ORIGINAL ARTICLE

Pseudorange error prediction for adaptive tightly coupled GNSS/IMU navigation in urban areas

Rui Sun[1] · Zixuan Zhang[1] · Qi Cheng[1] · Washington Yotto Ochieng[1,2]

Abstract

The integration of global navigation satellite systems (GNSS) and inertial measurement unit (IMU) with the Kalman filter is widely used to enhance the availability of positioning in urban areas for many intelligent transport system (ITS) applications. In the traditional Kalman filter, the GNSS measurement noise is fixed based on factors determined a priori, instead of reflecting the impact of the surrounding environment on the received GNSS signal. This has the effect of degrading position accuracy and the a posteriori quality indicators. To address this issue, we propose a new measurement noise covariance update scheme, with the adaptive indicator generated from pseudorange error prediction results, for a tightly coupled GNSS/IMU navigation system in urban areas. Specifically, the pseudorange errors are predicted by means of an ensemble bagged regression tree model accounting for signal strength, satellite elevation angle and coordinate information. The urban experimental results show that the proposed algorithm provides a 3D accuracy of 9.21 m, with an improvement of 55% and 15%, respectively, over the traditional fixed covariance extended Kalman filter (EKF)-based fusion and EKF-based fusion with pseudorange error correction.

Keywords Integrated navigation · Pseudorange error prediction · Adaptive kalman filter

Introduction

The level of stringency of positioning, navigation and timing (PNT) requirements is increasing with emerging mission-critical intelligent transportation system (ITS) applications in urban areas. As one of the main PNT sources, GNSS is widely used for vehicle positioning (Feng and Law 2002). GNSS signals, however, are prone to reflection, diffraction and blockage by surrounding buildings, resulting in multipath interference and non-line-of-sight (NLOS) reception in urban areas. This can result in pseudorange errors tens of meters in magnitude, leading to unacceptable positioning errors for ITS applications (MacGougan et al. 2002). To date, these effects are largely mitigated through hardware design (i.e., antenna design), signal processing and measurement-based modeling.

Among antenna design-based methods, choke-ring and dual-polarization antennas are commonly used to mitigate the effects of multipath and NLOS signals (Sun et al. 2021a; Li and Schwieger 2018). The cost and bulky size of the antennas limit the applications for ITS in urban environments, however. Signal processing methods can be used to reduce certain kinds of multipath error by improving the form of the receiver correlator structure design or correlator function, but they are not able to address NLOS effects (Yang 2016; Chen et al. 2020; He et al. 2020; Kumar and Singh 2020; Qi et al. 2021).

Measurement-based modeling methods improve the positioning accuracy in urban areas by combining GNSS measurements, observables and other information sources in the measurement domain to ameliorate the effects of NLOS reception and multipath. For example, data from additional sources, such as spatial information (i.e., 3D city models), inertial measurement units (IMU) and vision sensors, are used together with GNSS measurements to improve positioning accuracy in urban canyons (Sun et al. 2021c; Bhatti et al. 2007). The famous shadow matching method uses 3D

✉ Rui Sun
  rui.sun@nuaa.edu.cn

1 College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China

2 Center for Transport Studies, College of Civil and Environmental Engineering, Imperial College London, London SW7 2AZ, UK

city models to detect and exclude NLOS and therefore obtain improved urban positioning results (Groves 2011; Yozevitch and Ben Moshe 2015; Wang et al. 2015; Adjrad and Groves 2018; Xu et al. 2020; Zhang et al. 2020). Accurate positioning can also be achieved in urban areas by the weighted average of the candidate positions, which are generated by comparing the similarity between the simulated and actual pseudorange measurements with the assistance of the surrounding geospatial information (Miura et al. 2013, 2015; Hsu et al. 2016; Hsu 2017).

The extent to which the mentioned algorithms are able to improve positioning in urban areas by reducing the errors due to NLOS is constrained by some key issues, however. One of the main issues is the determination of the threshold of the variable (i.e., C/N_0, elevation angle, etc.) for the NLOS detection. One recent proposal to address this is to use machine learning algorithms, where a training process is conducted to extract the rules for the input variables from raw GNSS measurements, including signal strength, satellite elevation angle, satellite azimuth angle, pseudorange residuals, pseudorange rate, number of visible satellites and dilution of precision (DOP) or their combinations and the corresponding output signal reception classifications (i.e., line-of-sight (LOS)/NLOS or LOS/multipath/NLOS). The rules extracted from the corresponding input and output variables are applied to newly collected GNSS data to predict the signal reception classifications. The position is then calculated by excluding the predicted NLOS/multipath signals (Yozevitch et al. 2016; Guermah et al. 2018; Quan et al. 2018; Sun et al. 2019, 2020). The positioning accuracy achievable through this approach is constrained, however, since the classification accuracy is affected by errors introduced from the offline labeling process needed for the machine learning algorithms (e.g., a 3D city model or camera-assisted labeling).

Unlike the above methods, which ultimately rely on the accuracy of signal reception classification, we have sought to address the problem through a machine learning-based algorithm to predict the pseudorange errors and thence correct the observed pseudoranges by considering the signal strength, satellite elevation angle and pseudorange residuals (Sun et al. 2021b). The proposed positioning method avoids the errors and costs arising from additional geospatial information during the labeling phase. Positioning tests in various urban canyons demonstrated that, when in static mode, the algorithm can deliver a 70% improvement in positioning accuracy compared to the conventional C/N_0 and elevation angle-based multipath/NLOS exclusion and positioning algorithm.

In the dynamic mode, a Kalman filter is always used to integrate GNSS and IMU data to enhance the positioning accuracy, integrity, continuity and availability in urban environments. In general, the values of the filter parameters are fixed, but this may cause the filter to diverge in changing environments (Wang et al. 2006). Thus, a series of adaptive Kalman filters (AKF) have been developed to overcome the limitation of using a priori statistics to model errors that have time-varying characterized (Mohamed and Schwarz 1999; Hide et al. 2003; Xiao et al. 2010). The adaptive indicators may take on a range of roles, including adjustment of the covariance matrix of the state estimation vector, the covariance matrix of the process vector and the covariance matrix of measurement vector (Xia et al. 1994; Bian et al. 2006; Xian et al. 2013; Liu et al. 2018; Yan et al. 2020).

None of the adaptive indicators in the above fusion methods, however, have been adjusted specifically for the errors caused by multipath signals and NLOS that are common in urban areas. An effective method to update the measurement noise covariance in the filtering algorithm is therefore essential for integrated navigation performance in urban areas.

As demonstrated in our previous research, the pseudorange error can be used as an important indicator to evaluate measurement quality. Here, therefore, we propose a pseudorange error prediction-based adaptive Kalman filter (PEP-AKF) algorithm for GNSS/IMU integration in urban areas. The pseudorange error is predicted by an ensemble bagged regression tree model accounting for signal strength, satellite elevation angle and coordinate information. An adaptive indicator based on the predicted pseudorange errors is then proposed to update the measurement noise covariance in the tightly coupled integrated navigation system.

Algorithm framework

The framework of the proposed pseudorange error prediction-based adaptive tightly coupled GNSS/IMU integration algorithm (PEP-AKF) is presented in Fig. 1.

The main parts of the offline phase are variable selection, the labeling process and the ensemble bagged regression tree-based training process. First, offline reference positioning data are collected from the intended route using a high-grade GNSS/IMU integrated system with backward and forward post-processing. Every set of variables, from the raw GNSS measurement, which is obtained from the output data of the integrated system, including signal strength (C/N_0), satellite elevation angle (θ) and coordinate information (Lat, Lon), is labeled with the corresponding pseudorange errors. The ensemble bagged regression tree model is then used to fit the pseudorange errors through the offline training process, thereby obtaining the prediction rules, that is, the relationship between the input variables (C/N_0, θ, Lat, Lon) and the corresponding labeled pseudorange errors.

