Received: 1 February 2020 | Revised: 3 November 2020

DOI: 10.1002/navi.408

ORIGINAL ARTICLE

WILEY ION

Using dual-polarization GPS antenna with optimized adaptive neuro-fuzzy inference system to improve single point positioning accuracy in urban canyons

Rui Sun[1] | **Linxia Fu**[1] | **Guanyu Wang**[1] | **Qi Cheng**[1] | **Li-Ta Hsu**[2] | **Washington Yotto Ochieng**[1,3]

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

[2] Interdisciplinary Division of Aeronautical and Aviation Engineering, the Hong Kong Polytechnic University, Hong Kong

[3] Center for Transport Studies, Imperial College London, London SW7 2AZ, UK

Correspondence
Rui Sun, College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China
Email: rui.sun@nuaa.edu.cn

Funding information
National Natural Science Foundation of China, Grant/Award Numbers: 41974033, 41704022; Natural Science Foundation of Jiangsu Province, Grant/Award Number: BK20170780; Fundamental Research Funds for the Central Universities, Grant/Award Number: KFJJ20200703

Abstract

This paper builds on the machine learning research to propose two new algorithms based on optimizing the Adaptive Neuro Fuzzy Inference System (ANFIS) with a dual-polarization antenna to predict pseudorange errors by considering multiple variables including the right-hand circular polarized (RHCP) signal strength, signal strength difference between the left-hand circular polarized (LHCP) and RHCP outputs, satellites' elevation angle, and pseudorange residuals. The final antenna position is calculated following the application of the predicted pseudorange errors to correct for the effects of non-line-of-sight (NLOS) and multipath signal reception. The results show that the proposed algorithm results in a 30% improvement in the root mean square error (RMSE) in the 2D (horizontal) component for static applications when the training and testing data are collected at the same location. This corresponds to 13% to 20% when the testing data is from locations away from that of the training dataset.

KEYWORDS

ANFIS, dual-polarization antenna, firefly algorithm, genetic algorithm, GPS

1 | INTRODUCTION

Location-based services (LBS) underpinned by positioning, navigation, and timing (PNT) information are essential for the provision of services in urban areas. However, the de-facto system for most PNT applications', Global Navigation Satellite Systems (GNSS) such as GPS, performance could be significantly degraded in urban canyons, due to the effects of non-line-of-sight (NLOS) and multipath signal reception. NLOS and multipath can lead to code phase (or pseudorange) measurement errors in turn resulting in a single point positioning error of several tens of meters, unsuitable for many LBS applications (Groves & Jiang, 2013; Jiang & Groves, 2014). Several methods have been proposed to mitigate errors caused by NLOS and multipath effects including signal processing, antenna design, measurement-based modeling, and their combinations.

Signal processing methods exploit the fact that the characteristics of correlation functions of line-of-sight (LOS), NLOS, and multipath signals are different. Correlators are aimed at estimating the optimal approximation of the signal range (Groves, 2013; Heinrichs et al., 2004; Weill & Fisher, 2002). Some correlator-related and delay-locked loop (DLL) technologies have been proposed for

medium- to long-range multipath error mitigation, including the narrow correlator, high-resolution correlator, strobe correlator, shaping correlator, and multipath estimating delay lock loop (MEDLL) (Lee, 2002; McGraw & Braasch, 1999; Townsend & Fenton, 1994; Van Dierendonck et al., 1992; Van Nee et al., 1994; Weill & Fisher, 2002). However, short-range multipath error mitigation is difficult, and signal processing cannot mitigate NLOS effects.

Antenna design methods include using antenna array, choke-ring antenna, and other types of antennas to directly mitigate multipath effects at low elevation angles. Thornberg et al. (2003) combined multipath-limiting antenna with a high-zenith antenna to improve the rejection of ground multipath by the proposed antenna array system in dense urban areas. Choke-ring antenna can also attenuate those signals due to its special architecture (i.e., using a set of concentric rings around the antenna element). However, the approach does not work for indirect signals with high elevation angles (Groves et al., 2013). Besides, their bulky size and heavy mass make them inconvenient for applications.

Integrating GNSS observables, measurements, and satellite and signal information with other information sources is considered a measurement-based method. Soloviev et al. (2008) proposed an integrated algorithm of inertial measurement unit (IMU), laser scanner, and GPS to identify multipath affected GPS signals in the urban environment. Meguro et al. (2009) utilized an omnidirectional infrared camera with GNSS for environment detection to eliminate the invisible satellites caught by the receiver. However, visual sensors are also vulnerable to the impact of urban landscape obstacles and weather, resulting in reduced image recognition. Apart from sensor integration, recent research has explored the use of spatial information (e.g., 3D city models) to mitigate the effects of NLOS and multipath. Groves et al. (2011) have developed the shadow matching technique, in which a 3D city model is used to assist the detection of NLOS reception and then to improve positioning accuracy. The shadow matching technique is able to avoid the lack of available satellites caused by excluding NLOS signals and therefore to simulate the visibility of each satellite to find the best matching candidate location (Wang et al., 2013). Other research has addressed the calculation of an optimized user position based on the weighted average of the estimated candidate positions obtained from a comparison of the simulated and measured pseudoranges based on 3D city models (Gu et al., 2015; Hsu et al., 2016; Miura et al., 2015). These methods can improve the positioning accuracy by reducing the errors caused by NLOS to a certain extent with transferability and generic validity still to be demonstrated.

The quality of a signal received depends to a large extent on the physical environment proximate to the antenna. Combining the antenna-based with measurement-based methods could also be feasible to mitigate the effects of NLOS and multipath. Jiang and Groves (2014) used the variable signal strength difference, obtained from the left-hand circular polarized (LHCP) outputs and the right-hand circular polarized (RHCP) counterpart of the dual polarization antenna as a classifier for the NLOS/multipath signals. NLOS signals are determined based on the output value of the signal strength difference. The affected measurements are then excluded resulting (geometry permitting) in higher single point positioning accuracy. In this method, it is critical to determine the appropriate threshold for classification. Similarly, these single or multiple variables, including the carrier to noise ratio (C/N_0), elevation angle, pseudorange residuals, and the other derivatives or combinations, are used also to determine the status of pseudorange measurements in recent research (Deng, 2010; Hsu, 2017; Wang et al., 2015; Yozevitch et al., 2016).

Applying machine learning in the process of mitigating pseudorange measurement errors in GNSS positioning has been shown to have the potential to improve positioning accuracy by considering the factors that influence measurement errors. The machine learning methods used include support vector machine (SVM), decision tree, logistic regression, Naïve Bayes, Artificial Neuro Fuzzy Inference System (ANFIS), etc. (Phan et al., 2013; Socharoentum et al., 2016; Sun et al., 2020; Yozevitch et al., 2016). By considering the signal reception classification as a decision problem, Guermaha et al. (2018) used the dual-polarization antenna with the related variables to improve positioning accuracy. They proposed a decision tree-based GNSS signal classifier with the inputs of satellite elevation and signal strength difference obtained from RHCP and LHCP antennas. The paper claims, based on tests in urban environments, that the decision tree-based classifier employed delivers a classification accuracy of 99%.

However, to date, machine learning methods have been used for signal reception classification only in urban GNSS positioning applications (Hsu, 2017; Phan et al., 2013; Sun et al., 2020). Given that positioning accuracy is highly dependent on correct classification, it is important that the sources of error in the classification process are understood. This includes errors introduced from the other information sources used in the offline labeling phase of the machine learning algorithms (e.g., 3D city model, camera, or compass, etc.) (Hsu, 2017; Sun et al., 2019; Yozevitch et al., 2016). For example, the accuracy of the labeling with a camera or compass often depends on the cost (and thus quality) of the hardware, in addition to the fact

that camera-based approaches also have the weakness of attitude determination accuracy. Using the 3D city model in the offline labeling phase also has some weaknesses, including the: (i) need for significant computational resources for storage, updating, and offline processing; (ii) difficulty of generating accurate building borders for some special building shapes; and (iii) out-of-date mapping and omission of temporary signal obstructions, such as buses and lorries. The error introduced by using 3D city models, cameras, compasses, or other sensors in the labeling process for signal reception classification will inevitably result in final positioning errors. Clearly, in addition to geometry, a key issue in positioning accuracy is the ability to mitigate the corresponding measurement errors.

Building on the above, this paper develops new algorithms using dual-polarization GPS antenna with an optimized adaptive neuro-fuzzy inference system to improve single point positioning accuracy in urban canyons by employing the RHCP signal strength, signal strength difference between the LHCP and RHCP outputs, elevation angle, and pseudorange residuals as the algorithm's inputs. The dual-polarization antenna used in this paper is the ZYACP-L004 from Zhongyu Instrumentation Co., Ltd (specification in Appendix A to C). Unlike the traditional methods, which use machine learning algorithms to predict the signal reception classification (classification function of the machine learning), we use machine learning algorithms to predict the pseudorange errors (regression function of the machine learning). Single point positioning accuracy is then improved by applying the predicted corrections to pseudorange measurements before computing the position solution. The proposed positioning method avoids the errors arising from additional hardware or geospatial information during the labeling process and therefore addresses the limitations of current signal reception classification-based positioning approaches using machine learning. The contributions of the research are summarized as follows:

- Development of two improved ANFIS-based pseudorange error prediction models, including Genetic Algorithm-Adaptive Neuro-Fuzzy Inference System (GA-ANFIS) and Firefly Algorithm-ANFIS (FA-ANFIS), using RHCP signal strength, signal strength difference obtained from a RHCP and LHCP antenna, elevation angle, and pseudorange residual as the input variables.
- The two models are in turn used to improve positioning accuracy by applying the predicted corrections to pseudorange measurements before computing the position solution.

