



A Wasserstein distance-based technique for the evaluation of GNSS error characterization

Jinpei Chen^{1,2,3} · Wenyu Zhang² · Bingqing Feng⁴ · Nan Zhi^{2,5} · Yi Zhao² · Mingquan Lu¹

Received: 29 September 2023 / Accepted: 1 March 2024 / Published online: 25 March 2024
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2024

Abstract

The characteristics of residual errors in GNSS positioning are crucial for fault detection and integrity monitoring. Despite the wide use of the zero-mean Gaussian assumption in the navigation community, studies highlight non-Gaussian traits and heavy-tailed patterns in residual errors. The problem will be even more challenging for users in difficult environments where residual errors consist of a combination of multiple modes with high complexity and cannot be fitted with known distributions or empirical models. To address these issues, our work introduces a novel approach leveraging the Wasserstein distance for assessing the performance of error characterization and fault modeling. However, relying solely on the Wasserstein distance value for direct similarity assessment is hindered by its dependency on dimensionality. We propose a second-order Gaussian Wasserstein distance-based precision metric to offer a quantitative evaluation of GNSS error models in terms of both goodness-of-fit and underlying assumptions. We also establish a robust scoring criterion to distinguish between various GNSS error models, ensuring comprehensive evaluation. The proposed method is validated through a known high-dimensional Gaussian model, achieving a score of 99.95 over 100 with a sample size of 10,000. To demonstrate the capability in dealing with complexity, two multivariate complex GNSS models incorporating copula functions to capture intricate inter-dimensional correlations are established and assessed by our approach. Experimental results show that the method can effectively deliver the evaluation of goodness-of-fault models using the establishment of a universal criteria with different dimensions. It provides a quantitative measure on the goodness of fittings and enhances the modeling to reflect the reality, therefore solving the problems raised above. In addition, with this technique, the close-to-reality fault models can be chosen to generate simulated faulty datasets, thus benefiting algorithm testing and improvement. This is also beneficial to more accurate integrity risk assessment to avoid overbounding- or underbounding-resulted false or missed alert.

Keywords GNSS · Fault model evaluation · Wasserstein distance · Scoring criterion · Copula function

Introduction

As a critical piece of contemporary infrastructure, Global Satellite Navigation Systems (GNSS) are experiencing rapid advancements to cater to the exponential growth of GNSS user groups. However, there are numerous factors that hinder the optimal performance of these satellite navigation systems, which can be typically divided into several categories: satellite and signal errors, atmospheric anomalies, product (correction) errors, operating environment anomalies, and user-end errors (Du et al. 2021). The portions of the above errors that have not been completely eliminated by data processing collectively form the residual errors in the navigation system, named navigation residual estimation errors. Magnificent errors can lead to navigation fault/failure, including range/positioning failure, integrity failure, overbounding

✉ Wenyu Zhang
wenyu.zhang@wz-inc.com

¹ Department of Electronic Engineering, Tsinghua University, Beijing 100084, China

² Qianxun Spatial Intelligence Inc, Shanghai 200438, China

³ Qianxun Spatial Intelligence (Zhejiang) Inc, Deqing 313200, Zhejiang, China

⁴ State Grid Beijing Electric Power Company, Beijing 100031, China

⁵ School of Information Science and Technology, Fudan University, Shanghai 200438, China

failure, and statistical failure (Milner et al. 2016). Range/positioning failure happens when errors in range/position domain exceed a tolerable threshold. Integrity failure represents the event that the probability of positioning failure without timely alert exceeds the predefined integrity risk. Overbounding failure refers to the cumulative distribution function (CDF) curve of real data that lies outside the CDF of the assumed model. And statistical failure denotes the probability distribution function (PDF) of error distribution changes to a failure state by misdefined statistical parameters. Aiming at better navigation performance, a thorough comprehension of GNSS error characteristics is a foundation of handling all the above possible GNSS error sources and fault/failure modes.

A considerable amount of research has been devoted to the characterization of GNSS errors and faults (Anderson and Ellis 1971). Errors are approximated by simplified time-dependent mathematical functions, as demonstrated in Bhatti's (Bhatti et al. 2007) work, including step error, ramp error, random noise, random walk, oscillation, and bias. It should be noted that these approximations aim to depict temporal trends rather than provide precise models. Various common distributions are utilized to capture the CDF/PDF of errors for simplicity and intuitiveness. And Gaussian assumption is the most widely adopted in navigation community. Navigation residual errors are usually assumed to be normally distributed, and the probabilities of false alert and missed detection in GNSS integrity monitoring are thereby assumed to follow Chi-squared distributions (Panagiotakopoulos et al. 2014). However, the fact is error distributions are usually heavy-tailed in the presence of gross errors, blunders, and faults (Hsu 1979). And the behavior of the tails is significant, especially in integrity risk assessment, so we must treat it in a more 'pessimistic' way. Better descriptions such as Exponential, Laplace (Shively and Braff 2000), and Generalized Extreme Value (GEV) distributions (Panagiotakopoulos et al. 2008) have been suggested by researchers to deliver better descriptions to the heavy tails. To decide which is the best alternative to the Gaussian model, we need to estimate which distribution is the closest to the characteristics of real data, so we require a technique to fully validate the goodness-of-fit of each Gaussian alternative for the tails of error distributions.

The 'tails,' in other words, the faults, are where the characterization problem exactly lies in. The goodness-of-fit verification of GNSS fault modeling can be generally categorized into two types: mathematical and positional. From the perspective of mathematics, for faults that can be fitted with known distributions, conventional statistical metrics like correlation coefficients, root mean square error, skewness, and excess kurtosis (Joanes and Gill 1998) are generally utilized to gauge the similarity between fault models and actual data. Besides, several goodness-of-fit techniques (described

in Methodology section) originating from the fields of mathematical statistics and machine learning are available. While from the positional perspective, modeling accuracy is evaluated through positioning performance. For instance, when a machine learning technique is used to model multipath faults (Pan et al. 2023), any improvement in position accuracy is considered an indicator of the accuracy of GNSS fault modeling. However, there are several shortcomings in the current techniques for GNSS fault model evaluation. One problem stems from the limited evaluation capacity of the common statistical metrics. These metrics, on the one hand, often require parameter estimation and typically provide specific characteristics of a distribution and do not consider the overall similarity of distributions. On the Other hand, they are typically suitable for low-dimensional, single-factor fault situations with relatively simple mathematical distributions, but they are impractical for high-dimensional, multifactor fault types. Another issue arises with the growth of machine learning techniques is that there is an increasing number of novel fault models, particularly those generated by black box methods or those with extremely complex structures. There is a need for an evaluation method that can be applied to both new and established models, enabling us to determine which model is more suitable.