The part outside the dashed box in Fig. 1 is the online tightly coupled GNSS/IMU navigation structure. The newly collected online GNSS variables from raw measurements and INS, including C/N_0, θ, Lat and Lon, are used together

Fig. 1 Framework of the proposed algorithm
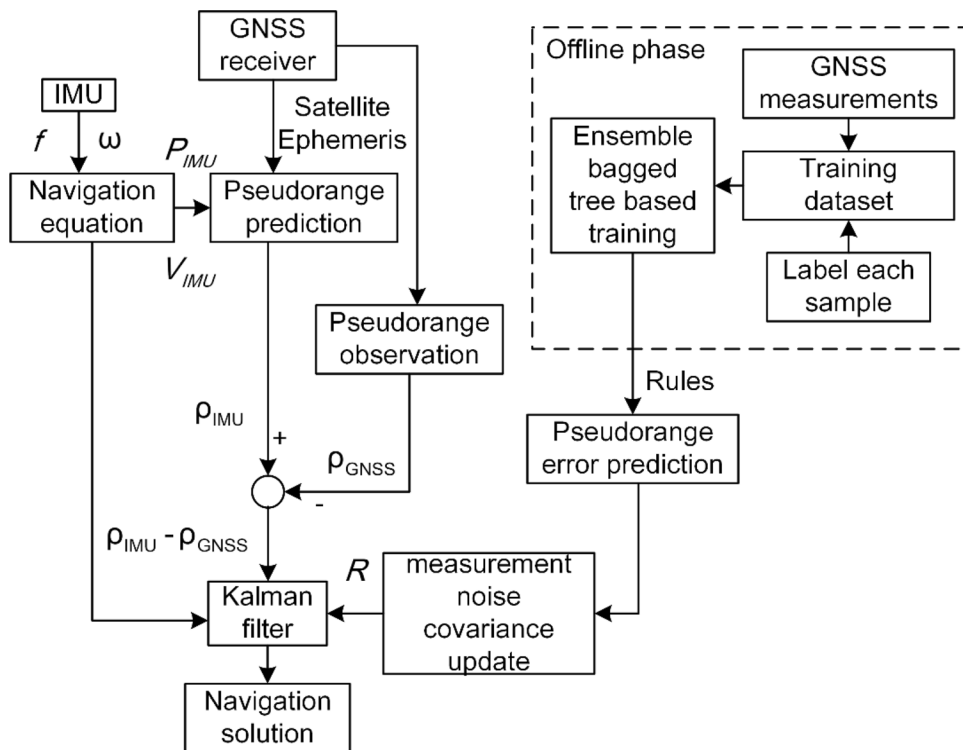
with the rules extracted from the offline phase to predict the pseudorange errors. The predicted pseudorange errors are then used to update the measurement noise covariance for the tightly coupled GNSS/IMU navigation.

Variable determination

The received GNSS signal contains various variables that can be used to predict the pseudorange error. Using too many variables, however, may increase the cost of computation. Following the trade-off between computational cost and training accuracy in Sun et al. (2019), we used four variables in this process: C/N_0, θ, Lat and Lon.

(1)   Signal strength (C/N_0)

The signal strength can be determined by the ratio of carrier power to noise power per unit of bandwidth in decibel-hertz (dB-Hz) of the received signal. The C/N_0 of NLOS/multipath signals is always lower than that of LOS signals, and since this can be used to measure the degree of multipath contamination of the signal, it represents for the pseudorange error prediction (Gu et al. 2016). In some cases, however, the receiver may output a high C/N_0, such as when the signal is reflected by materials like glass curtain walls. This means that other input variables must be used in tandem with C/N_0 in order to predict the pseudorange error accurately.

(2)   Satellite elevation angle (θ)

In general, the higher the elevation of a satellite, the less likely it is to be obstructed by objects, and the more likely it is to be received by a receiver in the form of LOS. This phenomenon, therefore, can also be used as a measure of the degree to which the signal has been contaminated by the multipath effect. Indeed, the technique of weighting measurements based on elevation angle to alleviate the multipath effect is widely used in positioning. The satellite elevation angle, θ, for the satellite (i) can be calculated by:

$$\theta^{(i)} = -\arcsin\left(\frac{z_r - z_s^{(i)}}{P_r^{(i)}}\right) \tag{1}$$

where z_r and $z_s^{(i)}$ are the z-coordinates of the receiver and satellite (i) in an earth-centered earth-fixed (ECEF) coordinate system and $P_r^{(i)}$ is the distance from the satellite (i) to the receiver position. The complexity of urban environments means, however, that reflected signals may also reach the receiver at a high elevation angle, so other features still need to be considered to assist the training rule extraction.

(3)   Latitude and longitude of the GNSS antenna (Lat and Lon)

We select latitude and longitude as two input variables. Since the NLOS/multipath repeats in similar conditions, taking longitude and latitude as two variables can contribute to

predict pseudorange errors more accurately (Margaria and Falletti 2016).

Pseudorange error labeling

The labeling of the pseudorange error is an important process in the offline phase. The pseudorange errors of the received signals can be calculated and labeled with their values from the offline reference data, obtained from the post-processing of a high-grade GNSS/IMU integrated system. The pseudorange ρ between the receiver and a satellite can be calculated as in the first term of (2) and (3):

$$\rho = P + c\left(\delta t_s^{(i)} - \delta t_r\right) + I + T + \varepsilon \tag{2}$$

$$P = \sqrt{\left(x_s^{(i)} - x_r\right)^2 + \left(y_s^{(i)} - y_r\right)^2 + \left(z_s^{(i)} - z_r\right)^2} \tag{3}$$

where P is the geometric range between the observed satellite (i) and the receiver; $\left(x_r, y_r, z_r\right)$ and $\left(x_s^{(i)}, y_s^{(i)}, z_s^{(i)}\right)$ are the coordinates of the receiver and the satellite (i) in an ECEF coordinate system; the constant c is the speed of light in a vacuum; $\delta t_s^{(i)}$ and δt_r are the clock offset of satellite (i) and the receiver, respectively; I and T are, respectively, the delays caused by the ionosphere and troposphere; and ε represents the other errors, mainly from NLOS/multipath errors.

As mentioned earlier, the reference position is determined during the offline phase. The calculation of the corrected pseudorange ρ^c could then be expressed by considering the reference position and the error models applied on the related error sources, as in (4):

$$\rho^c = P + c\left(\Delta\delta t_s^{(i)} - \Delta\delta t_r\right) + \Delta I + \Delta T + \varepsilon \tag{4}$$

where the geometric range, P, is calculated with the reference position and the position of the observed satellite from the broadcast ephemeris; $\Delta\delta t_s^{(i)}$ represents the residual of the satellite errors of satellite i after applying the satellite clock offset corrections from the broadcast ephemeris; $\Delta\delta t_r$ is the receiver clock offset after applying the calculated corresponding satellite and receiver clock error from the pseudorange positioning equations with the reference ground point position; and ΔI and ΔT are those residuals in the ionospheric and tropospheric delays that are not fully corrected by Klobuchar and Saastamoinen models, respectively. The pseudorange error $\Delta\rho$ can be further derived as in (5):

$$\Delta\rho = \rho^c - P = c\left(\Delta\delta t_s^{(i)} - \Delta\delta t_r\right) + \Delta I + \Delta T + \varepsilon. \tag{5}$$

