

Research Article

Mitigating GNSS Multipath Effects Using XGBoost Integrated Classifier Based on Consistency Checks

Dengao Li ^{1,2,3} Xiaoli Ma,^{1,2} Jumin Zhao ^{2,3} and Fanming Wu^{1,2}

¹College of Data Science, Taiyuan University of Technology, Taiyuan 030024, China

²Key Laboratory of Big Data Fusion Analysis and Application of Shanxi Province, Taiyuan 030024, China

³Intelligent Perception Engineering Technology Center of Shanxi, Taiyuan 030024, China

Correspondence should be addressed to Dengao Li; lidengao@tyut.edu.cn

Received 9 February 2022; Revised 7 June 2022; Accepted 17 June 2022; Published 1 August 2022

Academic Editor: Shobhit K. Patel

Copyright © 2022 Dengao Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Under the influence of urban building roads, especially interference from multipath effects, global navigation satellite system (GNSS) receiver-related output signal distortion can affect the robustness of the positioning system and the final positioning accuracy. To deal with the above problems, this paper proposes a two-layer consistency-checks (CC) positioning model based on eXtreme Gradient Boosting (XGBoost) integrated learner. First, the model excludes the abnormal values from the correlated output of the first layer by the classical statistical distribution test method. Then, the remaining available measurements are used as the second-layer input, and the measurements are used as learning data using an integrated machine learning method, XGBoost, to efficiently detect and identify non-line-of-sight (NLOS), LOS, and other reflective multipath signals. In order to better mitigate errors in the dynamic relative positioning process, the second-layer checking process uses dynamic pseudorange differencing technique (DPDT) and weighted least squares method (WLS) to smooth the output outcome of the receiver. In the experimental part, we compare and analyze the proposed method with the existing methods from different perspectives in this paper, respectively. The results show that the performance of the model is significantly improved after applying the CC method, in which the average classification accuracy of the multipath signals in the target feature set can reach 91.6%. According to the final positioning results, the proposed method shows a significant accuracy improvement compared to the existing research methods.

1. Introduction

With the implementation of global navigation satellite system (GNSS) overall networking and the further development of BeiDou navigation system, the accuracy requirements of location-based services are getting higher and higher. Despite the more comprehensive range and number of positioning satellites, the ionospheric and tropospheric delays in the atmosphere produce serious errors, which in turn affect the reception of satellite signals [1].

For urban environments, the simultaneous reflection of direct signals from satellites by the surfaces of buildings and obstacles can seriously disrupt the available signals in the target area. Among them, the reflected and non-line-of-sight (NLOS) are generated because of reflected and blocked

signal during transmission process. The reception delay caused by the reflected and diffracted signals can cause different degrees of distress for the position accuracy of the receiver, which are known as multipath propagation phenomena [2]. The presence of multipath phenomenon does not obtain more reliable and real-time information in studies, such as urban traffic path planning and pedestrian obstacle detection. Therefore, highly accurate discrimination of direct, reflected multipath signals or NLOS signals and effective exclusion are key aspects of the study to mitigate multipath effects on GNSS positioning errors.

So far, the four most classical types of NLOS and multipath detection techniques have been applied, including antenna-based, receiver-internal improvement techniques, navigation processor-internal, and machine learning-based

[3]. As antenna array technology, the current dual-polarized antenna implementation [4] is more mature, but it is difficult to implement considering the experimental cost and equipment limitations in universities [5]. Among the research results based on receiver signal processing techniques, solving the code loop distortion and separating the direct from reflected signals in two perspectives can only reduce the MP effect and improve the success rate of NLOS detection [6]. Therefore, [7, 8] proposed a method for NLOS discrimination using GNSS receiver output, but the method is difficult to guarantee the accuracy of the discrimination with the help of signal-to-noise ratio (SNR) only as a discrimination criterion. In order to increase the multipath detection probability, the use of triple frequency receivers was investigated in [9]. The difference in SNR across three frequencies was modeled as a threshold for determining multipath and assessing the robustness of the system. However, the validation performance of the method makes no sense when open areas with high SNR are encountered. Secondly, among the research results on the internal improvement of navigation processors, the application of CC is a breakthrough in the last years [10, 11]. They use empirical thresholds for test statistics of pseudorange residuals to determine the possible presence of NLOS or multipath signals from the perspective of measurements. The receiver autonomous integrity monitoring method (RAIM) for intermediate applications is more mature and is based on exhaustive and greedy detection and exclusion of “faulty” signals after the fault detection and exclusion (FDE) framework has been built [10]. In deep urban environments, exhaustive FDE is able to reduce the localization error by 8%. However, a single-layer consistency check is only able to find a consistent set of measurements, making the overall positioning solution error large [3]. Further, many papers have proposed the use of practical instruments, such as visible light, fisheye, and infrared laser scanning [12–15], to detect NLOS and determine the approximate satellite position with the help of a combination of sensors and navigation processors. However, these special sensors are easily affected by light and weather, and the limited measurement range of the scanners cannot be fully applied to special buildings in urban living areas. Therefore, methods to detect and correct NLOS errors in cities with the help of 3D map matching are proposed [16–19]. The ability of 3D models to predict satellite visibility was investigated in [20]. In most implementations, three-dimensional (3D) mapping was used to improve positioning accuracy through shadow matching (SDM), terrain height assistance, or LOS detection [21]. Although these methods are able to target multipath phenomena near intersections and streets in static environments, the acquisition of map-aided information is impractical and too costly. Considering the disclosure of limited map resources and the location changes in dynamic scenes, this will make methods ineffective for detecting and mitigating building occlusion applications.

As a result of the above analysis and comparison, it is not necessary to obtain additional reference information and the choice of an improved angle based on the internal receiver is more appropriate. Considered from that

perspective, machine learning techniques in the last two years have made great progress in the field of satellite navigation and positioning research [2, 22–27]. The first simple classifiers that applied machine learning techniques to achieve binary classification were decision trees (DT) and support vector machines (SVM). Reference [22] used DT to classify the received L1 signals into two types of LOS and MP signals, with a prediction accuracy of about 98% for the classifier. References [25, 26] applied SVM to classify the receiver-related signal output for the target. But only with the help of signal strength as feature prediction evaluation, it leads to overfitting and feature bias in the training process. To improve the accuracy of feature extraction and training, [2, 26] simulated an indoor virtual multipath environment and migrated a deep learning network to classify the correlated signal output with dimensional expansion processing and then compared the classification results with those of SVM, which was able to achieve an average classification accuracy of 94%. However, this environment is limited to indoor simulated occlusion environments and is not practical for application in outdoor environments.

In dense urban living areas with dynamic pedestrians or vehicles moving, the additional delay distance due to multipath and NLOS effects is an important cause of physical degradation of the on-board receiver [28]; it also generates pseudorange bias and carrier phase bias that are more difficult to compensate. For this reason, some studies have described the detection identification and compensation techniques for NLOS signals [29, 30]. However, in contrast to the previous static location scenarios, for pedestrians or vehicles, the relative differential dynamic positioning technology [31] can be used to effectively detect and eliminate NLOS and MP effects. Therefore, the CC method was migrated to this field. It is a method to determine whether the values measured are fitted consistently with empirical limits set in advance and then select the useful data retained. Among them, the GNSS pseudorange measurement CC method proposed by [32] tends to be mature, which helps receivers to detect and exclude faulty measurements autonomously by calculating the pseudorange residuals combined with the threshold judgment method in statistics, and finally excludes the detected NLOS and other kinds of multipath and then estimates the location of the target receiver.

In summary, the current application of machine learning algorithms for classifying correlated output signals is the most feasible and has room for improvement enhancements. First, previous studies in the literature have only been applied to a single classifier, the dimensionality of the feature vector selected before classification training is very limited, and the study of statistical distributions such as angular features and pseudodistance similarity features associated with the signal is not addressed. Therefore, we propose to apply encapsulated multiple classification models for processing and training of target features. Second, most of the receiver internal validation of multipath suppression methods are fixed in static scenes, while for our slow-moving pedestrians and vehicles, the received satellite signal dataset

will contain more outlier data present. Therefore, it requires us to choose an integrated classifier that can effectively handle the special sample data for optimisation. Third, in the application area of the consistency checking method mentioned in the previous section, [33] was used for the detection and exclusion of NLOS and multipath signals by this method in collaboration with other error mitigation methods, but it should be noted that their validation experimental data were only screened and judged by statistical methods without preprocessing and feature conversion, making the available measurement set data redundant and interfering, which in turn affects the final deviation of the positioning accuracy. Therefore, it is necessary to apply the principle of consistency checking method and combine it with a better integrated classification method to jointly detect and exclude the inconsistent NLOS signals with the internal output signal measurement set of the receiver in order to improve the robustness and generalization ability of the machine learning algorithm on the cooperative detection model.