Brief reviews on the methods of goodness-of-fit and the details of the Wasserstein distance-based scoring method are presented first. This is followed by the method validation using simulation and demonstration using real data from correction services and road tests. The conclusions are given in the last section.

Methodology

This part initially provides a concise overview of existing goodness-of-fit methodologies, laying the groundwork for the selection of the Wasserstein distance as the fundamental element within the proposed algorithm. We introduce the primary implementation of the Wasserstein distance, outlining its strengths and limitations. Recognizing the inadequacy of the raw Wasserstein distance in effectively measuring the similarity between real and modeled data across varied dimensions, we propose a novel Gaussian Wasserstein distance-based precision metric. This innovative metric aims to rectify evaluation disparities arising from diverse dimensions. Additionally, we establish a model scoring criterion, enhancing the interpretability of model assessments.

Goodness-of-fit

In scientific research, evaluating the compatibility between models and observed data, known as goodness-of-fit, is essential for validating a model's applicability. Several

techniques are proposed to check the fitting goodness of data and known distribution candidates, each with its unique strengths and weaknesses. The Chi-Square Goodness-of-Fit Test (Pearson 1992), suited for discrete distributions, is straightforward but less effective for small samples and continuous distributions. The Kolmogorov–Smirnov Test (Smirnov 1948) evaluates continuous distributions, especially normality, but loses accuracy with small samples. The Anderson–Darling Test (Anderson and Darling 1954), sensitive to tail fitting in continuous distributions, is ideal for small samples but involves complex calculations. The Cramér–von Mises Test (Cramér, 1928) also assesses continuous distributions and tail fitting, suitable for small samples but might miss non-normal distributions. The Lilliefors Test (Lilliefors 1967) checks goodness-of-fit to normal distributions, fitting small samples but limited to normality. Bayesian Goodness-of-Fit Tests (Rubin 1984) incorporate prior information, beneficial for low sample size or data-scarce cases but require careful prior selection and entail computational complexity.

The major limitation of traditional goodness-of-fit tests is that they often focus on specific distributional properties or assume parametric models, while they may provide misleading results if the data do not conform to known distributions. To tackle these, many sophisticated and flexible statistical tools are developed with the advancements in machine learning, which offer a more comprehensive assessment of goodness-of-fit and provide more accurate results. The popular methods are Kullback–Leibler (KL) divergence (Kullback and Leibler 1951; Dwass and Kullback 1960), Jensen–Shannon (JS) divergence (Jensen 1906; Shannon 1948), and Wasserstein distance (Vaserstein 1969). These methods significantly outperform traditional statistical metrics and goodness-of-fit techniques in that 1) they offer global comparison measures that take into account not only the central tendency (mean) and dispersion (variance) of distributions but also capture differences in their shapes; 2) they do not require prior assumptions on the distributions; and 3) they can serve as distance metrics, whereas traditional methods may not possess metric properties. While KL divergence is prone to an asymmetry issue when measuring the similarity of two distributions, this issue can be rectified by JS divergence. However, both KL divergence and JS divergence are ineffective when the two distributions do not intersect.

Wasserstein distance addresses this problem by providing an effective distance metric and gradient information even for two non-overlapping distributions in a high-dimensional space (Panaretos and Zemel 2019). Given these advantages, Wasserstein distance is employed here as a method of evaluating the precision of GNSS fault models.

Raw Wasserstein distance

Consider a fault type f with a dimension of $n(n \geq 1)$, with its true distribution denoted as P_{real} and its modeled distribution denoted as P_{simu} . Suppose that the sequence of k samples x_1, x_2, \dots, x_k is collected from the real datasets or sampled from P_{real} . Since each sample is n -dimensional, we denote the sequence of the sample points in i^{th} dimension as $X^{(i)}$ and rewrite it as

$$X^{(i)} = [x_1^{(i)}, x_2^{(i)}, \dots, x_k^{(i)}]^T, i = 1, 2, \dots, n \tag{1}$$

where $x_k^{(i)}$ is the i^{th} dimension of sample point k . Then the real sample points can be represented as

$$X = \{X^{(1)}, X^{(2)}, \dots, X^{(n)}\} \tag{2}$$

We denote the fault model of fault type f as M_f . A sequence of k samples y_1, y_2, \dots, y_k is collected from the fault model. Similarly, we denote the sequence of the simulated points in i^{th} dimension as $Y^{(i)}$ and rewrite it as

$$Y^{(i)} = [y_1^{(i)}, y_2^{(i)}, \dots, y_k^{(i)}]^T, i = 1, 2, \dots, n \tag{3}$$

where $y_k^{(i)}$ is the i^{th} dimension of simulated point k . Then the simulated sample points can be represented as

$$Y = \{Y^{(1)}, Y^{(2)}, \dots, Y^{(n)}\} \tag{4}$$

Both X and Y can be treated as a $k \times n$ matrix. p -order Wasserstein distance $W(P_{\text{real}}, P_{\text{simu}})^p$ P_{real} and P_{simu} can be calculated by (Vaserstein 1969)

$$W(P_{\text{real}}, P_{\text{simu}})^p = \left(\inf_{\gamma \sim \Pi(P_{\text{real}}, P_{\text{simu}})} E_{(x,y) \sim \gamma} [d(X - Y)^p] \right)^{1/p} \tag{5}$$

where $\Pi(P_{\text{real}}, P_{\text{simu}})$ represents the collection of all possible joint distributions between P_{real} and P_{simu} for the fault type f . γ denotes a possible joint distribution belonging to this collection. $d(X - Y)$ is the Euclidean distance between X and Y . $(x, y) \sim \gamma$ means sampling from the joint distribution γ . $E_{(x,y) \sim \gamma} [d(X - Y)^p]$ represents mathematical expectation of $d(X - Y)^p$. And \inf represents the infimum.

