{-N}}^{(k)}\rho(t_0,t_{-N}) \\ \sigma_{u,t_{-1}}^{(k)}\hat{\sigma}_{t_0}\rho(t_{-1},t_0) & {\sigma_{u,t_{-1}}^{(k)}}^2 & \cdots & \sigma_{u,t_{-1}}^{(k)}\sigma_{u,t_{-N}}^{(k)}\rho(t_{-1},t_{-N}) \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{u,t_{-N}}^{(k)}\hat{\sigma}_{t_0}\rho(t_{-N},t_0) & \sigma_{u,t_{-N}}^{(k)}\sigma_{u,t_{-1}}^{(k)}\rho(t_{-N},t_{-1}) & \cdots & {\sigma_{u,t_{-N}}^{(k)}}^2 \end{bmatrix} \tag{A.7}$$

where $\hat{y}$ is the posterior of $\hat{\beta}_{u,t_0}^{(k)}$, $y = [\beta_{u,t_{-1}}^{(k)}, \ldots, \beta_{u,t_{-N}}^{(k)}]^T$, $\rho\left(t, t'\right)$ is the temporal correlation of the UREs over time, and ν_{t_0} is the uncertainty of propagation.

Let

$$X^* = t_0 \tag{A.8}$$

$$X = [t_{-1}, \ldots, t_{-N}]^T \tag{A.9}$$

$$Y^* = \hat{f}\left(t_0 \mid \cdot\right) \tag{A.10}$$

$$Y = [\hat{f}\left(t_{-1} \mid \cdot\right), \ldots, \hat{f}\left(t_{-N} \mid \cdot\right)]^T \tag{A.11}$$

$$K\left(X, X'\right) = \begin{bmatrix} \sigma_{t_{-1}} \cdot \sigma_{t'_{-1}} & \cdots & \sigma_{t_{-1}} \cdot \sigma_{t'_{-N}} \cdot \rho\left(t_{-1}, t'_{-N}\right) \\ \vdots & \ddots & \vdots \\ \sigma_{t_{-N}} \cdot \sigma_{t'_{-1}} \cdot \rho\left(t_{-N}, t'_{-1}\right) & \cdots & \sigma_{t_{-N}} \cdot \sigma_{t'_{-N}} \end{bmatrix} \tag{A.12}$$

Eq. A.6 can be written as:

$$\begin{bmatrix} \hat{y} \\ y \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} Y^* \\ Y \end{bmatrix}, \begin{bmatrix} K\left(X^*, X^*\right) + \nu_{t_0} & K\left(X^*, X\right) \\ K\left(X, X^*\right) & K\left(X, X\right) \end{bmatrix}\right) \tag{A.13}$$

The conditional distribution of $\hat{y}|y$ is given by:

$$\mu\left(\hat{y}|y\right) = Y^* + K\left(X^*, X\right) K\left(X, X\right)^{-1}\left(y - Y\right) \tag{A.14}$$

$$\Sigma\left(\hat{y}|y\right) = K\left(X^*, X^*\right) - K\left(X^*, X\right) K\left(X, X\right)^{-1} K\left(X, X^*\right) + \nu_{t_0} \tag{A.15}$$

where $\mu\left(\hat{y}|y\right)$ is the mean and $\Sigma\left(\hat{y}|y\right)$ is the covariance of the posterior distribution of $\hat{\beta}_{u,t_0}^{(k)}$. Note that Eqs. A.14 and A.15 are simply the conditional distribution of

random variables with Gaussian distribution. They are identical to the standard measurement updates of the EKF.

The Gaussian Processes described above allow the user to apply the integrity information as pseudorange corrections within the framework of the EKF as shown in Eqs. 3.1, 3.2 and Eqs. 3.3, 3.4.

A.2 Choice of Design

Eq. A.6 is the framework for modeling the temporal relations in general. There are several parts that remain to be determined. These design choices within the Gaussian Processes and more study needs to be done to verify the robustness of these designs. The following designs are adopted, but other choices are possible depending on the requirements.

A.2.1 N: Total Time Steps of Past Information

N in Eqs. A.1 and A.2 is the number of past integrity information the Gaussian Processes use when predicting the current distribution of the UREs. One needs to determine how many time-steps of integrity information to keep in memory.

A.2.2 $\hat{f}(t_p|\cdot)$: The Prediction Model

The prediction model to be used for prediction is also a design choice. It can be of any type, such as linear, polynomial, or even Fourier models.

A.2.3 $\hat{\sigma}_{t0}$, ν_{t_0}: Priori of URE

The uncertainty ($\hat{\sigma}_{t0}$) of the prior estimate of the URE $\left(\hat{\beta}_{u,t_0}^{(k)} = \hat{f}\left(t_0 \,|\, \cdot\right)\right)$ can be assumed to be the nearest point of uncertainty.

$$\hat{\sigma}_{t0}^2 = {\sigma_{u,t_{-1}}^{(k)}}^2 \tag{A.16}$$

As for the uncertainty of propagation, ν_{t_0}, it is modeled as:

$$\nu_{t_0} = {\sigma^{(k)}_{u,t_{-1}}}^2 + Q_t \cdot (t_0 - t_{t-1})^2 \tag{A.17}$$

where Q_t is the process noise of the integrity information over time. It determines how far the user can trust the prediction from the last update of the integrity information. The baseline value ${\sigma^{(k)}_{u,t_{-1}}}^2$ ensures the uncertainty of the evaluation is not below the uncertainty of the data point at t_{-1}. However, given the assumptions that the satellite faults can be anything at any moment, the above model can only be treated as accuracy model.

A.2.4 $\rho(t, t')$: Correlation of Time

$\rho(t, t')$ describes the time correlation of the accuracy model. It controls how the corrections of the URE is affected over time.

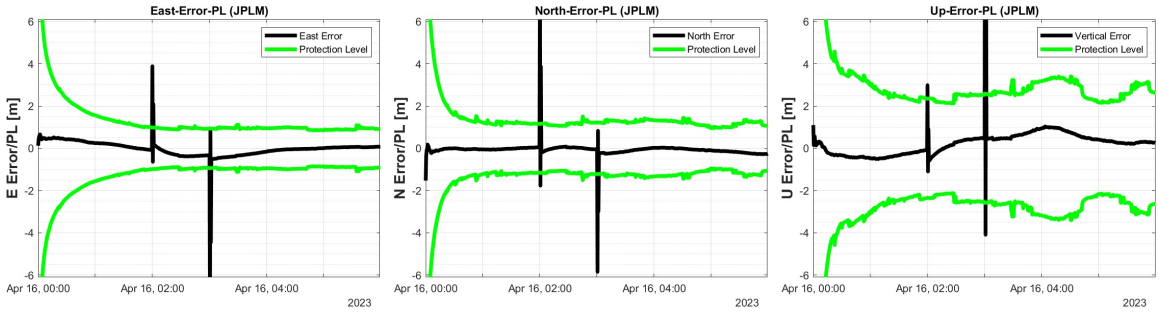
A common choice of correlation is the squared exponential function:

$$\rho(t, t') = exp\left(\frac{(t - t')^2}{2 \cdot l^2}\right) \tag{A.18}$$

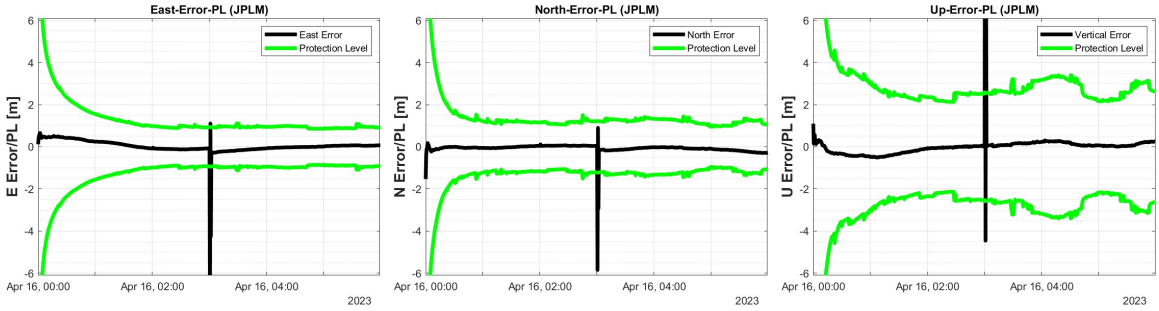
where l, the characteristic length scale, corresponds to how far t and t' are separated should the relation change significantly. In the context of time correlation, it determines how long over time until the integrity information is not relevant to the user. Another way of interpreting Eq. A.18 is considering it as a normal function. Then l is the standard deviation of a zero-mean normal distribution that determines how the variables are correlated in time.