To address the problems analyzed above, we apply a CC statistical method to improve the multiclassification learner by adding NLOS and multipath classification categories, so that the reflected signals and so on in the target output signal can be effectively filtered out. According to the previous training results, the proposed method can improve the degree of generalization of the data training process and prevent the overfitting phenomenon of parameters and narrow the range of accuracy deviation of the whole model.

The innovative points of this paper are as follows.

- (1) We build the process of software receiver loop output and decomposition into an overall consistency checks framework, combined with the DPDT, to collaboratively assist positioning.
- (2) In order to solve the problems of overfitting, the existence of missing sample values, and only serial processing in the classification process of single classifier methods in the field of machine learning, this paper preprocesses the output set of GNSS-related features and effective features transformation and applies the mature integrated classification method of the field of supervised machine learning to classify out NLOS and multipath signals.
- (3) Different from the previous acquisition of fixed satellite signals for different static scenes, this paper uses the proposed overall consistency checks method to a slow dynamic real environment.

The remainder of this paper is organized as follows: First, in Section 2, we introduce the traditional GNSS positioning principles and the more mature positioning techniques. Section 3 presents the specific architecture and implementation process of the two-layer consistency-checks positioning model based on the XGBoost classification method. For an implementation of the algorithmic framework of Section 3 in a practical context, see Section 4. Finally, we present the discussion and analysis in Section 5.

2. GNSS Positioning Methods

Considering the research background of this paper, the current bottleneck is the negative interference of multipath and NLOS in the deep urban environment in real reception environments, such as tall buildings and vegetation occlusion, as shown in Figure 1. Advancing slow-moving vehicles will be disturbed by a variety of signals, such as NLOS, reflection, and diffraction in LOS. So, it can bring serious distortion of the internal loop and signal delay and then affect the accuracy of the output position.

Before solving the above problems, the traditional GNSS receiver positioning process is solved according to 3D coordinates in the static position. And we use relative parameters to evaluate the impact of position accuracy in the presence of ideal noise based on the principle of least squares (LS) method. The pseudorange measurement equation between the satellite and the receiver is as follows:

$$\hat{\rho}^{(i)} = [\hat{\rho}_1^{(i)} \hat{\rho}_2^{(i)} \dots \hat{\rho}_n^{(i)}] = R^{(i)} + c(\delta t^r(i) - \delta t^{sv}) + I + T + \varepsilon^{\text{ref}(i)}, \quad (1)$$

where n is the number of observation satellites, i indicates satellite index, c is the speed of light, δt^{sv} denotes the satellite clock offset, δt^r is the offset time of the receiver, I is the ionospheric delay distance, and T is the tropospheric delay distance. Besides, the most critical $\varepsilon^{\text{ref}(i)}$ contains multipath error and noise error. In this paper, noise error is considered as zero-mean error. What is more, $R^{(i)}$ denotes the geometric distance between the i th visible satellite and the ground receiver. It can be calculated as follows:

$$R^{(i)} = \sqrt{(x_k - x)^2 + (y_k - y)^2 + (z_k - z)^2}, \quad (2)$$

where (x, y, z) is the position of the receiver. The satellite position (x_k, y_k, z_k) is tackled with the help of ephemeris and satellite clock offsets. We solve the unknowns in (1) to minimize by the LS method conventionally. The estimation is iterated step by step starting from the initial position solution. To better solve (2), it is transformed into a linearized form containing $(x + \Delta x, y + \Delta y, z + \Delta z)$ based on the first-order Taylor expansion. And we can solve the pseudorange observation equation to obtain the final positioning solution [33].

2.1. Weighted Least Squares Algorithm. In the process of positioning, based on the satellite signal propagation time, it is necessary to select at least four satellites to obtain the target distance. The positioning process requires the selection of at least four satellites to obtain the destination distance based on the satellite signal propagation time. The satellite coordinates are known and combined with Newtonian iteration for the prediction of the reference point and smoothing of the residual. In the iterative process, the solution vector of the initial estimated position is (x, y, z) , and the nonlinear objective equation f is set to

$$f(x, y, z) = 0. \quad (3)$$

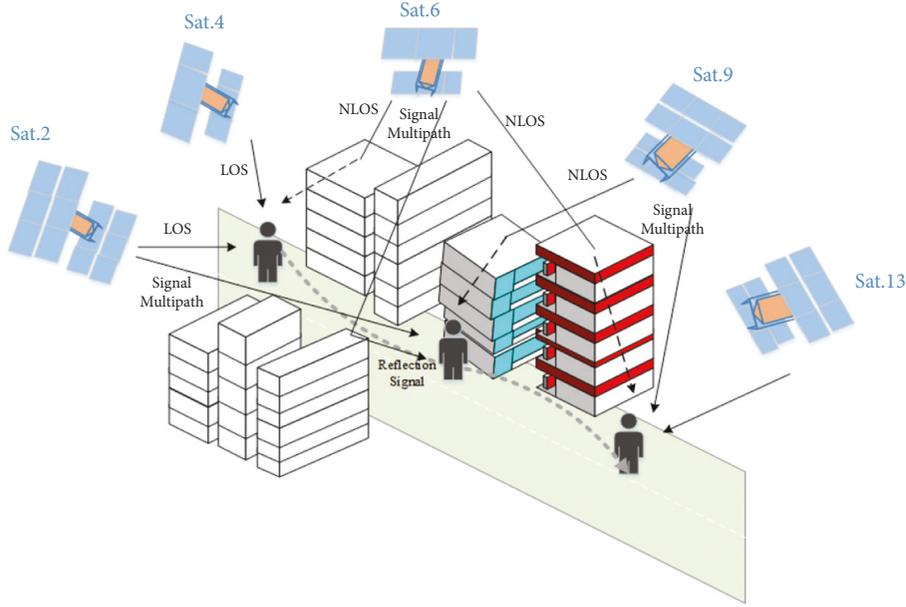


FIGURE 1: The multipath phenomena in the urban environment.

when the m th iteration of the initial position is carried out, the linearized Taylor expansion performed at this point is

$$f(x, y, z) \approx f(x_{m-1}, y_{m-1}, z_{m-1}) + \frac{\partial f(x_{m-1}, y_{m-1}, z_{m-1})}{\partial x} (x - x_{m-1}). \quad (4)$$

After the linear processing of (4), the linear equation solution for the next update moment m is obtained as follows:

$$(x_m, y_m, z_m) = (x_{m-1}, y_{m-1}, z_{m-1}) - \frac{f(x_{m-1}, y_{m-1}, z_{m-1})}{f'(x_{m-1}, y_{m-1}, z_{m-1})},$$

$$\Delta v = (x_m, y_m, z_m) - (x_{m-1}, y_{m-1}, z_{m-1}). \quad (5)$$

And so on, the iterative process is repeated for similar m moments and $(m-1)$ moments until the accuracy of the current moment reaches the criterion and then stop the update iteration. For the linearized solution of (3), the LS solution Δv_{LS} is solved with the help of the Jacobi matrix A , and the procedure is shown as in follows.

$$A \cdot \Delta v_{LS} = b, \quad (6)$$

$$\Delta v_{LS} = (A^T \cdot A)^{-1} A^T \cdot b. \quad (7)$$

In the above formula, A is the geometric unit vector matrix between the user observation satellite and the other satellites. Δv_{LS} is the state vector representation of the estimated user location. A^T represents the transpose of the matrix A , and b represents the pseudorange difference between the measured and predicted values.

Considering that the output estimated value of the original pseudorange has different error iterative influence,

each output measurement value needs to be assigned a pseudoweight to control the iterative error so that the negative effect of the low-elevation angle measurement environment between buildings can be reduced.

$$Y = \begin{bmatrix} w_1 & & & \\ & w_2 & & \\ & & \dots & \\ & & & w_n \end{bmatrix}. \quad (8)$$

Therefore, after setting the diagonal weight matrix Y in (8), multiply (7) by it. And we can get the error correlation between n different measurement values, and then effectively reduce the dominant measurement error in weak signal environment. The whole process is called WLS, and the principle formula is as follows:

$$\Delta v_{WLS} = (A^T Y A)^{-1} A^T Y b. \quad (9)$$

2.2. Double-Differenced Relative Kinematic Positioning Principle. Based on the dynamic urban mobile environment, the relative change in position between the mobile receiver and the base station generates real-time differential dynamic positioning. The phase correction values between different epochs are sent in real time, and another dynamic receiver also does the same deviation measurement and output processing to the observation satellite. The difference between the two stations is finally used to send the correction values to the user receiver in the dynamic environment in time.