Based on static field tests in urban canyons, the results show that the proposed algorithm results in a 30% improvement in the Root Mean Square Error (RMSE) in the horizontal when the training and testing data are collected in the same location. This corresponds to 13% to 20% when the testing data is from locations away from that of the training dataset.

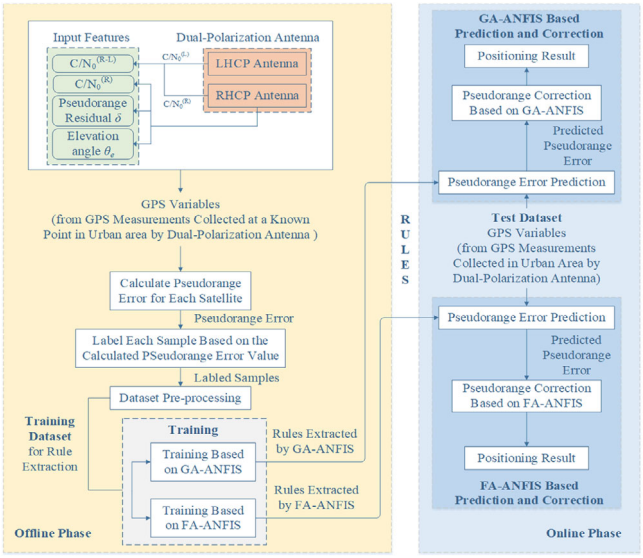

FIGURE 1 Algorithm framework of the proposed method [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com and www.ion.org]

2 | ALGORITHM DESIGN

2.1 | Algorithm framework

The framework of the proposed pseudorange correction algorithm with GA-ANFIS/FA-ANFIS and dual polarization antenna is presented in Figure 1. The framework includes an offline training phase and an online testing phase.

In the offline phase, the data used are GPS raw pseudorange measurements collected by the RHCP antenna of the dual-polarization antenna from a known point in the urban canyon. Some of the data from the known location in the urban canyon contain NLOS and/or multipath effects resulting in relatively large pseudorange errors. These errors are computed from the difference between the raw pseudoranges and the corresponding geometric ranges from the known station coordinates and satellite ephemeris.

Every set of variables at each epoch, including RHCP signal strength ($C/N_0^{(R)}$), signal strength difference obtained from the RHCP and LHCP antenna ($C/N_0^{(R-L)}$), elevation angle (θ_e), and pseudorange residual (δ), is then mapped to, or labeled with, the corresponding pseudorange error.

Here, the pseudorange error can be calculated by Equation (8) in Section 2.3. The details on training and testing datasets are given in Section 3. Training is used to extract the rules for different values of the input variables and the corresponding pseudorange errors accounting for the temporal changes in the visible satellites. The GA-ANFIS and FA-ANFIS algorithms are then used to fit the calculated pseudorange error by means of an offline dataset training process, thereby obtaining the rules, respectively, that is, the relationship between the input variables ($C/N_0^{(R)}$, $C/N_0^{(R-L)}$, θ_e, δ) and the corresponding labeled pseudorange errors. The main parts of the offline training process, including variable selection, details for the labeling process, and the GA-ANFIS and FA-ANFIS-based training process, are explained further in the subsequent sections.

In the online phase, new GPS variables from raw measurements of the dual polarization antenna in urban canyons, including $C/N_0^{(R)}$, $C/N_0^{(R-L)}$, elevation angle θ_e, and pseudorange residuals δ, are used together with the rules extracted from the offline phase to predict the pseudorange errors. Based on the predicted pseudorange errors, the positioning solutions are calculated based on the application of the predicted pseudorange errors as corrections to the new raw pseudoranges from the RHCP antenna.

2.2 | Variable determination

The raw measurements of GPS contain a variety of variables that can be used to determine the pseudorange error. Considering computational cost and training accuracy, we use the following four variables as the inputs for the proposed algorithms:

1. RHCP signal strength ($C/N_0^{(R)}$). The strength of the signal can be determined by the C/N_0 value of the signal received by RHCP with measurements affected by NLOS/multipath exhibiting lower signal strength compared to LOS signals (Gu et al., 2015). The GNSS satellite signal is RHCP with LOS signals received by RHCP antenna generally having high C/N_0 values. However, signals reflected by objects like glass walls also exhibit high C/N_0 values (Yozevitch et al., 2016). Therefore, more variables are needed to determine the pseudorange errors.
2. Signal strength difference between the LHCP and RHCP outputs ($C/N_0^{(R-L)}$). RHCP is more sensitive with LOS signals while LHCP is more sensitive with reflected ones. Therefore, reflected signals, in theory, have negative $C/N_0^{(R-L)}$ values. Jiang & Groves (2014) note that although it is possible for the NLOS/multipath signals to have positive $C/N_0^{(R-L)}$, this is usually lower than the required threshold. However, the probability of error at a higher elevation angle is low (Groves et al., 2013). Therefore, $C/N_0^{(R-L)}$ could also be used as an indicator for the pseudorange error prediction.
3. Elevation angle (θ_e). Reflected signals often are at a low or negative elevation angle (θ_e). According to Teunissen & Montenbruck (2017), θ_e is calculated by

$$\theta_{e(i)} = arctan \frac{\cos\alpha_{la}^R \cos\left(\alpha_{lon(i)}^S - \alpha_{lon}^R\right) - r^E/r_{(i)}^S}{\sqrt{1 - \left[\cos\alpha_{la}^R \cos\left(\alpha_{lon(i)}^S - \alpha_{lon}^R\right)\right]^2}}, \tag{1}$$

where $\alpha_{lon(i)}^S$ is the longtitude of satellite (i), α_{la}^R and α_{lon}^S are the latitude and longitude of receiver, respectively, r^E represents the earth's radius, and $r_{(i)}^S$ represents the orbital radius. In general, signals from satellites at a higher elevation are less likely to be blocked or reflected by buildings and are more likely to reach the receiver directly. Therefore, the elevation angle can be used as a feature to mitigate the NLOS/multipath. Weighting the measurements based on the elevation angle to reduce the effect of multipath is used widely for position determination. Therefore, the satellite elevation angle is adopted in this paper for pseudorange error prediction.
4. Pseudorange residual (δ). The relationship between the unknowns and pseudorange measurements is nonlinear. Hence, nonlinear least squares (NLSQ) estimation is used in which linearization is undertaken around approximate values of the unknowns ($\boldsymbol{x_o}$). The resulting lineriazed observation equation is expressed as $\boldsymbol{G\Delta x} = \boldsymbol{b} + \delta$ with $\Delta\boldsymbol{x}$ being the corrections applied to $\boldsymbol{x_o}$ to determine the final position and time. $\boldsymbol{G}$ is the design matrix, and $\boldsymbol{b}$ is the difference between observed and computed pseudoranges. The solution for $\Delta\boldsymbol{x}$ for unweighted measurements is expressed as

$$\Delta\boldsymbol{x} = (\boldsymbol{G}^T\boldsymbol{G})^{-1}\boldsymbol{G}^T\boldsymbol{b}. \tag{2}$$

The final position and time $\boldsymbol{x}$ is computed as

$$\boldsymbol{x} = \boldsymbol{x_o} + \Delta\boldsymbol{x}. \tag{3}$$

The measurement residual (δ), in effect representing the contribution of a measurement to the position solution is determined *a posteriori* as Equation (4):

$$\boldsymbol{\delta} = G\Delta\boldsymbol{x} - \boldsymbol{b}. \tag{4}$$

In urban canyons, most signals are contaminated by NLOS and/or multipath. In theory, the magnitude of the absolute value of the pseudorange residual is related

TABLE 1 Correlation strength between variables and pseudorange error

Variable	$C/N_0^{(R)}$	$C/N_0^{(R-L)}$	Elevation angle θ_e	Pseudorange residuals δ
C_{Spear}	−0.4934	−0.4318	−0.1523	0.4107
Correlation	Moderate	Moderate	Very Low	Moderate

to the degree of signal contamination. This property is exploited in this paper to augment those from the other three variables.

The strength of association between two variables can be measured according to the rank correlation coefficient. One of the most popular methods is the Spearman's correlation coefficient. It's not necessary for the Spearman's correlation coefficient to assume that the relationship between variables is linear (Hauke & Kossowski, 2011). Therefore, the correlation coefficient was chosen to measure the correlation between the input feature information and pseudorange errors.

$$C_{Spear} = \frac{\sum_{i=1}^{Q}(V_i - \bar{V})\left(\Delta\rho_i - \overline{\Delta\rho}\right)}{\sqrt{\sum_{i=1}^{Q}(V_i - \bar{V})^2 \sum_{i=1}^{Q}\left(\Delta\rho_i - \overline{\Delta\rho}\right)^2}}. \quad (5)$$

The Spearman's correlation coefficient of the four variables determined above can be calculated using Equation (5), where V represents the variable, $\bar{V}$ and $\overline{\Delta\rho}$ are the means of the variable and pseudorange errors, respectively, and Q represents the size of the whole dataset. The correspondence between its absolute value and the strength of association can be expressed as (Weir, 2016):

0.00-0.19: "Very Weak"
0.20-0.39: "Weak"
0.40-0.59: "Moderate"
0.60-0.79: "Strong"
0.80-1.00: "Very Strong"

The C_{Spear} results are shown in Table 1; a minus sign indicates a negative correlation while a plus sign indicates a positive correlation. The scatter plot with marginal distributions of the variables and pseudorange errors can be seen in Figure 2.