The smaller the Wasserstein distance, the lower the cost required to move the modeled distribution to the true distribution, indicating a higher similarity between the two distributions. This, in turn, demonstrates a higher accuracy of the established fault model.

Wasserstein distance exhibits many good properties which can be summarized as follows:

- The properties of the distributions are not tightly constrained. In other words, the fault model can be arbitrary

finite-dimensional, discrete, or continuous, generated by known distributions or black box methods.

- Even though the real faulty data and its fault model do not overlap in high dimensions, Wasserstein distance is still able to measure their similarities.
- Although the solution of Wasserstein distance in high-dimensional space can be very complex, a special case is second-order Gaussian Wasserstein distance, which has analytical solution, so we can take good advantage of it.

However, when utilizing Wasserstein distance for fault model evaluation, two problems need to be tackled:

- Models with different dimensions may have the same value of Wasserstein distance but different precision. The precision metric of models should be uncoupled with the dimensionality and more intuitive.
- Users cannot decide whether the model is good or not if merely given the numeric value of the Wasserstein distance.

Based on above discussions, an indicative criterion should be established for reference.

Gaussian Wasserstein distance-based precision metric

For two n-dimensional Gaussian distributions (each dimension is independently and identically distributed), $\mu_1 \sim N_1(m_1, C_1), \mu_2 \sim N_2(m_2, C_2)$, the second-order Gaussian Wasserstein distance, denoted as $W_2(\mu_1, \mu_2)^2$, has a closed-form solution, which is expressed as (Takatsu 2011)

$$W_2(\mu_1, \mu_2)^2 = |m_1 - m_2|^2 + \text{Tr}\left(C_1 + C_2 - 2(C_1^{1/2}C_2C_1^{1/2})^{1/2}\right) \tag{6}$$

where $(|\cdot|)$ denotes the Euclidean norm, $\text{Tr}(\cdot)$ is the trace of a matrix, and $C_1^{1/2}$ represents the square root of the covariance matrix C_1 .

More intuitively, assume μ_1 as a n-dimensional standard Gaussian distribution and μ_2 as a n-dimensional Gaussian distribution with a mean vector of zero and diagonal elements of the covariance matrix equal to σ^2 . In this case, (6) can be further simplified to

$$W_2(\mu_1, \mu_2)^2 = n \cdot (\sigma - 1)^2 \tag{7}$$

We denote the second-order Wasserstein distance between the true distribution and the generated distribution corresponding to the n-dimensional fault type f as W_f^2 . For fault model M_f , let W_f^2 be equal to $W_2(\mu_1, \mu_2)^2$ and $\sigma \geq 1$. Hence, σ can be calculated by

$$\sigma = \sqrt{\frac{W_f^2}{n}} + 1 \tag{8}$$

Since the dimension n is known to us, σ can serve as the precision metric of the fault model. The smaller σ is, the higher the accuracy of the model.

It can be derived from (8) that Wasserstein distance is related to the dimensionality of the model. Taking a 2D model and a 3D model as examples, even if both models have similar accuracy, the Wasserstein distance of the 2D model will be smaller than that of the 3D model due to the difference in dimensionality. By converting Wasserstein distance into the precision metric σ , it eliminates the evaluation discrepancy caused by different dimensions and facilitates the establishment of a unified model scoring criterion.

Model scoring criterion

For a given fault type f , there is a full database generated from real world. Three datasets with the same sample size N are built: two real datasets from the full database (denoted as set 1 and set 2) and one simulated dataset from the fault model (denoted as set 3). We suppose that one database exhibits the same characteristics.

The second-order Wasserstein distance $W_{1,2}^2$ between set 1 and set 2 is calculated by Eq. (5); let $W_{1,2}^2$ be equal to W_f^2 in Eq. (8), and we can derive σ between set 1 and set 2. Since set 1 and set 2 are real datasets from the same database, σ is treated as the baseline σ which serves as the benchmark value for model scoring, denoted as $\sigma_{\text{baseline}}^f$. Then, we sample N points from set 1 and set 2 to generate set 0, and the σ calculated between set 0 and set 3 is denoted as σ^f .

A scoring criterion is proposed to convert σ to model score. The percentage changes from σ^f to $\sigma_{\text{baseline}}^f$ can be derived by

$$\sigma_{\text{inc}} = \text{abs}\left(\sigma^f - \sigma_{\text{baseline}}^f\right) / \sigma_{\text{baseline}}^f (\%) \tag{9}$$

where $\text{abs}(\cdot)$ represents the absolute value. σ_{inc} is treated as the indicator of the loss of precision in the process of modeling. And the scoring criterion is expressed as

$$\text{score} = \begin{cases} 100(1 - \sigma_{\text{inc}}), & 0 < \sigma_{\text{inc}} < 1 \\ 0, & \sigma_{\text{inc}} \geq 1 \end{cases} \tag{10}$$

Based on user requirements and model scores, we can decide whether to select a particular model. For example, if the user's requirement is to select models with a score above

80, then models with scores below 80 are considered to have insufficient accuracy and not recommended to use.

Experiments and analysis

To validate the performance of the proposed evaluation method, three experiments are conducted and analyzed subsequently. Real-world samples of each fault type are collected and divided into training set and testing set in a 4:1 ratio. The evaluation process can be divided into four steps: 1) datasets construction, 2) fault modeling, 3) fault simulation, and 4) model evaluation. The first step is to extract required data from raw data and establish training and testing datasets. The second step is to exploit training datasets to generate fault models. The third step is to simulate faults with the models built in step two. The final step is to evaluate the performance of fault models within the proposed evaluation method on the testing set. A flowchart is demonstrated in Fig. 1 for better understanding.

Validation of the proposed method

In high-dimensional space, since analytical solutions are available only for second-order Gaussian Wasserstein distance, we have to calculate Wasserstein distance by numerical approximation techniques, and tolerable inaccuracy in final model score should be allowed. A numerical example is presented below to validate the proposed method.