A.3 Test Results

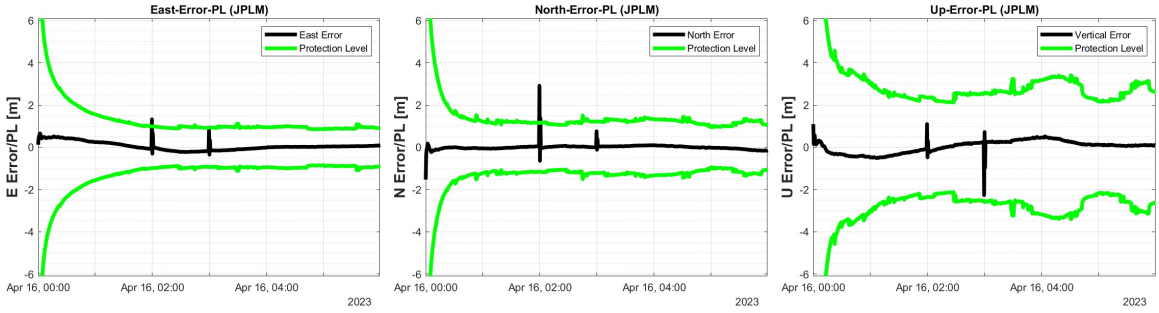
A.3.1 Latency: 30 seconds



(a) Position Errors and protection levels at JPLM during satellite clock step fault with integrity information.



(b) Position Errors and protection levels at JPLM during satellite clock ramp fault with integrity information.



(c) Position Errors and protection levels at JPLM during ephemeris step fault with integrity information.

Figure A.1: Position errors and protection levels at JPLM with integrity information under 30-second latency.

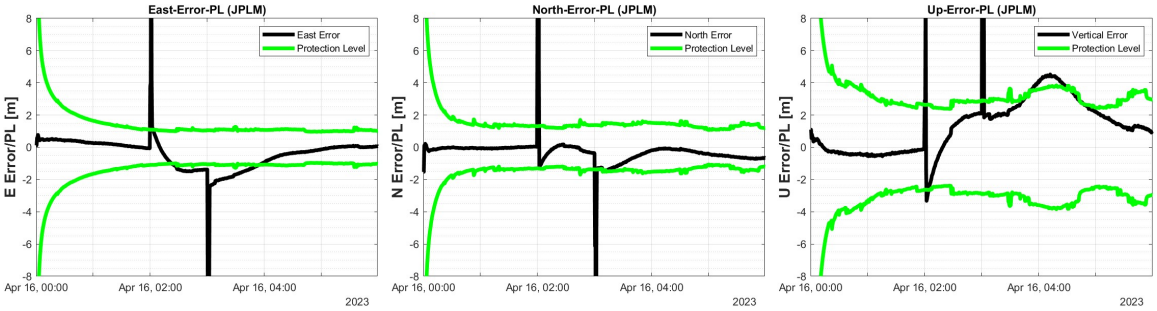
The data set and setup from Sec. 2.1.2 continued to be used. In this test, there is a 30 second latency between the integrity information and the user measurements. That is, the integrity information the user received from the fault-tolerant filter is 30 seconds later than the pseudorange measurements the user filter applies.

Fig. A.1 shows the position errors of the JPLM station under the satellite faults, but with 30 seconds of latency between the reception of integrity information and the user time. Due to the gap between the presence of the satellite fault and the correction of from the integrity information, for abrupt changes such as the 30 m clock step fault in Fig. A.1a, the user experienced a short period of position failure until the integrity information reflected the satellite clock fault. When the step clock fault disappears at 03:00:00, since the integrity information is 30 seconds behind the disappearance of the clock fault, the user still applies the URE correction as if the step fault exists.

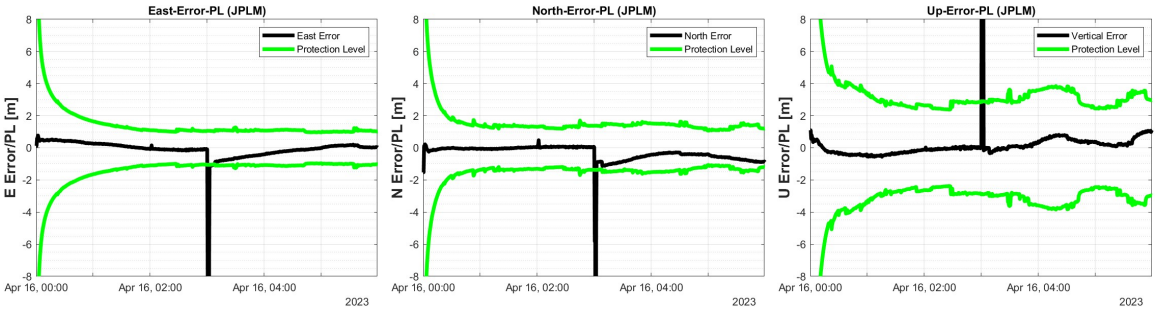
The sigmas of the position error also expand to 30∼40 cm due to the propagation of the uncertainty, Q_t. As a result of the inflation of the uncertainty, the protection levels are inflated to ∼1.5 m in the horizontal direction and ∼3 m in the vertical direction.

Similar to the beginning of the fault, the user loses its position for a short period until the latest update of the integrity information. This situation also occurs due to the disappearance of the ramp fault in Fig. A.1b, which has a similar effect to a step fault, and the step ephemeris fault in Fig. A.1c. Nevertheless, even with the presence of the 30 second latency, the user can still quickly recover from the satellites faults and maintain accuracy during and after the faults. In comparison, without the integrity information the user would struggle to recover from the faults as shown in Fig. 3.4.

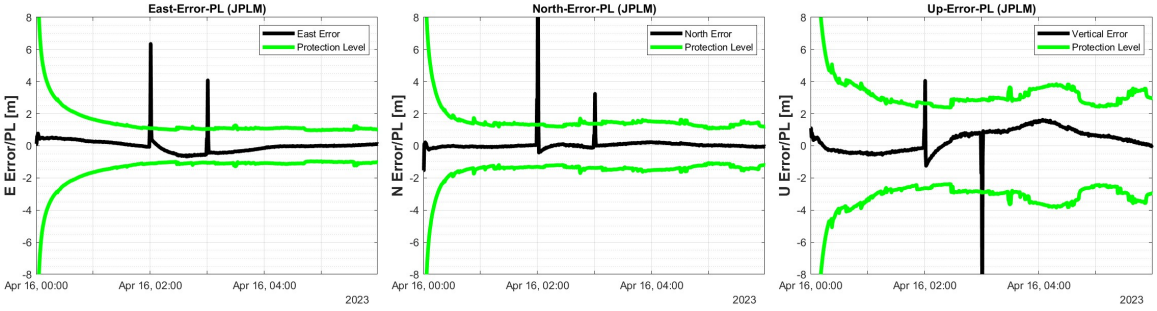
A.3.2 Latency: 60 seconds



(a) Position Errors and protection levels at JPLM during satellite clock step fault with integrity information.



