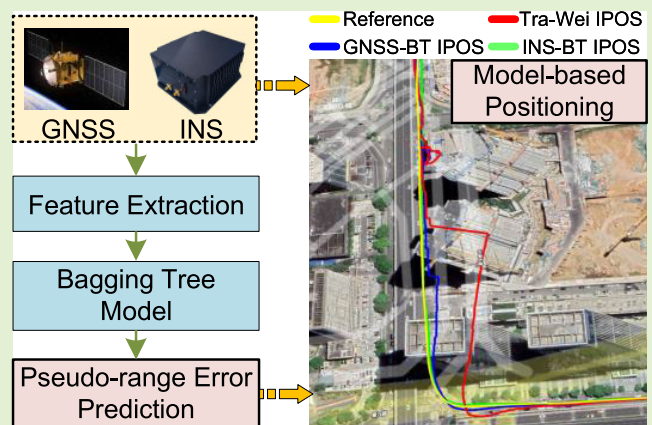


INS-Aided GNSS Pseudo-Range Error Prediction Using Machine Learning for Urban Vehicle Navigation

Tisheng Zhang^{ID}, Long Zhou^{ID}, Xin Feng^{ID}, Jinwei Shi^{ID}, Quan Zhang^{ID}, and Xiaoji Niu^{ID}

Abstract—Global navigation satellite system (GNSS) is being extensively applied in different navigation applications. However, GNSS direct signals are easily affected by multipath and non-line-of-sight (NLOS) signals, resulting in severe deterioration of positioning. GNSS receiver output information, such as carrier-to-noise ratio (C/N₀) and satellite elevation, cannot accurately reflect the pseudo-range quality, leading to a significant increase in positioning errors. This article proposes an inertial navigation system (INS)-aided GNSS pseudo-range error prediction approach based on machine learning for urban vehicle navigation. As an important feature, the pseudo-range residual estimated by INS is employed for model training, together with the C/N₀, satellite elevation, and pseudo-range rate consistency. The predicted model of the pseudo-range errors is obtained by an ensemble bagging decision tree learning method. Urban vehicle tests show that compared to GNSS single-point positioning (SPP) with C/N₀-based weighting, the horizontal accuracy in the form of CEP95 of SPP with model-based weighting improves 52.81%, and the GNSS/INS horizontal positioning error in the form of CEP95 is reduced from 21.23 to 5.02 m in deep urban environments.

Index Terms—Bagging decision tree, global navigation satellite system (GNSS) multipath, GNSS pseudo-range, inertial navigation system (INS), urban positioning.



I. INTRODUCTION

GLOBAL navigation satellite system (GNSS)-based sensors are widely used as localization sources. However, GNSS performance is easily affected by multipath and nonline-of-sight (NLOS) signals reception in urban environments. While in the case of multipath, the direct and the reflected signals are received simultaneously, in the case of NLOS, only reflected signals can be received [1]. These factors cause serious deterioration in the quality of GNSS pseudo-range observations, increasing positioning error and

affecting the GNSS application in urban canyons. To mitigate the effects of multipath and NLOS on GNSS positioning, scholars have done lots of research through antenna design, signal processing, observation processing, external information aiding, and machine learning.

High-quality survey antennas can improve GNSS signal reception. The choke coil antennas [2] have been widely adopted for multipath suppression [3]. Since the right-hand circular polarized (RHCP) signal would be transferred to the left-hand circular polarized (LHCP) signal when reflected, Jiang and Groves [4] applied dual-polarized antenna detecting reflected signals. However, survey antennas are expensive and bulky, unsuitable for consumer applications. Suzuki et al. [5] proposed a method to mitigate line-of-sight (LOS) and NLOS multipath errors by rotating the GNSS antenna arm horizontally at a certain angular velocity. The positioning error was decreased from 18.96 to 2.83 m by the rotating antenna, which requires an additional mechanical device. At the signal processing level, the narrowband correlators were proposed by Van Dierendonck et al. [6] to mitigate multipath, which was widely studied and applied [7]. In addition, specially designed

Manuscript received 6 December 2023; revised 16 January 2024; accepted 16 January 2024. Date of publication 24 January 2024; date of current version 14 March 2024. This work was supported in part by the National Natural Science Foundation of China under Grant 42374034 and Grant 41974024. The associate editor coordinating the review of this article and approving it for publication was Prof. Fuqiang Gu. (Corresponding authors: Xiaoji Niu; Quan Zhang.)

Tisheng Zhang, Long Zhou, Jinwei Shi, Quan Zhang, and Xiaoji Niu are with the GNSS Research Center, Wuhan University, Wuhan 430079, China (e-mail: zts@whu.edu.cn; zhoulong@whu.edu.cn; skingwei@whu.edu.cn; zhangquan@whu.edu.cn; xjniu@whu.edu.cn).

Xin Feng is with the School of Geodesy and Geomatics, Wuhan University, Wuhan 430079, China (e-mail: focusfeng@whu.edu.cn).

Digital Object Identifier 10.1109/JSEN.2024.3355705

delay-locked loops such as CADLL [8] and MEDLL [9] could also effectively mitigate multipath. However, these baseband processing methods can only suppress long-delay multipath, have limited effect on short-delay multipath, and cannot solve the NLOS problem. At the observation level, scholars utilized information such as the carrier-to-noise density ratio (C/N_0) and elevation to estimate the quality of GNSS observation. While the high C/N_0 signal can be classified as LOS, the low C/N_0 is classified as NLOS or multipath [10], [11]. The higher the elevation, the more likely the signal will be LOS. As the elevation decreases, the satellite is more likely to be blocked by buildings or other obstacles with the corresponding signal being NLOS [12], [13]. However, these observation features cannot accurately distinguish between multipath and NLOS signals. They cannot quantitatively represent the pseudo-range observation error, which has limited improvement in the positioning accuracy.

To effectively mitigate the impact of NLOS on GNSS positioning, scholars have investigated external information (cameras, lidar, 3-D models, etc.) assisting GNSS signal classification. Visible light cameras [14], fish-eye cameras [15], [16], and omnidirectional infrared cameras [17] pointed at the zenith can detect satellites blocked or not by buildings. However, the camera is greatly affected by the weather and lighting conditions. Real-time 3-D point cloud generated by LiDAR is also used to identify NLOS signals, but it has a limited measurement range [18], [19]. Three-dimensional city model is helpful for NLOS identification and positioning accuracy [20]. Shadow matching was proposed by Groves [21] and Lei et al. [22] to improve the accuracy of cross-street positioning in dense urban through 3-D building models. Hsu et al. [23], [24] and Obst et al. [25] utilized the 3-D map to perform ray tracing to simulate the signal-transmitting path among buildings and trees, determine the visibility of the signals, and even correct the NLOS errors. Kbayer and Sahmoudi [26] and Peyraud et al. [27] exploited 3-D city models and map information to assist in detecting and correcting NLOS. Three dimensional model assisting methods depend highly on model accuracy and have limited usefulness for multipath signals.

With the development of satellite navigation positioning systems, modernized satellite navigation systems modulate signals by binary offset carrier (BOC) or a variant of it, such as the B1C signal of Beidou using quadrature multiplexed BOC (QMBOC) modulation [28], the E5 signal of Galileo using alternative BOC (AltBOC) modulation, and so on [29]. The new signals reduce the multipath effect from the perspective of signal design and bring better performance. In the literature [30], the PCF algorithm is proposed that realizes unambiguous tracking by constructing pseudo-correlation functions. The PCF algorithm can completely eliminate the multiple peaks of BOC signals, and at the same time use the appropriate width factor to obtain narrower correlation peaks and better multipath resistance performance [31]. Wang et al. [32] using the noncoherent early-late processing track loop constructed the multipath error model and analyzed the performance of multipath suppression of time division AltBOC (TD-AltBOC) and

AltBOC signals for different front-end bandwidth of the receiver. The experimental results show that AltBOC (15, 10) has the best multipath suppression performance. Although the new signals can have some anti-multipath effect in design, they cannot completely solve the multipath. It is still necessary to apply other multipath suppression algorithms.

Traditional methods can only mitigate one of the effects of multipath and NLOS. Conventional integration systems can only use a limited number of partial features to model multipath signals, and there are limitations to the scenarios in which they can be used. Due to the stochastic nature of multipath signals, it is difficult to unite multiple observational information to obtain an effective prediction model, so more and more studies identified and mitigated GNSS multipath and NLOS by machine learning. Machine learning can fit relationships between variables that cannot be shown to be modeled, which makes it well suited to solve the problem of multipath. Phan et al. [33] used elevation and azimuth as key features of a support vector machine (SVM) to mitigate multipath effects, reducing the standard deviation (STD) of multipath errors by an average of 79%. Yozevitch et al. [34] classified LOS and NLOS based on a decision tree approach with C/N_0 , elevation, and pseudo-range variables as inputs, and the recognition accuracy is 77.60% for LOS signals and 87.20% for NLOS signals. Hsu [35] constructed a signal classifier based on the SVM algorithm using received signal strength, the change rate of RSS, pseudo-range residue, and difference between delta pseudo-range and pseudo-range rate as inputs, and the recognition accuracy is 75.40%. Sun et al. [36] utilized an artificial neuro fuzzy inference system (ANFIS) to classify LOS, multipath, and NLOS signals with nine variables such as HDOP, VDOP, etc., and the recognition rate was higher than 84% in static tests. Existing machine learning algorithms mainly focus on identifying and classifying multipath and NLOS signals [37], [38]. However, the multipath and NLOS classification cannot quantitatively reflect the pseudo-range observation error directly related to the positioning accuracy.