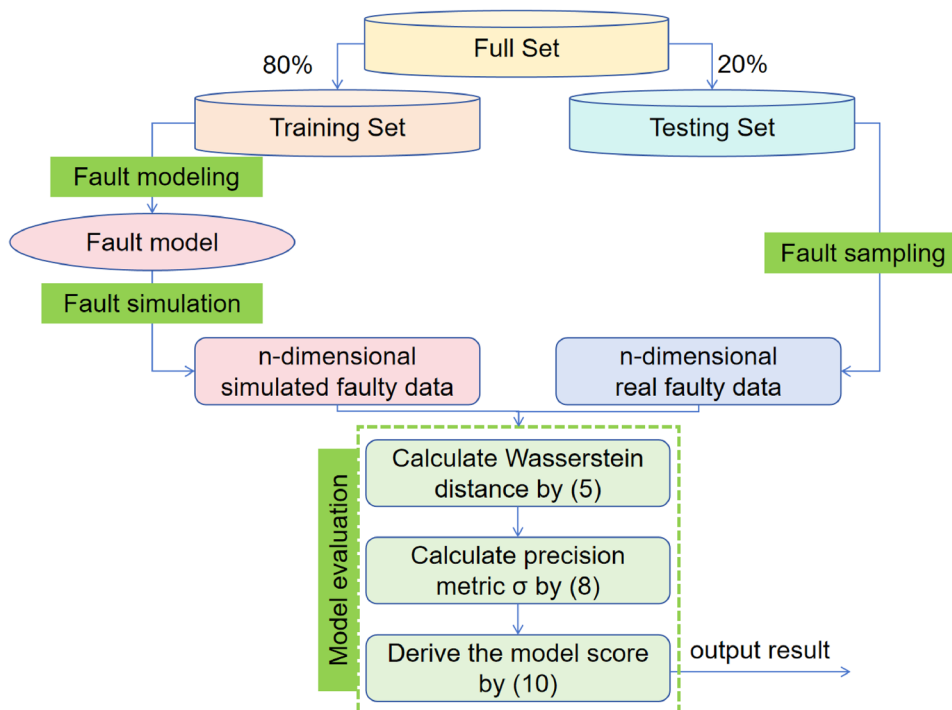
Given that second-order n-dimensional Gaussian Wasserstein distance has a closed-form formula, we assume that a Gaussian model is the fault model to be evaluated. Suppose that we have a fault model which follows a four-dimensional Gaussian distribution where each dimension has a mean of 0 and a standard deviation of 1, and apparently theoretical precision metric σ will be 1. Since we are sampling from a known distribution, the ideal model score should be 100.

Then, we simulated data following the distribution with sizes of 100, 1000, 10,000, 50,000, and 100,000. Wasserstein distances are calculated through numerical approximation method (with a convergence of 10–12) between these datasets and four-dimensional standard Gaussian distribution. Table 1 shows that the calculated values of σ are approximating to the theoretical value ‘1’ and the model scores progressively become more accurate as the sample size increases. And the relationship between sample size and model score is plotted in Fig. 2. We can conclude that when the sample size is relatively large (for example, $\geq 10,000$ in

Table 1 The model scores corresponding to different sample sizes

Sample size	Numeric value of precision metric	Model score
100	1.2335	76.65
1000	1.1128	88.72
10,000	1.0068	99.32
50,000	1.0044	99.56
100,000	0.9995	99.95

Fig. 1 Flowchart of evaluation process



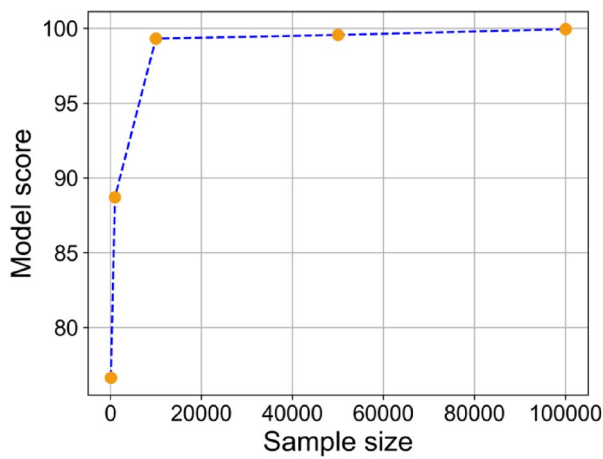


Fig. 2 Relationship between sample size and model score

this case), the precision loss is tolerable and can be ignored. Besides, even for the precise model, undersampling can introduce large inaccuracy in model scoring, which should be avoided in practical use.

Fault model evaluation using the proposed method

Skalar’s theorem (Skalar 1959) states that a n-dimensional multivariate probability distribution can be decomposed into two parts: n univariate marginal distributions and a copula (Oh 2014) which is solely a function of the dependence structure to capture the inter-correlations between the n dimensions. To validate the method’s ability for complex model evaluation, two copula-based multivariate models with correlated variables are built and scored in this part.

GNSS correction product fault model evaluation

GNSS correction product fault is one of fault sources in GNSS positioning. In this part, a Slant Troposphere Delay (STD) correction fault model is established and evaluated. In general circumstances, the precision of STD correction

products degrades under extreme troposphere conditions such as typhoon and heavy rain.

The real-world STD correction faults are generated from the grid STD correction products broadcast by Qianxun Spatial Intelligence (QXSI) company. With the hydrostatic components corrected by Saastamoinen model (Saastamoinen 1972), the generation is based on the difference between $STD_{wet,real}$ (true values of wet components of STD) obtained by monitoring stations and $STD_{wet,est}$ (estimated values of wet components of STD) obtained with troposphere correction, denoted as dSTD, as follows:

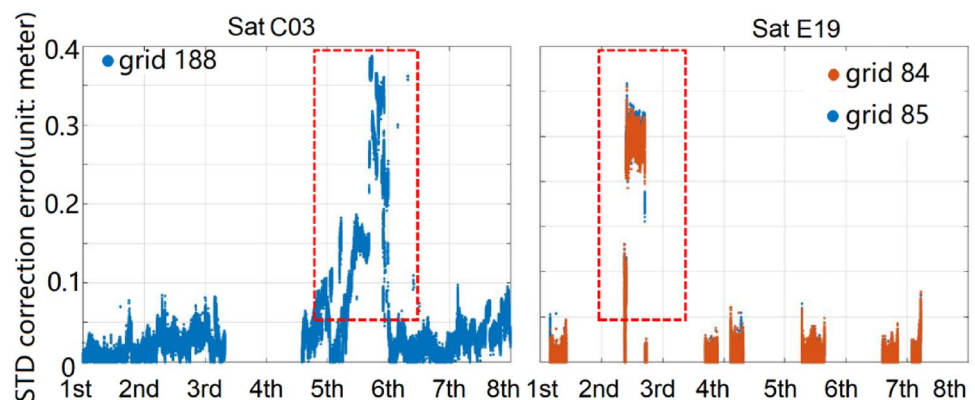
$$dSTD_i^s = STD_{wet,real\ i}^s - STD_{wet,est\ i}^s \tag{11}$$

where $dSTD_i^s$ represents the correction error on grid i of satellite s .