The calculated pseudorange error, $\Delta\rho$, is then labeled as its value. Every set of variables from the raw GNSS measurements, including C/N_0, θ, Lat and Lon, will then have the corresponding value of $\Delta\rho$ after the labeling process.
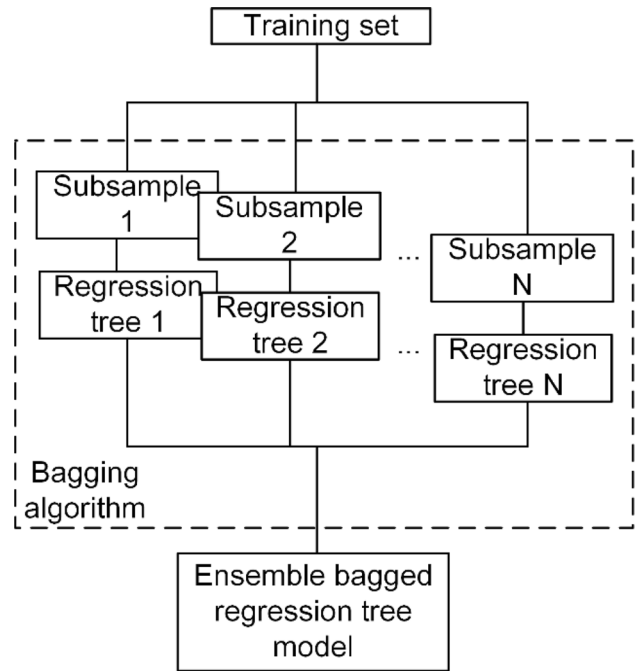

Fig. 2 Flowchart of the ensemble bagged regression tree algorithm

Ensemble bagged regression tree-based pseudorange error prediction

The ensemble decision tree is a popular machine learning method that is able to provide high accuracy prediction in many applications (Adler et al. 2011; Saeed et al. 2019). The ensemble bagged regression tree is composed of a combination of multiple regression decision trees established based on a subsample that comprises a training set. Thus, the ensemble model of multiple regression trees can bring a group of weak learners together to form a strong learner with the merit of not being sensitive to small perturbations or uncertainties in the training dataset. This makes it suitable for pseudorange error prediction. The process of establishing the ensemble bagged regression tree model is shown in Fig. 2.

The discussed problem can be defined as follows: Given the training set $\{x_i, \Delta\rho_i\}_1^I$ of known $(x, \Delta\rho)$ values, the objective is to determine a function that maps x to $\Delta\rho$, where $x_i = \left(C/N_{0i}, \theta_i, \text{Lat}_i, \text{Lon}_i\right)$, $i = 1, 2, 3, \cdots, I$. i is the sequence number of the sample, and I is the total number of the samples. $\Delta\rho_i$ is the corresponding labeled pseudorange error of x_i. The ensemble bagged regression tree model-based pseudorange error prediction is described as follows:

(1) Single regression tree-based model generation

In the input space of the subsample training set, each region is recursively divided into two subregions, and

the output value of each subregion is determined so as to construct the regression tree.

(a) Select the optimal segmentation variable j and the segmentation point s to solve:

$$\min_{j,s} \left[\min_{c_1} \sum_{x_i \in R_1(j,s)} (\Delta\rho_i - c_1)^2 + \min_{c_2} \sum_{x_i \in R_2(j,s)} (\Delta\rho_i - c_2)^2 \right]$$
(6)

$$c_1 = \text{ave}\left(y_i \big| x_i \in \text{Reg}_1(j,s) \right)$$
(7)

$$c_2 = \text{ave}\left(y_i \big| x_i \in \text{Reg}_2(j,s) \right)$$
(8)

where ave is the average function. Traverse the variable j, scan the segmentation point s for the fixed segmentation variable j and determine the (j, s) that minimizes formula (6).

(b) Divide the region with the selected (j, s) and determine the corresponding output value:

$$\text{Reg}_1(j,s) = \left\{ x \big| x^{(j)} \le s \right\}, \text{Reg}_2(j,s) = \left\{ x \big| x^{(j)} > s \right\}$$
(9)

$$\hat{c}_q = \frac{1}{N_q} \sum_{x_i \in R_q(j,s)} \Delta\rho_i, x \in R_q, \quad q = 1, 2$$
(10)

where $\text{Reg}_1(j,s)$ and $\text{Reg}_2(j,s)$ mean the divided regions and $\hat{c}_q$ is the corresponding output value. For each divide of the region, two new nodes will be created. A node is comprised of a sample of data and a decision rule.

(c) Continue with steps (a) and (b) for the two subregions until the iteration stop conditions are met: i.e., where the sample size of the node is less than a user-specified value. Here, from the sensitivity analysis, the defined value is set as 8.

(d) Divide the input space into L regions and generate a regression tree:

$$h(x) = \sum_{l=1}^{L} \hat{c}_l I\left(x \in \text{Reg}_l \right), \quad (l = 1, 2, \dots, L)$$
(11)

where $h(x)$ is the predicted value of the regression tree; l is the number of the region; and I is a characteristic function.

(2) Bagged regression tree-based training model generation

The final output result of the bagged regression tree model is the average value of the predicted values of all the individual regression trees. The bagged regression tree model is expressed as:

$$h_{\text{bag}}(x) = \frac{1}{N} \sum_{n=1}^{N} h_n(x), \quad (1 \le n \le N)$$
(12)

where $h_n(x)$ is the predicted value of the regression tree n for input x and n is the number of the regression trees.

(3) Pseudorange error prediction

The experiment in this paper uses only global positioning system (GPS) L1 data to train a bagged regression tree model. Since NLOS/multipath errors repeat in similar conditions of the surrounding environment and constellation, the rules from one constellation in the area may have a limited effectiveness on predicting pseudorange errors in other constellations in the same area. The pseudorange errors of the newly collected variables can be predicted based on the corresponding final predictor $h_{\text{bag}}(x)$. The input $x = \left(C/N_0, \theta, \text{Lat}, \text{Lon} \right)$ can be used together with the output rules of the $h_{\text{bag}}(x)$ to obtain the pseudorange errors.

Adaptive tightly coupled fusion scheme

The adaptive tightly coupled fusion scheme is based on the EKF structure. The state vector, X, for the EKF is defined in (13).

$$X = \left[\begin{array}{cccccccccc} \delta r_{3\times1}^T & \delta v_{3\times1}^T & \Psi_{3\times1}^T & b_{g3\times1}^T & b_{a3\times1}^T & \nabla_{g3\times1}^T & \nabla_{a3\times1}^T & t_b & \delta t_b \end{array} \right]^T$$
(13)

where the first 21 dimensions of X are built on the IMU estimation errors, including the three-axis navigation parameters of position error $\delta r_{3\times1}$, velocity error $\delta v_{3\times1}$, attitude error $\Psi_{3\times1}$, bias of gyroscope $b_{g3\times1}$, bias of the accelerometer $b_{a3\times1}$, scale factors of gyroscopes $\nabla_{g3\times1}$ and scale factors of accelerometers $\nabla_{a3\times1}$. The last two dimensions of X include the GNSS receiver clock bias t_b and GNSS receiver clock drift δt_b.

The state transition equation for the state vector X is expressed in (14):

$$\dot{X} = AX + w$$
(14)

where $\dot{X}$ is the first derivative of the state vector X, A is the IMU dynamic transition model and w is the system noise. Q is defined as the covariance matrix of w.