When there are few available LOS satellites, multipath, and NLOS signals need to participate in GNSS positioning, and it is very valuable to estimate pseudo-range errors to ensure positioning accuracy in signal-challenged environments. Sun et al. [39] proposed a gradient boosting decision tree (GBDT) based method to predict GNSS pseudo-range errors by taking signal strength, satellite elevation, and pseudo-range residual as input features. The 3-D positioning accuracy of the pseudo-range error correction-based positioning measured in terms of the root mean square (rms) error has an improvement of more than 70% over the conventional methods in urban environments. Zhang et al. [40] predicted the satellite visibility and pseudo-range errors based on deep learning networks with long short-term memory. They employed five input features, including elevation, azimuth angle, pseudo-range residual, C/N_0 , and root-sum-squares of pseudo-range residuals. The networks can predict the pseudo-range errors with an average difference of 4.90 m to the labeled errors. However, the pseudo-range residual is obtained based on GNSS positioning, constrained by GNSS positioning accuracy, and would not be credible in deep urban environments.

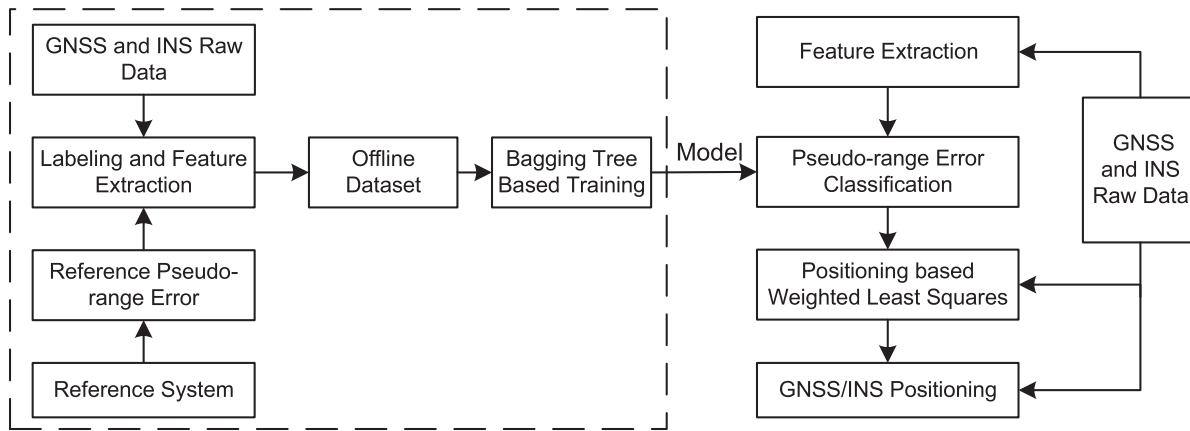


Fig. 1. Framework of the proposed pseudo-range error prediction method.

In addition, these methods are only suitable for static scenarios, and their performance would degrade under dynamic conditions. Sun et al. [41] constructed the random forest-based pseudo-range error prediction and correction models considering the C/N_0 , satellite elevation, and local positional information. The new research improves the kinematic positioning accuracy in urban environments. However, the location information used in this method is still obtained from GNSS positioning, which can be influenced in deep urban environments. In GNSS and inertial navigation system (INS)-integrated navigation systems, the INS can estimate the GNSS pseudo-range residual more accurately in GNSS challenged environments. Therefore, GNSS pseudo-range residual estimated by INS can be an excellent input feature of machine learning. This article predicts GNSS pseudo-range errors by a bagging decision tree learning method, with pseudo-range residual estimated by INS, C/N_0 , satellite elevation, and pseudo-range rate consistency as feature inputs. The predicted model is employed in GNSS single-point positioning (SPP) and GNSS/INS-integrated navigation to verify pseudo-range prediction and positioning performance. To ensure the effectiveness and generalization of the model, we collected nine sets of vehicle-based data from different environments for model training and testing. The main contributions are summarized as follows.

- 1) To improve the positioning accuracy of GNSS/INS coupled system in urban vehicle environments, the GNSS pseudo-range residual calculated by INS is considered an important input feature to train the GNSS pseudo-range error prediction model.
- 2) The INS-aided prediction model is trained by an ensemble bagging tree learning method, with multiple sets of urban vehicle testing data, including elevated roads, tree-lined roads, building clusters, and tunnels, to ensure its effectiveness and generalization ability.
- 3) Compared to SPP with C/N_0 -based weighting, the horizontal accuracy in the form of circular error probable (CEP) 95 of SPP with model-based weighting improves 52.81%, and the horizontal error (CEP95) GNSS/INS-integrated navigation is reduced from 21.23 to 5.02 m in urban environments.

The remainder of the article is organized as follows. Section II presents the overall framework for the pseudo-range error prediction method. Section III describes the characteristics and extractions of labels and features. Section IV discusses the machine learning algorithm of pseudo-range error prediction. Testing and analysis are presented in Section V. Section VI gives conclusions.

II. ALGORITHM FRAMEWORK

The proposed pseudo-range error prediction method is presented within the framework depicted in Fig. 1, which consists of an offline training phase (dashed box on the left) and an online model testing phase. During the offline training phase, multiple sets of vehicle data of the GNSS/INS-coupled system were collected in typical urban environments, including GNSS C/N_0 , satellite elevation, pseudo-range, Doppler, raw data from microelectromechanical system (MEMS) inertial measurement units (IMUs), and GNSS position, INS position, and integrated position. The collected data are preprocessed to extract input features, such as C/N_0 , elevation, pseudo-range rate consistency, and estimated pseudo-range residual. Meanwhile, a high-performance GNSS/INS-coupled system, as a reference system, is employed to obtain a pseudo-range error labeled dataset. The bagging tree trains the dataset to get the relationship model between the input features and pseudo-range error.

The online test phase, shown on the right side of Fig. 1, includes feature extraction, labeling process, and bagging tree algorithms. In this phase, a dataset not employed for training is for online model testing. First, the features, including C/N_0 , elevation, estimated pseudo-range residual, and pseudo-range rate consistency, are extracted for each epoch and input into the learning model, and obtain corresponding classification results of pseudo-range error. Then, the results are transformed into weight values of the least squares for GNSS positioning. Finally, the GNSS and INS are fused to obtain the integrated position.

The GNSS position can show the effect of the weight setting on pseudo-ranges, which reflects the pseudo-range error prediction accuracy of the learning model. The integrated

results demonstrate that the proposed prediction method can improve GNSS/INS-integrated navigation performance.

III. LABELING PROCESS AND FEATURE EXTRACTION

A. Labeling Process

Labeling GNSS pseudo-range errors is an important step of the model training. The reference of pseudo-range errors can be calculated using the vehicle position, base station, and satellite ephemeris. For satellite i , its pseudo-range observation equation can be expressed as

$$\rho^{(i)} = r^{(i)} + \delta t_u - \delta t^{(i)} + I^{(i)} + T^{(i)} + \varepsilon^{(i)}. \quad (1)$$

In which, $r^{(i)}$ is the geometric range between the satellite i and the vehicle, $\delta t^{(i)}$ is the satellite clock offset, δt_u is the receiver clock offset, $I^{(i)}$ is the ionospheric delay, $T^{(i)}$ is the tropospheric delay, and $\varepsilon^{(i)}$ is the pseudo-range error caused by environment multipath, thermal noise, etc.

Since the vehicle is close to the base station, it can be assumed that they are affected by the same atmospheric delays. We can eliminate the effects of the ionospheric, tropospheric, and satellite clock offset by the station difference. The result of the difference for satellite i is

$$\delta \rho^{(i)} = r^{(i)} - r_{\text{base}}^{(i)} + \delta t_u - \delta t_{\text{base}} + \varepsilon^{(i)} - \varepsilon_{\text{base}}^{(i)}. \quad (2)$$

In which, $r_{\text{base}}^{(i)}$ is the range between the satellite i and the base station, and δt_{base} is the receiver clock offset of the base station. The base station is located in open environments and is minimally affected by multipath, so its pseudo-range error $\varepsilon_{\text{base}}^{(i)}$ can be negligible. The satellite-vehicle range $r^{(i)}$ is obtained by the position of the reference system and the satellite

$$\begin{aligned} r^{(i)} &= \|\mathbf{x}^{(i)} - \mathbf{x}\| \\ &= \sqrt{(x^{(i)} - x)^2 + (y^{(i)} - y)^2 + (z^{(i)} - z)^2}. \end{aligned} \quad (3)$$

$\mathbf{x}^{(i)} = (x^{(i)}, y^{(i)}, z^{(i)})$ is the satellite position calculated using the satellite ephemeris, while $\mathbf{x} = (x, y, z)$ represents the position obtained by the high-precise reference system. The position of the reference system is obtained by post-processing integrated navigation of GNSS real-time kinematic positioning (RTK) and high-precision INS. And $r_{\text{base}}^{(i)}$ can be calculated in the same way.