With STD correction data of six months broadcast in Shanghai, China, $dSTD_i^s$ sequences are collected for faulty data extraction. With time series anomaly detection methods applied to detect STD correction faults, faulty $dSTD_i^s$ sequences are generated and added to training and testing datasets. Two examples of real faults detected in the dSTD sequences are demonstrated in Fig. 3, showing that STD correction temporal anomalies undergo a sudden rise–sustained–fall in value.

Based on the observations and analysis of faulty samples, STD correction faults can be abstracted as a trapezoid-like model, as depicted in Fig. 4. This ‘trapezoid’ is characterized by four parameters: (a) rising slope (unit: m/s) denoted as k_1 ; (b) descending slope (unit: m/s) denoted as k_2 ; (c) ramp duration (unit: s) denoted as Δt_1 ; and (d) step duration (unit: s) denoted as Δt_2 . Since rainfall events typically include several features with correlation such as the intensity, precipitation, and duration, it is straightforward that the model is multivariate, multidimensional model (Balistrocchi and Bacchi 2011). It is verified by Kendall coefficients that the aforementioned four parameters have rank correlations. A multivariate, four-dimensional trapezoid-like model is established by four marginal functions and one copula function

Fig. 3 Real faults (outlined with red dashed lines) detected in the dSTD sequences on grid 188 of Sat C03 (left) and on grid 84, 85 of Sat E19 (right) during July 1–8, 2023



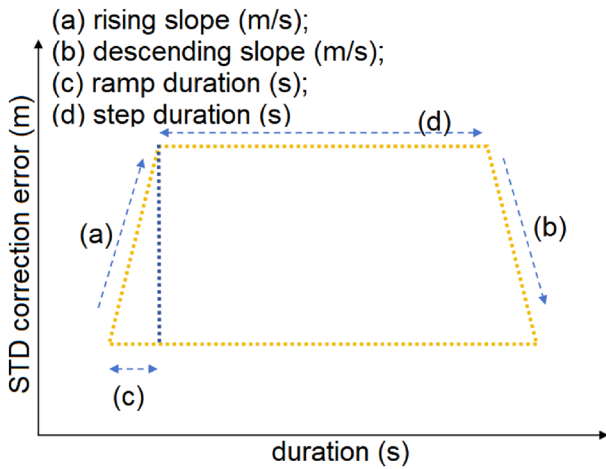


Fig. 4 Abstract trapezoidal model for STD correction faults

for STD correction faults characterization. The joint distribution $F(k_1, k_2, \Delta t_1, \Delta t_2)$ is expressed as

$$F(k_1, k_2, \Delta t_1, \Delta t_2) = C(F_1(k_1), F_2(k_2), F_3(\Delta t_1), F_4(\Delta t_2)) \tag{12}$$

where $C(\cdot)$ is the copula function, and $F_*(\cdot)$ denotes each marginal function. The distribution fittings of four marginal functions are shown in Fig. 5.

In the previous step, STD correction fault model is built on the training set. And the testing set is equally divided in two, named set 1 and set 2, and the count of the sample

points in each set is denoted as N . Then, we generate N points with four dimensions from the established STD correction fault model to build set 3. Thereby, we have three datasets with equal count of samples for model evaluation.

We obtained three sets of n -dimensional STD correction fault samples in fault simulation: Set 1 and set 2 are the real data testing sets, and set 3 is the simulated dataset. The sample size of each set is 10,000. Then, we use set 1 and set 2 to calculate baseline result and testing set and set 3 to calculate model result. The result is summarized in Table 2. Values of raw Wasserstein distance are derived by numerical approximation techniques, and numeric values of precision metric are calculated by (8). The model score of baseline result is set as 100. And the model score of the established model, computed by (9) and (10), yields a result of 96.80.

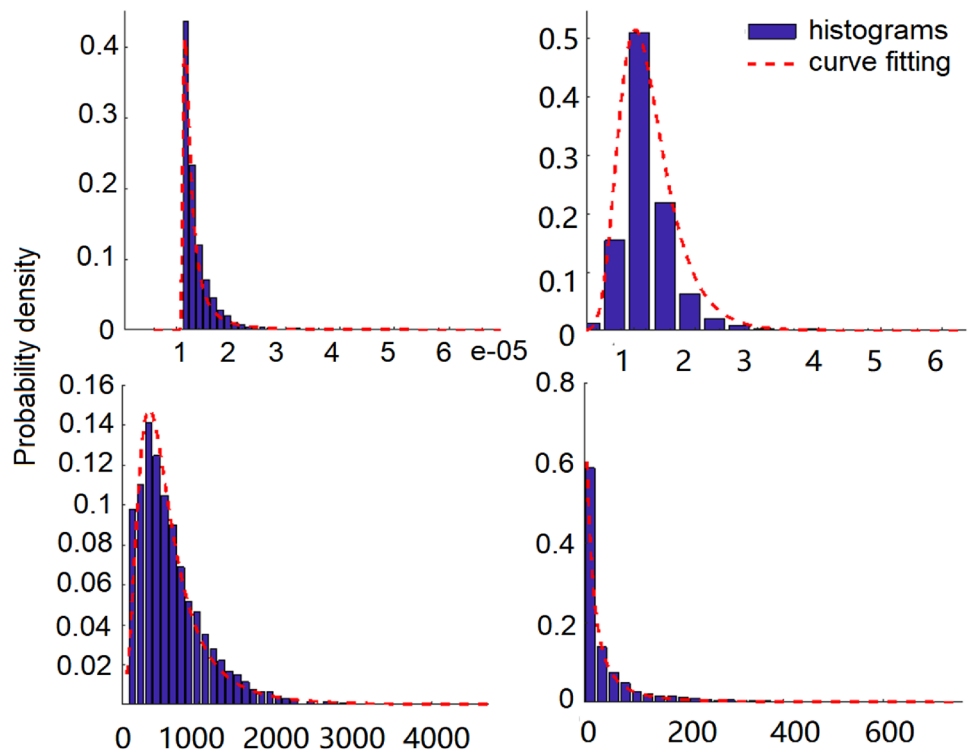
Multipath fault model evaluation

Multipath is a typical unmodeled fault source which dramatically deteriorates GNSS positioning performance. It arises