From the distributions in Figure 2, it is indicated that the correlation between any single variable and the pseudorange error is not very high and some 2D distributions even look bimodal. This may be due to the fast changing and scattering characteristics of multipath and NLOS in urban environments. Meanwhile, from the perspective of the correlation coefficient, the variable with the strongest correlation ($C/N_0^{(R)}$) is only "moderately correlated." According to the results, it is probably not reliable to use a single variable to predict the pseudorange error. Besides, as discuss earlier in the paper, the relationship between each variable and the actual type of signal received may also be inaccurate, which is inconsistent with the signal type obtained by numerical reasoning based on each single variable. Therefore, in order to address this problem and exploit any synergies and/or complementarities, this paper uses multiple variables to predict the pseudorange error.

2.3 | Pseudorange error labeling process

The labeling of the pseudorange error is critical in the offline training. The ranging errors result from the fact that the contaminated signal (i.e., multipath or NLOS) travels an additional route due to being reflected in the surrounding environment. These errors are typically a few tens of meters in urban canyons but can be larger if the signal is reflected by a remote tall building. With the knowledge of the ground truth, the pseudorange errors of the received signals can be calculated and the set of signal features or variables labeled with the psedorange error values.

The pseudorange observation equation $\tilde{\rho}_{(i)}$ for receiver R and satellite S(i) is given by

$$\tilde{\rho}_{(i)} = r_{(i)} + \left(\tau^R - \tau^S_{(i)}\right)c + D_{trop_{(i)}} + D_{iono(i)} + D_{orb(i)} + \rho_{sac(i)} + \varepsilon_i, \quad (6)$$

where the geometric range, $r_{(i)} = \sqrt{\left(X^S_{(i)} - X^R\right)^2 + \left(Y^S_{(i)} - Y^R\right)^2 + \left(Z^S_{(i)} - Z^R\right)^2}$;

$(X^S_{(i)}, Y^S_{(i)}, Z^S_{(i)})$ is the position of the satellite (i), and the (X^R, Y^R, Z^R) is the position of the GPS receiver. ε_i consists of the contribution to the range error of the effects of NLOS signal reception and multipath and observation noise. The receiver clock offset is calculated in the position solution as the fourth unknown. Satellite clock error can be corrected by using the corrections in the navigation message. ρ_{sac} is the satellite position error caused by earth's rotation, which is corrected using the Sagnac correction and its residual is negligible. D_{orb} reflects the error of broadcast ephemeris in the case of single point positioning. Ionospheric delay$D_{iono}$and tropospheric delay D_{trop} can be corrected by the Klobuchar model and Saastamoinen model, respectively; then Equation (6) is rewritten as follows:

$$\rho^c_{(i)} = r_{(i)} + \left(\Delta\tau^R - \Delta\tau^S_{(i)}\right)c + \Delta D^K_{trop} + \Delta D^S_{iono} + \Delta D_{orb} + \varepsilon_i, \quad (7)$$

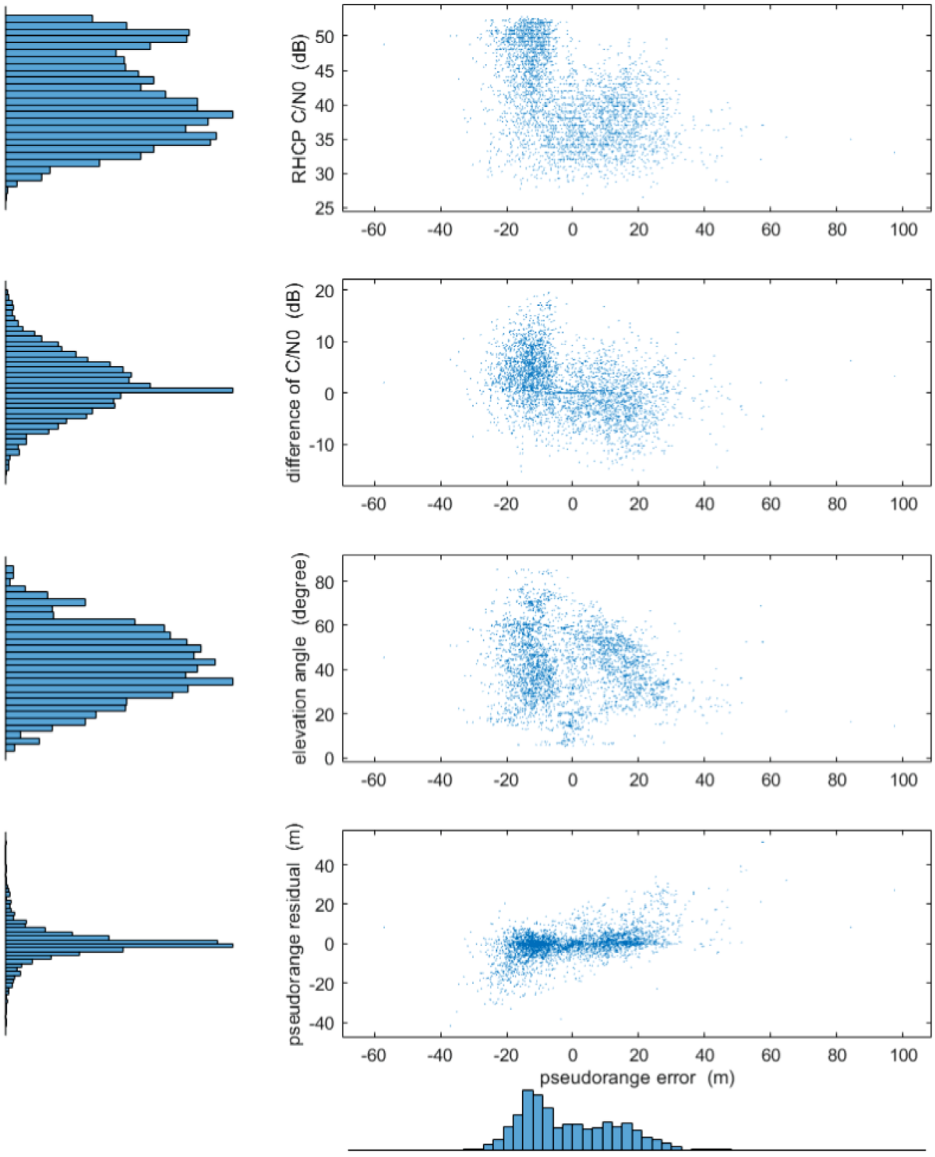

FIGURE 2 Scatter plot with marginal distributions [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com and www.ion.org]

where $\Delta\tau^R$, $\Delta\tau^S_{(i)}$, ΔD^K_{trop}, ΔD^S_{iono}, ΔD_{orb} are the errors that remain in turn included in pseudorange residuals. From Equation (7), pseudorange error $\Delta\rho$ can be expressed by

$$\Delta\rho_{(i)} = \rho^c_{(i)} - r_{(i)} = \left(\Delta\tau^R - \Delta\tau^S_{(i)}\right) c + \Delta D^K_{trop} + \Delta D^S_{iono} + \Delta D_{orb} + \varepsilon_i. \quad (8)$$

The rest of the terms in Equation (8) can be mitigated or reduced by using more accurate data and models (e.g., multiple frequencies and precise ephemeris). However, given the limitations of the current mitigation methods, the error caused by multipath/NLOS can reach tens of meters (Jiang & Groves, 2014) particularly in built environments, making it dominant.

Through an offline labeling process, we can then relate a pseudorange error $\Delta\rho$ to the corresponding set of variables $C/N_0^{(R)}$, $C/N_0^{(R-L)}$, elevation angle θ_e, and pseudorange residuals δ in the offline labeling phase.

2.4 | Positioning with ANFIS and pseudorange correction

The performance of the ANFIS method is critical for the pseudorange error prediction and therefore final positioning accuracy. ANFIS is the adaptive integration of neural network (NN) and fuzzy logic, which maintains the interpretability of fuzzy interference systems while enhancing the feedforward calculation and backpropagation learning

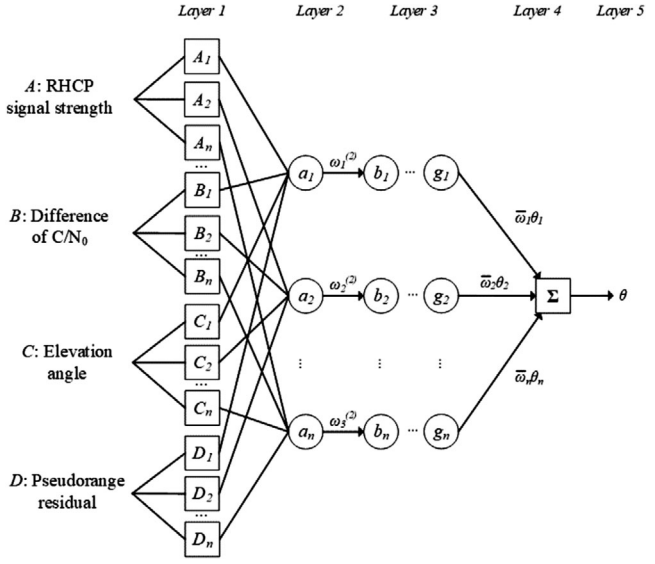

FIGURE 3 Architecture of ANFIS

capabilities of the system output (Ghomsheh et al., 2007; Jang et al., 1997). It has been demonstrated from our previous work that ANFIS could provide high accuracy in decision problems, such as signal reception classification (Sun et al., 2019). However, the nondeterministic design features of ANFIS make it difficult through a trial-and-error process to obtain a suitable architecture for a specific model, which can be resource intensive (Sun et al., 2020). Thus, combining ANFIS with other optimization algorithms such as Genetic Algorithm (GA) and Firefly Algorithm (FA) should enable a reduction in the processing time and increase prediction accuracy.