The measurement model Z is expressed in (15)

$$Z = HX + n$$
(15)

where Z is the measurement vector, H is the measurement mapping matrix and n is the measurement noise. R is defined as the covariance matrix of n. If the number of visible satellites is M, the measurement vector, Z, is formed as:

$$Z = \begin{bmatrix} \rho_{\text{IMU},1} - \rho_{\text{GNSS},1} \\ \vdots \\ \rho_{\text{IMU},m} - \rho_{\text{GNSS},m} \\ \vdots \\ \rho_{\text{IMU},M} - \rho_{\text{GNSS},M} \end{bmatrix}_{M \times 1} \tag{16}$$

where ρ_{IMU} is the pseudorange derived from the IMU and ρ_{GNSS} is the pseudorange decoded from GNSS observations; $1 \leq m \leq M$.

An adaptive indicator is proposed based on the pseudorange error prediction in order to estimate the GNSS measurement noise covariance parameters in urban areas. The adaptive indicator $f(\cdot)$ can be expressed in (17):

$$f(\cdot) = \left(\left| h_{\text{bag}}(\cdot) \right| + \eta \right)^2 \tag{17}$$

where $f(\cdot)$ is the proposed adaptive indicator function; $h_{\text{bag}}(\cdot)$ is the designed ensemble bagged regression tree model; and η is the perturbation coefficient. Here, $\eta = 0.1$. The measurement noise covariance matrix, R, at epoch k can then be constructed based on the adaptive indictor function at epoch k, as expressed in (18):

$$R_k = \begin{bmatrix} f(x_{p_1}) & & & & \\ & \ddots & & & \\ & & f(x_{p_m}) & & \\ & & & \ddots & \\ & & & & f(x_{p_M}) \end{bmatrix}_k \tag{18}$$

where $f(x_{p_m})$ is the predicted results of the adaptive indicator function for the satellite p_m, $x = (C/N_0, \theta, \text{Lat}, \text{Lon})$, and $p_m (m = 1 \dots M)$ is the mth satellite used for the positioning.

The final fusion results can then be obtained based on the following Kalman filter procedures.

Prediction stage:

$$\hat{X}_{k,k-1} = \Phi_{k,k-1} \hat{X}_{k-1} \tag{19}$$

$$P_{k,k-1} = \Phi_{k,k-1} P_{k-1} P_{k,k-1}^T + Q_{k-1} \tag{20}$$

Update stage:

$$K_k = P_{k,k-1} H_k^T \left(H_k P_{k,k-1} H_k^T + R_k \right)^{-1} \tag{21}$$

$$P_k = (I - K_k H_k) P_{k,k-1} (I - K_k H_k)^T + K_k R_k K_k^T \tag{22}$$

$$\hat{X}_k = \hat{X}_{k,k-1} + K_k \left(Z_k - H_k \hat{X}_{k,k-1} \right) \tag{23}$$

where, $\hat{X}_k$: state vector estimates at time epoch k, Φ_k: state transition matrix at time epoch k, P_k: error covariance matrix at time epoch k, Q_k: system noise covariance matrix at time epoch k, R_k: measurement noise covariance matrix at time epoch k, H_k: measurement matrix at time epoch k, K_k: Kalman gain matrix at time epoch k, $\blacksquare_{k,k-1}$: matrix/vector $\blacksquare$ propagation from time epoch $k-1$ to k.

Experiment and results

In order to validate the proposed method, a vehicle-based field test was carried out in the urban areas in Taipei city of Taiwan, China. The training dataset was collected from about 10:00 in the morning in the Beijing time with the time duration of 2–3 h for consecutive 3 days. The total training dataset was collected for about seven hours at the experimental site. The data, including the raw GNSS measurement and fused positioning results, were collected using a high-grade GNSS/IMU integrated system (i.e., NovAtel SPAN-LCI). The variables extracted from the GPS L1 measurements were then used as the inputs for the training, and the post-processed GNSS/IMU positioning results were used for the pseudorange error labeling.

The testing data were collected around 11:00 in the morning in the Beijing time after one week, at the same experimental site, where the training data were collected. The testing trajectory is shown in Fig. 3, and the onboard navigation sensors used included: (1) a micro-electro-mechanical system (MEMS) IMU, STIM-300, with a sampling rate of 250 Hz. The corresponding IMU parameters are given in Table 1; (2) a GNSS receiver, NovAtel ProPak6, with a sampling rate of 1 Hz; and (3) a high-end GNSS/IMU integrated system, i.e., NovAtel SPAN-LCI, for high accuracy navigation solutions in the experiment. The corresponding IMU parameters are given in Table 2. The reference positions in the test were generated by the SPAN-LCI outputs with post-processing kinematic mode of the tightly coupled GNSS/IMU using NovAtel Inertial Explore software. A
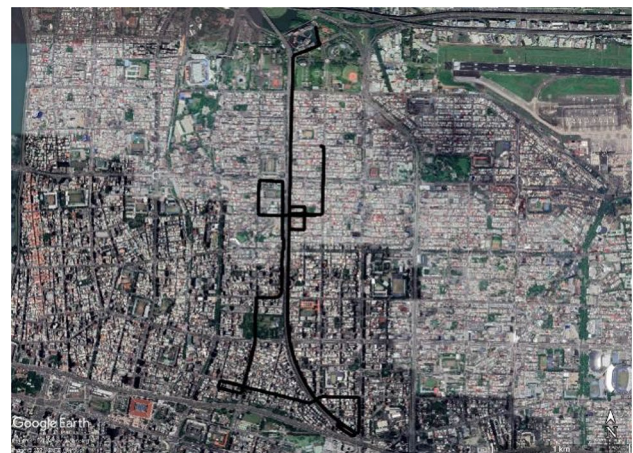

Fig. 3 Vehicle trajectory

Table 1 Sensor parameters of STIM-300

Accelerometer	Bias instability	0.05 mg
	Random walk noise	0.06 m/s/sqrt(hr)
Gyroscope	Bias instability	0.5 deg/hr
	Random walk noise	0.15 deg/sqrt(hr)

Table 2 Sensor parameters of NovAtel SPAN-LCI

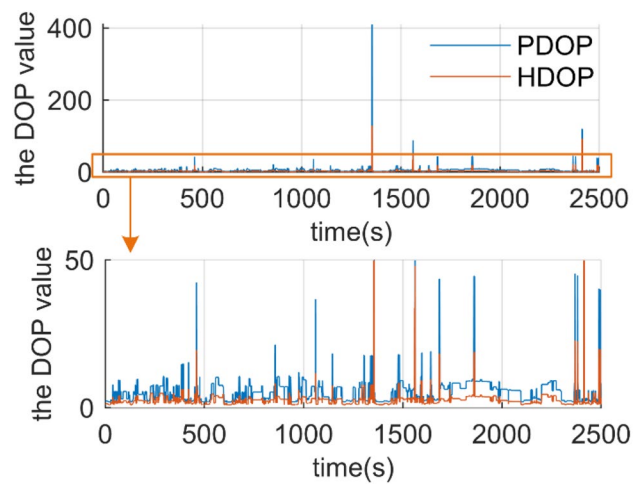| Accelerometer	Bias instability	0.1 mg
	Random walk noise	0.1 mg/sqrt(hr)
Gyroscope	Bias instability	0.05 deg/hr
	Random walk noise	0.012 deg/sqrt(hr)

Table 3 Summary of the datasets

	Training dataset	Testing dataset
Collected location	Urban canyon	Urban canyon
Sample size	25,953	6109

Fig. 4 Number of visible GPS satellites
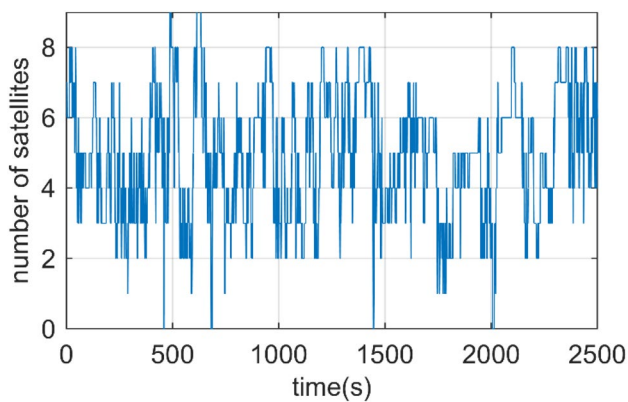
Fig. 5 PDOP and HDOP values

summary of the training dataset and testing dataset is shown in Table 3. The number of visible GPS satellites is depicted in Fig. 4, and the corresponding position dilution of precision (PDOP) and horizontal dilution of precision (HDOP) values are depicted in Fig. 5.