(b) Position Errors and protection levels at JPLM during satellite clock ramp fault with integrity information.



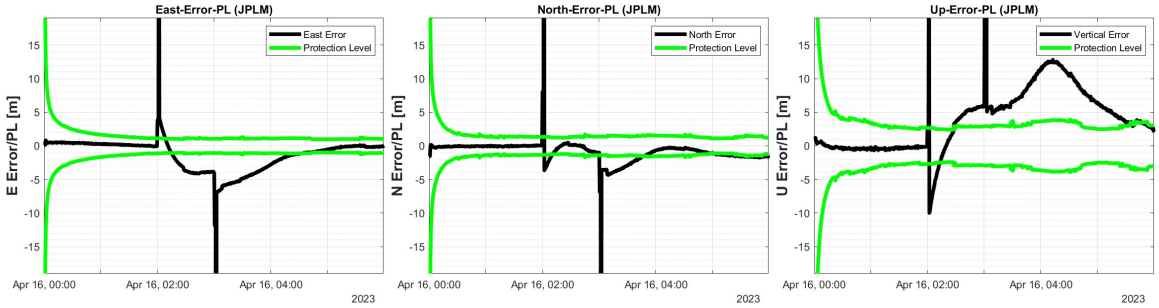
(c) Position Errors and protection levels at JPLM during ephemeris step fault with integrity information.

Figure A.2: Position errors and protection levels at JPLM with integrity information under 60-second latency.

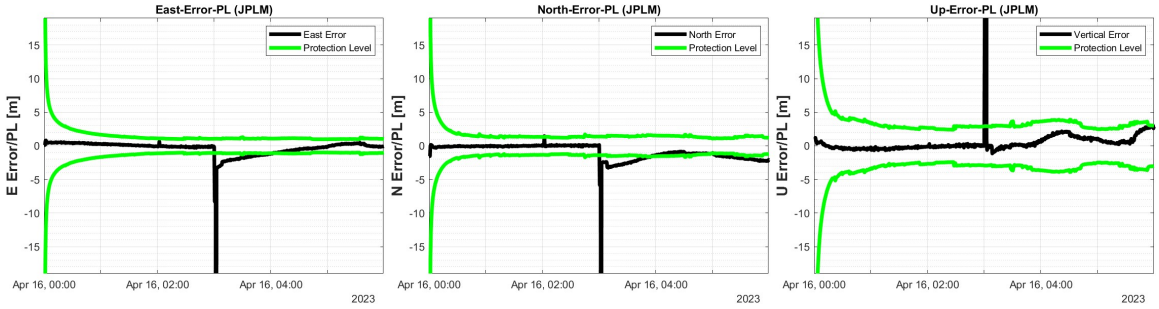
Fig. A.2 shows the user position errors in three different fault scenarios with integrity information under 60-second latency. In Fig. A.2a, the 30 m step fault has caused significant impact on the integrity of the users. Even when the step fault is corrected

by the integrity information, the position estimates still could not recover from the initial exposure to the step faults. The cut-off of the step faults even causes the position errors to exceed protection levels.

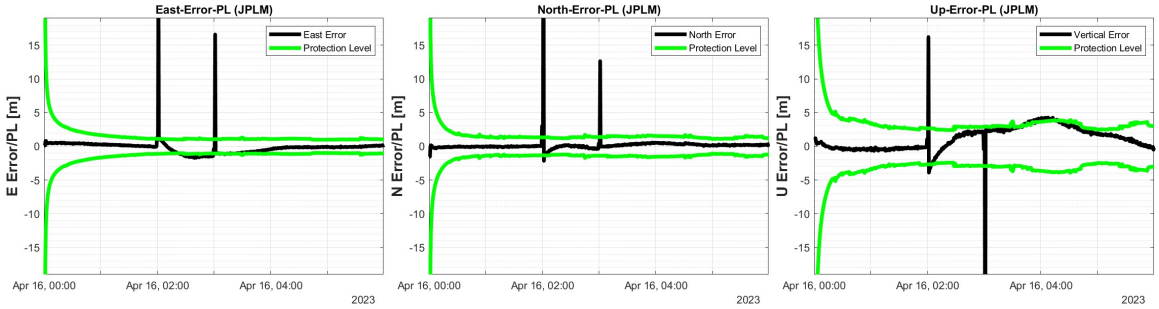
A.3.3 Latency: 90 seconds



(a) Position Errors and protection levels at JPLM during satellite clock step fault with integrity information.



(b) Position Errors and protection levels at JPLM during satellite clock ramp fault with integrity information.



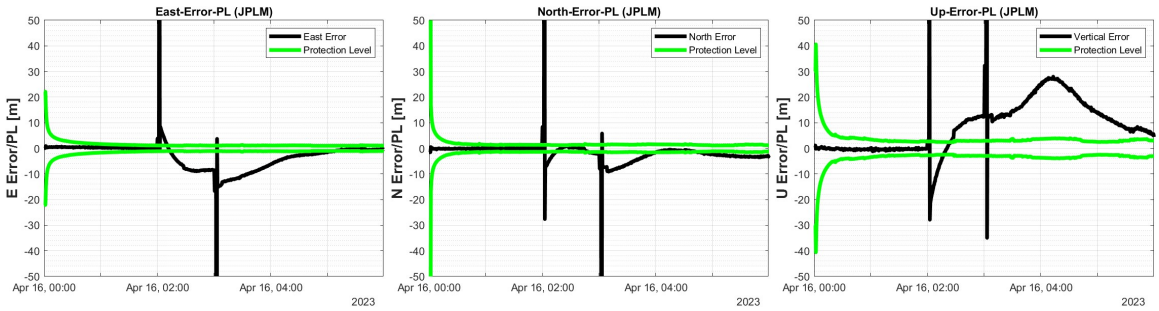
(c) Position Errors and protection levels at JPLM during ephemeris step fault with integrity information.

Figure A.3: Position errors and protection levels at JPLM with integrity information under 90-second latency.

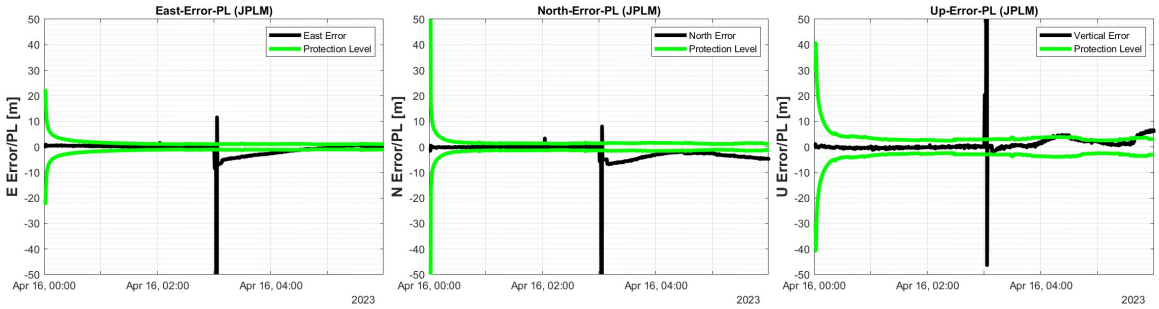
Fig. A.3 shows the position errors under 90-second latency with the integrity information. The user would immediately lose the integrity of its position estimates. The position errors reach as much as 13 m in the vertical direction as shown in Fig.

A.3a. As for the ramp fault , the user can still maintain its accuracy and integrity at the beginning until the sudden cut-off at the end of the fault. The ephemeris step fault shown in Fig. A.3 exhibits similar behavior as the 30 m step clock fault. The sudden changes at the start and end of the faults corrupt the estimator and eventually result in loss of the integrity.