GA is a heuristic algorithm employing a search technique to arrive at optimal solutions by simulating the natural evolution process. Recently, the combination of genetic with machine learning algorithms has been applied in a wide range of applications, such as medical image registration and prediction of map expansion, modeling laser brazing (Goldberg & Holland, 1988; N'Diaye et al., 2017; Rong et al., 2016). The firefly algorithm is a swarm intelligence algorithm that mimics the flashing behaviour of fireflies. Chahnasir et al. (2018) combined FA with the support vector machine (SVM) for the shearing capacity estimation of angular shearing connectors. The work showed that FA could be used to optimize the prediction results of SVM.

With these potential benefits, GA-ANFIS- and FA-ANFIS-based algorithms for classification and pseudorange error prediction are proposed in this paper for the GPS single point positioning. The proposed GA/FA-ANFIS-based training structure is presented in Figure 3, where $[A_1A_2\cdots A_n]$, $[B_1B_2\cdots B_n]$, $[C_1C_2\cdots C_n]$, and $[D_1D_2\cdots D_n]$ are the inputs of each variable, n is the number of semantic labels (e.g., high, medium, and low). a_h, b_h, g_h represent the neurons of different layers, where $h = 1, 2, \ldots, n$. Layer 1 is the input layer, and the input training sample is represented as $I_l = (C/N_0^{(R)}{}_l, C/N_0^{(R-l)}{}_l, \theta_{el}, \delta_l)$, where $l = 1, 2, \ldots N$, N is determined by the size of the input training dataset. The training dataset labeled with the corresponding pseudorange errors is represented as $S = [(I_1, \Delta\rho_{(1)}), (I_2, \Delta\rho_{(2)}), \ldots, (I_n, \Delta\rho_{(n)})]$. The input training data is to be fuzzed in Layer 1, the input layer. In this layer, each input dimension is split by fuzzy membership functions (MF), which are usually Gaussian. The shape parameters of MFs are called the antecedent parameters of ANFIS (Ghomsheh et al., 2007). Layer *5* is the output layer - the total output of all input features information is de-fuzzed to obtain the exact output value in this layer. Layers 2 to 4 are the middle layers (also referred to as hidden layers) of ANFIS. In these layers, the rules are extracted from the fuzzed input data, and the corresponding firing strength of each rule is calculated and normalized. The conclusion parameters consist of each rule given to birth in these layers. $\omega_1^{(2)}$ is the output vector of neuron a_1, the subscript denotes the neuron, and the superscript represents the number of layers, e.g., $^{(2)}$ means Layer 2. θ_l is the output of the extracted rule while $\bar{\omega}_l$ is the corresponding normalized firing strength. The product of θ_l and $\bar{\omega}_l$ of each rule can be added to obtain $\sum$. The output of ANFIS θ corresponding to the input feature information is then obtained after the de-fuzzing process. Using a back-propagation algorithm, ANFIS can adjust the antecedent parameters and the conclusion parameters while learning. An intelligent algorithm, such as FA and GA, can be utilized to optimize the parameter tuning process (Ghomsheh et al., 2007). Therefore, the optimization problem of FA and GA for ANFIS rule extraction is also transformed into the optimization problem of ANFIS parameter set tuning. The flow of the two algorithms proposed are:

1. Step 1: Training dataset pre-processing. In order to facilitate the calculation of proposed algorithms and improve the accuracy of prediction, it is necessary to pre-process the dataset. In this paper, we use the variable normalization method. Each variable is normalized to have zero mean and unit variance. The function normalized is

$$w' = \frac{w - \mu}{\sigma}, \tag{9}$$

   where w is a variable vector, and μ and σ are the corresponding mean and variance.
2. Step 2: Generate initial fuzzy rules using the Fuzzy C-Means (FCM) clustering method. The advantage of FCM is the ability to model complex systems with

limited data (Dabbagh & Yousefi, 2019). Compared to the traditional cluster analysis, which is to strictly divide each element into a specific class, the fuzzy clustering method treats each cluster as a fuzzy set and determines the clustering relationship through the degree of membership, which is more flexible and accurate. The objective function of FCM clustering is

$$f_{obj} = \sum_{t=1}^{N}\sum_{s=1}^{k}\left[d\left(x_t, C_s\right)\right]^2 \times \left(M_{st}\right)^{\sigma}, \tag{10}$$

where N is determined by the size of the dataset, k is the number of cluster centers, $d(x_t, C_s)$ denotes the distance from t-th data x_t to the s-th cluster center C_s. Cluster centers are determined by finding the optimal data points in input data to define cluster centers based on the density of surrounding data points. All data points within the cluster influence range of a center are removed in order to determine the next data cluster and its center. This process is repeated until all of the data points fall within the influence range of a cluster center, then we can determine the magnitude of k. M_{st} is the membership matrix, which determines the degree of membership, and σ is the corresponding weighted index. Besides, the membership matrix satisfies the following conditions:

$$\sum_{s=1}^{k} M_{st} = 1. \tag{11}$$

The membership matrix can be obtained by using the Lagrange method and the constraint condition (11):

$$M_{st} = \frac{1}{\sum_{m=1}^{k}\left[\frac{d(x_t, C_s)}{d(x_m, C_s)}\right]^{-\frac{2}{\sigma-1}}}. \tag{12}$$

Letting the the partial derivative f_{obj} with respect to cluster center C_s be $\frac{\partial f}{\partial C_s}$ =0, the cluster center can be obtained using

$$C_s = \frac{\sum_{t=1}^{N}\left(M_{st}\right)^{\sigma} x_t}{\sum_{t=1}^{N}\left(M_{st}\right)^{\sigma}}. \tag{13}$$

The objective function then represents the weighted sum of squares from each data point to each cluster center.

3. Step 3: Rules extraction based on GA-ANFIS/FA-ANFIS. In this step, GA and FA are separately used to optimize the initial rules obtained in the previous step for best fuzzy inference rules, i.e., obtained the tuned parameter set.

1. GA-ANFIS

(1) Generate initial population according to the parameter set of the initial Fuzzy Inference System (FIS) generated in step 2. Based on the input sample, give every individual a position and evaluate its cost by the cost functions f_{obj}, which are obtained from initial FIS, as shown in Equation (14). The concept of "the initial population" is similar to the one in the particle swarm optimization algorithm (PSO). And the concept of "individual" is similar to the one of a particle in the PSO. Position here refers to the location of individuals in the search space, whose meaning is different from the concept of the position obtained by GNSS positioning. Represent the i-th individual in the population as $\vec{pop}(i) = (position_{pop(i)}, cost_{pop(i)})$, where $i = 1, 2, \ldots, M$, M is the size of population. Each individual of the population is given a randomly assigned position (denoted by $position_{pop(i)}$) according to the search space, which is determined by the parameter bounds of the initial FIS.

$$cost_{pop(i)} = f_{obj}\left[position_{pop(i)}\right]. \tag{14}$$

Save the cost value of each individual denoted by ***cost***, and sort the individuals in ascending order based on the cost value.

$$\mathbf{cost} = \left(\text{cost}_{pop(1)}, \text{cost}_{pop(2)}, \ldots, \text{cost}_{pop(i)}, \ldots, \text{cost}_{pop(M)}\right). \tag{15}$$

Let the first individual $pop(1)$ be the current optimal solution, denoted by $bestsol$, and the worst cost is from the last individual $pop(n)$, that is

$$bestsol = pop\,(1) \tag{16}$$

$$worstcost = f_{obj}\left[position_{pop(M)}\right] = cost_{pop(M)}, \tag{17}$$

where M is the quantity of individuals.

(2) Calculate the fitness value by

$$V_{fit_{pop(i)}} = -\frac{\beta \times cost_{pop(i)}}{worstcost}, \quad \boldsymbol{V}_{fit} = \left(V_{fit_{pop(1)}}, \ldots, V_{fit_{pop(n)}}\right), \tag{18}$$

where $\vec{V}_{fit}$ is the fitness value of the population, and β is the selection pressure parameter. Then normalize the vector $\boldsymbol{V}_{fit}$.

(3) Select parents using the principle of roulette based on the calculated fitness value. Performing a single crossover operation will select two individuals from the original population. The number of selected parents $M_c$depends on the crossover percentage p_c:

$$M_c = 2 \times \text{round}\left(\frac{M \times p_c}{2}\right). \tag{19}$$

The basic idea of the roulette principle is that the probability P_i of the individual being selected is proportional to its fitness value $V_{fit_{pop(i)}}$:

$$P_i = \frac{V_{fit_{pop(i)}}}{\left|\boldsymbol{V}_{fit}\right|}. \tag{20}$$

Calculate the cumulative probability according to the order of individuals. Simulate the selection process of a roulette wheel by generating a random number of ranges from zero to one, which can be regarded as a pointer to the roulette wheel.

