Journal of Global Positioning Systems (2025)
Vol. 21, Issue 1: 104-106

Theory and methods of testing and integrity monitoring for multiple gross errors

PhD Candidate: Yangkang Yu
Email: yangkang.yu@outlook.com
Supervisors: Assoc. Prof. Ling Yang and Prof. Yunzhong Shen
University: Tongji University
Defense date: March, 2025

ABSTRACT

In the fields of mathematical statistics, geodetic surveying, and satellite navigation, the handling of outliers has always been a crucial aspect in ensuring the accuracy of models and the reliability of systems. An outlier is defined as an observation that significantly differs from other observations in terms of scale, frequency, or generation mechanism. Outliers typically have a significant impact on parameter estimation results based on the least squares method, leading to biased parameter estimates, inaccurate confidence intervals, and reduced model prediction capabilities. Consequently, this directly affects the precision and reliability of coordinate calculations, deformation monitoring, and frame maintenance in practical applications, and even poses a threat to the integrity and safety of systems in fields such as aerospace, transportation, and disaster monitoring. For measurement data, outliers are mainly caused by gross errors. Therefore, the key to addressing outlier issues in measurement data lies in in-depth research and effective management of gross errors.

With the continuous development of modern observation technology, diverse observation methods and complex environments have introduced greater challenges in outlier handling. The increasing scale of diverse observation data has led to a significant increase in the number of gross errors, rendering traditional methods for handling individual or small numbers of gross errors insufficient. As observation environments become more complex, the types of gross errors have increased, making traditional methods targeting only simple gross errors no longer applicable. Moreover, with the widespread adoption of automated and intelligent processing workflows, not only is precise detection of gross errors required, but also effective monitoring of system integrity. To address these challenges, this thesis conducts a systematic investigation into theoretical frameworks and practical methods for handling multidimensional gross errors, bridging Classical and Bayesian statistical approaches. The paper is divided into seven chapters and is structured as follows:

Chapter 1 introduces the research background and significance, providing a comprehensive review of the current state of outlier processing in mathematical statistics, geodesy, and satellite navigation. It defines the fundamental concepts of outliers and gross errors, analyzing their impact on parameter estimation and system safety. The chapter

outlines the limitations of existing methods in the face of massive, complex, and heterogeneous observation data, highlighting the necessity for researching multiple gross errors. It establishes the research objectives, which focus on developing precise detection methods for multiple error types and robust integrity monitoring frameworks, and presents the overall technical route of the thesis.

Chapter 2 systematically reviews the fundamental theories and methods for outlier processing and integrity monitoring. It begins by discussing outlier handling in mathematical statistics, contrasting classical frequency-based methods with Bayesian approaches and influence function diagnostics. In the context of geodesy, it details the principles of Baarda's w-test, Data Snooping (DS), Likelihood Ratio Tests (LRT), and the Detection, Identification, and Adaptation (DIA) procedure. Furthermore, it examines integrity monitoring in satellite navigation, specifically Receiver Autonomous Integrity Monitoring (RAIM) and Advanced RAIM (ARAIM), analyzing their fault detection mechanisms and protection level calculations. This chapter provides the theoretical basis for the subsequent methodological improvements proposed in the thesis.

Chapter 3 constructs a statistical model for gross errors, termed the Bernoulli-Gaussian (BG) model, and proposes a parameter estimation method based on the Expectation-Maximization (EM) algorithm. The BG model conceptualizes gross errors as the product of a Bernoulli variable, representing the occurrence state, and a Gaussian variable, representing the error magnitude. This modeling approach allows for the unification of various outlier generation mechanisms, effectively explaining common observation models such as the mean-shift and variance-inflation models by adjusting model parameters. The chapter details the derivation of the EM algorithm for estimating these parameters in linear observation models, employing a sequential form to unify data processing for both static and dynamic scenarios. Numerical experiments involving linear fitting and GNSS positioning verify the algorithm's effectiveness, establishing the BG model as a guiding framework that distinguishes between Classical Statistical methods (where outliers are deterministic biases) and Bayesian Statistical methods (where outliers possess prior distributions).