In GNSS-challenged environments, it is difficult to estimate the receiver clock offset during GNSS positioning accurately. Therefore, it is necessary to eliminate the influence of the receiver clock offset through satellite difference. Selecting satellite j as the reference, the result of station difference for satellite j is expressed as

$$\delta \rho^{(j)} = r^{(j)} - r_{\text{base}}^{(j)} + \delta t_u - \delta t_{\text{base}} + \varepsilon^{(j)} - \varepsilon_{\text{base}}^{(j)}. \quad (4)$$

The reference satellite is chosen by a cluster-based algorithm proposed by Weng et al. [42]; its pseudo-range error $\varepsilon^{(j)}$ can be negligible compared to the satellite's pseudo-range error affected by multipath or NLOS. Therefore, by subtracting (2) and (3), the receiver clock offset can be eliminated as shown in (5). As discussed above the variables $\varepsilon_{\text{base}}^{(i)}$ and $\varepsilon_{\text{base}}^{(j)}$ can be neglected while $r^{(i)}$, $r^{(j)}$, $r_{\text{base}}^{(i)}$ and $r_{\text{base}}^{(j)}$ can be

TABLE I
LABEL CATEGORIZATION STRATEGY

Pseudo-range error(m)	Label
0-4	1
4-10	2
10-40	3
≥ 40	4

calculated by (3). So the pseudo-range error $\varepsilon^{(i)}$ can be get by (6)

$$\begin{aligned} \delta \rho^{(i)} - \delta \rho^{(j)} &= r^{(i)} - r_{\text{base}}^{(i)} + \varepsilon^{(i)} \\ &\quad - \varepsilon_{\text{base}}^{(i)} - r^{(j)} - r_{\text{base}}^{(j)} + \varepsilon^{(j)} - \varepsilon_{\text{base}}^{(j)} \end{aligned} \quad (5)$$

$$\varepsilon^{(i)} = \delta \rho^{(i)} - \delta \rho^{(j)} - r^{(i)} + r_{\text{base}}^{(i)} + r^{(j)} - r_{\text{base}}^{(j)}. \quad (6)$$

According to the typical ranges of the pseudo-range errors in different scenarios, we classify the pseudo-range errors into four categories as shown in Table I. The classifications are the labels for machine learning.

Based on experience, the pseudo-range errors of direct signals are generally within 4 m. Therefore, errors between 0 and 4 m are classified into the first category, and the labels are uniformly set as 1. When the satellite signals are affected by short multipath, their pseudo-range errors are almost between 4 and 10 m, and the errors within this range are classified as the second category, with a uniform label of 2. When satellite signals suffer from long multipath or NLOS, the pseudo-range errors are generally within 10–40 m, which are classified as the third category and labeled as 3. The pseudo-range errors of some NLOS signals or close to lose lock signals may exceed 40 m, which are classified as the fourth category and labeled as 4.

B. Feature Extraction

The GNSS/INS-integrated navigation system can output not only raw data, including GNSS C/N₀, elevation, pseudo-range, and Doppler for each satellite, and IMU raw data, but also GNSS position, INS position, and integrated position. By pre-processing these data, four types of information related to GNSS pseudo-range errors can be extracted: C/N₀, elevation, pseudo-range rate consistency, and estimated pseudo-range residual. This article utilizes these information as input features for the training set. The characteristics and extraction methods of the four features are described as follows.

- 1) *C/N₀*: C/N₀ is the ratio of carrier power and the noise power of the received signal in unit Hz bandwidth, expressed in dB·Hz. It represents signal quality and the extent of signal attenuation. A lower C/N₀ is generally associated with a signal subject to multipath or NLOS. The C/N₀ can somewhat reflect the quality of the pseudo-range measurement. It can be directly obtained from the GNSS receiver and is commonly adopted for weighting in GNSS positioning in GNSS challenged environments.
- 2) *Satellite Elevation, θ* : The satellite elevation θ refers to the angle between the satellite and the horizon at the

receiver's location. Signals of low-elevation satellites are more likely to be affected by high-rise buildings and cause multipath and NLOS compared with those from satellites at high elevations. The elevation can be directly obtained from the receiver and is often employed in open sky environments for GNSS positioning weighting [43].

- 3) *Pseudo-range Rate Consistency, ζ* : Pseudo-range rate consistency ζ refers to the consistency between delta pseudo-range and pseudo-range rate. Delta pseudo-range is the difference between the pseudo-range of the last epoch and that of the current epoch. Pseudo-range rate is the change of the pseudo-range calculated by the Doppler. In urban environments, compared with pseudo-range, Doppler error is smaller and can be employed to detect pseudo-range gross errors. Pseudo-range rate consistency is obtained by preprocessing data through the following formula:

$$\zeta = (\rho_i - \rho_{i-1}) - (-f_d \cdot \lambda \cdot T). \quad (7)$$

ρ_i is current epoch pseudo-range, λ denotes the wavelength of the signal, f_d is the Doppler, and T is the inter-epoch interval.

- 4) *Estimated Pseudo-range Residual, η* : Estimated pseudo-range residual η is obtained by the pseudo-range observations and the predicted pseudo-range calculated based on satellite ephemeris and vehicle position. The accuracy of the estimated pseudo-range residual is decided by the predicted pseudo-range, which is constrained by the vehicle position error. GNSS positioning result is utilized to calculate the predicted pseudo-range in existing literature [39]. In this article, the predicted pseudo-range is calculated by the INS predicted position for the GNSS/INS-integrated navigation system.

From (1), we know that if $r^{(i)}$, $\delta t^{(i)}$, $I^{(i)}$ and $T^{(i)}$ are all known, the pseudo-range error $\epsilon^{(i)}$ can be calculated. But we can't calculate the precise position, we can only get the estimated pseudo-range residual $\eta^{(i)}$ with the effect of positioning by the following equation:

$$\eta^{(i)} = \rho^{(i)} - r^{(i)} - \delta t_u + \delta t^{(i)} - I^{(i)} - T^{(i)}. \quad (8)$$

The satellite-vehicle range r is obtained by (3), here $\mathbf{x} = (x, y, z)$ represents the vehicle position obtained by GNSS SPP or GNSS/INS-integrated navigation. As we do not have a base station, the satellite clock offset can only be obtained by satellite ephemeris

$$\delta t^{(i)} = \Delta t^{(i)} + \Delta t_r - T_{GD}. \quad (9)$$

In which, $\Delta t^{(i)}$ is the satellite clock offset given by the first data block of the satellite navigation message. Δt_r is the relativistic correction. T_{GD} is group delay correction. the receiver clock offset is eliminated through satellite difference. Selecting satellite j as the reference. After removing the satellite-vehicle range and satellite clock offset from the observations, the residual information of the two pseudo-ranges can be expressed as

$$\begin{cases} \rho_c^{(i)} = \delta t_u + I^{(i)} + T^{(i)} + \eta^{(i)} \\ \rho_c^{(j)} = \delta t_u + I^{(j)} + T^{(j)} + \eta^{(j)}. \end{cases} \quad (10)$$

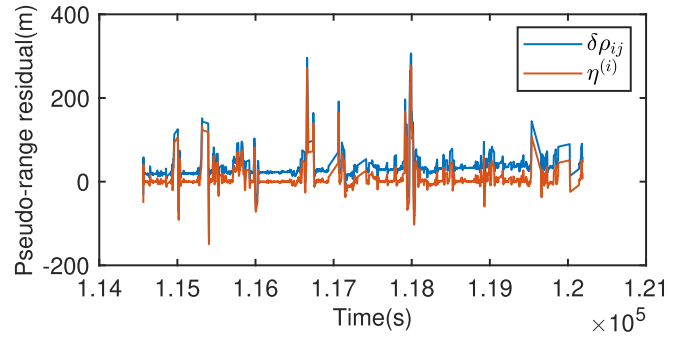


Fig. 2. Pseudo-range residual before and after removing the ionosphere and troposphere effects.

As mentioned in the labeling process section, the reference satellite's pseudo-range residual can be negligible. Therefore, by subtracting the above two equations, the receiver clock offset can be eliminated, and the equation can be written as follows:

$$\begin{aligned} \eta^{(i)} &= \delta \rho_{ij} - \delta IT \\ \delta IT &= I^{(i)} - I^{(j)} + T^{(i)} - T^{(j)}. \end{aligned} \quad (11)$$

$\delta \rho_{ij}$ includes not only pseudo-range residual caused by environments but also the impact of ionosphere and troposphere on both the reference satellite and satellite i . $\eta^{(i)}$ can be obtained once removing the effects of the ionospheric and tropospheric. Under vehicle dynamic conditions, compared to the GNSS pseudo-range residual $\eta^{(i)}$, the ionospheric and tropospheric delay can be regarded as slowly varying terms. Therefore, the δIT can be estimated by robust locally weighted scatterplot smoothing (RLOWESS) filtering to $\delta \rho_{ij}$. $\eta^{(i)}$ can be obtained by subtracting the estimated δIT from $\delta \rho_{ij}$. Fig. 2 presents the $\delta \rho_{ij}$ and the $\eta^{(i)}$ in urban vehicle environments. The fluctuation consistency of the two curves shows that δIT can be estimated and removed through RLOWESS filtering.