The reason why the differential positioning technology weakens or eliminates the noise signal is that the traditional static single-point positioning method has been unable to

solve the signal redundancy and interference in the complex environment. So, the existing ionospheric delay, clock delay, and clock error can be eliminated as much as possible under the condition of multireference base station. Equation (1) is the basic pseudorange observation equation, and the parameter $\varepsilon^{ref(i)}$ is considered as the multipath error, which is also the target to be studied in this paper. Figure 2 illustrates the process of pseudorange differential positioning.

Suppose that GNSS receiver n in the figure is the dynamic mobile receiver used and m is a fixed receiving base station of the nearby rooftop. According to the signal frequency and time delay of satellites p and q , D_{nm}^{pq} denotes the observed value of the distance difference. The calculation principle is as follows:

$$D_{nm}^{pq} = (\vec{e}^p - \vec{e}^q) \cdot \Delta \vec{r}_{nm} + [(\varepsilon_n^q - \varepsilon_m^q) - (\varepsilon_n^p - \varepsilon_m^p)], \quad (10)$$

where \vec{e}^p and \vec{e}^q represent the visible satellite p and q unit vector distances, respectively, and the relative matrix vector between the reference base station m and the receiver n is represented by $\Delta \vec{r}_{nm}$. Because the relative distance between the two receivers is much smaller than the actual measured pseudoranges, the measured pseudoranges between n and m and satellites can be considered as the same. The common difference between satellite p and the base station and receiver can be eliminated by the difference calculation (10). Among them, the common ephemeris error and clock deviation of the DD calculation process are eliminated, leaving the more difficult error to deal with. The position vector $\Delta \vec{r}_{nm}$ is obtained using (7) to help solve.

$$\Delta \vec{r}_{nm} = (A^T \cdot A)^{-1} A^T \cdot D_{nm}. \quad (11)$$

From the experimental environment studied in this paper, the MP effect in the real environment will greatly affect the pseudorange differential relative positioning technology, which needs to be effectively detected and eliminated before this and then combined with DPDT for receiver position estimation finally.

Therefore, we introduce the method of XGBoost integrated classifier to classify satellite signals to detect available signals and eliminate interference signals in Section 3 and how to apply it in a CC model to achieve the final pseudorange positioning.

3. Proposed Method

In the real scene of Figure 1, the multipath and NLOS have the most significant influence on the positioning results of the receiver. And the absolute positioning solution can no longer be obtained only by pseudorange single-point difference and basic LS method. Based on this [6], then WLS method is applied to a real environment containing MP and NLOS signals, and healthy satellite signals are selected for iterative weighting to improve the accuracy of prediction. In addition, 3D shadow matching (SDM) method used a matching function to determine the reference threshold and narrow the candidate range of the destination receiver and

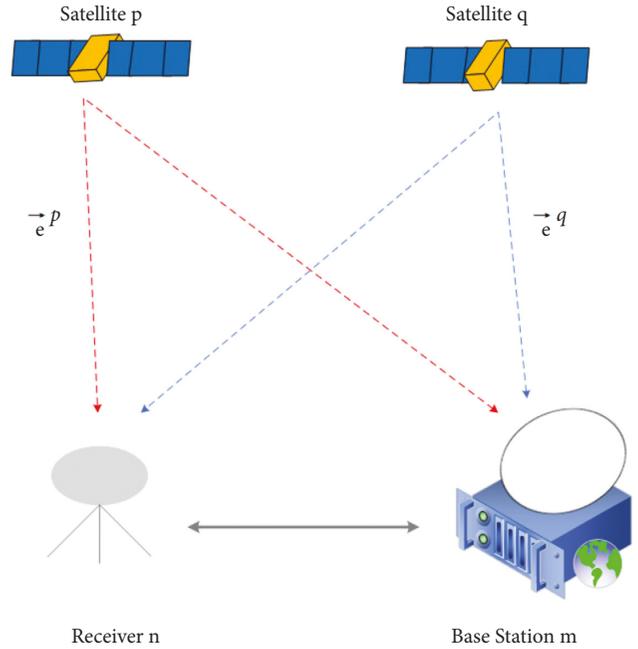


FIGURE 2: The differential positioning diagram between reference base station m and receiver n .

exclude unavailable signals as much as possible [16]. However, SDM is suitable for cross-street scenarios at crossroads, and the positioning scene along the street cannot achieve the ideal effect.

Considering that the construction of a specific 3D model requires precision instruments and detailed databases, but meanwhile, we want to achieve the purpose of mitigating multipath effect under precise positioning. After comparing and analyzing various methods, a feasible solution is to check and deal with the consistency of pseudorange measurements before WLS. Therefore, this paper plans to apply a feasible solution to check and process the two-layer consistency of the pseudorange measurements before WLS after comparison and analysis of different methods. And we weight the evaluation of the residuals of measurements between fixed scenes and continuous epochs.

Therefore, Figure 3 shows the method framework built in this paper. After the GNSS receiver obtains original measurement values through the first layer of preprocessing to screen the effective data, the process detects the signal interference with the help of adaptive threshold residual judgment of LS and carrier-to-noise ratio (CNR). During uniform dynamic acquisition, the DD measurement residuals of the front and back epochs are retained as reference feature to engage with the second layer of the consistency-checks process (CC2). This paper feeds the remaining measurements from the first layer into the XGBoost classifier in the second layer. After features training and testing, the classification learning classifier output results: among them, the NLOS and the presence of obvious outliers are eliminated and retain LOS and part of the corrected multipath reflection signal set. The retained signal measurements are combined with the residual to perform a fit test to obtain the

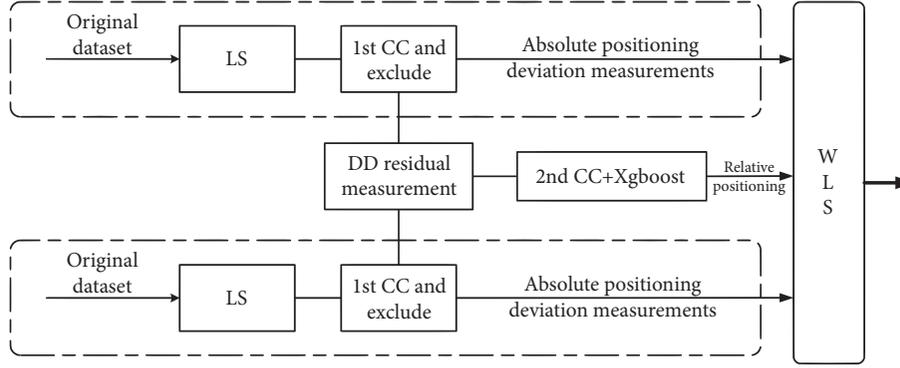


FIGURE 3: The proposed positioning model framework.

processed pseudorange positioning measurements. Finally, in order to further improve the accuracy of position estimation, the WLS method is used to calculate and fit the smooth pseudorange residual.

3.1. The First-Layer Consistency-Check Based on LS. The purpose of the LS-based CC1 is to evaluate and detect the presence of signal interference with specific parameter thresholds at the current specific real-time ephemeral moment t . In order to observe the consistency between the pseudorange measurements of the target, we use the pseudorange residual values to evaluate the fitness of the consistency. Once the inconsistent interference is detected, the least square fitting is performed to help reduce the positioning error. The principle of evaluation is as follows:

$$\varepsilon_{LS} = \overrightarrow{D}_{nm} - \overrightarrow{A} \cdot \Delta \overrightarrow{r}_{nm}. \quad (12)$$

In (12), the pseudorange error variation between the measured and predicted values is denoted by \overrightarrow{D}_{nm} . And $\Delta \overrightarrow{r}_{nm}$ represents the current mobile user receiver state vector, so the residual relationship is constructed with the help of the unit vector matrix \overrightarrow{A} . Due to data samples in the deep urban environment are complex and highly disturbed, a single pseudorange residual cannot mitigate the fitting error accurately. Therefore, the normalization in equation (12) is used to perform the consistency assessment, and the obtained pseudorange residual threshold for the reference SSE_{LS} is expressed as follows:

$$SSE_{LS} = \widehat{\varepsilon}_{LS}^T \cdot \widehat{\varepsilon}_{LS}, \varepsilon \in N(0, Y^{-1}), \quad (13)$$

where Y^{-1} is the inverse matrix form of the diagonal matrix of (8). In this paper, we set ε as the variance when it satisfies the normal distribution. The process of consistency checking algorithm has been mentioned in the study [10]; once a large SSE_{LS} is obtained, it indicates that the pseudorange measurement of the path in the region is inconsistent and may be disturbed by anomalous signals such as multipath and NLOS, and further signal detection and classification processes are needed. If a lower SSE_{LS} is obtained, it means that the signals collected at this point are healthy signal set and could not need to be screened out. In the process of repeated

CC, the partial elimination of the first step will optimise the next repeated evaluation operation and screen and exclude the invalid measurement set until it reaches the range of fitting test standard [10], which can be used as the input reference of the next layer. The exclusion criteria use the classical statistical method—the chi-square goodness-of-fit test—and determine the threshold value $Loss_x$ [32]. If $SSE_{LS} < Loss_x$ and there is a probability of nearly 99.99% false alarm, the next level of classification and exclusion can be done.