(4) Execute a crossover process on the selected parents and create new individuals.

$$position^{c}_{pop(i1)} = \vec{\alpha}\Delta position^{+}_{pop(i1)} \left(1 - \vec{\alpha}\right) position_{pop(i2)} \tag{21}$$

$$position^{c}_{pop(i2)} = \vec{\alpha}\Delta position^{+}_{pop(i2)} \left(1 - \vec{\alpha}\right) position_{pop(i1)}, \tag{22}$$

where $\vec{pop}(i1)$ and $\vec{pop}(i2)$ are the selected parents, and $\vec{\alpha}$ is an array of random numbers chosen from the continuous uniform distribution on the interval from $-\gamma$ to $\gamma + 1$, where γ is the crossover index. Then evaluate the cost value of crossover individuals.

(5) Randomly select individuals to execute the mutation process and create new individuals. All individuals in a population are judged to be mutated with a predetermined probability, the mutation percentage p_m.
The total number of mutated individuals can be obtained using Equation (23):

$$M_m = round\left(Mp_m\right). \tag{23}$$

Based on p_m and the bounds of initial FIS parameters, the mutation process is performed by changing several randomly chosen elements of the selected individual $position_{pop(i3)}$.

$$position^{m}_{pop(i3)}(x) = position^{(x)}_{pop(i3)} + s * rand\left(size\left(x\right)\right) \tag{24}$$

$$s = 0.1 \times \left(B_{max} - B_{min}\right). \tag{25}$$

x represents the array made up by the ordinal numbers of the randomly selected elements. $rand(size(x))$ is used to return an array that contains values drawn from the standard normal distribution. The size of the array is subject to the quantity of the selected elements. B_{max} and B_{min} are the upper and lower bounds of initial FIS parameters. Then evaluate the cost value of mutated individuals.

(6) Merge population. Merge new individuals from crossover and mutation with the original population, as shown in Equation (26). According to the cost value, sort the individuals in $newpop$ from smallest to largest. Then update the $worstcost$ (cost value of the last individual that has the largest value among the new population) after sorting.

$$newpop = \left[pop\ pop^{c}\ pop^{m}\right], \tag{26}$$

where pop^c and pop^m are the new individuals produced by the crossover and mutation process, respectively. At the sorted new population $sortedpop$, the number of individuals M' is

$$M' = M + M_c + M_m. \tag{27}$$

(7) Truncate the $sortedpop$ to the original amount M and update the $bestsol$ (the individual that has the smallest cost value, i.e., $sortedpop(1)$)for the current iteration.

(8) Repeat steps ② to ⑦ until the iteration ends. The iteration number is set to 300 due to experience.

2. FA-ANFIS

(1) Generate the initial fireflies (which is similar to the "individual" in GA) according to the parameter set of the initial FIS generated in step 2. Based on the input sample, give every firefly a random position and evaluate its cost by the cost functions f_{obj}, which is obtained from initial FIS (see GA-ANFIS). The position and cost value of the j-th firefly in the m-th

iteration is denoted as $position_j(m)$ and $cost_j(m)$, respectively, $j = 1, 2, \ldots, M$. Let the cost value of the initial optimal solution *bestsol* be infinite, which is denoted by *bestcost*.

(2) Calculate the distance between fireflies. For example, the distance from the *i*-th firefly to the *j*-th firefly is

$$r_{ij} = \left\| position_i - position_j \right\|. \tag{28}$$

(3) Attraction coefficient calculation.

$$\beta = \beta_0 \times e^{-\Gamma \cdot r_{ij}^2}, \tag{29}$$

where β_0 is the attraction coefficient base value, and Γ is the light absorption coefficient.

(4) If $cost_j^{(}m) < cost_i^{(}m)$, then: update the firefly's position using

$$x_i'(m) = x_i(m) + \beta \times \left(x_j(m) - x_i(m)\right) + \varepsilon. \tag{30}$$

ε depends on the bounds of FIS parameters and can be obtained using Equation (31):

$$\varepsilon = 0.05r \times (B_{max} - B_{min}). \tag{31}$$

r is a random number chosen from the continuous uniform distribution on the interval from -1 to 1.

(5) Calculate the corresponding cost value $y_j'(m)$ and compare it with the current best cost. If $y_j'(m) < bestcost$, then

$$bestcost = y_j'(m). \tag{32}$$

(6) Repeat steps ② to ⑤ until the iteration ends. The iteration number is set to 300 from experience. Set the ultimate *bestcost* to be the *bestsol*.

The *bestsol*, consisting of a series of parameters, of the population (GA-based) or firefly (FA-based), is chosen as the ultimate parameter set to create the optimized FIS model for the purpose of pseudorange error prediction. The setting of GA parameters is determined by parameter testing, which, for example, is made by manually specifying the value of β to obtain the most suitable one. Many parameters in the GA are also obtained by this method, so are the γ, p_c, and p_m. The FA parameters, Γ and β_0, are also determined by this way.

4. Step 4: pseudorange error prediction. Once the final predictor $f_M(\boldsymbol{x})$ (representing the FIS model optimized by the GA/FA-ANFIS method) is obtained, the corresponding pseudorange errors of the newly collected variables from GPS measurements can be predicted.

$$f_M(\boldsymbol{x}) = \theta, \tag{33}$$

where the definition of θ is already given in Section 2.4. The input $\boldsymbol{x} = (C/N_0^{(R)}, C/N_0^{(R-L)}, \theta_e, \delta)$ is used together under the rules to predict the pseudorange errors for each observed satellite.

5. Step 5: position calculation with pseudorange correction. The newly collected pseudorange measurements are corrected in Equation (33) by subtracting the predicted pseudorange error by GA-ANFIS or FA-ANFIS in step 4.

$$\rho_{(i)}^c = \rho_{(i)} - \Delta\rho_{(i)}, \tag{34}$$

where $\rho_{(i)}^c$ is the corrected pseudorange of the ith signal, and $\Delta\rho_{(i)}$ is the predicted pseudorange error of the *i*-th signal. With the corrected pseudorange measurements, the NLSQ is used to compute the antenna position (see Section 2.2).

3 | TEST AND RESULT ANALYSIS

To test and validate the proposed algorithm, we collected 24 hours of data using two NovAtel OEM6 geodetic receivers with a Zhongyu ZYACF-L004 dual polarization antenna in urban canyons in Hong Kong for a static case. During the labeling process, the pseudorange errors are calculated using Equation (8) and labeled with their values. In particular, the pseudorange errors within 5m are considered as small and therefore from LOS signals. The others are considered as large pseudorange errors, mainly from NLOS/multipath signals. The 5m threshold is determined from experience. The 24-hour data (noted as D_0) was collected from 13:13 on 16 May to 14:11 on 17 May (Beijing Time), containing a total of 388,022 samples. The testing dataset (noted as $D_{testing}$), containing 46,805 samples with large pseudorange errors and 8,689 with small pseudorange errors within the dataset captured from 14:00 to 17:00 on 16 May, were extracted from D_0. With $D_{testing}$ excluded, the training dataset (noted as $D_{training}$), containing randomly chosen 30,000 samples with large pseudorange errors and 12,000 samples with small pseudorange errors, was extracted from D_0.

Based on the pseudorange errors predicted by GA-ANFIS and FA-ANFIS with the dual polarization antenna (noted as GADP and FADP in the following tables and figures), the positioning results from the proposed methods are compared with the following four methods with the abbreviations noted in the following brackets, respectively:

TABLE 2 Comparison of positioning accuracy in Scenario 1

RMSE (m)	E	N	U	2D	3D
CSPP	**33.46**	**28.57**	**112.14**	**44.00**	**120.46**
CSPP-LR	**27.31**	**31.54**	**143.49**	**41.72**	**149.44**
FAR	**29.86**	**21.61**	**95.21**	**36.86**	**102.10**
GAR	**29.09**	**21.90**	**98.52**	**36.42**	**105.03**
FADP	**26.90**	**19.22**	**82.92**	**33.06**	**89.27**
GADP	**25.29**	**16.65**	**76.19**	**30.28**	**81.99**

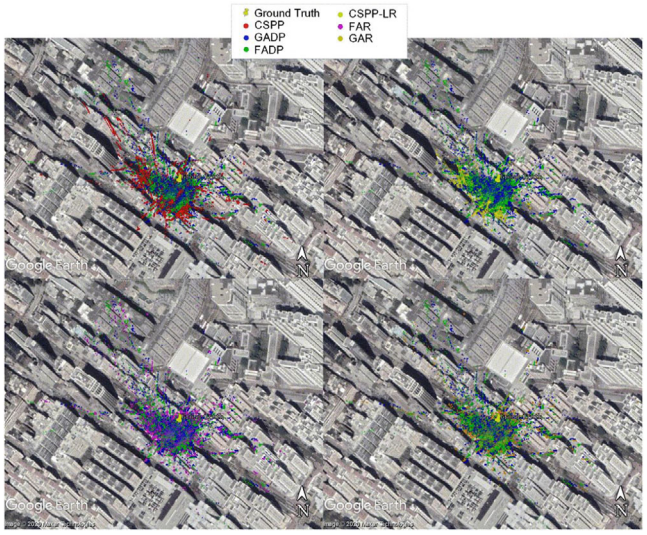

FIGURE 4 Positioning results of Scenario 1 [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com and www.ion.org]

- Conventional single point positioning (CSPP) method, i.e., positioning with outlier detection and exclusion from Biagi & Caldera (2013), which uses the efficient leave one block out (ELOBO) approach to identify outliers and exclude them from the positioning process.
- Conventional single point positioning using the LHCP-RHCP C/N_0 difference and satellite elevation angle to select and weight the measurements. (CSPP-LR)
- FA-ANFIS using RHCP measurement data only. (FAR)
- GA-ANFIS using RHCP measurement data only. (GAR)

RMSE is used as the metric to evaluate the positioning accuracy of positioning methods proposed; see Table 2 and Figure 4.