- a) *Estimated Pseudo-Range Residual Based on GNSS, η_{GNSS}* : By utilizing the GNSS positioning result and satellite position, the pseudo-range residual caused by environments can be estimated. The satellite-vehicle range is obtained by the GNSS position of the target system and the satellite, expressed as

$$r_{GNSS}^{(n)} = \|\mathbf{x}^{(n)} - \mathbf{x}_{GNSS}\|. \quad (12)$$

Here, $\mathbf{x}^{(n)} = (x^{(n)}, y^{(n)}, z^{(n)})$ is the satellite position calculated by satellite ephemeris, while $\mathbf{x}_{GNSS} = (x_{GNSS}, y_{GNSS}, z_{GNSS})$ is the vehicle position obtained through GNSS positioning. Other processing is as explained above. Since the effect of the satellite clock, the receiver clock, δIT , and the observation noise of the reference satellite on the pseudo-range residual estimation can be ignored, the error of GNSS positioning is the main factor. In urban environments, the available satellites are limited, and most of them are affected by

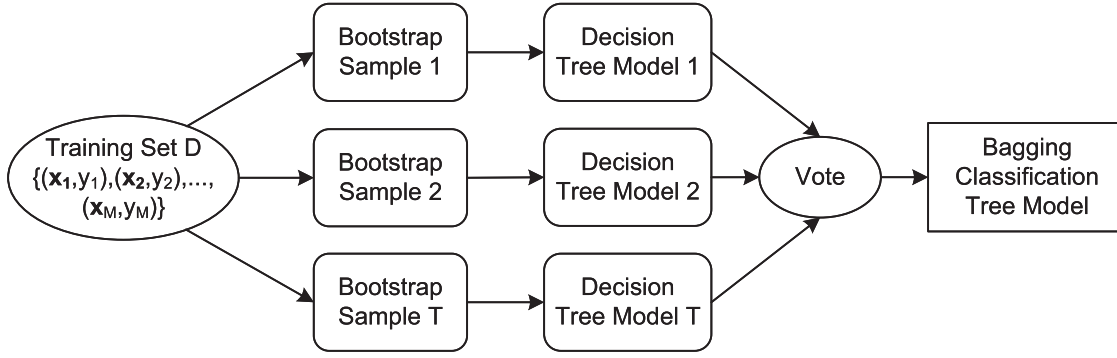


Fig. 3. Flowchart of the ensemble bagging classification tree algorithm.

multipath and NLOS, leading to severe degradation of GNSS positioning accuracy. Therefore, η_{GNSS} cannot accurately reflect the pseudo-range error in GNSS challenged scenarios.

- b) *Estimated Pseudo-Range Residual Based on INS*, η_{INS} : In GNSS/INS-integrated navigation systems, INS-predicted position can be utilized to calculate the satellite-to-vehicle range. INS-predicted position refers to the position calculated by INS mechanization based on the previous epoch GNSS/INS-integrated result. Therefore, the satellite-vehicle range $r_{\text{INS}}^{(n)}$ based on INS-predicted positions can be expressed as

$$r_{\text{INS}}^{(n)} = \|\mathbf{x}^{(n)} - \mathbf{x}_{\text{INS}}\|. \quad (13)$$

$\mathbf{x}_{\text{INS}} = (x_{\text{INS}}, y_{\text{INS}}, z_{\text{INS}})$ is the predicted position by INS. Another calculation process for the pseudo-range residual η_{INS} is the same as η_{GNSS} . Since the INS predicted position is not affected by the GNSS of the current epoch, the estimated pseudo-range residual based on INS can be more accurate than η_{GNSS} in GNSS-challenged environments.

To verify the effect of INS aiding, this article will carry out training using η_{GNSS} and η_{INS} , respectively, and generate two models for performance comparison. The training set that uses η_{GNSS} is referred to as the GNSS-based training set, while the corresponding learning model is termed the GNSS-based bagging tree model (GNSS-BT). Similarly, the training set based on η_{INS} is called the INS-based training set, and the corresponding model is named the INS-based bagging tree model (INS-BT).

IV. BAGGING TREE CLASSIFIER

The decision tree is a widely employed classification algorithm in machine learning that effectively segments data to enhance the clarity of complex datasets [43]. However, a single decision tree is prone to overfitting, resulting in low accuracy, and ensemble methods are generally utilized to combine multiple decision trees to improve prediction performance [44], [45]. The bagging algorithm is one ensemble method that trains decision trees using a sampling approach with replacement. By introducing sample perturbations and

increasing the randomness, it effectively reduces variance [46]. Therefore, the bagging tree classification model performs well on training sets that are easily disturbed by samples, making it suitable for predicting and classifying GNSS pseudo-range errors in urban environments.

The bagging classification tree is based on the principle illustrated in Fig. 3. We can obtain a training set $D = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_i, y_i)\}$ by combining the features with their corresponding labels, where $\mathbf{x}_i = (C/N_{0i}, \theta_i, \zeta_i, \eta_i)$, y_i is the classified label of pseudo-range error corresponding to \mathbf{x}_i , and $i = 1, 2, 3, \dots, M$, is the sample index, M is the total number of input samples. The input is the training set D , and multiple sample sets are obtained using the bootstrap sampling method. For each sample set, a decision tree is trained as the base learner. Subsequently, these base learners are combined by a simple voting method to obtain the final classification model [47], [48], [49]. The specific implementation process is as Fig. 3.

- 1) *Bootstrap Sampling*: Given a dataset $D = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_M, y_M)\}$ containing M samples, the bootstrap sampling method randomly selects one sample to be included in a sample set. The selected sample is then returned to the initial dataset, allowing it to be chosen again in subsequent sampling rounds. After T rounds of random sampling, we obtain T sample sets, each containing M samples, denoted as $D_t, t = 1, 2, 3, \dots, T$.
- 2) *Single Decision Tree Classification Model*: The process of generating a classification decision tree for each subsampled training set D_t is as follows.

- a) Computing the Gini index for the input features $\{C/N_0, \theta, \zeta, \eta\}$ on the dataset D_t . For each feature A , we calculate the Gini index for $A = a$ for each possible value a . The Gini index for a given sample set D_t can be defined as follows:

$$\text{Gini}(D_t) = 1 - \sum_{k=1}^K \left(\frac{|C_k|}{|D_t|} \right)^2. \quad (14)$$

C_k represents the subset of samples that belong to the k th class, and k is the number of classes. For pseudo-range error training, k is 4. According to whether the feature A can take the value a , D_t is

divided into two parts, D_1 and D_2

$$\begin{aligned} D_1 &= \{(x, y) \in D_t | A(x) = a\} \\ D_2 &= D_t - D_1. \end{aligned} \quad (15)$$

Under the condition of feature A , the Gini index of set D_t is defined as

$$\begin{aligned} \text{Gini}(D_t, A) &= \frac{|D_1|}{|D_t|} \text{Gini}(D_1) \\ &+ \frac{|D_2|}{|D_t|} \text{Gini}(D_2). \end{aligned} \quad (16)$$

- b) Among all possible features A and their corresponding splitting point a , the feature with the smallest Gini index and its corresponding splitting point are selected as the optimal feature and optimal split point. Using the optimal feature and split point, two child nodes are created from the parent node and the training set is assigned to the child nodes based on their features.
 - c) Recursively applying steps a) and b) to the two child nodes until either the number of samples in the node is less than the predetermined threshold or the Gini index of the samples is less than the threshold.
 - d) Once the stopping criteria are met, a decision tree model for the sub-training set D_t is generated and denoted as $h_t(x)$.
- 3) *Bagging Tree Classifier*: After obtaining all base learners, they are combined by a voting method, represented as

$$H(x) = \arg \max_{y \in K} \sum_{t=1}^T I(h_t(x) = y). \quad (17)$$

The classification model $H(x)$ is obtained through the above steps. Its inputs are the four features C/N_0 , satellite elevation θ , pseudo-range rate consistency ζ , and estimated pseudo-range residual η . The outputs are pseudo-range error prediction results and scores for each category. A higher score indicates a higher likelihood of predicting that particular category. When a new set of data $x = \{C/N_0, \theta, \zeta, \eta\}$ is available, the model can generate the corresponding predicted pseudo-range error.

Only four classification results may lead to pseudo-range errors with large differences corresponding to the same weight for GNSS SPP. Therefore, the output scores of the machine learning model are employed for more refined weighting. The weight of the pseudo-range can be obtained by multiplying the median of the error range represented for each category and the output score. Assuming the model output scores are $S = [S_1 \ S_2 \ S_3 \ S_4]$, the pseudo-range weight P for GNSS SPP is given by

$$P = \frac{1}{S_1 \cdot 2 + S_2 \cdot 7 + S_3 \cdot 25 + S_4 \cdot 60}. \quad (18)$$

The above weighting scheme can avoid the unreasonable weights caused by not discrete enough classification, especially for pseudo-ranges at the boundary of two classes. It can compensate for the losses caused by misclassification simultaneously.