3.2. The Second-Layer Consistency Check Based on XGBoost Classifier. The purpose of this section: due to the complexity of the types and interference of multipath in urban environment, the residual iterative exclusion with only one layer cannot detect and exclude invalid signals in low CNR occlusion environment. Considering the difficulty of obtaining coordinate data using ray tracing by previous study, we avoid the map limitation of city modeling and the complex calculation of database. Therefore, we choose to apply a second layer for consistency check with weighted assessment. Before repeating CC2, the retained signal data need to be effectively classified in order to obtain a high-precision signal source so that we can distinguish NLOS and direct signals in the LOS in the low CNR environment. And then, we can assist the receiver output solving and achieving the goal of improving the positioning accuracy by correcting the existing pseudorange residual.

3.2.1. The Principle of Applying XGBoost to Multiclassification. For the multiclassification problem of the first layer of the remaining measurement signal set, the training structure built based on the research background of this paper is shown in Figure 4. The overall structure is divided into three parts: the initially screened dataset is used as the input of XGBoost classifier. The target signal labels and the final classification results are output after the forward-backward gradient boosting.

XGBoost is an engineered implementation of a strong classifier in machine learning, based on algorithmic enhancements and optimisations of GBDT. The underlying framework is an integrated thought model formed by

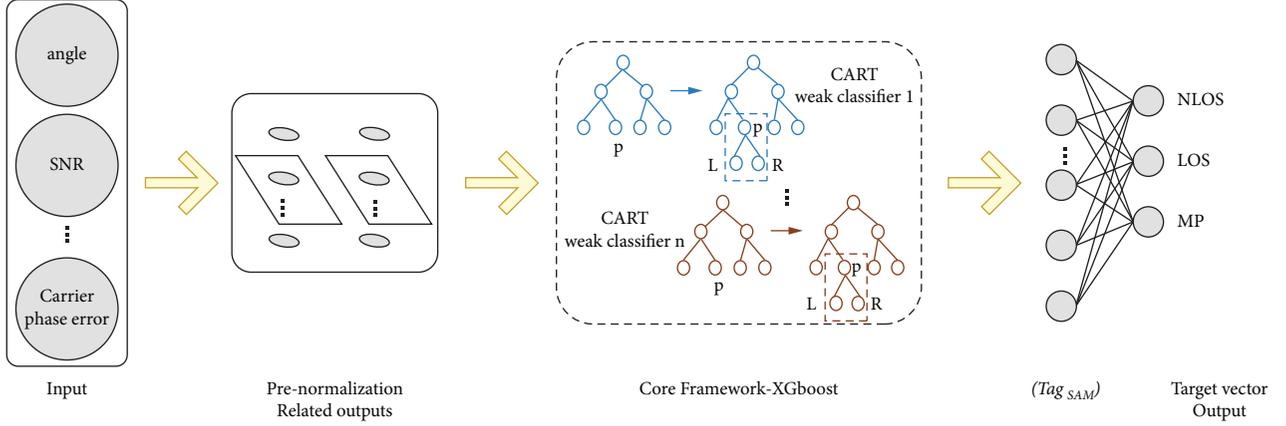


FIGURE 4: The XGBoost multisignal classification structure diagram.

forward gradient iterations on the structure of multiple DT [34]. Before training and predicting the sample features, we need to know the core parts of XGBoost. It is divided into the following parts.

(1) Prenormalization Process

The software receiver can receive the acquired real data source and prepare for output processing. And the distorted mixed signals need to be correlated for feature preprocessing and simple denoising to obtain the remaining measurements as the input to XGBoost. The correlation set I_k of k signals under n features at 1 s ephemeris is expressed as follows:

$$I_k = \{Q_{1,k}, Q_{2,k} \dots Q_{n,k}\}, \quad (14)$$

where $Q_{n,k}$ denotes the form of each data set of the k th signal corresponding to the n th dimensional feature. Secondly, we consider the specificity of satellite correlation data form. For example, the received pseudorange measurements' order is 10^7 and that of the carrier phase observations are 10^3 . In order to ensure the order unity, the i th correlation feature variable m_i needs to be normalized and preprocessed into a new reference variable f_i of equal order of magnitude before input. The process is calculated as follows. The iterative convergence speed of the objective function is improved by means of the ratio of the difference with the mean $\text{mean}(m_i)$ and the standard deviation $\text{std}(m_i)$.

$$f_i = \frac{m_i - \text{mean}(m_i)}{\text{std}(m_i)}. \quad (15)$$

(2) Target Function

Before establishing the objective function, it is necessary to determine that the number of classified output tags is 3 under supervised learning, that is, the NLOS, the direct signal in LOS (DS(LOS)) and the MP signals, including the reflected signals. Based on

the principle of XGBoost, the predicted value $\widehat{y}_{m,i}$ of a signal sample of the signal source collected in this paper is expressed as follows:

$$\sum_{i=1}^3 \widehat{y}_{t,i} = \sum_{k=1}^K \{f_k(x_i), \|I_k\|\} \begin{cases} m = 1, & \text{NLOS,} \\ m = 2, & \text{DS(LOS),} \\ m = 3, & \text{MP,} \end{cases} \quad (16)$$

where m represents the number of signal classes: 1, 2, 3 and $\|I_k\|$ is the characteristic parameter output of the correlator. The prediction result of each type of signal is based on the weighted sum of the residuals of K weak classifiers f_k . The loss function under n signal samples is represented by the relationship between pseudorange prediction value \widehat{y}_i and the true value $y_{m,i}$.

$$\sum_{m=1}^3 L = \sum_{i=1}^n l(y_{m,i}, \widehat{y}_{m,i}). \quad (17)$$

After determining the loss function, it needs to add the objective function to the next step of judgment. As the selected loss function (17) represents the bias, and the variance size needs to be controlled by the regularization term. It is shown as the right half of

$$\sum_{m=1}^3 \text{Tag} = \sum_{i=1}^n l(y_{m,i}, \widehat{y}_{m,i}) + \sum_{i=1}^t \Omega(f_{m,i} | I_{m,i}). \quad (18)$$

At time t of training, the predicted value of the previous moment ($t-1$) and the current moment need to update learning value $f_t(x_{m,i})$. It is clear from (18) that the optimisation objective is to solve $f_t(x_{m,i})$. In order to simplify the objective function, we find the constant term in the function and the regularization term that affects the model complexity with the help of the Taylor expansion series and second-order derivative process. After the expansion of (18), it is shown as follows:

$$\sum_{m=1}^3 \text{Tag} \approx \sum_{i=1}^n \left[l(y_{m,i}, \hat{y}_{m,i}^{t-1}) + g_{m,i} \cdot f_t(x_{m,i}) + \frac{1}{2} h_{m,i} \cdot f_t^2(x_{m,i}) \right] + \sum_{i=1}^t \Omega(f_{m,i}), \quad (19)$$

where $g_{m,i}$ is the first-order derivative of the predicted value at the previous moment and $h_{m,i}$ is the second-order derivative. Because $y_{m,i}$ at moment t is known, the first term of the above equation can be regarded as a constant term C . So, the objective function can be simplified as follows:

$$\sum_{m=1}^3 \text{Tag} \approx \sum_{i=1}^n \left[C + g_{m,i} \cdot f_t(x_{m,i}) + \frac{1}{2} h_{m,i} \cdot f_t^2(x_{m,i}) \right] + \sum_{i=1}^t \Omega(f_{m,i} | I_{m,i}). \quad (20)$$

According to (20), it is only necessary to solve for the first- and second-order derivatives of the loss function and iteratively calculate the optimal objective value to obtain the GNSS signal objective function $f(x)$ for predictive classification and finally obtain the output of the feature labels 0, 1, 2.