The results of CSPP, FADP, GADP, CSPP-LR, FAR, and GAR are depicted in Figure **4**. It is clear that the positioning results for the proposed algorithms are much closer to the ground truth than the conventional positioning results in the testing case.

It can be seen from the results shown in Figure **5** that, in partial epochs, the positioning results of the CSPP-LR (orange) are even worse than the one of CSPP (red). This may be due to the insufficient number of satellites in the urban environment, which make the CSPP-LR not work well. The interplay between measurement error and geometry built environments may also cause this problem. FAR and GAR performed a little better than CSPP, but are inferior to the multi-feature methods. In general, the GADP-based one performed best. Figure **6** shows the number of satellites received by each dual-polarized antenna. In most epochs, the number of satellites received in the LHCP antenna is less than the RHCP antenna. But it is not small, which implies that the degree of contamination of satellite signals by multipath effects is serious in this environment; the C/N_0 curves of RHCP and LHCP shown in Figure **8** can also illustrate this.

However, in terms of vertical positioning accuracy, almost all the RMSE results of the methods exceed 50m. According to the document on "Explanatory Notes on Geodetic Datums in Hong Kong" issued by the Lands Department of Hong Kong, the difference of semi-major axis between HK80 and WGS84 is more than 200m, and the coordinate origin selected by the two is also different. Considering that GPS takes WGS84 as the reference ellipsoid, the reason for the unsatisfactory elevation positioning results may be related to the different reference ellipsoid and elevation datum.

In the case of single feature information input (i.e., RHCP only), GA and FA showed similar performance in positioning accuracy optimization. However, GA performs better under multiple features, which may be caused by the defect of the FA algorithm. According to Equation (30), when two fireflies get too close, a large number of fireflies will gather in the local optimum and easily miss the global optimum. Therefore, it may lead to the consequence that the FA had an inferior effect on FIS parameter optimization than the GA when the feature dimension is increased.

From the results, GADP and FADP can deliver overall 2D positioning accuracies (RMSE) of about 30m and 33m, respectively, corresponding to the improvement of 31.2% and 24.9% on the conventional positioning method. It can also be seen from Figure **7** that the PDOP value is generally higher than the HDOP one, which also indicates that the accuracy of the 3D positioning result will be relatively weaker than that of the 2D. In particular, due to the street direction and tall buildings on the north part of the testing point, the improvement in the northing is significantly higher than the easting, which indicates that the impact of the physical environment is accounted for by the proposed algorithms.

For this testing dataset, positioning improvements using the proposed algorithm are also shown in Figure **10**. Using the proposed GA-ANFIS algorithm, most of the positioning results are concentrated around 15m, while the results using CSPP are mostly above 27m. Although CSPP-LR

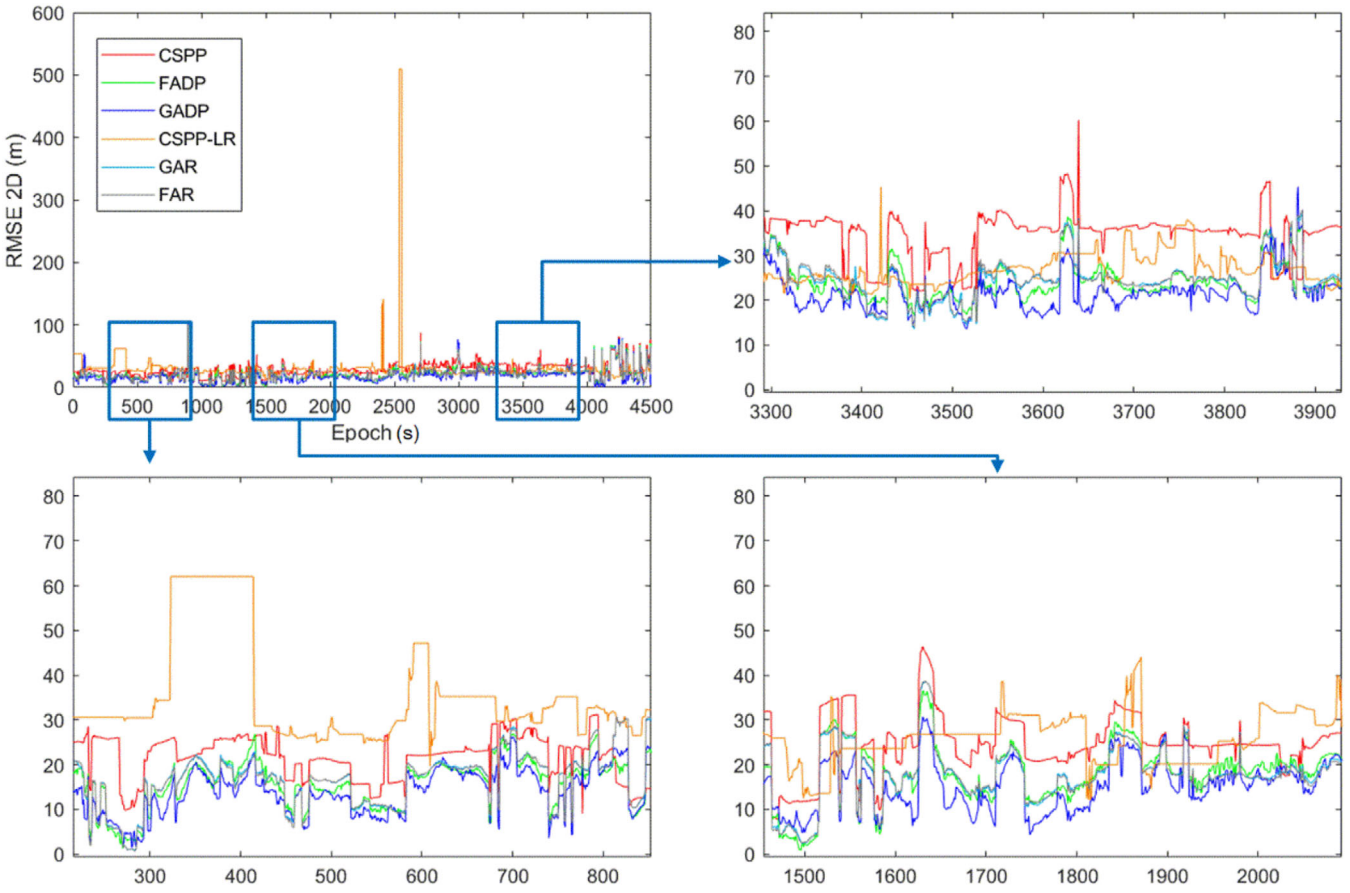

FIGURE 5 Example of 2D positioning error of the candidate methods [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com and www.ion.org]

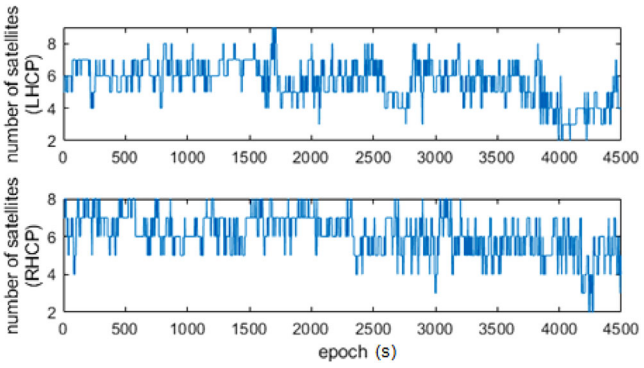

FIGURE 6 Number of satellites [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com and www.ion.org]

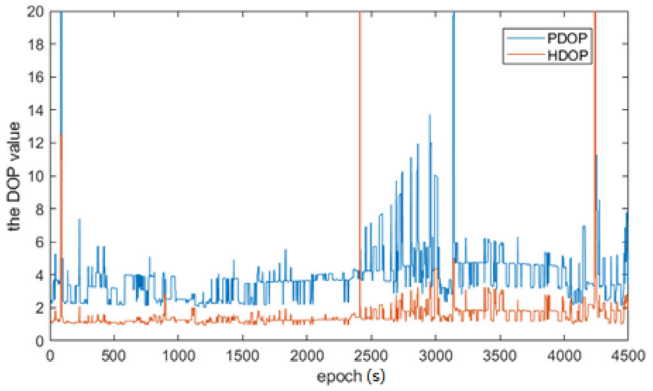

FIGURE 7 HDOP and PDOP values of Scenario 1 [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com and www.ion.org]

reduces the proportion of errors over 36m, the error is still concentrated at about 30m. The improvement from FADP is similar to that of GADP, but slightly inferior to that of GADP. FAR and GAR methods, although more primitive than the rest, compared with the conventional methods, their results are still significantly improved. Overall, the results positioning accuracy is improved by the proposed methods compared with the conventional positioning methods, especially the GADP.