TABLE II
VEHICLE TESTING RAW DATASETS

	Location	Sampling duration
group 1	Wuhan University	47min31s
group 2	Wuhan University	46min17s
group 3	Wuhan University	53min
group 4	Greenland Center	56min3s
group 5	Greenland Center	44min15s
group 6	Shahu-Donghu	75min3s
group 7	Shahu-Donghu	99min50s
group 8	Moshan- Shahu	62min29s
group 9	Moshan- Shahu	66min41s

TABLE III
SENSOR PARAMETERS OF LEADER PPOI-A15

Accelerometer	Bias instability	15mGal
	Random walk noise	0.03 m/s/sqrt(hr)
Gyroscope	Bias instability	0.027deg/hr
	Random walk noise	0.003 deg/sqrt(hr)

TABLE IV
SENSOR PARAMETERS OF ADIS16465

Accelerometer	Bias instability	50mGal
	Random walk noise	0.1 m/s/sqrt(hr)
Gyroscope	Bias instability	50deg/hr
	Random walk noise	0.1 deg/sqrt(hr)

V. FIELD TEST AND RESULTS ANALYSIS

A. Experiment Setup

To ensure the accuracy and generalization of the model, we collected nine sets of raw data based on vehicle testing in typical urban areas. Table II provides the location and duration of each group of raw data. The total duration of the above tests exceeds 10 h, including open sky, high-rise buildings, forests, elevated roads, and tunnels. Groups 1–3 are typical campus environments with buildings and trees. Groups 4–5 represent dense high-rise environments. Groups 6–7 are typical urban road environments, including elevated roads, high-rise buildings, and tunnels. Groups 8–9 encompass both mountainous and urban environments, including tree-lined roads, tunnels, and high-rise buildings.

The vehicle tests were equipped with a ground reference device and a collected device of datasets, which shared the same GNSS antenna as shown in Fig. 4. The ground reference device was a high-precision GNSS/INS-integrated navigation system—Leader PPOI-A15, which is composed of a surveying GNSS receiver and a ring laser gyro IMU. The corresponding IMU parameters are given in Table III. The centimeter-level positions can be provided by post-processing the data of PPOI-A15, which meets the reference requirement.

The raw data collected device used a low-cost GNSS/INS-integrated navigation system INS-Probe, which was independently developed by ourselves. It has a GNSS receiver module u-blox M8P and a MEMS inertial sensor ADIS16465. The corresponding IMU parameters are given in Table IV.



Fig. 4. Experimental equipment.

INS-Probe collects raw data such as GNSS C/N_0 , elevation, pseudo-range, Doppler from GPS L1 and BDS B1 satellites, and raw data from the inertial sensor. We used the GNSS positioning software RTKLIB to calculate GNSS positioning results. Our team's open-source GNSS/INS-integrated navigation software KF-GINS was utilized to calculate the inertial predicted position and the integrated navigation results.

B. Classification Results

The training and testing sets were generated by performing feature extraction and labeling on the nine raw datasets. Two sets were obtained from one raw dataset based on the two pseudo-range residual estimating modes discussed above. These sets were then employed to train or test the GNSS-BT model and the INS-BT model, respectively. It should be noted that open-field data accounts for a higher proportion compared to other scenarios, and its proportion was appropriately reduced in the data preprocessing stage. When a dataset was selected for testing, the other 8 datasets were for model training. Different datasets were selected as testing sets to verify the model's generalization ability.

This article adopts two indicators, prediction accuracy (η_{Acc}) and jump class ratio (η_{JC}), to evaluate the model's performance. The accuracy represents the percentage of correctly predicted instances in the test set, which reflects the overall classification performance of the model. However, when the test set has a severe class imbalance, the overall accuracy may not accurately reflect the model's performance. In such cases, it is necessary to further compare the accuracy of each class. So we use the $\eta_{Acc-all}$ to represent the overall accuracy and η_{Acc-ec} to represent the accuracy of each class.

The jump class ratio refers to the proportion of data where the predicted class differs from the true class by two or more levels. Assume N_{ij} represents the number of instances where the actual error belongs to i th class and the model predicts it as the j th class, and N_{total} represents the total number of instances in the dataset. Since we only classify four categories, i and j are maximized to 4. The expression of the jump class ratio is

$$\eta_{JC} = \frac{N_{13} + N_{14} + N_{24} + N_{31} + N_{41} + N_{42}}{N_{Total}} * 100\%. \quad (19)$$

The misclassification between adjacent categories causes a relatively small deviation between the predicted category and

TABLE V
MODEL CLASSIFICATION PERFORMANCE

		INS-BT				GNSS-BT			
$\eta_{Acc-all} \text{ (%)}$		88.70				82.80			
group 1		1	2	3	4	1	2	3	4
	$\eta_{Acc-ec}(\%)$	89.30	81.70	80.90	86.90	84.00	69.60	62.60	74.30
	$\eta_{JC} \text{ (%)}$	1.45				2.17			
$\eta_{Acc-all} \text{ (%)}$		87.50				87.60			
group 6		1	2	3	4	1	2	3	4
	$\eta_{Acc-ec} \text{ (%)}$	88.50	67.50	74.00	71.90	89.30	59.60	57.90	51.50
	$\eta_{JC} \text{ (%)}$	1.97				2.35			
$\eta_{Acc-all} \text{ (%)}$		92.20				86.90			
group 8		1	2	3	4	1	2	3	4
	$\eta_{Acc-ec} \text{ (%)}$	92.90	69.40	83.60	95.60	88.10	55.60	61.80	73.50
	$\eta_{JC} \text{ (%)}$	0.74				1.81			

the truth, while a jump in classification results in a significantly unreliable estimate. Therefore, a higher jump class ratio means a higher probability of large deviations between the predicted and true pseudo-range errors, unfavorable for the weight allocation in GNSS positioning.

Three datasets (groups 1, 6, and 8) collected from different typical environments were selected as test sets, and the results are shown in Table V. While the jump class ratio of GNSS-BT and INS-BT are less than 2.5%. It indicates that both models have strong usability and will assign little unreasonable weights to GNSS pseudo-ranges. In addition, the η_{JC} of the INS-BT model is lower than that of the GNSS-BT model. While the overall accuracy of the GNSS-BT model is higher than 82% in the three scenarios, the accuracy of the INS-BT model exceeds 87%. As mentioned before, the overall accuracy cannot accurately reflect the model's performance for a testing dataset where one class of data has a significantly higher proportion. For example, the overall accuracy of the two models is similar in group 6 data. When comparing the accuracy of each class, it is found that GNSS-BT can only estimate class 1 (small pseudo-range errors) accurately, the accuracy of other classes is significantly worse than that of INS-BT. This is because a significant proportion of the data in the test set belongs to class 1. The results of the three scenarios all show that INS-BT can get high overall accuracy and for



Fig. 5. Overall test route.

each category accuracy simultaneously. The INS-BT model predicts the pseudo-range errors more accurately, especially for large pseudo-range errors that significantly negatively impact positioning.

The above results demonstrate that both models accurately predict pseudo-range errors and have strong generalization abilities. INS-BT outperforms GNSS-BT in classifying large pseudo-range errors, which is more valuable for improving the positioning performance in challenged environments. Next, the error-predicted results will be employed to assign weights to pseudo-ranges for positioning, verifying the effectiveness of the models.

C. Positioning Performance

Data from group 6 was collected in typical urban road environments, including open areas, high-rise buildings, overpasses, and tree-lined roads, and the test route is shown in Fig. 5. The GNSS pseudo-range errors for this route are shown in Fig. 6, which shows that there are many parts of the test route where the GNSS observations deteriorate. To fully validate the effectiveness of the proposed pseudo-range error prediction method, the classification results were employed to allocate weights to the pseudo-range observations in GNSS positioning. Furthermore, the performance of GNSS/INS-integrated navigation was tested based on the above GNSS positioning. The performance of the model was evaluated by comparing the positioning performance of the traditional, GNSS-BT and INS-BT weighting methods. As the traditional weighting method, the weighting model of elevation and signal-to-noise ratio from reference [50] was adopted. The GNSS-BT based positioning employed the predicted pseudo-range errors from the GNSS-BT to provide weights to the weighted least-squares estimator in GNSS positioning. The INS-BT-based positioning assigned weights to the pseudo-range observations in the positioning based on the predicted pseudo-range errors from the INS-BT.

1) *Vehicle Positioning Performance Analysis for Urban Environments*: Accuracy statistics of GNSS SPP with different weighting methods are shown in Table VI. A comparison with GNSS positioning with traditional weighting (Tra-Wei POS) reveals that, both GNSS positioning with INS-BT weighting (INS-BT POS) and that with GNSS-BT weighting (GNSS-BT POS) significantly improve the positioning accuracy. GNSS-BT and INS-BT improve the horizontal position accuracy

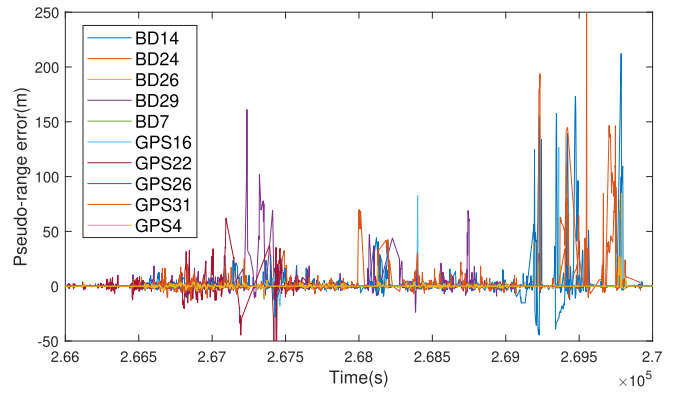


Fig. 6. Pseudo-range errors in test route.