Since the predicted pseudorange error affects the performance of the final positioning solutions, the evaluation of the fitting performance of the proposed ensemble bagged regression tree-based pseudorange error prediction compared with the other popular used fitting algorithms becomes essential. Here, the three other candidate fitting algorithms are the ensemble boosted trees model, the fine Gaussian support vector machine (SVM) model and the squared exponential Gaussian process regression model. The parameters

used to evaluate the fitting performance are root mean square error (RMSE) of the residuals, R-squared and training time. The RMSE of the residuals is used to evaluate the accuracy of fitting results. The smaller the RMSE of the residuals, the closer the fitting result is to the reference pseudorange errors. The R-squared is another informative parameter for evaluating the fitting results. It can be any value between 0 and 1, with a higher value indicating a better fit for the regression model (Chicco et al. 2021). To prevent overfitting for some candidate machine learning algorithms, fivefold cross-validation was used for all models. The comparison results for the fitting algorithms are shown in Table 4.

It is clear that the ensemble bagged regression tree model outperforms the other fitting methods based on the evaluation of the overall fitting performances. In particular, the RMSE of the ensemble bagged regression tree-based fitting results was 2.14 m, which is much smaller than the other fitting models, whose overall performance ranged from 2.36 to 3.01 m. In addition, the two highest R-squared values were the ensemble bagged regression tree model and the squared exponential Gaussian process regression model, with values of 0.70 and 0.64, respectively. The training time consumption for the squared exponential Gaussian process regression model was 1031.40 s, however, which was 22 times higher than that for the ensemble bagged regression tree model.

In order to evaluate the final algorithm positioning performance, the proposed PEP-AKF-based fusion algorithm was compared with two other candidate fusion algorithms: (1) traditional EKF-based fusion and (2) EKF-based fusion with pseudorange error correction. Also, the GPS L1 measurements are used for the candidate fusion algorithms. The detail steps of the candidate algorithms are shown in Table 5, and the fusion results are shown in Fig. 6.

In Figs. 6 and 7, it is obvious that the positioning results of the proposed algorithm are closer to the reference than the

Table 4 Comparison of ensemble bagged regression trees and other candidate models

Model	RMSE of the residuals (m)	R-Squared	Training time (s)
Ensemble bagged regression tree	2.14	0.70	46.01
Ensemble boosted tree	2.75	0.51	39.50
Fine Gaussian SVM	3.01	0.41	284.84
Squared exponential Gaussian process regression	2.36	0.64	1031.40

Table 5 Summary of the algorithms

Algorithm	Algorithm description	Color
Traditional EKF-based fusion	GNSS and IMU fusion in the tightly coupled mode with EKF	Red
EKF-based fusion with pseudorange error correction	Step 1: Using the trained ensemble bagged regression trees model to predict the pseudorange error and then applying this for the pseudorange error correction Step 2: Using the corrected pseudorange to form the measurement vector Z in the EKF of the tightly coupled fusion model	Green
PEP-AKF-based fusion (proposed)	Step 1: Using the trained ensemble bagged regression trees model to predict the pseudorange error Step 2: Construction of an adaptive indicator based on the predicted pseudorange error so as to adjust the measurement covariance matrix R in the Kalman filter of the tightly coupled fusion model	Blue

other fusion algorithms, especially in densely built environment areas. The positioning accuracy for the candidate algorithms and the corresponding improvements compared to the traditional EKF-based fusion are shown in Table 6. The comparisons of the positioning error of the fusion results for the algorithms are depicted in Figs. 8 and 9.

Table 6 indicates that the proposed fusion algorithm can provide a 3D position accuracy RMSE of 9.21 m, while it is 10.85 m for the EKF-based fusion with pseudorange error correction and 20.76 m for the traditional EKF-based fusion, corresponding to improvements of 15% and 55%. Meanwhile, the improvement in the horizontal positioning accuracy for the proposed algorithm compared to the EKF-based fusion with pseudorange error correction is also significant, with an improvement value of about 32%.

In the period between 1000 and 1100 s in Fig. 8, the surrounding urban environment is characterized by moderately dense development, with the received GPS signals experiencing medium multipath and NLOS. The surrounding urban environment is shown in Fig. 7, Part 1. In this condition, the measurement noise covariance matrix was adjusted by the generated adaptive indicator, alleviating the fused positioning error caused by the contaminated measurement arising from the multipath and NLOS. The use of the adaptive indicator, therefore, led to substantial improvements in the horizontal positioning results compared to the EKF-based fusion with pseudorange error correction model and

the traditional EKF-based fusion model by 34% and 64%, respectively.

In the period between 1500 and 1600 s in Fig. 8, the environment was a dense urban area with taller and more closely packed buildings than in the moderately dense urban areas. This led to more significant NLOS and multipath effects, as shown in Fig. 7, Part 2. Strikingly, the improvement in performance offered by the proposed algorithm in this environment was even greater than in the moderately dense urban areas. Specifically, there was a 40% improvement in 3D positioning compared to the EKF-based fusion with pseudorange error correction model and a 76% improvement compared to the traditional EKF-based fusion model. Correspondingly, the maximum positioning error is also reduced to 16.30 m from 71.42 and 34.44 m.