(3) Feature Splitting and Filtering

For the data features in this paper, the final set of features identified for the multidimensional variables are mostly closely related to the signal type, but a few redundant features are also present. Before applying this classifier, it is crucial to determine the appropriate features. In order to distinguish valid signals, the statistical contribution calculation method with reference to the previous literature [35] is used to select the target features. The following feature sets are used as the final reference values: elevation angle, azimuth angle, SNR, carrier phase error, pseudorange deviation amplitude, multipath amplitude ratio, the mean time delay variance, mean delay distance, root mean square error value, and so on.

To reduce the impact of the selected features on the training time, we need to evaluate the importance of the features after the model is trained. XGBoost's submodel is a DT construction, which relies on node recursive splitting to achieve tree generation. To find the optimal node for the next branching step, the size of the split gain needs to be calculated. As shown in the core tree splitting part in Figure 4, and the split L and R nodes are compared to do the difference.

$$\text{Gain} = \frac{1}{2} \left[\frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} - \frac{(G_L + G_R)^2}{H_L + H_R + \lambda} \right] - \gamma. \quad (21)$$

In the process, we define the total GNSS signal sample set as $S(i)$, where the weighted sums of the

first-order derivative and second-order derivative under the total sample set are G and H , respectively. And G_L and G_R represent the weighted values of the left and right subtrees. H_L and H_R represent the weighted values of the second-order derivatives of the left and right subtrees. λ and γ are the hyperparameters of the training process as control factors. After the continuous computation of (21), the current split target feature is determined as the significant feature until the gain reaches the threshold of optimisation. Therefore, it is used as the splitting node of the tree. The results of the weights evaluated from the sample set $S(i)$ calculation are shown in Figure 5, and it is obvious to obtain the importance of features, such as elevation angle, azimuth angle, SNR, carrier phase error, and pseudorange deviation amplitude for signal classification.

3.2.2. Consistency Checking under Dynamic Relative Positioning. After the classification process in the previous subsection, NLOS signals are excluded and the reflected signals in the multipath phenomenon are corrected. Repeating the cardinality fit threshold test process, the data is processed for dynamic relative check of the results twice after high-precision classification, and the weighted residuals are used for iterative calculation to finally obtain the reference trajectory position with higher precision.

In the measured environment selected in this paper, there are more vegetation shielding and taller buildings on both sides of the D route in Figure 6, as a more typical urban environment. We get the process reference trajectory using a combined INS/GNSS navigation unit for slow dynamic homogeneous acquisition. During dynamic relative positioning acquisition, abnormal NLOS measurement signals that exceed empirical thresholds are excluded with the help of CC1 after the initial processing and calculation of the LS method.

For the next step of reference and comparison, the remaining set of measurements after CC1 is used as the training target data for the XGBoost classifier during the second layer of checks. This enables the dynamic differential positioning process to remove inconsistent measurements and achieve a refinement of the multipath signal species in the LOS, which in turn helps to smooth errors introduced by the positioning process.

Based on the above process, this paper combines the relative positioning process with the consistency checking principle, and the specific processing process is as follows.

Assume that the comparison position set matrix of the uniform dynamic moving process is shown as follows:

$$\mathbf{V}_{\mathbf{a}-\mathbf{b}} = \sum_{i=0}^n (\vec{r}_a, \vec{r}_b) \cdot \omega_{\mathbf{a}-\mathbf{b}}. \quad (22)$$

Among them, $\mathbf{V}_{\mathbf{a}-\mathbf{b}}$ is the matrix vector in which the results of two position errors are compared under the same path, and its value is used to measure the positioning effect of the consistency checking model under dynamic relative

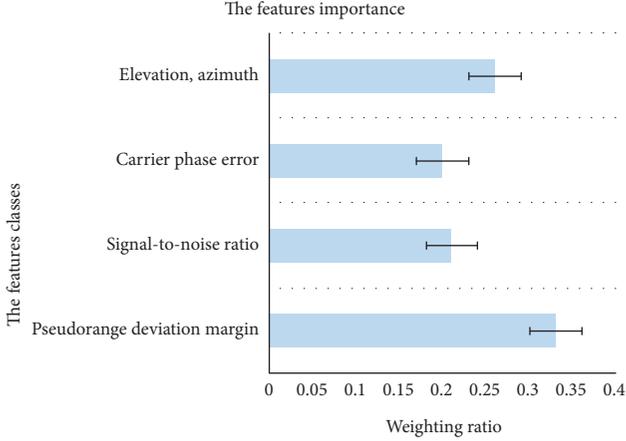


FIGURE 5: The features importance.

positioning. \vec{r}_a represents the predicted location set after the first CC, and \vec{r}_b represents the final set of valid locations after optimal classification by the XGBoost classifier and CC2. To verify the accuracy of the offset of \mathbf{V}_{a-b} , T_S is borrowed as the corresponding statistic, and ω_{a-b} represents the normalized weighting process after the CC2 optimisation. The process is as follows:

$$T_S = \sqrt{\frac{\text{SSE}_{\text{LS}-a}}{n_a}}, \quad (23)$$

$$\omega_{a-b} = \frac{1/T_S}{\sum_{i=0}^n T_S}. \quad (24)$$

where $\text{SSE}_{\text{LS}-a}$ is the sum of the squared errors in (13) and n_a is the amount of the remaining data set after the first layer of filtering. T_S is the average calculated amount of the measured pseudorange residuals, as shown by (23). The lower the results, the smaller the deviation of CC2 against the real trajectory and the more accurate the positioning results obtained later.

4. Experimental Results and Analysis

In this part, we introduce the details of the parameters set and the environment built. The validation of the experimental part is divided into two parts. Firstly, we compare the classification results of XGBoost integrated classification learning with other supervised learning classification methods to evaluate their performance. Secondly, the system after adding the two-layer consistency check is compared with the previously proposed CC model with DT method to verify the performance of the proposed method in this paper.

4.1. Experiments

4.1.1. Data Collection. We used a laptop computer and a centric microchip for B1 IF signal acquisition in BDS. An electric vehicle is selected as an auxiliary vehicle tool to record information at a speed of 5 m/s and 1 s interval.

According to the number and height of buildings in the nearby living area, we select data for 5 min each at an interval of 2 hours in each area with a sampling frequency of 20 MHz.

Figure 7 shows the relevant experimental scenarios and satellite distribution. Among them, Figure 7(a) shows the overhead trajectory of the experimental area near the selected school, divided into four routes A, B, C, and D. In addition, in order to fully verify the positioning performance of the proposed method, the E region is selected as the new test scenario. Figure 7(b) intercepts the distribution of all satellites in the whole area during a single 5 min sampling. For the visible number of some satellites, it can be seen from Figure 7(b) that the highest visible number of satellites reaches 28 when driving to the open area. However, only a few 9 visible satellites can transmit signals in the high-building obscured environment.

4.1.2. Classifier Parameter Settings. During the training of the classifier, the fixed important parameter settings are shown in Table 1. Before training, we divide the pre-processed sample data into training set, verification set, and test set according to the ratio of 8:1:1, and update the gradient weight parameters with the help of SAM optimizer.

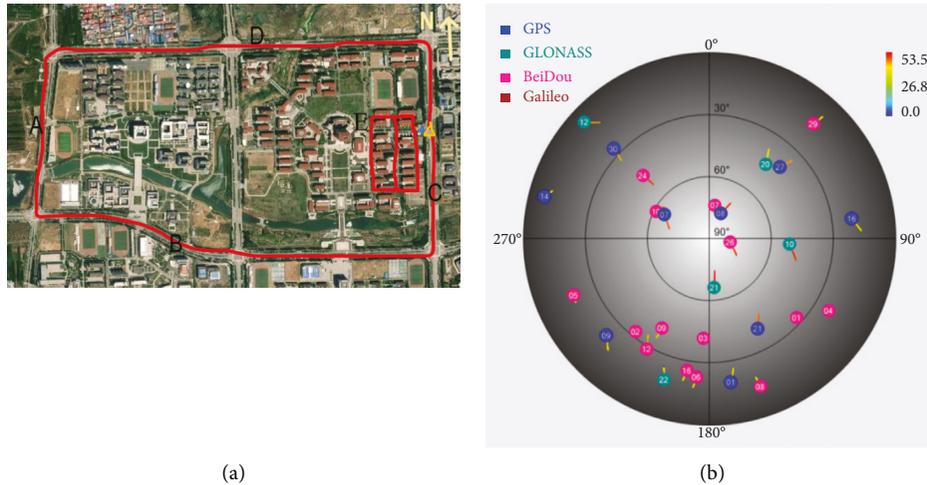
4.2. Classification Results. In this part, we analyse the classification performance of XGBoost and compare it with several commonly used supervised learning classifiers. Considering the multipath differences effects in the selected experimental scenarios, training data under four scenarios are collected in order to make the method more robust. The training data collected in five scenarios are used to enhance the reliability of the evaluation. The datasets of the four low multipath regions of A, B, C, and D are grouped and repeatedly evaluated, and the remaining more complex E-region datasets are used to verify the average classification rates.