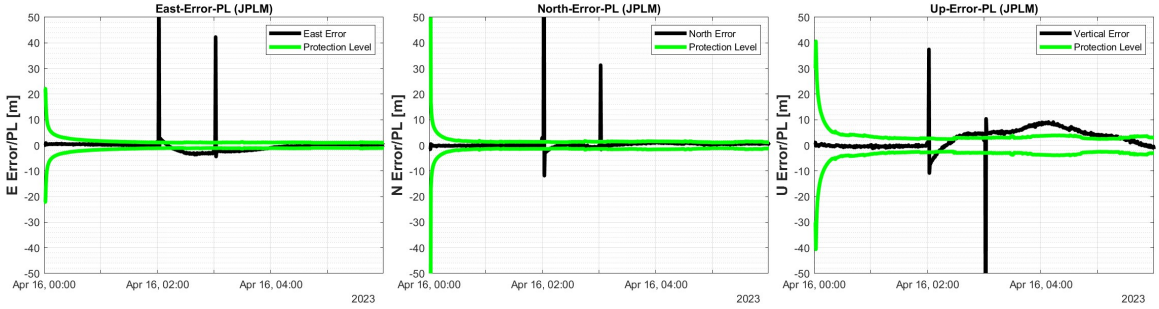
A.3.4 Latency: 120 seconds



(a) Position Errors and protection levels at JPLM during satellite clock step fault with integrity information.



(b) Position Errors and protection levels at JPLM during satellite clock ramp fault with integrity information.



(c) Position Errors and protection levels at JPLM during ephemeris step fault with integrity information.

Figure A.4: Position errors and protection levels at JPLM with integrity information under 120-second latency.

Fig. A.4 shows the position errors under 120-second latency. The user in all three fault scenarios loses integrity completely due to the corruption of the estimator.

Appendix B

Horizontal Protection Levels

This appendix includes the performance of the integrity monitor in terms of their horizontal protection levels (HPL).

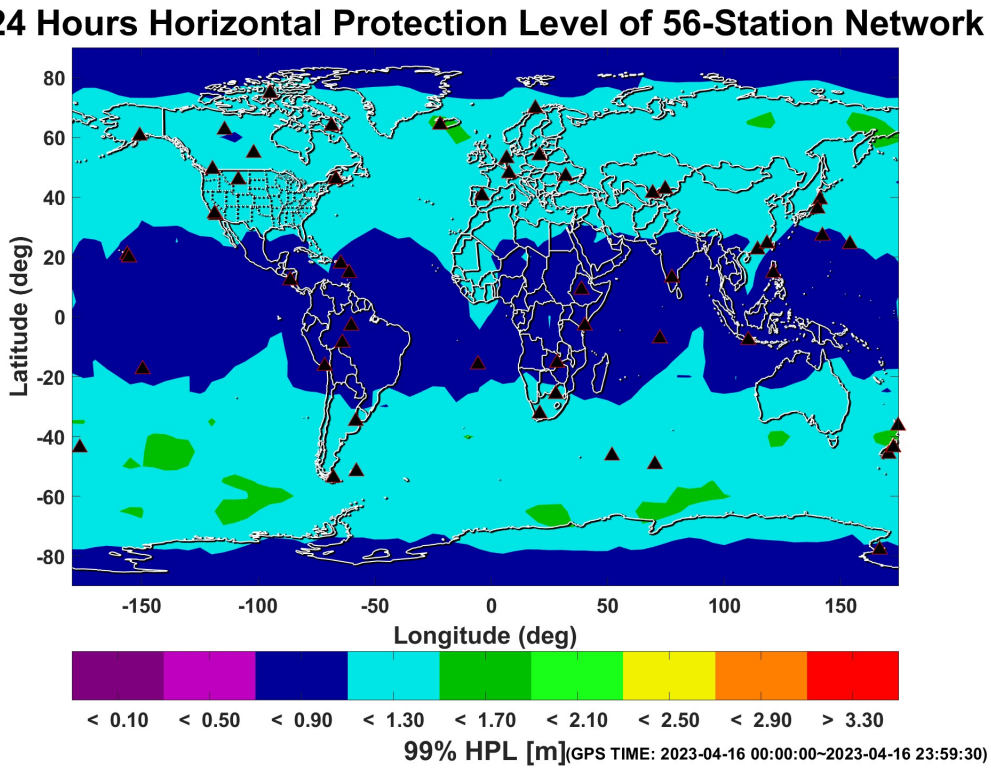


Figure B.1: 99% Availability steady-state HPL from partial simulation.

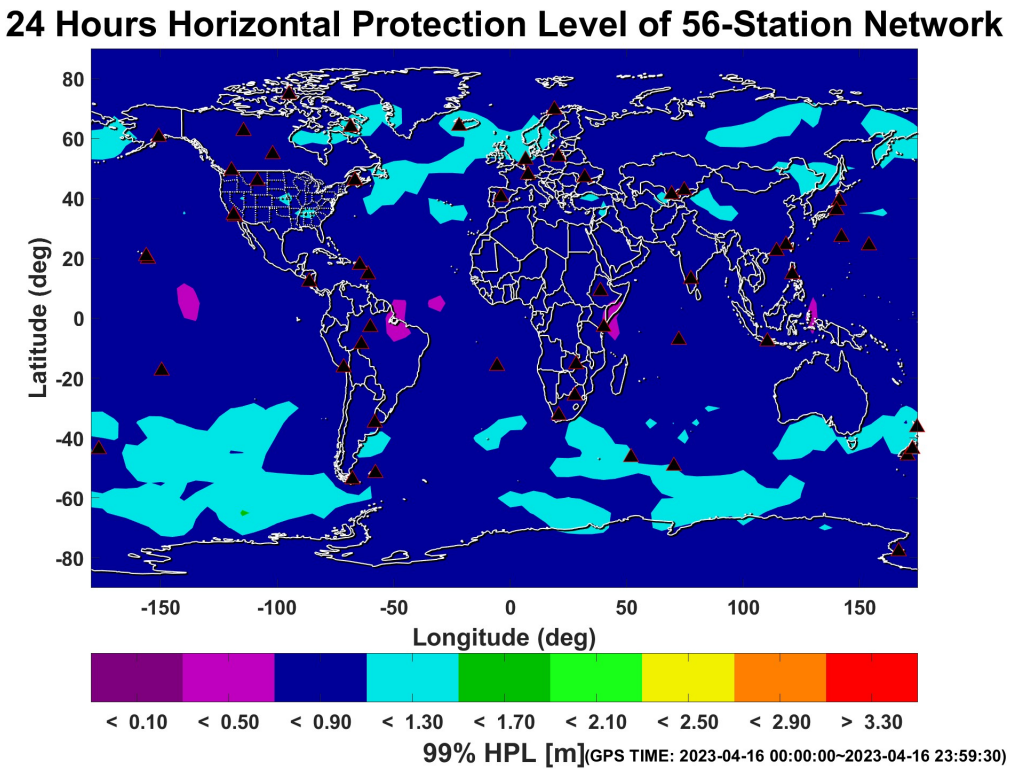


Figure B.2: 99% Availability steady-state HPL from full simulation.