Table 2 The precision metric and model scores of STD correction fault model

	Numeric value of raw Wasserstein distance	Numeric value of precision metric	Model score
Baseline result	56.674	29.337	100
Model result	58.55	30.275	96.80

Fig. 5 Distribution fittings of the four parameters k_1 (top left), k_2 (top right), Δt_1 (bottom left), and Δt_2 (bottom right) of STD correction fault model



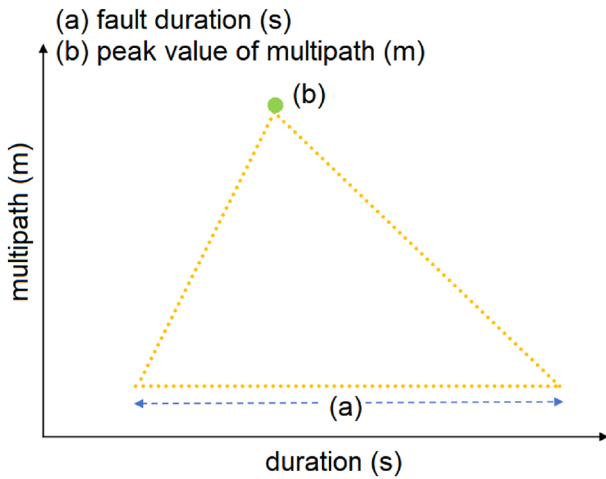


Fig. 6 Abstract triangle model for code multipath faults on frequency i

when the satellite signals received by GNSS receivers travel not only directly but also indirectly via reflections of nearby surfaces such as skyscrapers and terrain. A double-frequency code multipath fault model is constructed and estimated in this section.

Real-world multipath faults are generated from drive test data collected by QXSI in March 2022. And undifferenced and uncombined precise point positioning model is used to extract code multipath, which can be expressed as

$$v_{p_i} = MP_{p_i} + \epsilon_{p_i} = P_{i,r}^s - \rho_r^s - c(dt_r - dt^s) - T_r^s - I_{i,r}^s \tag{13}$$

where s, r, i denote the satellite, receiver, and frequency index, respectively; c represents the velocity of light in vacuum; $P_{i,r}^s$ is the pseudorange measurement from satellite s to receiver r on frequency i ; ρ_r^s is the receiver-to-satellite geometric distance; dt_r and dt^s denote the clock error of receiver r and satellite s , respectively; T_r^s is the tropospheric delay, and $I_{i,r}^s$ is the ionospheric delay.

Since the receivers' accurate coordinates can be derived by map-matching technique, satellite-end and atmosphere-end errors can be corrected by QXSI correction products, and code multipath datasets are thereby established based on drive test datasets.

Temporal code multipath faults on frequency i can be characterized as a triangle-like model, abstracted in Fig. 6, where the 'base' represents fault duration, and 'height' represents the peak value of faults. Besides, as is shown in Fig. 7, SNR (signal–noise ratio) measured by receivers usually descends when the magnitude of code multipath faults grows.

A double-frequency code multipath fault is described by four parameters: (a) fault duration (unit: s) denoted as Δt ; (b) the peak value of multipath on the first frequency denoted

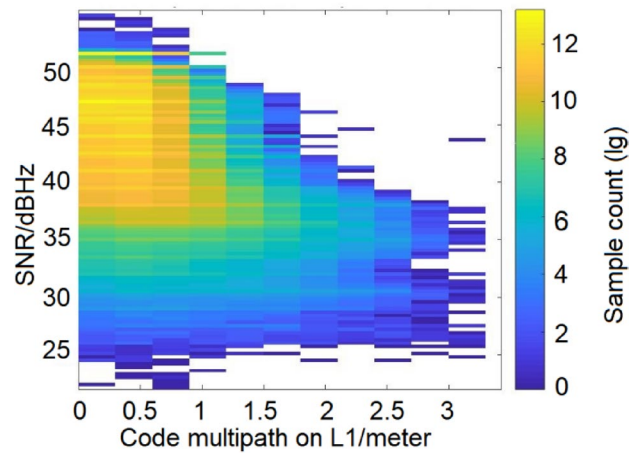


Fig. 7 Relationship between code multipath on frequency L1 and SNR value

as $MPL1$; (c) the peak value of multipath on the second frequency denoted as $MPL2$; and (d) delta SNR value denoted as ΔSNR . It is verified by Kendall coefficients that the four parameters are correlated. Hence, A four-dimensional double-frequency code multipath model is established by four marginal functions (the distribution fittings are shown in Fig. 8) and one copula function, and the joint distribution $F(\Delta t, MP_{L1}, MP_{L2}, \Delta SNR)$ is written as

$$F(\Delta t, MP_{L1}, MP_{L2}, \Delta SNR) = C(M_1(\Delta t), M_2(MP_{L1}), M_3(MP_{L2}), M_4(\Delta SNR)) \tag{14}$$

where $C(\cdot)$ is the copula function, and $M_k(\cdot)$ denotes each marginal function.

The procedure of fault simulation is the same as the simulation of STD product fault model. We obtained three sets of n-dimensional double-frequency code multipath fault samples in fault simulation: Set 1 and set 2 are the real data testing sets, and set 3 is the simulated dataset. The sample size of each set is 10,000. As is shown in Table 3, the model score of baseline result is set as 100. And the model score of the established model, computed by Eq. (9) and Eq. (10), yields a result of 97.15.