Table 2 shows the comparison between the proposed classification method and the existing supervised machine learning classification algorithms from the point of view of classification accuracy, recall rate, and F_1 score. The remaining measurements of satellite B09 after CC1 and exclusion are selected as the data training set. The average classification accuracy values of the reflected signals (RS) in the presence of NLOS and LOS are used to measure the level of multipath (i.e., MP in the table). The results in the table show that the classification accuracy of Decision Tree method for NLOS and DS (LOS) is better than that of other methods, in which the F_1 score also reaches 94.2%, while the classification accuracy of KNN method for three types of signals is approximately the same. However, for the classification results of the multilayer perceptron method (MLP), the classification accuracy of DS (LOS) is as high as 94.2%, which is the best among the three types of signals.

Due to the strong interference from NLOS and MP, the only available signals are DS (LOS) and RS. The results of



FIGURE 6: The original measurement trajectory under route D.



(a)

(b)

FIGURE 7: The relevant experimental scenarios and satellite distribution. (a) The overhead view of the four experimental routes. (b) Satellite distribution over a 20 min collection.

TABLE 1: Parameters for XGBoost. The learning rate and eta are hyperparameters in the training process.

Classification method	Input	Output	Learning rate	Max-depth	eta
XGBoost	Features set	3	0.4	10	0.8

TABLE 2: The results comparison of classical classification methods.

Algorithms	NLOS			DS (LOS)			MP		
	Acc	Rec	F_1	Acc	Rec	F_1	Acc	Rec	F_1
KNN	0.913	0.833	0.924	0.956	0.869	0.897	0.894	0.833	0.864
SVM	0.833	0.875	0.857	0.964	0.824	0.874	0.827	0.875	0.851
DT	0.925	0.846	0.942	0.953	0.837	0.92	0.881	0.86	0.87
MLP	0.875	0.839	0.893	0.942	0.903	0.862	0.893	0.831	0.862
LightGBM	0.918	0.914	0.955	0.935	0.89	0.943	0.901	0.826	0.924
XGBoost	0.936	0.817	0.961	0.984	0.826	0.928	0.916	0.817	0.956

the analysis show that the proposed XGBoost classification method could achieve the best classification rate among several methods for the three main classes. Among them, the F_1 score of the NLOS signal is also as high as 96.1%, the classification accuracy of the direct signal in the LOS signal is as high as 98.4%, and the classification accuracy of the MP signal is 91.6%. The F_1 function index is also higher, reaching 0.956, which proves the high performance of the classifier.

In the field of GNSS positioning, the impact of MP interference signals in an urban environment on the accuracy of a specific solution can be directly shown with the

help of LS estimation. Figure 8 shows the average positioning error of the receiver after interference in the original environment.

4.3. Performance Evaluation of Multipath Mitigation. The data in Figure 9 show that the positioning error of the interfered E original measurement is close to 25 m, and the multipath effect error is obviously lower than that of the four surrounding routes. In order to mitigate the effects of interference above, a double-layer consistency check is performed in conjunction with the signal classification

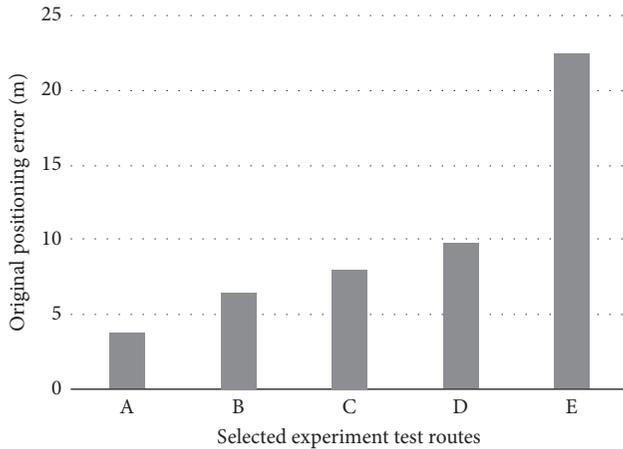


FIGURE 8: The raw errors disturbed by multipath effects.

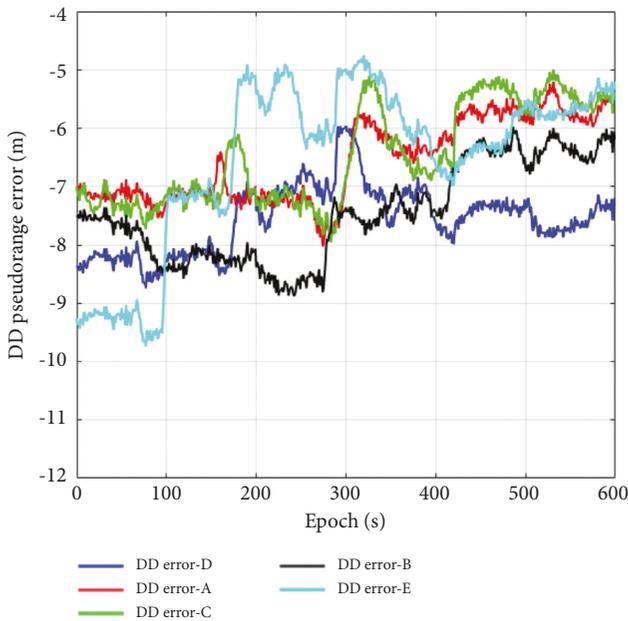


FIGURE 9: The trend in pseudorange error of MP under DD measurements.

method in the first half of the paper. This paper performs verification of multipath mitigation performance from two perspectives.

(1) *The Effect of DD Pseudorange Observations of MP Signals*

Figure 9 shows the variation of the pseudorange error under interference from MP signals for the five scenes. Because the selected multipath environment is not typical, most of the delays are short, about 1.2 m; while in the more complex *E* scene, the delay of DD pseudorange is larger than that of other scenarios, which is within the fluctuation range of 42–52 m.

(2) *The Multipath Mitigation Performance of the Proposed Method*

As an experimental scenario with strong multipath effects, Figure 10 compares the position errors in both horizontal and altitude dimensions under the *D* route and *E* route with conventional LS method and after using the improved CC method in this paper.

In the left figure, it can be seen that there is an obvious signal fluctuation jump around 100–200 s. The position error after the improved model still reaches an anomalous 5 m neighborhood. So, it can be inferred that the distortion of the receiver output signal is most severe during the period. The changing trend of the orange curve under *D* proves the feasibility of this method. Although the signals of some epoch segments are out of lock, after the optimisation of the proposed method, the error of the latter half of the epoch converges uniformly near the 2–4 m, which is about 5 m less than the original green existence error curve. The comparison trend of the error variation in the right figure clearly shows that near the *E* route, which is more deeply affected by the MP effect, resulting in the horizontal direction using the consistency checking method to optimise. And there are still many signals beating, such as fluctuations around 60 s, 125 s, and 200 s, but gradually become stable in the later stage, converging within the 5–10 m. Secondly, according to the changing trend of altitude error, the positioning error before consistency checking fluctuates around 22 m. After the consistency checking operation, the error variation is obviously smaller than the green curve, and the stable fluctuation is about 5 m.

On the other hand, we assess the performance of the proposed method from the comparison of ablation studies. For any of the five experimental location scenes, we compare the average error values of the three methods using LS iterations and give the percentage improvement in accuracy relative to the previous method improvement. The results in Table 3 indicate that the optimisation of *A* and *B* is not significant, while the relatively more complex *D* has a 43.1% improvement in the percentage of reduced positioning error after CC2's optimising. And after the optimisation of the proposed method and WLS fitting, the positioning error of the supplementary *E* scene is controlled at 4.17 m, which is about 33% higher than that before fitting. Thus, validating the feasibility of the XGBoost-based integrated classification method of CC model proposed in this paper can be considered to apply the dynamic driving scenarios to improve target positioning accuracy.

(3) *Comparison with Research Methods in the Same Field*

Compared to the previous, some scholars have used 3D maps to assist in obtaining datasets and using a CC method for DT classified data signals [33]. This method is a newer approach in the direction of navigation processors in the field of multipath mitigation. We refer to this method as CC + DT(3DMA).