TABLE VI
ACCURACY STATISTICS OF GNSS SPP WITH
DIFFERENT WEIGHTING METHODS

Method	Direction	CEP95(m)	RMS(m)	STD(m)
Tra-Wei POS	East	8.37	5.40	4.52
	North	9.68	5.64	4.59
	Up	26.42	15.17	12.99
GNSS-BT POS	East	5.20	3.50	2.71
	North	6.61	4.03	3.04
	Up	10.88	8.48	6.16
INS-BT POS	East	4.08	2.32	1.35
	North	5.04	2.86	1.71
	Up	9.69	5.68	2.92

in the form of rms by 31.63% and 52.75%, respectively. It indicates that machine learning models predict pseudo-range errors more accurately, leading to more reasonable weights and superior GNSS positioning accuracy. Meanwhile, the INS-BT can improve the horizontal position accuracy in the form of CEP95 by 52.81%, and in the form of STD by 70.32%, respectively.

Compared with the GNSS-BT, the INS-BT shows superior horizontal positioning accuracy in terms of the CEP95. Meanwhile, compared with GNSS-BT, the rms and STD of INS-BT improved by 30.90% and 52.33%, respectively. The improvement in GNSS positioning accuracy indicates that the pseudo-range errors predicted by INS-BT are more accurate, and can mitigate the negative impacts of large pseudo-range errors on positioning. Therefore, the INS-BT can assess the pseudo-range quality more accurately than GNSS-BT and traditional methods. The improvement of GNSS positioning accuracy demonstrates the advantage of INS-BT in estimating the pseudo-range errors. However, compared to GNSS-BT, INS-BT needs INS information, and it is unfair to evaluate their effects on positioning via GNSS positioning. It is necessary to compare the GNSS/INS-integrated positioning results, which are obtained by fusing INS and above GNSS positioning results, respectively.

The statistics of the GNSS/INS integration results are shown in Table VII. The GNSS/INS-integrated navigation results

TABLE VII
ACCURACY STATISTICS OF GNSS/INS INTEGRATION RESULTS
BASED ON DIFFERENT GNSS POSITIONING

Method	Direction	CEP95(m)	RMS(m)	STD(m)
Tra-Wei IPOS	East	6.09	3.34	2.48
	North	7.98	4.31	3.30
	Up	10.85	6.28	3.80
GNSS-BT IPOS	East	4.57	2.34	1.43
	North	5.96	3.36	2.24
	Up	9.13	5.26	2.62
INS-BT IPOS	East	3.90	2.15	1.27
	North	4.98	2.73	1.61
	Up	9.16	5.35	2.69



Fig. 7. Challenged road map.

demonstrate that the performance of integrated positioning based on INS-BT weighting (INS-BT IPOS) surpasses that based on GNSS-BT weighting (GNSS-BT IPOS) and based on traditional weighting (Tra-Wei IPOS). In terms of horizontal direction, compared with the Tra-Wei IPOS, the INS-BT-assisted integrated positioning accuracy in the form of CEP95 has an improvement of 37.85%, and the rms and STD also improve by 36.15% and 55.93%, respectively. Compared with the GNSS-BT IPOS, the horizontal positioning accuracy in the form of CEP95, rms, and STD of the INS-BT IPOS improves by 18.03%, 14.91%, and 28.15%, respectively. It indicates that accurately estimating GNSS pseudo-range error can effectively improve the accuracy of GNSS/INS-integrated navigation. The more challenging the environments, the more important it is to accurately estimate pseudo-range errors. Next, the positioning performance of challenged scenarios will be analyzed.

2) GNSS/INS Positioning Performance of Challenged Scenarios: The challenged roads in the Group6 test is shown in Fig. 7, where the elevated road (Part1) and high-rise building road (Part2) are highlighted in circles.

The GNSS/INS-integrated positioning error in the horizontal direction is shown in Fig. 8. With the assistance of INS-BT, the positioning error (CEP95) is reduced from 21.23 to 5.02 m, which shows an accuracy improvement of 76.35%. Compared with the GNSS-BT method, the accuracy also improves by 52.24% in the form of CEP95. This indicates that INS-BT can significantly improve the integrated positioning accuracy

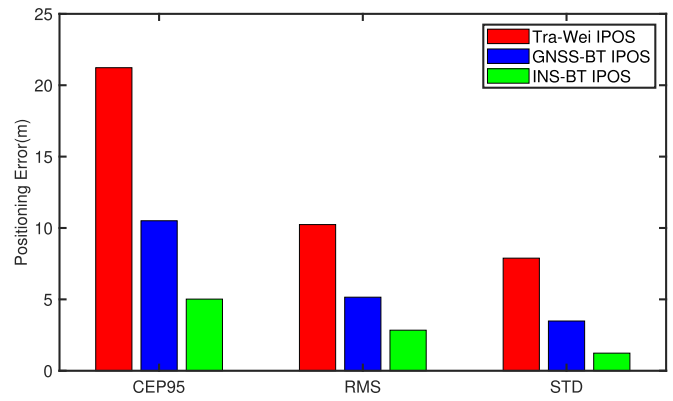


Fig. 8. Horizontal-integrated positioning error.

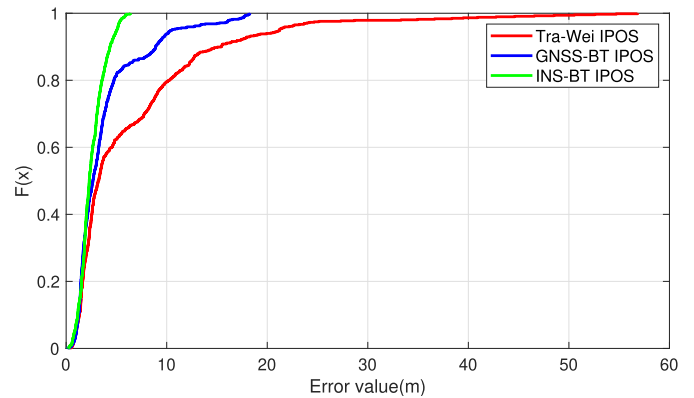


Fig. 9. CDF curve of integrated positioning errors on challenged roads.

in challenged environments. Additionally, compared with the traditional method, the rms of the positioning errors was reduced by 72.27%, and the STD was reduced by 84.41%. This demonstrates that INS-BT predicts large pseudo-range errors more effectively. When the large pseudo-range errors are estimated accurately, the corresponding pseudo-range observations can be assigned to reasonable weights, which helps improve positioning accuracy.

The cumulative distribution function (cdf) curve of the integrated positioning errors on the challenged roads is shown in Fig. 9. The horizontal axis of the cdf figure represents the positioning error in meters, and the vertical axis is the probability of points with errors smaller than the corresponding horizontal axis value. The cdf curve provides a visual representation of the distribution of positioning errors. The faster the value of the vertical axis reaches 1, the smaller the overall error of the curve. As can be seen from the figure, the INS-BT IPOS curve has the largest value on the vertical axis at the same value of error on the horizontal axis, indicating that its error is concentrated in a smaller range. With INS-BT, 94.89% of horizontal positioning errors are within 5 m. Comparatively, with GNSS-BT assistance, the proportion is 81.35%, and with the traditional method, only 62.56% is below 5 m. While only 79.55% errors of the traditional method are within 10 m, the overall horizontal positioning errors of the INS-BT aided method are less than 10 m. The improvement in positioning accuracy with INS-BT can be visualized very well through the cdf curve.

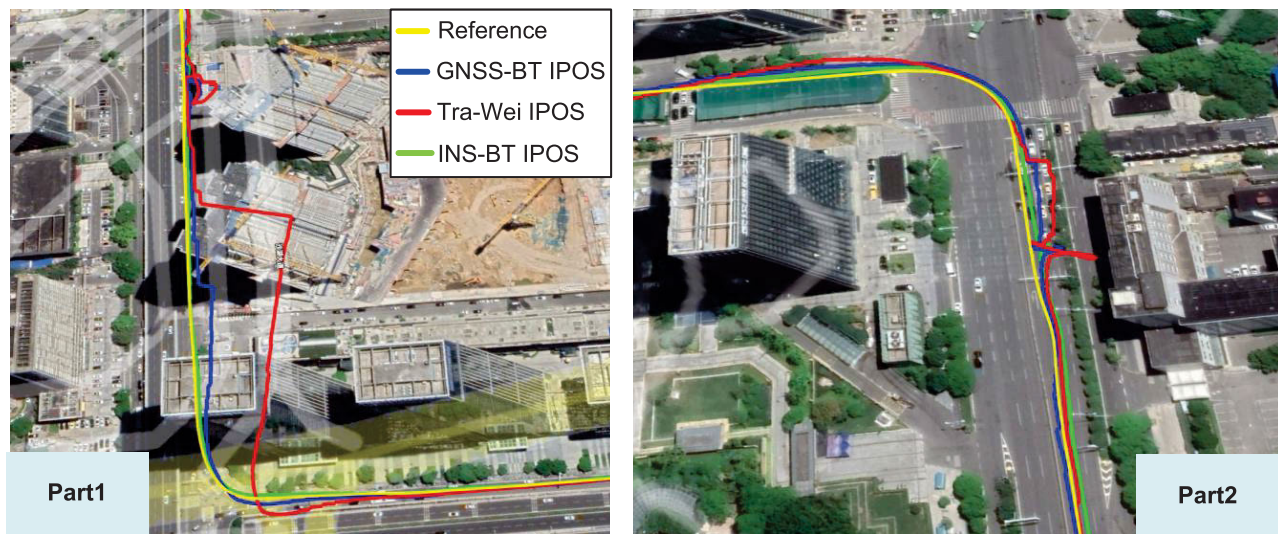


Fig. 10. GNSS/INS positioning results of typical scenarios.