In order to investigate the effectiveness of the proposed adaptive indicator further, the relationship for the adaptive indicator and the corresponding pseudorange errors was analyzed. An example of the typical adaptive indicator, and the corresponding pseudorange error for the satellite SV30, is shown in Table 7. This shows that, at 647 s, the satellite SV30 had a pseudorange error of 0.92 m, and the corresponding adaptive indicator was 0.799. At 1015 s, meanwhile, SV30 had a predicted pseudorange error of 18.95 m, and the corresponding adaptive indicator became 298.856. This means that the weighting of the pseudorange measurement for SV30 was significantly reduced by the severity of
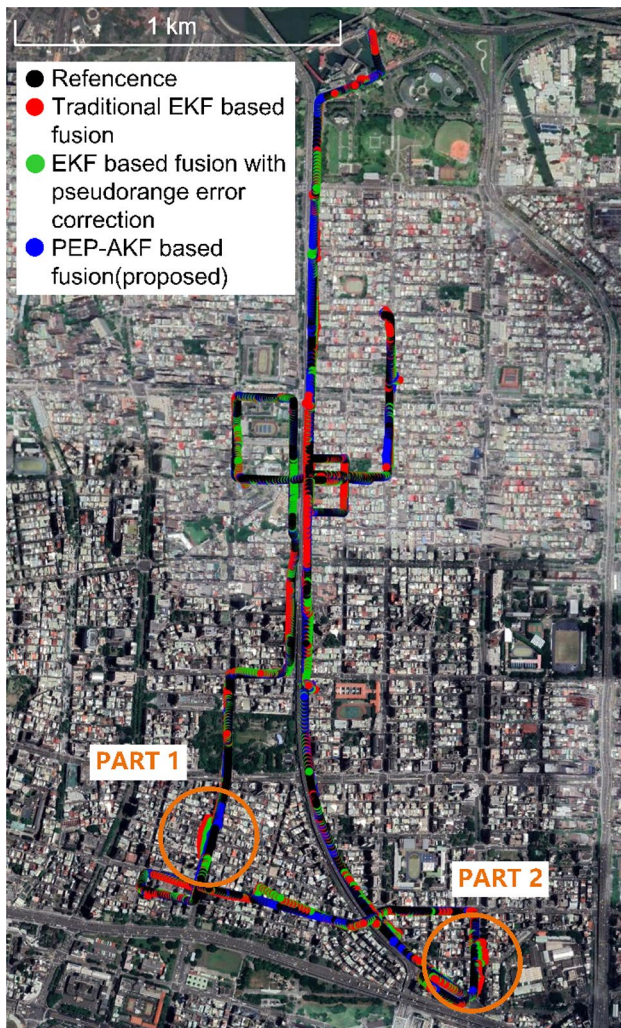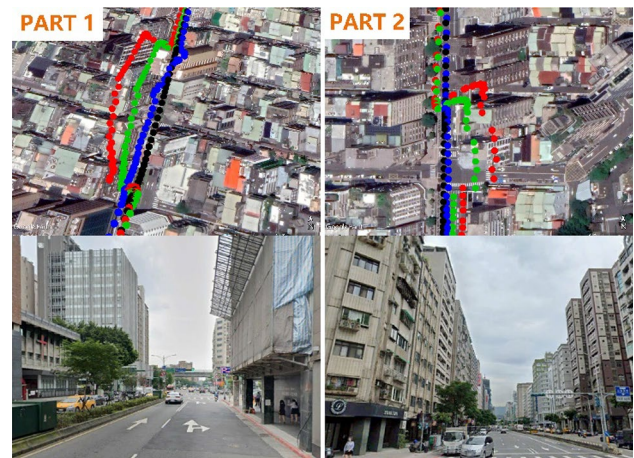
Fig. 6 Fusion results of the algorithms

Fig. 7 Partial experimental results and experimental scenes

the pseudorange error, resulting in that the NLOS and multipath will have less effect on the fusion results.

The improvement in the 3D positioning results arising from the use of the proposed algorithm is also depicted in Fig. 10. This shows that the positioning results from the conventional EKF-based fusion and EKF-based fusion with pseudorange error correction comprise only about 0.7% and 3% of the epochs within 2 m. With the proposed PEP-AKF-based fusion model, however, this proportion increases significantly to 13%. In addition, with the proposed fusion model, the proportion of epochs with a positioning accuracy in the range of 2–4 m is much higher than that of the conventional EKF-based fusion and EKF-based fusion with pseudorange error correction. The EKF-based fusion with pseudorange error correction can improve the positioning accuracy to a certain extent, however. Due to the error present in the pseudorange error prediction for some epochs, using the incorrectly corrected pseudorange in the fusion

will introduce errors in the positioning results. For the proposed algorithm, meanwhile, even when an incorrect pseudorange error is predicted for some epochs, the fusion algorithm is less sensitive to those errors because of the use of the corresponding adaptive indicator in the filter. This is why the proposed algorithm will return a smaller fusion error than the EKF-based fusion with pseudorange error correction model, even with the same predicted pseudorange error.

Table 8 shows a comparison of the positioning accuracy analysis according to each epoch. This reveals that the proposed PEP-AKF-based fusion improved the horizontal and 3D positioning results in, respectively, 83.7% and 88.5% of the epochs, compared to the traditional EKF-based fusion. Although a deterioration in positioning accuracy was seen in about 13% of the epochs, this may have been due to the incorrect pseudorange error prediction results, and the algorithm is clearly still effective for most of the epochs. Compared to traditional EKF-based fusion, meanwhile, the EKF-based fusion with pseudorange error correction algorithm improved the horizontal and 3D positioning results in, respectively, 78.1% and 84.2% of the epochs. For both algorithms, the worse epochs for 3D positioning were fewer than the epochs for the horizontal positioning, since height errors tended to be more substantially corrected as the pseudorange errors were corrected. This is because NLOS and multipath signals affect height positioning results more severely than horizontal ones in urban areas.

Conclusion

We have developed a pseudorange error prediction-based adaptive Kalman filter algorithm for tightly coupled GNSS/IMU integration in urban areas. An adaptive indicator generated using a pseudorange error predicted by means of an
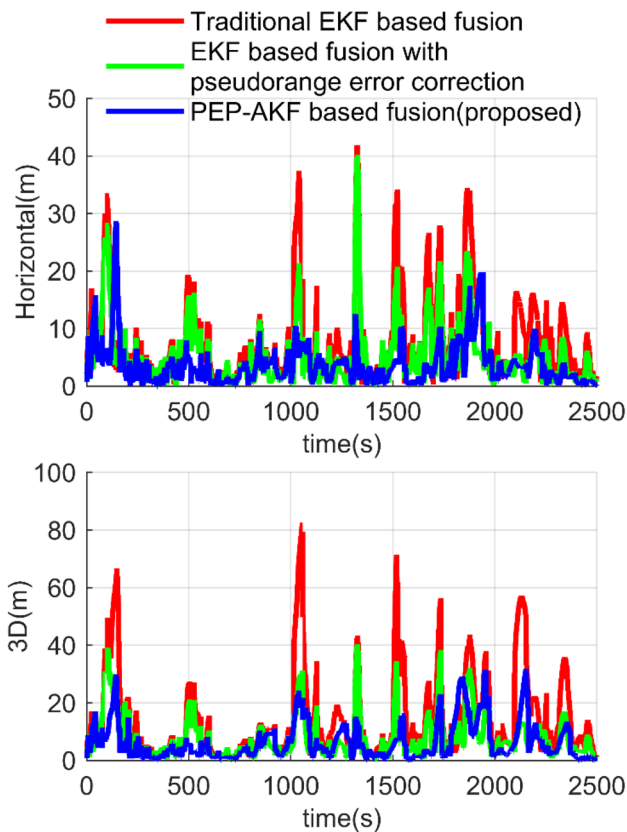
Table 6 RMSE results comparison