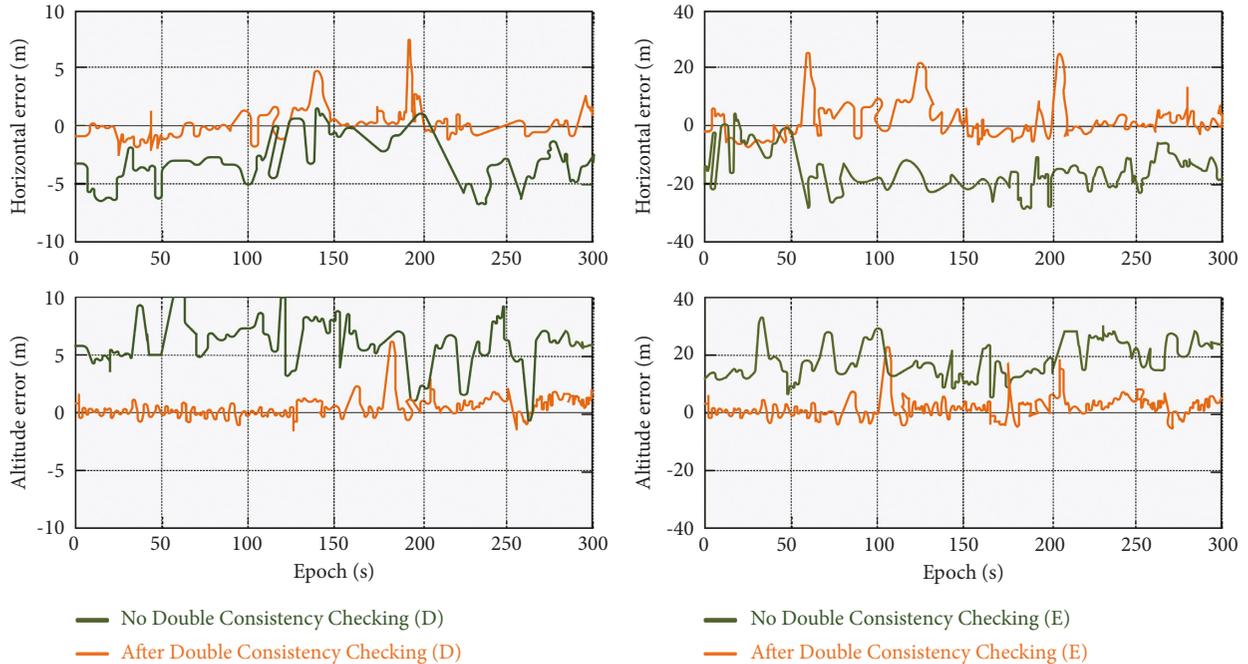


FIGURE 10: The comparison of positioning errors before and after the application of CC in scene *D* and *E*.

TABLE 3: Positioning error results for the selected experimental scenarios.

Experimental scenarios	CC1 + LS	CC2 + XGBoost (%)	CC + WLS (%)
<i>A</i>	1.4	1.32 (5.7%)	1.2 (9.1%)
<i>B</i>	2.3	2.15 (6.5%)	1.93 (10.23%)
<i>C</i>	5.4	4.75 (12%)	4.2 (11.6%)
<i>D</i>	8.52	4.85 (43.1%)	3.8 (21.7%)
<i>E</i>	15.1	6.3 (58.3%)	4.17 (33.8%)

In this paper, we obtain better results for changes in the experimental scenario and improvements in the internal algorithm. In this section, two evaluation metrics, root mean square error (RMSE) and coefficient of determination (*R*-square), are introduced. They are chosen to reflect the superiority of our proposed method.

Figure 11 shows the error trend of the fit performance between the original data before improvements and two current research methods. As can be seen from the blue dashboard trend, the performance of CC + DT(3DMA) has improved considerably compared to the previous error profile of the ground truth. After the addition of the XGBoost classification method, the RMSE value is approximately 1.8 m, and the final RMSE value after WLS smoothing is improved by nearly 0.14 m from the previous one.

Secondly, the orange dashboard trend shows that *R*-square values are fitted around 0.5 after the optimisation of the model in this paper, which clearly reduces the level of multipath interference within the receiver. Therefore, based

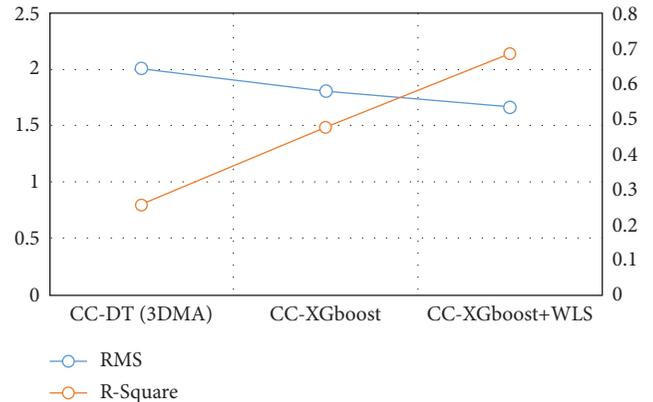


FIGURE 11: The trends in evaluation indicators under different methods.

on the fitting trends of the two types of evaluation metrics, it can be concluded that the consistent measurement set obtained with the aid of the XGBoost classification method improves the receiver position accuracy better than CC + DT(3DMA). After the second-layer checking, the final positioning enhancement is achieved by applying WLS, which optimises the positioning accuracy by almost 17% compared to CC + DT(3DMA). Compared to the previous, some scholars have used 3D maps to assist in obtaining datasets and using a CC method for DT classified data signals. This method is a newer approach in the direction of navigation processors in the field of multipath mitigation. We refer to this method as CC + DT(3DMA). In this paper, we obtain better results for changes in the experimental scenario and improvements in the internal algorithm.

5. Conclusion

The double-layer consistency checking model based on XGBoost algorithm proposed in this paper effectively mitigates strong interference of multipath effects in urban environments. In this method, the cardinality fit test and traditional LS estimation of position are used in the first layer to exclude inconsistency-checked measurements. And the remaining valid measurement dataset is used as the input for the second layer of consistency checking, applying the preprocessed dataset and debugged XGBoost classifier to classify signals into typically three classes. The classification results show that the XGBoost classifier is able to improve the classification accuracy of NLOS to 93.6% compared to several different supervised learning classification methods. After we achieve the classification results, it is equally important to combine DPDT to solve for the pseudorange measurements between satellite and receiver to eliminate systematic errors. Finally, two results are fitted based on the WLS smoothing method, and the results show that the RMSE value after iteration reduces to 1.668 m. It is an improvement of about 0.34 m compared to the CC + DT (3DMA) method for mitigating multipath.

Although the method proposed in the paper has a good improvement, there are still problems. For example, the level of multipath interference with the help of an electric vehicle simulating a dynamic reception scenario is not typical. Therefore, we could select a more complex cellular dense building environment to verify the feasibility of the method in the next step. In addition, due to the slow movement of the electric vehicle and the limited amount of training data collected, we suggest converting typical features of GNSS signals into two-dimensional or multidimensional segmentation for more accurate evaluation with the help of deep learning methods.

Abbreviations

GNSS: Global navigation satellite system
 NLOS: Non-line-of-sight
 DPDT: Dynamic pseudorange differencing technique
 XGBoost: eXtreme gradient boosting.

Data Availability

The data that support the findings of this study are available from the corresponding author, Dengao Li, upon reasonable request.

Conflicts of Interest

The authors declare that they have no financial and personal relationships with other people or organizations that can inappropriately influence their work.

Authors' Contributions

Dengao Li contributed to the conception of the study; Xiaoli Ma performed the experiment and contributed significantly to analysis and manuscript preparation; and Jumin Zhao helped perform the analysis with constructive discussions.

Acknowledgments

This work was supported by the National Key R&D Project under Grant 2018YFB2200900; The General Object of National Natural Science Foundation under Grant 61772358; and The Key Technology R&D Program of Jinzhong under Grant Y201024.