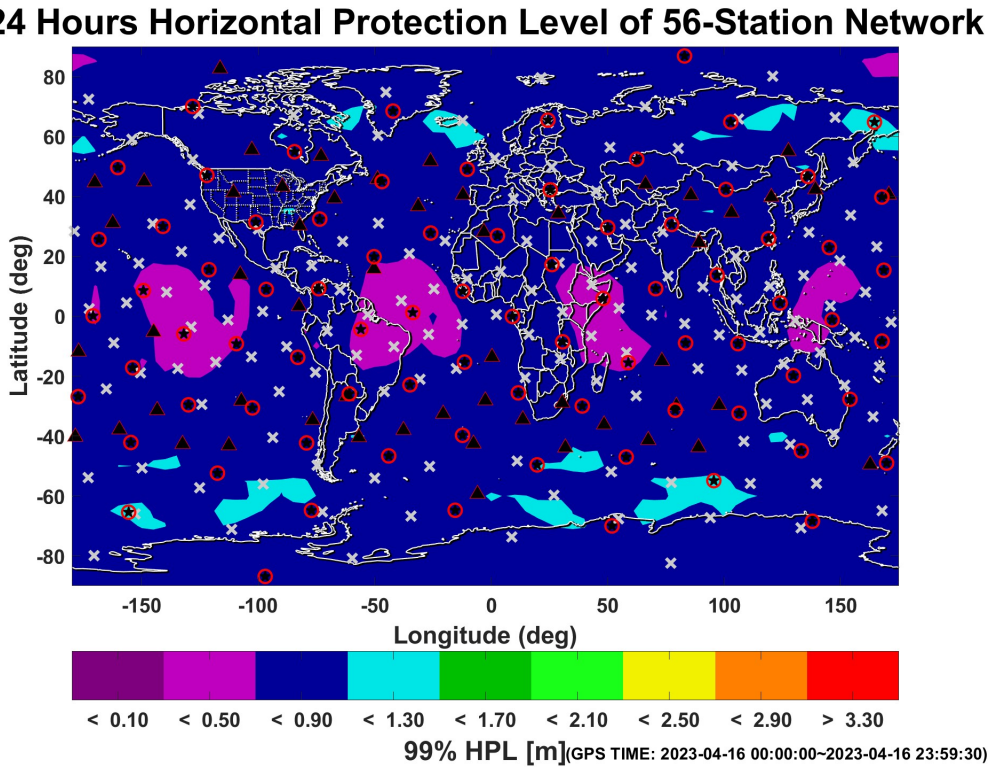


Figure B.3: 99% Availability steady-state HPL from 200 evenly spread-out candidate receivers.

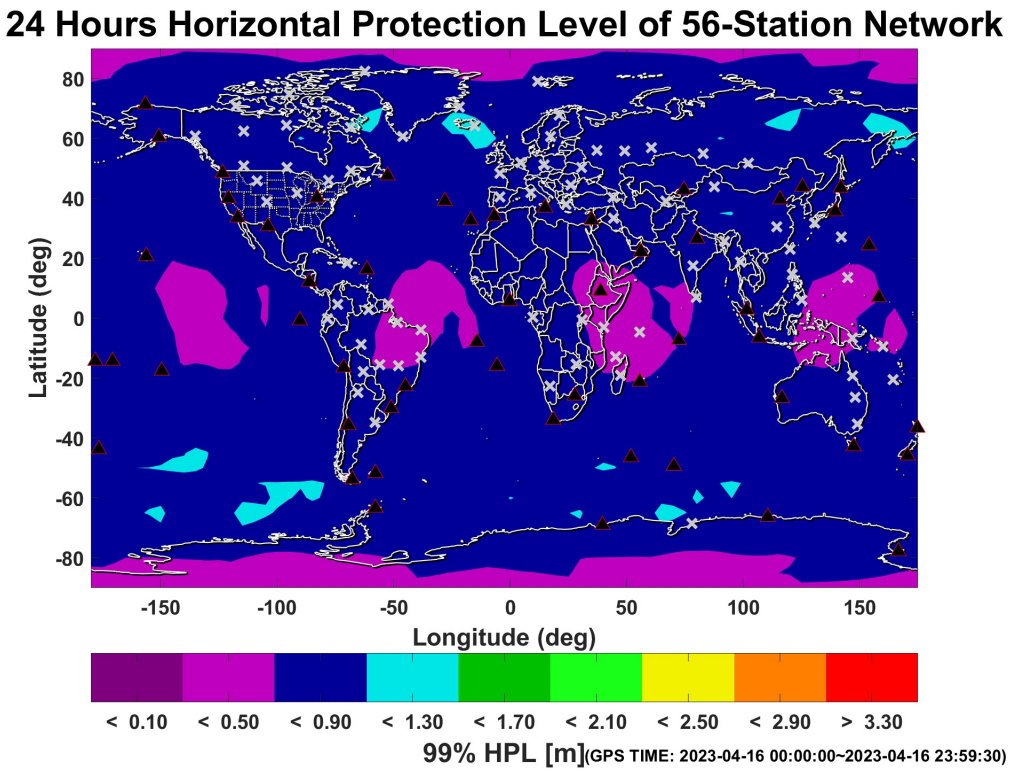


Figure B.4: 99% Availability steady-state HPL from IGS candidate receivers.

Appendix C

Paired Bounds Ad-hoc Algorithm

The ad-hoc algorithm numerically searches for the upper and lower bound to the QQ-plot of stations in a group. The condition for distributions to be considered bounded is defined as follows:

$$B_{lo}(\beta,\sigma) \leq Q_i(X) \leq B_{up}(\beta,\sigma)\ \forall i \in \text{Station}\{1,...,N\} \tag{C.1}$$

where Q_i is the QQ-plot of station i in the group, X is a random variable representing tropospheric delay residuals, β is bias, and σ is standard deviation to the paired-bound B.

The first and last probability points are ignored in the distribution since the cumulative probability distribution will exceed the bound on both ends, any bound satisfying Eq.(C.1) is a permissible bound, and there are infinite combinations of β and σ. In order to narrow down the bounding solutions, three extra criteria are added and formulated in the process of searching for the bound as optimization problems.

C.1 Minimum Bias

$$\min_{\beta,\sigma} \beta \tag{C.2}$$
$$\text{s.t. } B(-\beta,\sigma) \leq Q_i(X) \leq B(\beta,\sigma)\ \forall i \in \text{Station}\{1,...,N\}$$

where the bias is minimized by selecting a pair of bias and standard deviation.

C.2 Constant σ

$$\begin{gathered} \min_{\beta} \beta \\ \text{s.t. } B(-\beta,\sigma) \leq Q_i(X) \leq B(\beta,\sigma)\ \forall i \in \text{Station}\{1,...,N\} \\ \sigma = \sigma_0 \end{gathered} \tag{C.3}$$

where β is minimized with a constant standard deviation.

The reason for calling the ad-hoc algorithm is the solution to the above optimization problems is computed through a series of numerical operations that are specifically designed and only applicable to the search of a paired-bound for the tropospheric QQ-plot presented here. The algorithm is as follows:

Algorithm 1 Determine if residuals are inside Paired-Bound B

```
1: procedure INSIDE(Q,β,σ)
2:     For Q_i in Q
3:         IF B(−β,σ) ≥ Q_i OR B(β,σ) ≤ Q_i
4:             return False
5:     return True
```

Algorithm 2 Find minimal bias given constant standard deviation

procedure FINDBIAS($\boldsymbol{Q}$,β_{init},σ,ϵ)
  Parameters
  $\beta_{up} \leftarrow \beta_{init}$
  $\beta_{lo} \leftarrow \beta_{init}$
  $\beta_{return} \leftarrow \phi$
  $\gamma \in (1.1, 2)$
  For Q_i **in** $\boldsymbol{Q}$
    IF INSIDE($Q_i, \beta_{init}, \sigma$)
      UNTIL INSIDE(Q_i, β_{lo}, σ) **IS** FALSE **DO** $\beta_{lo} \leftarrow \beta_{lo}/\gamma$
    ELSE
      UNTIL INSIDE(Q_i, β_{up}, σ) **IS** FALSE **DO** $\beta_{up} \leftarrow \beta_{up} \cdot \gamma$
    WHILE $\|\beta_{up} - \beta_{lo}\| \geq \epsilon$
      $\beta_{mid} \leftarrow \frac{\beta_{lo}+\beta_{up}}{2}$
      IF INSIDE(Q_i, β_{mid}, σ)
        $\beta_{up} \leftarrow \beta_{mid}$
      ELSE
        $\beta_{lo} \leftarrow \beta_{mid}$
    $\beta_{return} \leftarrow$MAX($\beta_{return}$,$\beta_{up}$)
  return β_{return}