Chapter 4 addresses the limitations of classical detection methods by proposing an extended w-test and a novel iterative detection strategy. It analyzes the susceptibility of Baarda's *w*-test and Data Snooping to "masking" and "swamping" effects when multiple gross errors are present. To mitigate these issues, the chapter introduces an extended form of the w-test accompanied by rigorous reliability measures. It theoretically reinterprets Iterative Data Snooping (IDS) and proposes a new method called Data Refining (DR) along with its iterative counterpart, Iterative Data Refining (IDR). Unlike IDS, which iteratively selects outliers from a presumed normal dataset to populate an anomalous set, IDR operates via a reverse mechanism: it iteratively validates and retrieves normal observations from a potentially contaminated dataset back into the normal set. Theoretical analysis and experiments demonstrate that IDR effectively alleviates masking and swamping, offering superior robustness compared to IDS in multiple outlier scenarios.

Chapter 5 extends the detection framework to handle diverse error types simultaneously by proposing Generalized Data Snooping (GDS) and Generalized Data Refining (GDR). It establishes an extended Gauss-Markov model that encompasses all potential gross error sources, facilitating the construction of multiple hypotheses corresponding to different error types. Based on the Likelihood Ratio Test (LRT) criterion, the chapter derives test statistics and decision rules for these multiple hypotheses. GDS and GDR are developed to generalize the logic of DS and DR, allowing them to process different types and quantities of gross errors concurrently. GNSS positioning results indicate that GDS is optimal for scenarios with favorable observation conditions demanding high estimation precision, while GDR demonstrates superior performance in complex environments where estimation robustness against multiple gross errors is paramount.

Chapter 6 proposes a class of Bayesian Detection, Identification, and Adaptation (DIA) methods based on Bayesian point estimation theory. Recognizing the value of prior information, this chapter utilizes the BG model to construct prior probability distributions for gross errors. It establishes two specific approaches: DIA based on Maximum A Posteriori estimation (DIA-MAP) and DIA based on Bayesian Hypothesis Testing (DIA-BHT). Furthermore, the chapter introduces Probability Level and Confidence Level as novel quality indicators for Bayesian DIA. The Probability Level predicts the likelihood of making a correct decision beforehand, while the Confidence Level assesses the credibility of a decision post-factum. GNSS positioning experiments confirm that these Bayesian DIA methods effectively detect and identify gross errors of varying types, quantities, and magnitudes, ensuring robust parameter estimation even under challenging conditions.

Chapter 7 extends the Bayesian framework from point estimation to interval estimation to address the critical requirement of system integrity, proposing the Bayesian Receiver Autonomous Integrity Monitoring (BRAIM) method. BRAIM is designed to not only estimate parameters but also rigorously assess their precision and integrity. The method utilizes the BG model to construct prior distributions, subsequently deriving the full posterior distribution of the position parameters. For point estimation, a Minimum Mean Square Error (MMSE) estimator is adopted to maximize positioning accuracy. For integrity monitoring, Bayesian interval estimation theory is applied to calculate the Protection Level (PL) based on a specified alarm risk. Simulation and real-world GNSS experiments demonstrate that, compared to traditional RAIM and ARAIM methods, BRAIM significantly reduces both positioning errors and Protection Levels, thereby simultaneously enhancing the accuracy and availability of the navigation system.

The summary and outlook sections mainly summarize the main research content of this paper, point out the limitations of the current research, plan the next research work, and give an outlook for future research work, such as optimizing the initialization of IDR for real-time efficiency, extending Bayesian methods to kinematic positioning modes, and developing adaptive methods for setting BG model parameters in scenarios lacking historical data.

Key words: Outliers, Gross Errors, Least Squares, GNSS, BG Model, Classical Statistics, Bayesian Statistics, *w*-test, IDS, IDR, Likelihood Ratio Test, DIA, Integrity Monitoring, ARAIM, BRAIM