In order to further investigate the performance of the proposed algorithms, the positioning accuracy compared with the conventional method is analyzed according to each epoch in Table 3. This shows that about 85% and

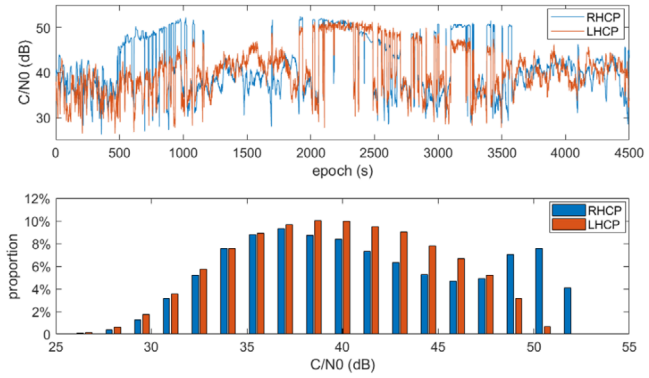

FIGURE 8 C/N_0 distribution of Scenario 1 [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com and www.ion.org]

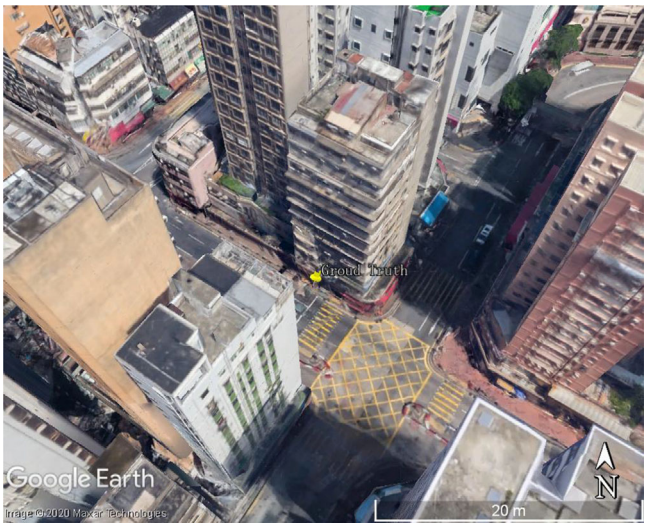

FIGURE 9 The test environment of Scenario 1 (3D city model) [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com and www.ion.org]

77% of the epochs are improved in the 2D- and 3D-positioning results based on the proposed algorithms. The general improvement from GA-ANFIS is better than FA-ANFIS. For 2D positioning, although around 14% to 21% of the epochs deteriorated due to the GA-ANFIS prediction errors, the proposed algorithm is still effective for most of the epochs. In addition, from the experiments, we find that the computation time of FA-ANFIS is longer than GA-ANFIS (see Table 4). This is due to the weakness of the firefly algorithm as the objective function information may not be fully used during the optimization due to the uncontrolled moving distance of fireflies. Moreover, FA-ANFIS need a shorter time in conducting pseudorange error prediction of a single epoch, only one-third of the one required by GA-ANFIS. If other algorithms are used to optimize FA and improve its training speed in future research, then this may become an advantage of FA-ANFIS in practical application.

TABLE 3 Algorithm performance evaluation with proportion of epochs (Scenario 1)

Proportion of epochs (%)	3D		2D	
	Better	Worse	Better	Worse
FADP	76.56	23.44	84.67	15.33
GADP	78.61	21.39	85.58	14.42

TABLE 4 Time required for training and testing

Execution content	Time consumed /s
GA-ANFIS Training	1058.552
GA-ANFIS Testing (Single Sample)	0.018
FA-ANFIS Training	1648.236
FA-ANFIS Testing (Single Sample)	0.005

To analyze the spatial transferability of the proposed method away from the training location, data from two different locations in the urban environment of Hong Kong were captured and processed. The environments for Scenario 2 and 3 are shown in Figures 13 and 17, respectively. The corresponding positioning results are shown in Figures 14 and 18. The DOP values are shown in Figures 11 and 15. C/N_0 distribution are shown in Figures 12 and 16.

As can be seen from the 3D model of the city in Figures 13 and 17, both Scenario 2 and Scenario 3 are typical urban environments. In Scenario 2, occlusion exists on all sides, but the height of the occlusion is obviously lower than the counterpart in Scenario 3. In Scenario 3, there is a wide driveway in the east direction, so occlusion in the north direction is more severe.

From Figure 19, it can be seen that the distance between Scenarios 1 and 2 are closer than the one between Scenarios 1 and 3. Under the environments of Scenario 2, the 2D positioning accuracy improvement is around 15% while 3D is around 24%. As for Scenario 3, it is 12% and 5%. Compared to CSPP-LR, the proposed algorithm is worse in the U-axis, resulting in a worse performance in 3D positioning. However, in terms of 2D RMSE improvement, the proposed algorithm is better than the other candidate algorithms in both scenarios. It is notable that the CSPP-LR has a much smaller vertical error than the other methods in Scenarios 2 and 3, but a significantly larger vertical error in Scenario 1. The CSPP-LR is essentially based on weighting based on C/N_0 and the elevation angle. In general, in open sky, the higher the elevation angle, the smaller the pseudorange error. Therefore, positioning accuracy is improved by considering the elevation angle as a variable. However, in the city areas, the relationship between the elevation angle

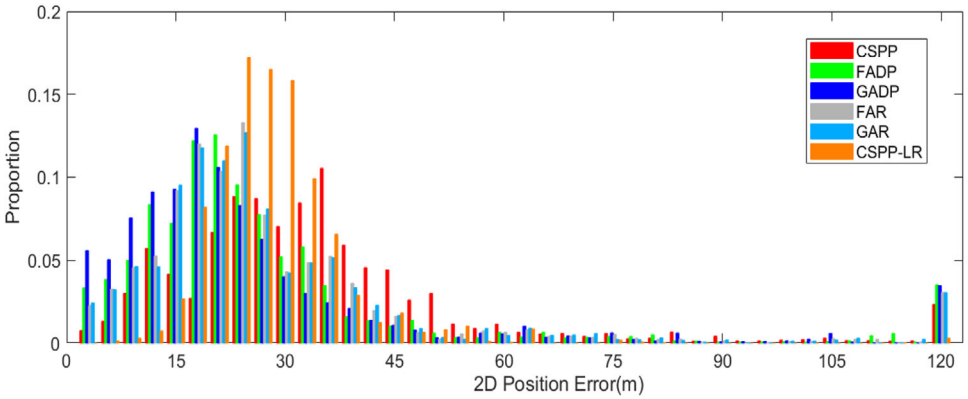

FIGURE 10 Positioning accuracy histogram of Scenario 1 [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com and www.ion.org]

TABLE 5 Comparison of positioning accuracy for Scenario 2

RMSE (m)	E	N	U	2D	3D
CSPP	**40.92**	**17.90**	**79.01**	**44.67**	**90.76**
CSPP-LR	**27.57**	**23.92**	**24.12**	**36.50**	**43.75**
FAR	**38.30**	**16.21**	**59.38**	**41.59**	**72.50**
GAR	**36.64**	**15.89**	**59.86**	**39.94**	**71.96**
FADP	**36.27**	**15.11**	**56.98**	**39.30**	**69.21**
GADP	**33.13**	**13.70**	**58.61**	**35.85**	**68.96**

and the corresponding pseudorange error is less deterministic. A decrease of a 5-degree elevation angle could introduce a pseudorange error of more than 100m, due to the NLOS in the urban areas. In addition, the relationship between C/N_0 and the pseudurorange error is not deterministic either in urban areas. The range error due to signal reflection could result in large errors. Therefore, weighting based on the combination of C/N0 and the elevation angle is less effective in heavily built urban environments. In Scenario 1, the obstruction due to the tall buildings in a narrow street is the most severe of the three scenarios, reflecting the most severe NLOS signals. From Figures 7, 11, and 15, the VDOP (estimated from PDOP and HDOP) is the worst for Scenario 1. Therefore, the vertical error estimated from CSPP-LR is the largest of the three scenarios.

In addition, the performance of FAR and GAR is inferior to FADP and GADP on both 2D and 3D. Although in these scenarios, the positioning accuracy is not improved as much as in Scenario 1, overall positioning accuracy is better nonetheless. This shows that the proposed algorithm can improve positioning accuracy in the cases where the locations for training and testing are different. In the future work, we will consider integrating the weighted method with machine learning and complementing each other to obtain more accurate positioning solutions.