Algorithm 3 Find minimal standard deviation given constant bias

procedure FINDSIGMA($\boldsymbol{Q}$,β,σ_{init},ϵ)
    Parameters
    $\sigma_{up} \leftarrow \sigma_{init}$
    $\sigma_{lo} \leftarrow \sigma_{init}$
    $\sigma_{return} \leftarrow \phi$
    $\gamma \in (1.1, 2)$
    For Q_i **in** $\boldsymbol{Q}$
        UNTIL INSIDE(Q_i, σ_{up}, σ) **IS** True **DO** $\sigma_{up} \leftarrow \sigma_{up}/\gamma$
        $\sigma_{lo} \leftarrow \sigma_{up}$
        UNTIL INSIDE(Q_i, σ_{lo}, σ) **IS** FALSE **DO** $\sigma_{lo} \leftarrow \sigma_{lo} \cdot \gamma$
        WHILE $\|\sigma_{up} - \sigma_{lo}\| \geq \epsilon$
            $\sigma_{mid} \leftarrow \frac{\sigma_{lo}+\sigma_{up}}{2}$
            IF INSIDE(Q_i, β, σ_{mid})
                $\sigma_{up} \leftarrow \sigma_{mid}$
            ELSE
                $\sigma_{lo} \leftarrow \sigma_{mid}$
        $\sigma_{return} \leftarrow$MAX($\sigma_{return}$,$\sigma_{up}$)
    return σ_{return}

Algorithm 4 Find minimal bias

procedure FINDMINBIAS($\boldsymbol{Q}$,β_{init}, σ_{init},ϵ,N)
  Parameters
  $\sigma_{return} \leftarrow \phi$
  $\beta_{return} \leftarrow \phi$
  $\boldsymbol{\sigma_{sample}} \leftarrow \phi$
  $\boldsymbol{\beta_{sample}} \leftarrow \phi$
  For Q_i **in** $\boldsymbol{Q}$
    $\boldsymbol{\sigma_{sample}} \leftarrow$ linspace(0,σ_{init},N) ▷ linspace creates a equally spaced vector ranged from 0 to σ_{init} with size N
    WHILE$\|MAX(\boldsymbol{\sigma_{sample}}) - MIN(\boldsymbol{\sigma_{sample}})\| \geq \epsilon$
      For σ_i **in** $\boldsymbol{\sigma_{sample}}$
        $\beta_i \leftarrow$FINDBIAS($Q_i, \beta_{init}, \sigma_i$)
      $I \leftarrow argmin(\boldsymbol{\beta_{sample}})$
      IF $I \leq N/2$
        $\boldsymbol{\sigma_{sample}} \leftarrow$linspace(MIN($\boldsymbol{\sigma_{sample}}$),$\boldsymbol{\sigma_{sample}}[ceil(N/2)]$,N)
      ELSE
        $\boldsymbol{\sigma_{sample}} \leftarrow$linspace($\boldsymbol{\sigma_{sample}}[floor(N/2)]$,MAX($\boldsymbol{\sigma_{sample}}$),N)
      $\beta_{min} \leftarrow \boldsymbol{\beta_{sample}}[I]$
      $\sigma_{min} \leftarrow \boldsymbol{\sigma_{sample}}[I]$
    $\beta_{return} \leftarrow MAX(\beta_{return}, \beta_{min})$
    $\sigma_{return} \leftarrow MAX(\sigma_{return}, \sigma_{min})$
  return (β_{return},σ_{return})

Algorithm 5 Search Paired-Bound

procedure SEARCHPAIREDBOUND($\boldsymbol{Q}$, α, β, σ, ϵ, N)
  Parameters
  $\beta_{return} \leftarrow \beta$
  $\sigma_{return} \leftarrow \sigma$
  SWITCH α
    Minimal Bias:
      (β_{return},σ_{return})$\leftarrow$FINDMINBIAS($\boldsymbol{Q}$,β, σ,ϵ,N)
    Constant Bias:
      σ_{return} $\leftarrow$FINDSIGMA($\boldsymbol{Q}$,β,σ,ϵ)
    Constant Sigma:
      β_{return} $\leftarrow$FINDBIAS($\boldsymbol{Q}$,β,σ,ϵ)
  return ($\beta_{return}, \sigma_{return}$)

Algorithm (2) computes paired-bound bias given a constant standard deviation. $\boldsymbol{Q}$ are the residuals in the QQ-plot, β_{init} is the initial value for bias, σ is the constant standard deviation, and ϵ is the tolerance for iterations. The algorithm first searches the upper and lower bound of the bias. The upper bound bias is guaranteed to bound the residual, and the lower bound will not. A binary search method is used to narrow down the upper and lower bound biases until the difference is less than the tolerance. The searching process is performed on each station residual, and the maximal bias is returned. Similarly, Algorithm (3) minimizes the standard deviation that bounds the residuals given a constant bias. β is the constant bias specified by the user, and σ_{init} is the initial standard deviation for iterations. The σ_{init} should be sufficiently large such that the residual is not bounded initially. The subsequent line search would determine the upper bound of σ by reducing σ_{init}. Once a feasible upper bound is found, the line search continues to reduce σ until the residual is not bounded to obtain the lower bound σ. With upper and lower bound σ, a binary search is applied to minimize σ. Finally, Algorithm (4) searches for a paired-bound that has minimal bias. To overcome the difficulties of multivariate minimization, a vector of standard deviations is first created, that are equally spaced from 0 to σ_{init} with N elements,

called $\boldsymbol{\sigma_{sample}}$, then samples of minimal bias are found, $\boldsymbol{\beta_{sample}}$, associated with $\boldsymbol{\sigma_{sample}}$ via Algorithm (2), and finally a binary search is conducted to find the region containing the lowest value of bias in $\boldsymbol{\beta_{sample}}$. The process is repeated until $\boldsymbol{\sigma_{sample}}$ converges, and the maximal β and σ are reported from each residual computation. This multivariate minimization process assumes the minimal σ to minimal β relations to be unimodal for all residual distributions. The computation result shows it is a sufficient assumption and can yield minimal bias on all IGS residual distributions in the last ten years.

BIBLIOGRAPHY

[1] MultiMedia IGS. International gnss service, 2026.

[2] Y Jade Morton, Frank van Diggelen, James J Spilker Jr, Bradford W Parkinson, Sherman Lo, and Grace Gao. *Position, navigation, and timing technologies in the 21st century: Integrated satellite navigation, sensor systems, and civil applications, volume 1.* John Wiley & Sons, 2021.

[3] Marek Grzegorzewski, J Cwiklak, Henryk Jafernik, and Andrzej Fellner. Gnss for an aviation. *TransNav: International Journal on Marine Navigation and Safety of Sea Transportation*, 2(4), 2008.

[4] Marco Porretta, David Jimenez Banos, Massimo Crisci, Giorgio Solari, and Alessandra Fiumara. Gnss evolutions for maritime: An incremental approach. *Inside GNSS*, pages 54–62, 2016.

[5] Paweł Zalewski, Andrzej Bąk, and Michael Bergmann. Evolution of maritime gnss and rnss performance standards. *Remote Sensing*, 14(21):5291, 2022.