GNSS/IMU tightly coupled fusion algorithms	RMSE (m)				
	North	East	Down	Horizontal	3D
EKF-based fusion	6.83	8.20	17.80	10.67	20.76
EKF-based fusion with pseudorange error correction	5.27	5.38	7.81	7.53	10.85
Improvement	22.75%	34.46%	56.12%	29.43%	47.72%
PEP-AKF-based fusion (proposed method)	3.60	3.63	7.66	5.11	9.21
Improvement	47.26%	55.78%	56.98%	52.11%	55.64%
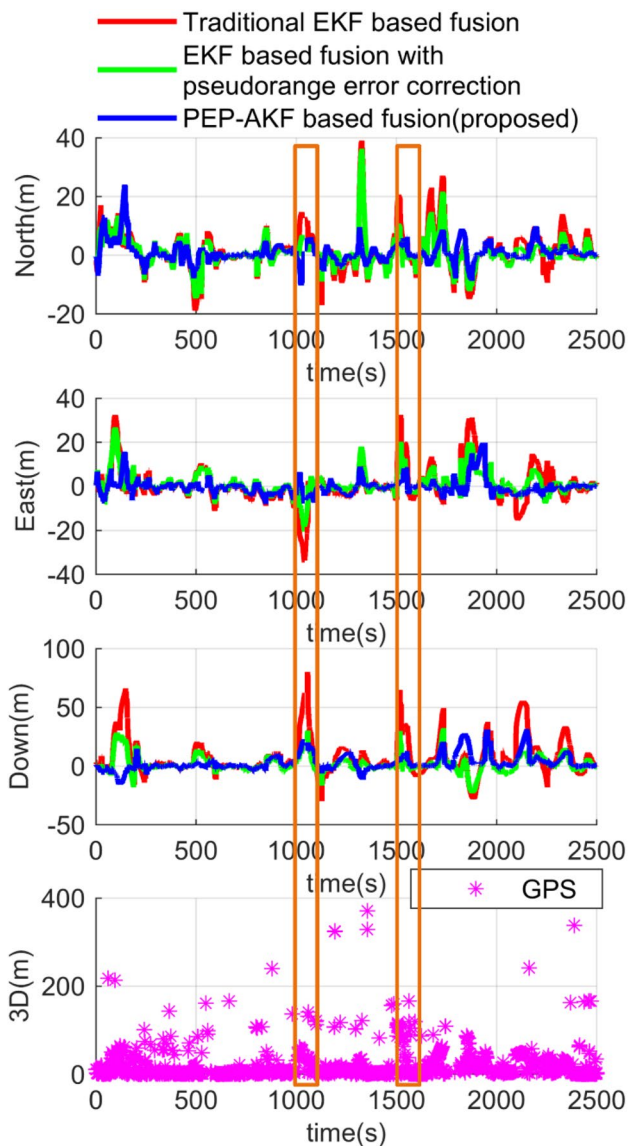
**Fig. 8** Positioning error in terms of the local-level coordinate system

Fig. 9 Horizontal and 3D positioning error in terms of the local-level coordinate system

Table 7 Examples of the typical adaptive indicator and the corresponding pseudorange error for satellite SV30

Time (s)	Pseudorange error (m)	Adaptive indicator
633	0.5538	0.296
647	0.9201	0.799
1015	18.9460	298.856
1071	10.2257	102.421

ensemble bagged regression tree algorithm is used to optimize the measurement noise covariance parameters in the Kalman filter with the characteristics of GNSS measurement noise properties in urban areas. The experimental results show that the proposed PEP-AKF-based fusion algorithm delivers a better positioning performance in urban areas than other candidate fusion algorithms, providing a horizontal

Fig.10 3D positioning error distribution within 40 m
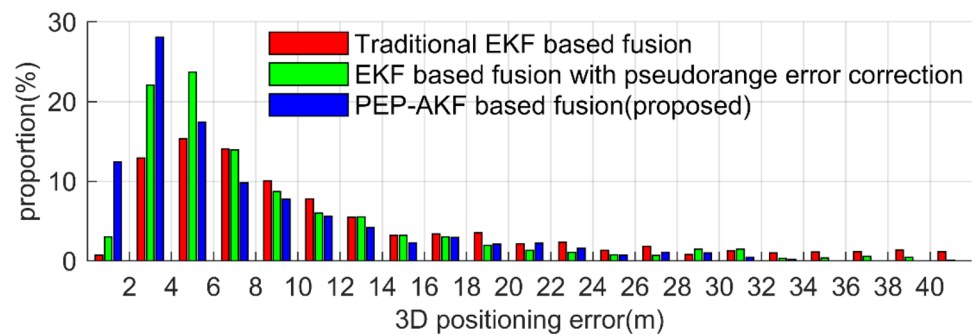
Table 8 Algorithm performance compared to the traditional EKF-based fusion results with proportion of epochs

Proportion of epoch (%)	Horizontal		3D	
	Better	Worse	Better	Worse
EKF-based fusion with pseudor-ange error correction	78.1	21.9	84.2	15.8
PEP-AKF-based fusion (proposed)	83.7	16.3	88.5	11.5

RMSE accuracy of 5.11 m, which represents a 52% improvement on traditional EKF-based fusion (RMSE of 10.67 m) and a 32% improvement on EKF-based fusion with pseudor-ange correction (RMSE of 7.53 m). The proposed algorithm can be used for the local urban positioning enhancement in the future, especially to provide the positioning accuracy improvement for incoming vehicles with low-cost onboard GNSS and IMU sensors. Further research is also needed to explore the added value of various aspects of multiple GNSS constellations.

Data availability The datasets analyzed during the current study are available from the corresponding author on reasonable request.

References

Adjrad M, Groves PD (2018) Intelligent Urban positioning: integration of shadow matching with 3D-mapping-aided GNSS ranging. J Navig 71:1–20. https://doi.org/10.1017/S0373463317000509

Adler W, Potapov S, Lausen B (2011) Classification of repeated measurements data using tree-based ensemble methods. Comput Stat 26:355–369. https://doi.org/10.1007/s00180-011-0249-1

Bhatti UI, Ochieng WY, Feng S (2007) Integrity of an integrated GPS/INS system in the presence of slowly growing errors. Part I Crit Rev GPS Solut 11:173–181. https://doi.org/10.1007/s10291-006-0048-2

Bian H, Jin Z, Wang J, Tian W (2006) The innovation-based estimation adaptive Kalman filter algorithm for INS/GPS integrated navigation system. J Shanghai Jiaotong Univ 40:1000–1003, 1009

Chen X, He D, Pei L (2020) BDS B1I multipath channel statistical model comparison between static and dynamic scenarios in dense urban canyon environment. Satell Navig 1:26. https://doi.org/10.1186/s43020-020-00027-7

Chicco D, Warrens MJ, Jurman G (2021) The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. Peer J Comput Sci. https://doi.org/10.7717/peerj-cs.623

Feng S, Law CL (2002) Assisted GPS and its impact on navigation in intelligent transportation systems. In: Proceedings. The IEEE 5th international conference on intelligent transportation systems. pp 926–931. doi:https://doi.org/10.1109/ITSC.2002.1041344

Groves PD (2011) Shadow matching: A New GNSS positioning technique for urban canyons. J Navig 64:417–430. https://doi.org/10.1017/S0373463311000087

Gu Y, Hsu L, Kamijo S (2016) GNSS/onboard inertial sensor integration with the Aid of 3-D building map for lane-level vehicle self-localization in urban canyon. IEEE Trans Veh Technol 65:4274–4287. https://doi.org/10.1109/TVT.2015.2497001

Guermah B, El Ghazi H, Sadiki T, Guermah H (2018) A robust GNSS LOS/multipath signal classifier based on the fusion of information and machine learning for intelligent transportation systems. In: IEEE vehicular technology conference proceedings. pp. 94–100

He C, Lu X, Guo J, Su C, Wang M (2020) Initial analysis for characterizing and mitigating the pseudorange biases of BeiDou navigation satellite system. Satell Navig 1:3. https://doi.org/10.1186/s43020-019-0003-3

Hide C, Moore T, Smith M (2003) Adaptive kalman filtering for low-cost INS/GPS. J Navig 56:143–152. https://doi.org/10.1017/S0373463302002151

Hsu L (2017) Analysis and modeling GPS NLOS effect in highly urbanized area. GPS Solut. https://doi.org/10.1007/s10291-017-0667-9

Hsu L, Gu Y, Kamijo S (2016) 3D building model-based pedestrian positioning method using GPS/GLONASS/QZSS and its reliability calculation. GPS Solut 20:413–428. https://doi.org/10.1007/s10291-015-0451-7