Conclusions

GNSS errors often exhibit heavy tails, non-Gaussian behavior, and are nearly impossible to completely eliminate. Gaussian assumption, although widely applied in the field of GNSS, fails to capture the heaviness of tails which are typically attributed to the projection of measurement faults to position domain. Error tails and faults should be more precisely modeled to deliver improved navigation performance. Researchers not only need precise models but also want to quantify how 'precise' the model is. And it

Fig. 8 Distribution fittings of the four parameters Δt (top left), MP_{L1} (top right), MP_{L2} (bottom left), and ΔSNR (bottom right) of the double-frequency code multipath fault model

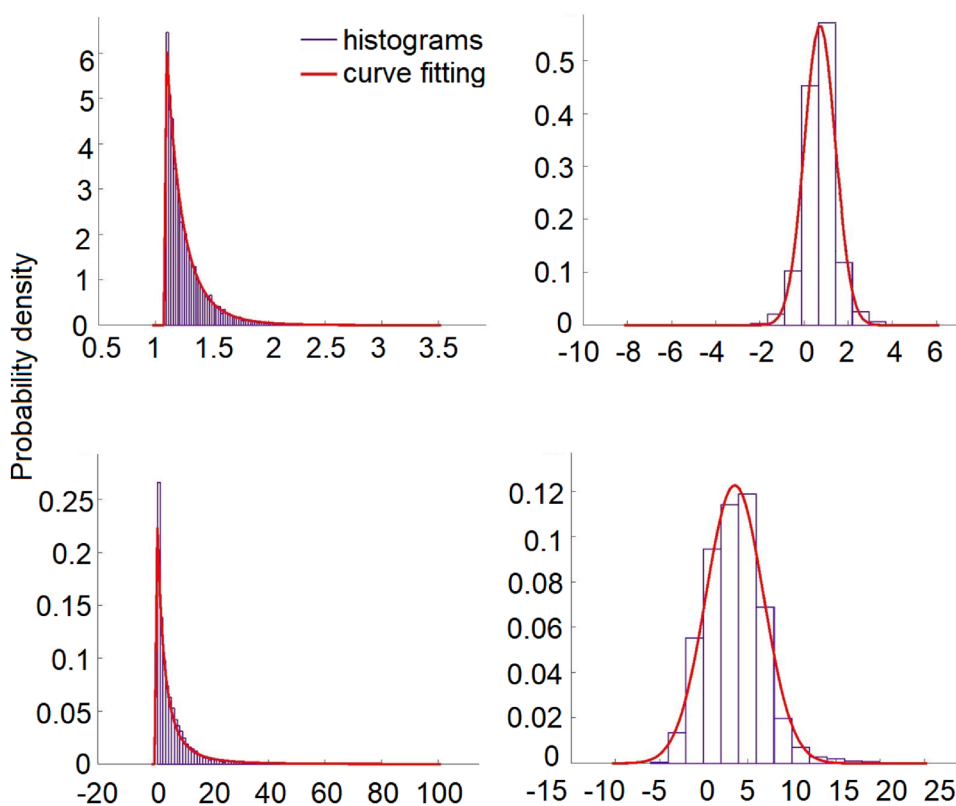


Table 3 The precision metric and model scores of double-frequency code multipath fault model

	Numeric value of raw Wasserstein distance	Numeric value of precision metric	Model score
Baseline result	27.716	14.858	100
Model result	28.564	15.282	97.15

is straightforward that universal goodness-of-fit evaluation criteria are required.

This study has focused on the goodness-of-fit evaluation of GNSS fault characteristics and models, by exploiting Wasserstein distance, a machine learning method to measure the similarities between the two distributions. Considering that numeric values of raw Wasserstein distance cannot reflect model precision intuitively (models with different dimensions may have the same value of Wasserstein distance but different precision), a second-order Gaussian Wasserstein distance-based precision metric σ is presented to convert the cost between n-dimensional real distribution and fault model to the cost between two n-dimensional Gaussian distributions. A scoring criterion is thereby established based on σ to provide reference opinions to users. The contributions of the proposed method can be summarized as follows:

- (1) The method inherits the good properties of Wasserstein distance, without imposing any constraints on the models' dimensionality or specific distribution, so we can evaluate fault models as long as we can sample from them.
- (2) The precision metric proposed uncoupled dimensionality with modeling precision, thus unifying the evaluation criteria among different models with different dimensions.
- (3) The scoring criterion further provides a more humane way for GNSS practitioners to select fault models according to their individual needs and contributes to the validation of the ubiquitous non-Gaussian problems.

From the experimental results, the proposed method is first validated through known n-dimensional Gaussian distributions, indicating that adequate sampling should be conducted to maintain the model scoring inaccuracy resulting from undersampling within an acceptable range. A STD product fault model and a double-frequency code multipath model are given as examples to apply the proposed method to fault model evaluation. Detailed implementation processes have been presented to show that this method is able to evaluate and score GNSS fault models. In future work, fault models can be scored, compared, and selected more conveniently within the method introduced in this study.

The computed model score can be used as the indicator of model iteration. Advanced fault models can provide sufficient faulty simulated measurements for large-scale system robustness testing and positioning algorithm refinement, as well as significantly reducing the cost of data acquisition. Furthermore, the proposed technique is expected to have broader applications such as assisting in the validation of non-Gaussian phenomena and identification of Gaussian alternatives, incorporating in novel integrity algorithms and helping derive tighter protection levels to improve availability.

Author contributions JC proposed the general idea of this contribution and completed the evaluation and was a major contributor in writing the manuscript. ML is the supervisor who modified this paper. WZ, BF, NZ, and YZ made suggestions for the framework and participated in the modification of this paper. All the authors read and approved the final manuscript.

Funding The research is sponsored by the Department of Science and Technology of Zhejiang Province (2020R01012) and the Ministry of Science and Technology China (2021YFA0717300).

Data availability The GNSS data were collected by real tests.

Declarations

Conflict of interest The authors declare that they have no competing interests.