[6] Calogero Cristodaro, Fabio Dovis, Gianluca Falco, and Marco Pini. Gnss receiver performance in urban environment: Challenges and test approaches for automotive applications. In *2017 International Conference of Electrical and Electronic Technologies for Automotive*, pages 1–6. IEEE, 2017.

[7] Niels Joubert, Tyler GR Reid, and Fergus Noble. Developments in modern gnss and its impact on autonomous vehicle architectures. In *2020 IEEE Intelligent Vehicles Symposium (IV)*, pages 2029–2036. IEEE, 2020.

[8] Farzaneh Zangenehnejad and Yang Gao. Gnss smartphones positioning: Advances, challenges, opportunities, and future perspectives. *Satellite navigation*, 2(1):24, 2021.

[9] Michael A Lombardi and Michael A Lombardi. An evaluation of dependencies of critical infrastructure timing systems on the global positioning system (gps). 2021.

[10] Rebecca Wang and Todd Walter. Characterization and comparison of galileo and gps anomalies. In *Proceedings of the 2023 International Technical Meeting of The Institute of Navigation*, pages 597–610, 2023.

[11] Rebecca Wang, Yu-Fang Lai, Juan Blanch, Dennis Akos, Eric Phelts, and Todd Walter. Characterization of recent gps satellite clock jumps. In *Proceedings of the 2024 International Technical Meeting of The Institute of Navigation*, pages 226–240, 2024.

[12] Michael H Rivers. 2 sops anomaly resolution on an aging constellation. In *Proceedings of the 13th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GPS 2000)*, pages 2547–2550, 2000.

[13] Frederick Blume, Henry T Berglund, Ignacio Romero, and Elisabetta D'Anastasio. Effects of the april 1 st, 2014 glonass outage on gnss receivers. In *AGU Fall Meeting Abstracts*, volume 2014, pages G13A–0508, 2014.

[14] Thomas Etchells, Karen L Aplin, Lucy Berthoud, Ariana Kalavana, and Andrew Larkins. Extreme space weather impacts on gnss timing signals for electricity grid management. *Space Weather*, 22(10):e2023SW003770, 2024.

[15] Yu-Fang Lai, Juan Blanch, Todd Walter, Erin Kahr, Edmond Leahy, Pedro Silva, and Cameron Ellum. Prototyping integrity monitors for ppp fault detections. In *Proceedings of the 36th international technical meeting of the satellite division of the institute of navigation (ION GNSS+ 2023)*, pages 2592–2605, 2023.

[16] Alexandru Ene, Juan Blanch, and Todd Walter. Galileo-gps raim for vertical guidance. In *Proceedings of the 2006 National Technical Meeting of The Institute of Navigation*, pages 432–440, 2006.

[17] Sam Pullen and Mathieu Joerger. Gnss integrity and receiver autonomous integrity monitoring (raim). *Position, navigation, and timing technologies in the 21st century: integrated satellite navigation, sensor systems, and civil applications*, 1:591–617, 2020.

[18] Juan Blanch, Todd Walker, Per Enge, Young Lee, Boris Pervan, Markus Rippl, Alex Spletter, and Victoria Kropp. Baseline advanced raim user algorithm and possible improvements. *IEEE Transactions on Aerospace and Electronic Systems*, 51(1):713–732, 2015.

[19] Todd Walter, Karl Shallberg, Eric Altshuler, William Wanner, Chris Harris, and Robert Stimmler. Waas at 15. *Navigation*, 65(4):581–600, 2018.

[20] Roland Braun, Marc Boyer, Eric Robert, Mickael Dall'Orso, Nicolas Giron, and Carlos López de Echazarreta. Egnos v3 legacy service–preliminary performances. In *Proceedings of the 35th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS+ 2022)*, pages 752–771, 2022.

[21] Atsushi Shimamura. Msas (mtsat satellite-based augmentation system) project status. *Air & Space Europe*, 1(2):63–67, 1999.

[22] Thilantha Dammalage, DN De Silva, and Chalermchon Satirapod. Performance analysis of gps aided geo augmented navigation (gagan) over sri lanka. *Engineering journal*, 21(5):305–314, 2017.

[23] Daehee Won, Chulhee Choi, Eunsung Lee, Hantae Cho, Dongik Jang, Eunok Jang, Heetaek Lim, Ho Sung Lee, Jungja Kim, and Joohap Choi. Development of ground monitoring and control system for korea augmentation satellite system. *Journal of Positioning, Navigation, and Timing*, 12(2):185–200, 2023.

[24] Attila Sipos. Icao standards and recommended practices (sarps). In *International Aviation Law: Regulations in Three Dimensions*, pages 203–231. Springer, 2024.

[25] RTCA. Waas minimum operational performance standards (mops) (rtca/do-229e), 2016.

[26] JF Zumberge, MB Heflin, DC Jefferson, MM Watkins, and Frank H Webb. Precise point positioning for the efficient and robust analysis of gps data from large networks. *Journal of geophysical research: solid earth*, 102(B3):5005–5017, 1997.

[27] Kaori Kawate, Yuki Igarashi, Hideki Yamada, Kyohei Akiyama, Masato Okeya, Hiroshi Takiguchi, Masaya Murata, Toshitaka Sasaki, Saya Matsushita, Shoh Miyoshi, Motoyuki Miyoshi, and Satoshi Kogure. Madoca: Japanese precise orbit and clock determination tool for gnss. *Advances in Space Research*, 71(10):3927–3950, 2023.

[28] Ignacio Fernandez-Hernandez, Adrián Chamorro-Moreno, Simón Cancela-Diaz, J David Calle-Calle, Paolo Zoccarato, Daniel Blonski, Tommaso Senni, F Javier de Blas, Carmelo Hernández, Javier Simón, et al. Galileo high accuracy service: initial definition and performance. *GPS solutions*, 26(3):65, 2022.

[29] Gary Johnston, Anna Riddell, and Grant Hausler. The international gnss service. *Springer handbook of global navigation satellite systems*, pages 967–982, 2017.

[30] Xiaoming Chen, Timo Allison, Wei Cao, Kendall Ferguson, Simon Grunig, Victor Gomez, Adrian Kipka, Jan Kohler, Herbert Landau, Rodrigo Leandro, et al. Trimble rtx, an innovative new approach for network rtk. In *Proceedings of the 24th international technical meeting of the satellite division of the institute of navigation (ION GNSS 2011)*, pages 2214–2219, 2011.

[31] Kevin Sheridan, Paul Alves, Sara Masterson, Jennifer Busser, Zafer Sadeque, Paul Verlaine Gakne, Stavros Melachroinos, and Robert Martin. Terrastar x: Precise point positioning with fast convergence and integrity. In *Proceedings of*

the 32nd International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS+ 2019), pages 1916–1943, 2019.

[32] Ulrich Weinbach, Markus Brandl, Xiaoming Chen, Herbert Landau, Fabian Pastor, Nico Reussner, and Carlos Rodriguez-Solano. Integrity of the trimble® centerpoint rtx correction service. In *proceedings of the 31st international technical meeting of the satellite division of the institute of navigation (ION GNSS+ 2018)*, pages 1902–1909, 2018.

[33] U.S. Space Force. Interface specification is-gps-200, navstar gps space segment/navigation user segment interfaces. Revision n, U.S. Coast Guard Navigation Center, 2022.

[34] Steve Hilla. The extended standard product 3 orbit format (sp3-d) 21 february 2016, 2016.

[35] J Paul Collins and Richard Brian Langley. *A tropospheric delay model for the user of the wide area augmentation system*, volume 20. Department of Geodesy and Geomatics Engineering, University of New Brunswick ..., 1997.

[36] Arthur E Niell. Global mapping functions for the atmosphere delay at radio wavelengths. *Journal of geophysical research: solid earth*, 101(B2):3227–3246, 1996.