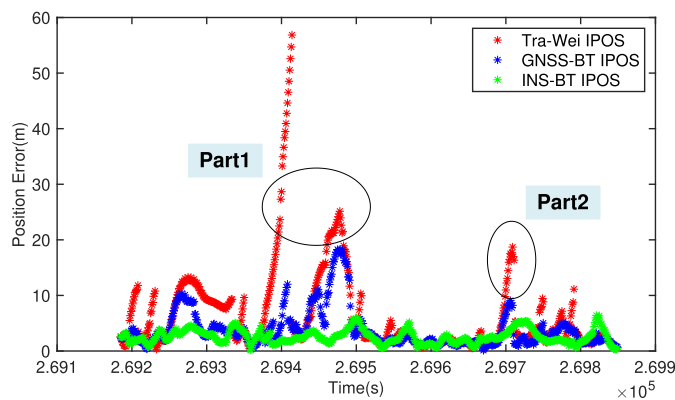


Fig. 11. Horizontal-integrated positioning errors curves on challenged roads.

The GNSS/INS positioning results of typical scenarios for Tra-Wei IPOS, GNSS-BT IPOS, and INS-BT IPOS are depicted in Fig. 10. The horizontal positioning error curve is shown in Fig. 11. The two parts with large positioning errors indicated in the figure correspond to the typical cases mentioned in Fig. 7. Fig. 6 shows that the pseudo-range errors are relatively large in these typical scenarios. Part 1 is an elevated road, and the live scenario is shown on the left of Fig. 10. Its left side is the viaduct, and high-rise buildings block the right side. There are a lot of NLOS satellites, resulting in large positioning errors for the traditional method. With INS-BT assistance, the positioning errors are kept small. While the positioning trajectories of the other two methods have already exhibited significant deviations compared to the reference trajectory, the positioning with INS-BT is almost unaffected. Part 2 corresponds to a scenario where both sides are obstructed by high-rise buildings. The positioning trajectory is shown on the right of Fig. 10. The trajectory of positioning with INS-BT assistance also has the best consistency with the reference trajectory. The above results of challenged roads demonstrate that INS-BT can accurately predict pseudo-range errors, and provide more reasonable weighting to improve positioning performance.

VI. CONCLUSION

This article proposes an INS-assisted GNSS pseudo-range error prediction method based on machine learning. It is intended to address the problem of inaccurate estimation of GNSS pseudo-range errors, which seriously affect the availability of GNSS in urban challenged environments. It utilizes pseudo-range residual estimated by INS as an important input feature, together with GNSS C/N_0 , satellite elevation, and pseudo-range rate consistency. A bagging decision tree learning method is employed to train the model. By evaluating the model with multiple sets of typical urban vehicle data, the tests demonstrate that the model achieves a high accuracy rate of over 87.00% in predicting pseudo-range errors. When applied to weight GNSS pseudo-range observations, the model significantly improves the horizontal positioning accuracy in the form of CEP95 by 52.81% compared to the conventional weighting method. Meanwhile, the horizontal error (CEP95) of the GNSS/INS-integrated navigation in deep urban areas is reduced from 21.23 to 5.02 m. The proposed GNSS pseudo-range error prediction method can significantly improve the performance of GNSS pseudo-range error estimation in urban vehicle environments. It is not only beneficial for positioning accuracy in GNSS-challenged environments, but also provides more accurate GNSS credibility information for unmanned systems.

REFERENCES

- [1] S.-H. Kong, "Statistical analysis of urban GPS multipaths and pseudo-range measurement errors," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 47, no. 2, pp. 1101–1113, Apr. 2011.
- [2] J. M. Tranquilla, J. P. Carr, and H. M. Al-Rizzo, "Analysis of a choke ring groundplane for multipath control in global positioning system (GPS) applications," *IEEE Trans. Antennas Propag.*, vol. 42, no. 7, pp. 905–911, Jul. 1994.
- [3] S. Daneshmand, A. Broumandan, N. Sokhandan, and G. Lachapelle, "GNSS multipath mitigation with a moving antenna array," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 49, no. 1, pp. 693–698, Jan. 2013.
- [4] Z. Jiang and P. D. Groves, "NLOS GPS signal detection using a dual-polarisation antenna," *GPS Solutions*, vol. 18, no. 1, pp. 15–26, Jan. 2014.