References

- Anderson TW, Darling DA (1954) A test of goodness of fit. *J Am Stat Assoc* 49(268):765–769. <https://doi.org/10.1080/01621459.1954.10501232>
- Anderson EW, Ellis DM (1971) Error Distributions in Navigation. *J Navig* 24(4):429–442. <https://doi.org/10.1017/S0373463300022281>
- Balistrocchi M, Bacchi B (2011) Modelling the statistical dependence of rainfall event variables through copula functions. *Hydrol Earth Syst Sci* 15(6):1959–1977. <https://doi.org/10.5194/hess-15-1959-2011>
- Bhatti UI, Ochieng WY, Feng S (2007) Integrity of an integrated GPS/INS system in the presence of slowly growing errors. Part II: *Anal GPS Solut* 11(3):183–192. <https://doi.org/10.1007/s10291-006-0049-1>
- Cramér H (1928) On the composition of elementary errors. *Scand Actuar J* 1928(1):13–74. <https://doi.org/10.1080/03461238.1928.10416862>
- Du Y, Wang J, Rizos C, El-Mowafy A (2021) Vulnerabilities and integrity of precise point positioning for intelligent transport systems: overview and analysis. *Satell Navig* 2(1):3. <https://doi.org/10.1186/s43020-020-00034-8>
- Dwass M, Kullback S (1960) Information theory and statistics. *Am Math Mon* 67(3):310. <https://doi.org/10.2307/2309734>
- Hsu DA (1979) Long-tailed distributions for position errors in navigation. *Appl Stat* 28(1):62. <https://doi.org/10.2307/2346812>
- Jensen JLWV (1906) Sur les fonctions convexes et les inégalités entre les valeurs moyennes. *Acta Math* 30(1):175–193. <https://doi.org/10.1007/BF02418571>
- Joanes DN, Gill CA (1998) Comparing measures of sample skewness and kurtosis. *J R Stat Soc: Ser D (stat)* 47(1):183–189. <https://doi.org/10.1111/1467-9884.00122>
- Kullback S, Leibler RA (1951) On information and sufficiency. *Ann Math Stat* 22(1):79–86. <https://doi.org/10.1214/aoms/1177729694>
- Lilliefors HW (1967) On the Kolmogorov–Smirnov test for normality with mean and variance unknown. *J Am Stat Assoc* 62(318):399–402. <https://doi.org/10.1080/01621459.1967.10482916>
- Milner C, Macabiau C, Thevenon P (2016) Bayesian inference of GNSS failures. *J Navig* 69(2):277–294. <https://doi.org/10.1017/S0373463315000697>
- Oh DH (2014) Copulas for high dimensions: models, estimation, inference, and applications.
- Pan Y, Möller G, Soja B (2023) Machine learning-based multipath modelling in spatial-domain: a demonstration on GNSS short baseline processing. <https://doi.org/10.21203/rs.3.rs-2555284/v1>
- Panagiotakopoulos D, Majumdar A, Ochieng WY (2014) Extreme value theory-based integrity monitoring of global navigation satellite systems. *GPS Solutions* 18(1):133–145. <https://doi.org/10.1007/s10291-013-0317-9>
- Panagiotakopoulos D, Ochieng W, Niemman P (2008) Robust statistical framework for monitoring the integrity of space-based navigation systems [PhD Thesis]. Imperial College London.
- Panaretos VM, Zemel Y (2019) Statistical aspects of Wasserstein distances. *Annu Rev Stat Appl*. <https://doi.org/10.1146/annurev-statistics-030718-104938>
- Pearson K (1992) On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. In: Kotz S, Johnson NL (eds) *Breakthroughs in statistics: methodology and distribution*. Springer, New York, pp 11–28
- Rubin DB (1984) Bayesianly justifiable and relevant frequency calculations for the applied statistician. *Ann Stat* 12(4):1151–1172. <https://doi.org/10.1214/aos/1176346785>
- Saastamoinen J (1972) Contributions to the theory of atmospheric refraction. *Bull Geod* 46:279–298. <https://doi.org/10.1007/BF02521844>
- Shannon CE (1948) A mathematical theory of communication. *Bell Syst Tech J* 27(3):379–423
- Shively CA, Braff R (2000) An overbound concept for pseudorange error from the LAAS ground facility, pp 661–671. <http://www.ion.org/publications/abstract.cfm?jqp=p&articleID=820>
- Sklar M (1959) Fonctions de repartition a n dimensions et leurs marges. <https://www.semanticscholar.org/paper/Fonctions-de-repartition-a-n-dimensions-et-leurs-Sklar/d017e0a6a9b93dcab7cdcd44b17544c04f7678d2>
- Smirnov N (1948) Table for estimating the goodness of fit of empirical distributions. *Ann Math Stat* 19(2):279–281. <https://doi.org/10.1214/aoms/1177730256>
- Takatsu A (2011) Wasserstein geometry of Gaussian measures. *Osaka J Math* 48(4):1005–1026
- Vaserstein LN (1969) Markov processes over denumerable products of spaces, describing large systems of automata. *Probl Pereg-dachi Inf* 5(3):64–72

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Jinpei Chen is a Ph.D. candidate at the Department of Electronic Engineering, Tsinghua University, Beijing, China. His research interests include GNSS augmentation services, applications of big data, and artificial intelligence in GNSS.

Wenyu Zhang received a master's degree in Transportation Engineering from Nanjing University of Aeronautics and Astronautics, Nanjing, Jiangsu province, China, and is an engineer at QianXun, China. Her research interest include GNSS augmentation services and integrity monitoring.

Bingqing Feng is a senior engineer at the State Grid Beijing Electric Power Company. Her research interests focus on monitoring the performance of the grid and risk control via new technologies including the applications of GNSS, autonomous systems, and Artificial Intelligence.

Nan Zhi received a Ph.D. degree in theoretical computer science from the University of Liverpool, United Kingdom, and is a researcher at QianXun, China. His research activities involve optimization algorithms, machine learning in geodesy, and GNSS applications.

Yi Zhao received the M.S. and Ph.D. degrees in precision instrument and machinery from Southeast University, Nanjing, China in 2005 and 2009. His research interests include multi-GNSS integration algorithms and GNSS augmentation services.

Mingquan Lu is a professor and director of the PNT Research Center at Tsinghua University. He is a Fellow of the USA ION and has been active in research and education of BDS/GNSS for two decades. His most significant contributions are BDS-3 signal design and BDS-3/GNSS interoperable receiver development. The signals he designed are being transmitted by all BDS-3 satellites today. Dr. Lu received the prestigious Thurlow Award in 2020.