References

- [1] N. Vagle, A. Broumandan, and G. Lachapelle, "Multi-antenna GNSS and INS/odometer coupling for robust vehicular navigation," in *Proceedings of the International Technical Symposium on Navigation and Timing (ITSNT)*, Toulouse, France, November 2017.
- [2] H. Z. W. J. Liu, "Indoor NLOS and multipath detection using deep learning approach," *GPS Solutions*, vol. 23, no. 75, 2021.
- [3] J. Zidan, E. Adegoke, E. Kampert, S. Birrell, and M. Higgins, "Gnss vulnerabilities and threats," *IEEE Access*, vol. 1, 2020.
- [4] S. R. W. Guangyu and C. Qi, "GPS-based dual-polarisation antenna for positioning in urban canyons," *IEEE Access*, vol. 4, no. 68, pp. 13–19, 2020.
- [5] Z. Jiang and P. D. Groves, "NLOS GPS signal detection using a dual-polarisation antenna," *GPS Solutions*, vol. 18, no. 1, pp. 15–26, 2014.
- [6] M. Adjrad and P. Groves, "3d-mapping-aided GNSS exploiting galileo for better accuracy in dense urban environments," in *Proceedings of the Navigation Conference*, pp. 108–118, IEEE, Lausanne, Switzerland, May 2017.
- [7] T. Suzuki, N. Kubo, and A. Yasuda, "The possibility of the precise positioning and multipath error mitigation in the real-time," in *Proceedings of the 2004 International Symposium on GNSS/GPS*, pp. 21–24, Long Beach, CA, USA, September 2004.
- [8] H. Tokura, H. Yamada, N. Kubo, and S. Pullen, "Using multiple GNSS constellations with strict quality constraints for more accurate positioning in urban environments," *Positioning*, vol. 05, no. 04, pp. 85–96, 2014.
- [9] P. R. R. Strode Strode and P. D. Groves, "GNSS multipath detection using three-frequency signal-to-noise measurements," *GPS Solutions*, vol. 20, no. 3, pp. 399–412, 2016.
- [10] L. T. Hsu, H. Tokura, N. Kubo, Y. Gu, and S. Kamijo, "Multiple faulty GNSS measurement exclusion based on consistency check in urban canyons," *IEEE Sensors Journal*, vol. 17, no. 6, pp. 1909–1917, 2017.
- [11] G. Zhang, W. Wen, and L. T. Hsu, "A novel GNSS based V2V cooperative localization to exclude multipath effect using consistency checks," in *Proceedings of the IEEE/ION Position, Location and Navigation Symposium (PLANS)*, April 2018.
- [12] J. Marais, M. Berbineau, and M. Heddebaut, "Land mobile GNSS availability and multipath evaluation tool," *IEEE Transactions on Vehicular Technology*, vol. 54, no. 5, pp. 1697–1704, 2005.
- [13] T. Suzuki and N. Kubo, "NLOS GNSS signal detection using fish-eye camera for vehicle navigation in urban environments," in *Proceedings of the 27th International Technical Meeting of the Satellite Division of the Institute of Navigation, ION GNSS 2014*, vol. 3, pp. 1897–1906, Tampa, FL, USA, September 2014.
- [14] D. Maier and A. Kleiner, "Improved GPS sensor model for mobile robots in urban terrain," in *Proceedings of the 2010 IEEE International Conference on Robotics and Automation*, pp. 4385–4390, Singapore, December 2010.

- [15] W. Wen, G. Zhang, and L. Hsu, "Exclusion of GNSS NLOS receptions caused by dynamic objects in heavy traffic urban scenarios using real-time 3D point cloud: an approach without 3D maps," in *Proceedings of the 2018 IEEE/ION Position, Location and Navigation Symposium (PLANS)*, pp. 158–165, Monterey, CA, USA, April 2018.
- [16] P. D. Groves and C. Qi, "Shadow matching: a new GNSS positioning technique for urban canyons," *Journal of Navigation*, vol. 64, no. 3, pp. 417–430, 2011.
- [17] L. T. Hsu, Y. Gu, and S. Kamijo, "3D building model-based pedestrian positioning method using GPS/GLONASS/QZSS and its reliability calculation," *GPS Solutions*, vol. 20, no. 3, pp. 413–428, 2016.
- [18] T. Suzuki, "Integration of GNSS positioning and 3D map using particle filter," *GPS Solutions*, vol. 2, pp. 1296–1304, 2016.
- [19] P. D. Groves and M. Adjrak, "Likelihood-based GNSS positioning using LOS/NLOS predictions from 3D mapping and pseudoranges," *GPS Solutions*, vol. 21, no. 4, pp. 1805–1816, 2017.
- [20] L. Wang, P. D. Groves, and M. K. Ziebart, "Multi-constellation GNSS performance evaluation for urban canyons using large virtual reality city models," *Journal of Navigation*, vol. 65, no. 3, pp. 459–476, 2012.
- [21] M. Adjrak and P. D. Groves, "Enhancing least squares GNSS positioning with 3D mapping without accurate prior knowledge," *Navigation*, vol. 64, no. 1, pp. 75–91, 2017.
- [22] R. YozevitchYozevitch, B. B. Moshe, and A. Weissman, "A robust GNSS LOS/NLOS signal classifier," *Navigation*, vol. 63, no. 4, pp. 429–442, 2016.
- [23] T. Suzuki, Y. Nakano, and Y. Amano, "NLOS multipath detection by using machine learning in urban environments," in *Proceedings of the 30th International Technical Meeting of The Satellite Division of the Institute of Navigation (ION GNSS+ 2017) Institute of Navigation*, pp. 3958–3967, Portland, Oregon, USA, September 2017.
- [24] J. Xu and H. Jia, "Intelligent GPS L1 LOS/multipath/NLOS classifiers based on correlator-, RINEX- and NMEA-level measurements," *Remote Sensing*, vol. 11, no. 16, p. 1851, 2019.
- [25] Z. Lyu and Y. Gao, "An SVM based weight scheme for improving kinematic GNSS positioning accuracy with low-cost GNSS receiver in urban environments," *Sensors*, vol. 20, no. 24, p. 7265, 2020.
- [26] Y. Quan, L. Lau, G. W. Roberts, X. Meng, and C. Zhang, "Convolutional neural network based multipath detection method for static and kinematic GPS high precision positioning," *Remote Sensing*, vol. 10, no. 12, p. 2052, 2018.
- [27] J. Fan and A. S. Awan, "Non-Line-of-Sight identification based on unsupervised machine learning in ultra wideband systems," *IEEE Access*, vol. 7, Article ID 32471, 2019.
- [28] P. D. Groves, *Principles of GNSS, Inertial, and Multi-Sensor Integrated Navigation Systems* Artech House, MA, USA, 2015.
- [29] J. S. W. Chong and W. Zhenjie, "The Correction method for non-line-of-sight signal detection in urban environments," *Journal of Navigation and Positioning(in Chinese)*, vol. 22, no. 2, pp. 40–44+49, 2018.
- [30] Y. Xu, *The Research on NLOS Identification and Compensation Techniques in High Accuracy Ranging Systems*, Xi'an University of Electronic Science and Technology, 2020, in Chinese.
- [31] L. X. S. Maojun, Z. Lijiang, "The Research on BDS dynamic pseudoranges differential localization algorithm based on Kalman filtering method," *Global Positioning Systems(in Chinese)*, vol. 44, no. 3, pp. 69–73, 2019.
- [32] P. D. Groves and Z. Jiang, "Height aiding, C/N0 weighting and consistency checking for GNSS NLOS and multipath mitigation in urban areas," *Journal of Navigation*, vol. 66, no. 5, pp. 653–669, 2013.
- [33] J. Zidan, O. Al-Luhaibi, E. Adegoke, E. Kampert, and C. Ford, "3D mapping methods and consistency checks to exclude GNSS multipath/NLOS effects," in *Proceedings of the 2020 International Conference on UK-China Emerging Technologies (UCET)*, August 2020.
- [34] G. C. Chen, "XGBoost: a scalable tree boosting system," 2016, <https://arxiv.org/abs/1603.02754>.
- [35] R. Kan and M. Wang, "Genetic algorithm-based optimization of SVM for LOS and NLOS classification for self-calibrating indoor positioning," *Modern Electronic Technology(in Chinese)*, vol. 44, no. 576, pp. 57–60, 2021.
- [36] J. Soubielle Soubielle, I. Fijalkow Fijalkow, and A. DuvautBibaut, "GPS positioning in a multipath environment," *IEEE Transactions on Signal Processing*, vol. 50, no. 1, pp. 141–150, 2002.
- [37] T. Suzuki and Y. Amano, "NLOS multipath classification of GNSS signal correlation output using machine learning," *Sensors*, vol. 21, no. 7, p. 2503, 2021.
- [38] J. Zhang Hxsz and M. Lu, "Space and frequency diversity characterization of mobile GNSS receivers in multipath fading channels," *Tsinghua Science and Technology*, vol. 25, 2019.