Kumar A, Singh AK (2020) A novel multipath mitigation technique for GNSS signals in urban scenarios. IEEE Trans Veh Technol 69:2649–2658. https://doi.org/10.1109/TVT.2019.2962919

Li Z, Schwieger V (2018) Investigation of a L1-optimized choke ring ground plane for a low-cost GPS receiver-system. J Appl Geodesy 12:55–64. https://doi.org/10.1515/jag-2017-0026

Liu Y, Fan X, Lv C, Wu J, Li L, Ding D (2018) An innovative information fusion method with adaptive Kalman filter for integrated INS/GPS navigation of autonomous vehicles. Mech Syst Signal Process 100:605–616. https://doi.org/10.1016/j.ymssp.2017.07.051

MacGougan G, Lachapelle G, Klukas R, Siu K, Garin L, Shewfelt J, Cox G (2002) Performance analysis of a stand-alone

high-sensitivity receiver. GPS Solut 6:179–195. https://doi.org/10.1007/s10291-002-0029-z

Margaria D, Falletti E (2016) The local integrity approach for urban contexts: definition and vehicular experimental assessment. Sensors. https://doi.org/10.3390/s16020154

Miura S, Hsu L, Chen F, Kamijo S (2015) GPS error correction with pseudorange evaluation using three-dimensional maps. IEEE Trans Intell Transp Syst. https://doi.org/10.1109/TITS.2015.2432122

Miura S, Hisaka S, Kamijo S (2013) GPS Multipath detection and rectification using 3D maps. In: IEEE international conference on intelligent transportation systems-ITSC. Pp. 1528–1534

Mohamed AH, Schwarz KP (1999) Adaptive kalman filtering for INS/GPS. J Geodesy 73:193–203. https://doi.org/10.1007/s001900050236

Qi Y, Yao Z, Lu M (2021) General design methodology of code multi-correlator discriminator for GNSS multipath mitigation. IET Radar Sonar Navig. https://doi.org/10.1049/rsn2.12088

Quan Y, Lau L, Roberts GW, Meng X, Zhang C (2018) Convolutional neural network based multipath detection method for static and kinematic GPS high precision positioning. Remote Sens 10:2052. https://doi.org/10.3390/rs10122052

Saeed MS, Mustafa MW, Sheikh UU, Jumani TA, Mirjat NH (2019) Ensemble bagged tree based classification for reducing non-technical losses in multan electric power company of Pakistan. Electronics. https://doi.org/10.3390/electronics8080860

Sun R, Hsu L, Xue D, Zhang G, Ochieng WY (2019) GPS signal reception classification using adaptive neuro-fuzzy inference system. J Navig 72:685–701. https://doi.org/10.1017/S0373463318000899

Sun R, Wang G, Zhang W, Hsu L, Ochieng WY (2020) A gradient boosting decision tree based GPS signal reception classification algorithm. Appl Soft Comput. https://doi.org/10.1016/j.asoc.2019.105942

Sun R, Fu L, Wang G, Cheng Q, Hsu LT, Ochieng WY (2021a) Using dual-polarization GPS antenna with optimized adaptive neuro-fuzzy inference system to improve single point positioning accuracy in urban canyons. Navigation 68:41–60. https://doi.org/10.1002/navi.408

Sun R, Wang G, Cheng Q, Fu L, Chiang K, Hsu L, Ochieng WY (2021b) Improving GPS code phase positioning accuracy in urban environments using machine learning. IEEE Internet Things J 8:7065–7078. https://doi.org/10.1109/JIOT.2020.3037074

Sun R, Wang J, Cheng Q, Mao Y, Ochieng WY (2021c) A new IMU-aided multiple GNSS fault detection and exclusion algorithm for integrated navigation in urban environments. GPS Solutions. https://doi.org/10.1007/s10291-021-01181-4

Wang W, Liu Z, Xie R (2006) Quadratic extended Kalman filter approach for GPS/INS integration. Aerosp Sci Technol 10:709–713. https://doi.org/10.1016/j.ast.2006.03.003

Wang L, Groves PD, Ziebart MK (2015) Smartphone Shadow Matching for Better Cross-street GNSS Positioning in Urban Environments. J Navig 68:411–433. https://doi.org/10.1017/S0373463314000836

Xia Q, Rao M, Ying Y, Shen X (1994) Adaptive fading Kalman filter with an application. Automatica 30:1333–1338. https://doi.org/10.1016/0005-1098(94)90112-0

Xian Z, Hu X, Lian J (2013) Robust innovation-based adaptive kalman filter for INS/GPS land navigation. Chinese Automation Congress. pp 374–379

Xiao Z, Zhao P, Li S (2010) Adaptive fuzzy Kalman filter based on INS/GPS integrated navigation system. J Chin Inert Technol 18(195–198):203

Xu H, Angrisano A, Gaglione S, Hsu L (2020) Machine learning based LOS/NLOS classifier and robust estimator for GNSS shadow matching. Satell Navig 1:15. https://doi.org/10.1186/s43020-020-00016-w

Yan F, Li S, Zhang E, Chen Q (2020) An intelligent adaptive Kalman filter for integrated navigation systems. IEEE Access 8:213306–213317. https://doi.org/10.1109/ACCESS.2020.3040433

Yang C (2016) Sharpen the correlation peak: a novel GNSS receiver architecture with variable IF correlation navigation. J Inst Navig 63:249–265. https://doi.org/10.1002/navi.147

Yozevitch R, Ben Moshe B (2015) A robust shadow matching algorithm for GNSS positioning navigation. J Inst Navig 62:251. https://doi.org/10.1002/navi.117

Yozevitch R, Ben Moshe B, Weissman A (2016) A robust GNSS LOS/NLOS signal classifier navigation. J Inst Navig 63:429–442. https://doi.org/10.1002/navi.166

Zhang G, Wen W, Xu B, Hsu L (2020) Extending shadow matching to tightly-coupled GNSS/INS integration system. IEEE Trans Veh Technol 69:4979–4991. https://doi.org/10.1109/TVT.2020.2981093

Rui Sun is currently an associate professor in the College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing, Jiangsu Province, China. She has a Ph.D. in Intelligent Transport Systems (ITS) from the Imperial College Engineering Geomatics Group (ICEGG) at Imperial College London and holds a M.Sc. in Satellite Positioning Technology from the University of Nottingham. Her research interests include seamless positioning and navigation in challenging environments. She was elected as a Fellow of the Royal Institute of Navigation (FRIN) and Editorial Advisory Board (EAB) member of the Journal of Navigation in 2021.

Zixuan Zhang received a B.S. degree in Logistics Engineering from Civil Aviation Flight University of China, Guanghan, Sichuan, China, in 2020. She is currently pursuing a master's degree in Transportation Engineering at Nanjing University of Aeronautics and Astronautics. Her research interest includes GNSS sensor fusion algorithms for intelligent transportation system applications.

Qi Cheng received a B.S. degree and master's degree in Transportation Engineering from Nanjing University of Aeronautics and Astronautics, Nanjing, Jiangsu, China, in 2017 and 2020, respectively. His research interest includes GNSS sensor fusion algorithms for intelligent transportation system applications.

Washington Yotto Ochieng received a Ph.D. degree in Space Geodesy from the University of Nottingham, UK, in 1993. He is the Head of the Centre for Transport Studies, the Co-Director of the Institute for Security Science and Technology, and the Chair in Positioning and Navigation Systems, Imperial College London, London, UK. His research interests are in infrastructure resilience and security, user-centric intermodal mobility and the design of positioning and navigation systems for land, sea, air and space applications. Dr. Ochieng is a Fellow of the Royal Academy of Engineering.