[37] Elizabeth J Petrie, Manuel Hernández-Pajares, Paolo Spalla, Philip Moore, and Matt A King. A review of higher order ionospheric refraction effects on dual frequency gps. *Surveys in geophysics*, 32(3):197–253, 2011.

[38] M Mainul Hoque and Norbert Jakowski. New correction approaches for mitigating ionospheric higher order effects in gnss applications. In *Proceedings of the 25th International Technical Meeting of the Satellite Division of the Institute of Navigation (ION GNSS 2012)*, pages 3444–3453, 2012.

[39] Neil Ashby. Relativity in the global positioning system. *Living Reviews in relativity*, 6(1):1–42, 2003.

[40] Geoffrey Blewitt. Carrier phase ambiguity resolution for the global positioning system applied to geodetic baselines up to 2000 km. *Journal of Geophysical Research: Solid Earth*, 94(B8):10187–10203, 1989.

[41] Horst Hartinger and Fritz K Brunner. Variances of gps phase observations: the sigma- model. *GPS solutions*, 2(4):35–43, 1999.

[42] Gérard Petit and Brian Luzum. Iers conventions (2010). Technical report, 2010.

[43] Jiun-Tsong Wu, Sien C Wu, George A Hajj, Willy I Bertiger, and Stephen M Lichten. Effects of antenna orientation on gps carrier phase. *Manuscripta geodaetica*, 18(2):91–98, 1993.

[44] Tobias Kersten and Steffen Schön. Gps code phase variations (cpv) for gnss receiver antennas and their effect on geodetic parameters and ambiguity resolution. *Journal of Geodesy*, 91(6):579–596, 2017.

[45] William G Melbourne. The case for ranging in gps based geodetic systems. In *Proceedings 1st International Symposium on Precise Positioning with the Global Positioning System, US Department of Commerce, Rockville, Maryland, 1985*, 1985.

[46] G Wubbena. Software developments for geodetic positioning with gps using ti 4100 code and carrier measurements. In *Proceedings 1 st International Symposium on Precise Positioning with the Global Positioning System, US Dept. of Commerce, Rockville, Maryland, 1985*, 1985.

[47] Geoffrey Blewitt. An automatic editing algorithm for gps data. *Geophysical research letters*, 17(3):199–202, 1990.

[48] Kazuma Gunning. *Safety critical bounds for precise positioning for aviation and autonomy*. Stanford University, 2021.

[49] RTCA. Minimum performance operational standards (mops) for global positioning system/satellite-based augmentation airborne equipment. rtca document do-229d, Dec. 2006.

[50] Mingqiang Xie, Ang Liu, Zishen Li, Ningbo Wang, Ahmed El-mowafy, Liang Wang, Zhiyu Wang, and Dongyan Wei. Quality monitoring of real-time precise satellite orbit and clock corrections for generating health indicators. *Measurement*, 247:116707, 2025.

[51] Yu Xue, Fu Zheng, Guangyu Zhou, Wenyue Chao, Dong Zhang, and Chuang Shi. A new quality monitoring method for real-time precise orbit and clock products. *GPS Solutions*, 28(3):125, 2024.

[52] Liuqi Wang, Liang Li, Min Li, Ningbo Wang, Weitianhua Ma, and Ruijie Li. Integrity monitoring of the real-time precise satellite orbit and clock correction products in the range domain. *Measurement Science and Technology*, 36(5):056305, 2025.

[53] Run Ji, Xinyuan Jiang, Xinghan Chen, Huizhong Zhu, Maorong Ge, and Frank Neitzel. Quality monitoring of real-time gnss precise positioning service system. *Geo-Spatial Information Science*, 26(1):1–15, 2023.

[54] Kazuma Gunning, Juan Blanch, Todd Walter, Lance de Groot, and Laura Norman. Design and evaluation of integrity algorithms for ppp in kinematic applications. In *Proceedings of the 31st International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS+ 2018)*, pages 1910–1939, 2018.

[55] Juan Blanch, Todd Walter, and Per Enge. Raim with optimal integrity and continuity allocations under multiple failures. *IEEE Transactions on Aerospace and Electronic Systems*, 46(3):1235–1247, 2010.

[56] Tomoji Takasu and Akio Yasuda. Development of the low-cost rtk-gps receiver with an open source program package rtklib. In *International symposium on GPS/GNSS*, volume 1, pages 1–6. International Convention Center Jeju Korea Seogwipo-si, Republic of Korea, 2009.

[57] BD Anderson. 0. and moore, jb (1979) optimal filtering. *Englewood Cliffs, NJ: P rentice-Hall*, 1979.

[58] Shau-Shiun Jan, Wyant Chan, and Todd Walter. Matlab algorithm availability simulation tool. *GPS solutions*, 13(4):327–332, 2009.

[59] Siddharth Joshi and Stephen Boyd. Sensor selection via convex optimization. *IEEE Transactions on Signal Processing*, 57(2):451–462, 2008.

[60] Nicholas Metropolis and Stanislaw Ulam. The monte carlo method. *Journal of the American statistical association*, 44(247):335–341, 1949.

[61] Fuzhen Zhang. *The Schur complement and its applications*. Springer Science & Business Media, 2006.

[62] Inc. CVX Research. CVX: Matlab software for disciplined convex programming, version 2.0. https://cvxr.com/cvx, August 2012.

[63] M. Grant and S. Boyd. Graph implementations for nonsmooth convex programs. In V. Blondel, S. Boyd, and H. Kimura, editors, *Recent Advances in Learning and Control*, Lecture Notes in Control and Information Sciences, pages 95–110. Springer-Verlag Limited, 2008. http://stanford.edu/~boyd/graph_dcp.html.

[64] Anton Semechko. Suite of functions to perform uniform sampling of a sphere. https://github.com/AntonSemechko/S2-Sampling-Toolbox, 2026. Retrieved May 10, 2026.

[65] Todd Walter Juan Blanch, Yu-Fang Lai. Evaluation of tropospheric and mixed single - dual frequency error models in advanced raim. *Proceedings of ION ITM*, 2023.

[66] Johannes Böhm, Gregor Möller, Michael Schindelegger, Gregory Pain, and Robert Weber. Development of an improved empirical model for slant delays in the troposphere (gpt2w). *GPS solutions*, 19(3):433–441, 2015.

[67] Daniel Landskron and Johannes Böhm. Vmf3/gpt3: refined discrete and empirical troposphere mapping functions. *Journal of Geodesy*, 92(4):349–360, 2018.

[68] EUROCAE. Ed-259a: minimum operational performance specification for galileo/global positioning system/satellite-based augmentation system airborne equipment, v0.16 (draft), Dec. 2023.

[69] Gary A McGraw. Tropospheric error modeling for high integrity airborne gnss navigation. In *Proceedings of the 2012 IEEE/ION Position, Location and Navigation Symposium*, pages 158–166. IEEE, 2012.

[70] JL Davis, TA Herring, II Shapiro, AEE Rogers, and Gunnar Elgered. Geodesy by radio interferometry: Effects of atmospheric modeling errors on estimates of baseline length. *Radio science*, 20(6):1593–1607, 1985.

[71] HD Black and A Eisner. Correcting satellite doppler data for tropospheric effects. *Journal of Geophysical Research: Atmospheres*, 89(D2):2616–2626, 1984.

[72] Willy Bertiger, Yoaz Bar-Sever, Angie Dorsey, Bruce Haines, Nate Harvey, Dan Hemberger, Michael Heflin, Wenwen Lu, Mark Miller, Angelyn W Moore, et al. Gipsyx/rtgx, a new tool set for space geodetic operations and research. *Advances in space research*, 66(3):469–489, 2020.

[73] Bruce E Vollmer, D Ostrenga, A Savtchenko, J Johnson, J Wei, W Teng, and I Gerasimov. Making earth science data records for use in research environments (measures) projects data and services at the ges disc. Technical report, 2011.