- [5] T. Suzuki, K. Matsuo, and Y. Amano, "Rotating GNSS antennas: Simultaneous LOS and NLOS multipath mitigation," *GPS Solutions*, vol. 24, no. 3, pp. 1–13, Jul. 2020.
- [6] A. J. Van Dierendonck, P. Fenton, and T. Ford, "Theory and performance of narrow correlator spacing in a GPS receiver," *Navigation*, vol. 39, no. 3, pp. 265–283, Sep. 1992.
- [7] M. S. Braasch, "Performance comparison of multipath mitigating receiver architectures," in *Proc. IEEE Aerosp. Conf.*, Big Sky, MT, USA, Mar. 2001, pp. 3–1309.
- [8] X. Chen, F. Dovis, S. Peng, and Y. Morton, "Comparative studies of GPS multipath mitigation methods performance," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 49, no. 3, pp. 1555–1568, Jul. 2013.
- [9] M. Z. H. Bhuiyan, E. S. Lohan, and M. Renfors, "Code tracking algorithms for mitigating multipath effects in fading channels for satellite-based positioning," *EURASIP J. Adv. Signal Process.*, vol. 2008, no. 1, Dec. 2007, Art. no. 863629.
- [10] R. Yozevitch, B. Ben Moshe, and H. Levy, "Breaking the 1 meter accuracy bound in commercial GNSS devices," in *Proc. IEEE 27th Conv. Electr. Electron. Eng. Isr.*, Nov. 2012, pp. 1–5.
- [11] P. D. Groves and Z. Jiang, "Height aiding, C/N0 weighting and consistency checking for GNSS NLOS and multipath mitigation in urban areas," *J. Navigat.*, vol. 66, no. 5, pp. 653–669, Sep. 2013.
- [12] A. R. Amiri-Simkooei, P. J. G. Teunissen, and C. C. J. M. Tiberius, "Application of least-squares variance component estimation to GPS observables," *J. Surveying Eng.*, vol. 135, no. 4, pp. 149–160, Nov. 2009.
- [13] H.-J. Euler and C. C. Goad, "On optimal filtering of GPS dual frequency observations without using orbit information," *Bull. Géodésique*, vol. 65, no. 2, pp. 130–143, Jun. 1991.
- [14] J. Marais, M. Berbineau, and M. Heddebaut, "Land mobile GNSS availability and multipath evaluation tool," *IEEE Trans. Veh. Technol.*, vol. 54, no. 5, pp. 1697–1704, Sep. 2005.
- [15] T. Suzuki and N. Kubo, "N-los gnss signal detection using fish-eye camera for vehicle navigation in urban environments," in *Proc. 27th Int. Tech. Meeting Satell. Division Inst. Navigat. (ION GNSS+)*, Sep. 2014, pp. 1897–1906.
- [16] J. Moreau, S. Ambellouis, and Y. Ruichek, "Fisheye-based method for GPS localization improvement in unknown semi-obstructed areas," *Sensors*, vol. 17, no. 12, p. 119, Jan. 2017.
- [17] T. Suzuki, M. Kitamura, Y. Amano, and T. Hashizume, "Multipath mitigation using omnidirectional infrared camera for tightly coupled GPS/INS integration in urban environments," in *Proc. 24th Int. Tech. Meeting Satell. Division Inst. Navigat. (ION GNSS)*, Sep. 2011, pp. 2914–2922.
- [18] D. Maier and A. Kleiner, "Improved GPS sensor model for mobile robots in urban terrain," in *Proc. IEEE Int. Conf. Robot. Autom.*, May 2010, pp. 4385–4390.
- [19] W. Wen, G. Zhang, and L.-T. 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 *Proc. IEEE/ION Position, Location Navigat. Symp. (PLANS)*, Apr. 2018, pp. 158–165.
- [20] Xu, Jia, Luo, and Hsu, "Intelligent GPS II LOS/multipath/NLOS classifiers based on correlator-, RINEX- and NMEA-level measurements," *Remote Sens.*, vol. 11, no. 16, p. 1851, Aug. 2019.
- [21] P. D. Groves, "Shadow matching: A new GNSS positioning technique for urban canyons," *J. Navigat.*, vol. 64, no. 3, pp. 417–430, Jul. 2011.
- [22] W. Lei, P. D. Groves, and M. Ziebart, "Urban positioning on a smartphone: Real-time shadow matching using GNSS and 3D city models," in *Proc. ION GNSS*, Nashville, TN, USA, Sep. 2013, pp. 1606–1619. [Online]. Available: <http://www.ion.org/publications/abstract.cfm?jp=p&articleID=11339>
- [23] L.-T. Hsu, "Analysis and modeling GPS NLOS effect in highly urbanized area," *GPS Solutions*, vol. 22, no. 1, p. 7, Nov. 2017.
- [24] 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, Jul. 2016.
- [25] M. Obst, S. Bauer, P. Reisdorf, and G. Wanielik, "Multipath detection with 3D digital maps for robust multi-constellation GNSS/INS vehicle localization in urban areas," in *Proc. IEEE Intell. Vehicles Symp.*, Jun. 2012, pp. 184–190.
- [26] N. Kbayer and M. Sahmoudi, "Performances analysis of GNSS NLOS bias correction in urban environment using a three-dimensional city model and GNSS simulator," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 54, no. 4, pp. 1799–1814, Aug. 2018.
- [27] S. Peyraud et al., "About non-line-of-sight satellite detection and exclusion in a 3D map-aided localization algorithm," *Sensors*, vol. 13, no. 1, pp. 829–847, Jan. 2013.
- [28] *BeiDou navigation Satellite System Signal in Space Interface Control Document, Open Service Signal b2a (Version 1.0)*, China Satellite Navigation Office, document BDS-SIS-ICD-B2b-1.0, 2017.
- [29] T. Fan, T. Zhang, H. Zhang, J. Mo, and X. Niu, "A double sideband combined tracking method for Galileo e5 AltBOC signals," *Satell. Navigat.*, vol. 4, no. 1, p. 27, Dec. 2023.
- [30] S. Kim, S. Yoon, and S. Yong Kim, "A novel multipath mitigated side-peak cancellation scheme for BOC(kn, n) in GNSS," in *Proc. 9th Int. Conf. Adv. Commun. Technol.*, vol. 2, Feb. 2007, pp. 1258–1262.
- [31] Z. Yao, M. Lu, and Z. Feng, "Unambiguous technique for multiplexed binary offset carrier modulated signals tracking," *IEEE Signal Process. Lett.*, vol. 16, no. 7, pp. 608–611, Jul. 2009.
- [32] Y. Wang, W. Zhao, Y. Shen, and S. Yu, "Research on multipath suppression performance of multiple boc signals of navigation satellite," *IOP Conf. Series: Mater. Sci. Eng.*, vol. 793, no. 1, p. 012044, Mar. 2020.
- [33] Q.-H. Phan, S.-L. Tan, and I. McLoughlin, "GPS multipath mitigation: A nonlinear regression approach," *GPS Solutions*, vol. 17, no. 3, pp. 371–380, 2013.
- [34] R. Yozevitch, B. B. Moshe, and A. Weissman, "A RobustGNSS-LOS/NLOS signal classifier," *Navigation*, vol. 63, no. 4, pp. 429–442, Dec. 2016.
- [35] L.-T. Hsu, "GNSS multipath detection using a machine learning approach," in *Proc. IEEE 20th Int. Conf. Intell. Transp. Syst. (ITSC)*, Oct. 2017, pp. 1–6.
- [36] R. Sun, L.-T. Hsu, D. Xue, G. Zhang, and W. Y. Ochieng, "GPS signal reception classification using adaptive neuro-fuzzy inference system," *J. Navigat.*, vol. 72, no. 3, pp. 685–701, May 2019.
- [37] A. A. Abdallah and Z. M. Kassas, "Deep learning-aided spatial discrimination for multipath mitigation," in *Proc. IEEE/ION Position, Location Navigat. Symp. (PLANS)*, Portland, OR, USA, Apr. 2020, pp. 1324–1335.
- [38] D. Min, M. Kim, J. Lee, and J. Lee, "Deep neural network based multipath mitigation method for carrier based differential GNSS systems," in *Proc. ION Pacific PNT*, May 2019, pp. 451–466.
- [39] R. Sun et al., "Improving GPS code phase positioning accuracy in urban environments using machine learning," *IEEE Internet Things J.*, vol. 8, no. 8, pp. 7065–7078, Apr. 2021.
- [40] G. Zhang, P. Xu, H. Xu, and L.-T. Hsu, "Prediction on the urban GNSS measurement uncertainty based on deep learning networks with long short-term memory," *IEEE Sensors J.*, vol. 21, no. 18, pp. 20563–20577, Sep. 2021.
- [41] R. Sun, L. Fu, Q. Cheng, K.-W. Chiang, and W. Chen, "Resilient pseudorange error prediction and correction for GNSS positioning in urban areas," *IEEE Internet Things J.*, vol. 10, no. 11, pp. 9979–9988, Jun. 2023.
- [42] D. Weng, Z. Hou, Y. Meng, M. Cai, and Y. Chan, "Characterization and mitigation of urban GNSS multipath effects on smartphones," *Measurement*, vol. 223, Dec. 2023, Art. no. 113766.
- [43] G. P. Gerdan, "A comparison of four methods of weighting double difference pseudorange measurements," *Austral. Surveyor*, vol. 40, no. 4, pp. 60–66, Dec. 1995.
- [44] S. R. Safavian and D. Landgrebe, "A survey of decision tree classifier methodology," *IEEE Trans. Syst. Man, Cybern.*, vol. 21, no. 3, pp. 660–674, Jun. 1991.
- [45] C. D. Sutton, "Classification and regression trees, bagging, and boosting," in *Handbook of Statistics (Data Mining and Data Visualization)*, vol. 24, C. R. Rao, E. J. Wegman, and J. L. Solka, Eds. Amsterdam, The Netherlands: Elsevier, Jan. 2005, pp. 303–329.
- [46] T. Hothorn and B. Lausen, "Bagging tree classifiers for laser scanning images: A data- and simulation-based strategy," *Artif. Intell. Med.*, vol. 27, no. 1, pp. 65–79, Jan. 2003.
- [47] L. Breiman, "Bagging predictors," *Mach. Learn.*, vol. 24, no. 2, pp. 123–140, Aug. 1996.
- [48] R. Sun, Z. Zhang, Q. Cheng, and W. Y. Ochieng, "Pseudorange error prediction for adaptive tightly coupled GNSS/IMU navigation in urban areas," *GPS Solutions*, vol. 26, no. 1, p. 28, Jan. 2022.
- [49] X. Niu, Y. Dai, T. Liu, Q. Chen, and Q. Zhang, "Feature-based GNSS positioning error consistency optimization for GNSS/INS integrated system," *GPS Solutions*, vol. 27, no. 2, p. 89, Apr. 2023.
- [50] Z. Zhang, B. Li, Y. Shen, Y. Gao, and M. Wang, "Site-specific unmodeled error mitigation for GNSS positioning in urban environments using a real-time adaptive weighting model," *Remote Sens.*, vol. 10, no. 7, p. 1157, Jul. 2018.



Tisheng Zhang received the B.Sc. and Ph.D. degrees in communication and information system from Wuhan University, Wuhan, China, in 2008 and 2013, respectively.

From 2018 to 2019, he was a Postdoctoral Researcher with Hong Kong Polytechnic University, Hong Kong. He is an Associate Professor with the GNSS Research Center, Wuhan University. His research interests focus on the fields of global navigation satellite system (GNSS) receiver and multisensor deep integration.



Jinwei Shi received the B.Eng. degree in electronic information engineering and the M.E. degree in navigation, guide, and control from Wuhan University, Wuhan, China, in 2020 and 2023, respectively.

His primary research interests include deeply coupled global navigation satellite system (GNSS) receiver.



Long Zhou received the B.Eng. degree in electronic information engineering from Wuhan University, Wuhan, China, in 2022, where he is currently pursuing the master's degree in navigation, guide, and control with the GNSS Research Center.

His primary research interests include deeply coupled global navigation satellite system (GNSS) receiver.



Quan Zhang received the B.S. degree in geomatics engineering from the Shandong University of Science and Technology, Qingdao, China, in 2009, and the Ph.D. degree from Wuhan University, Wuhan, China, in 2015.

From 2017 to 2018, he was a Postdoctoral Researcher with the Digital Photogrammetry Research Group (DPRG), Lyles School of Civil Engineering, Purdue University, West Lafayette, IN, USA. He is currently an Associate Professor with the GNSS Research Center, Wuhan University.

His research interests include vehicle-mounted global navigation satellite system (GNSS)/inertial navigation system (INS) integration technology and the integrity of multisensor-integrated navigation for intelligent driving.



Xin Feng received the B.E. and M.E. degrees from Wuhan University, Wuhan, China, in 2018 and 2021, respectively, where he is pursuing the Ph.D. degree in geodesy and survey engineering with the School of Geodesy and Geomatics.

His primary research interests include global navigation satellite system (GNSS) receiver and multisensor ultra-tightly integrations.



Xiaoji Niu received the bachelor's and Ph.D. degrees from the Department of Precision Instruments, Tsinghua University, Beijing, China, in 1997 and 2002, respectively.

He was a Postdoctoral Researcher with the University of Calgary, Calgary, AB, Canada, and worked as a Senior Scientist with SiRF Technology Inc., Shanghai, China. He is now a Professor with the GNSS Research Center, Wuhan University, Wuhan, China. He leads a multisensor navigation group focuses on global

navigation satellite system (GNSS)/inertial navigation system (INS) integrations, low-cost navigation sensor fusion, and its new applications. He has authored more than 90 academic articles and own 28 patents.