TABLE 6 Comparison of positioning accuracy for Scenario 3

RMSE (m)	E	N	U	2D	3D
CSPP	**36.45**	**55.27**	**40.52**	**66.21**	**77.62**
CSPP-LR	**20.51**	**60.92**	**25.84**	**64.28**	**69.28**
FAR	**34.87**	**53.06**	**37.02**	**63.49**	**73.49**
GAR	**34.66**	**53.43**	**37.36**	**63.69**	**73.84**
FADP	**32.65**	**48.89**	**42.03**	**58.79**	**72.27**
GADP	**31.73**	**47.92**	**48.38**	**57.48**	**75.12**

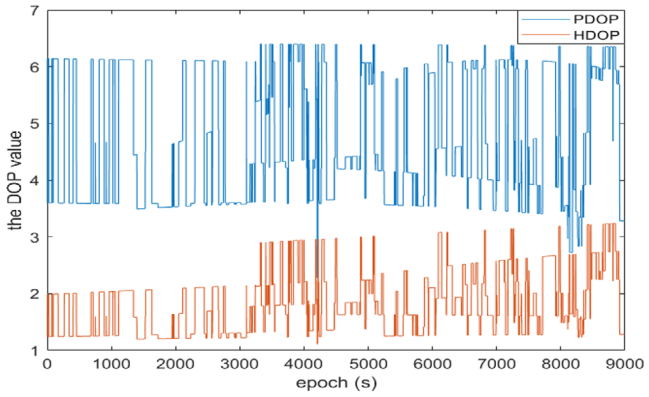

FIGURE 11 HDOP and PDOP Values of Scenario 2 [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com and www.ion.org]

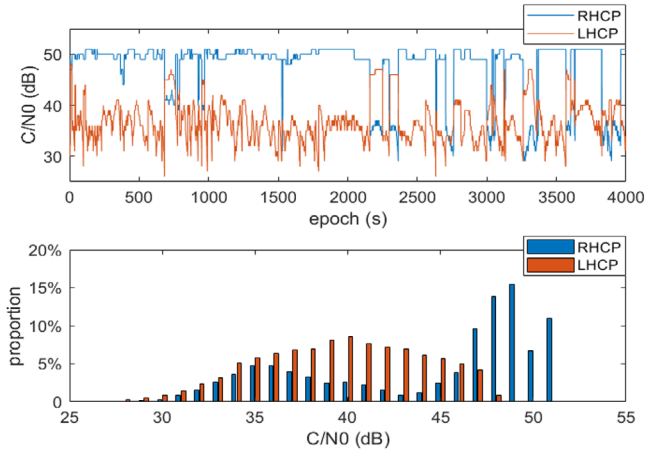

FIGURE 12 C/N_0 Distribution of Scenario 2 [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com and www.ion.org]

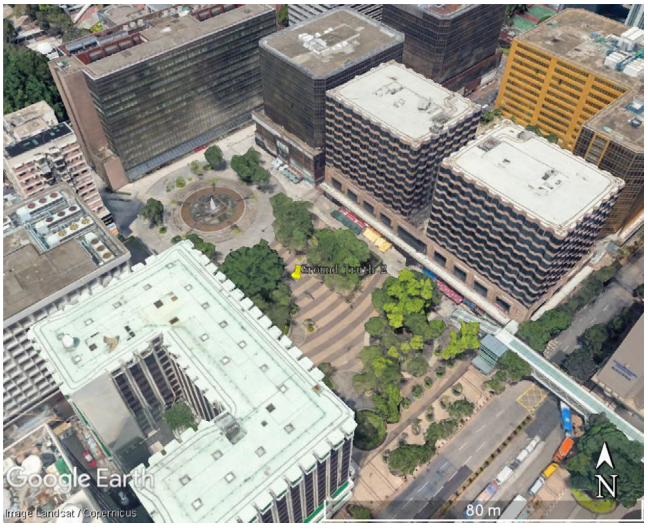

FIGURE 13 Experimental environment of Scenario 2 [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com and www.ion.org]

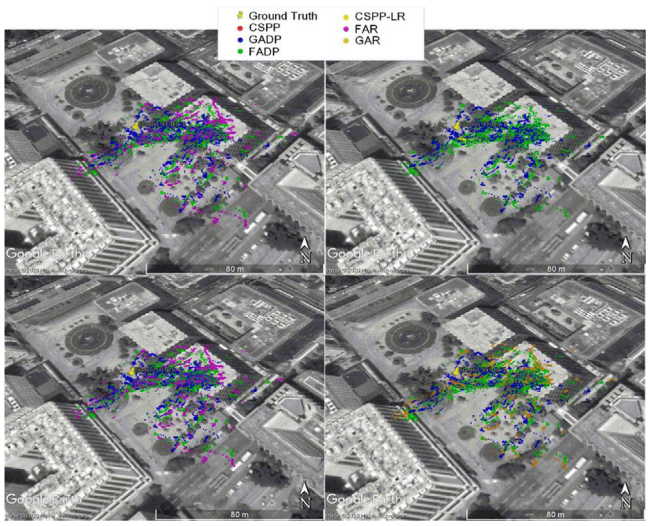

FIGURE 14 Positioning results for Scenario 2 [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com and www.ion.org]

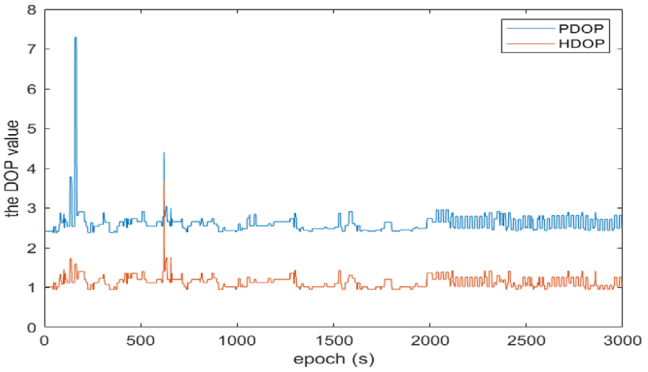

FIGURE 15 HDOP and PDOP values of Scenario 3 [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com and www.ion.org]

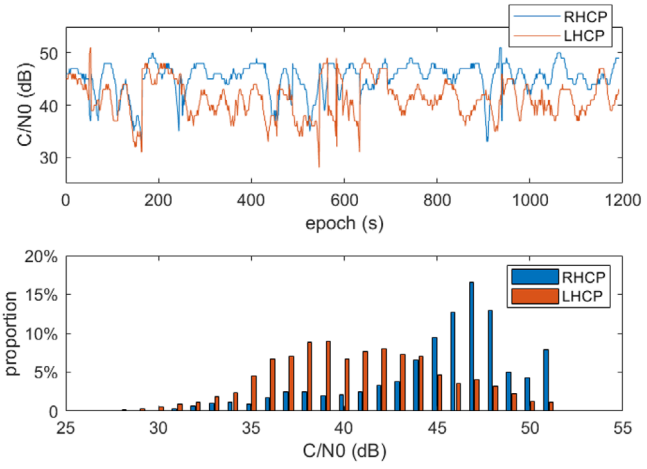

FIGURE 16 C/N_0 Distribution of Scenario 3 [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com and www.ion.org]

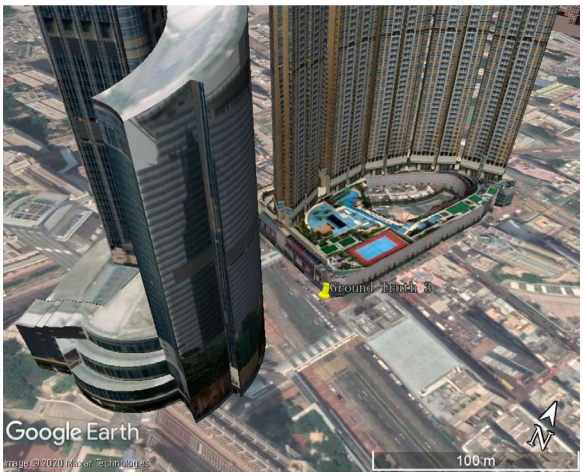

FIGURE 17 Experimental environment of Scenario 3 [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com and www.ion.org]

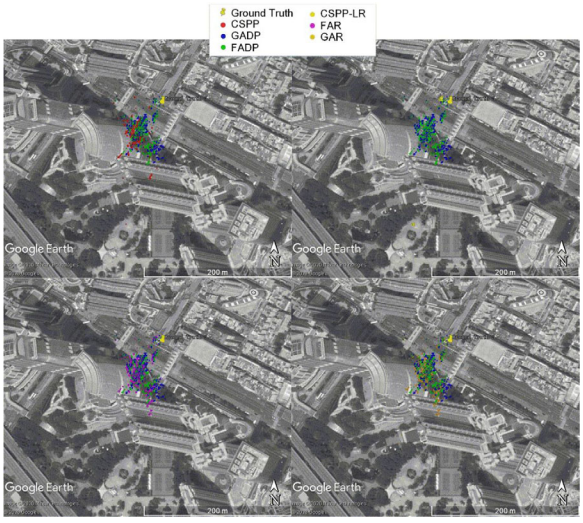

FIGURE 18 Positioning results for Scenario 3 [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com and www.ion.org]

4 | CONCLUSIONS

This paper has proposed a pseudorange error correction algorithm based on GA-ANFIS/FA-ANFIS using the dual-polarization antenna to improve positioning accuracy in built environments. The pseudorange error features or variables of signal strength, signal strength difference between RHCP and LHCP, elevation angle, and pseudorange residual are used to classify signal reception and predict pseudorange errors. The predicted errors are in turn employed as corrections to new measurements to improve positioning accuracy. The results show that the algorithms proposed outperform the conventional least squares-based all-in-view solutions. The results also show an overall improvement in the 3D-positioning accuracy (RMSE) from about 120 m with conventional positioning approaches to about 89m using the proposed FA-ANFIS-based pseudorange correction approach, an improvement about of 26%. The corresponding values for GA-ANFIS are 120m to 82m, an improvement of more than 30%. The 2D positioning accuracy (RMSE) for the pseudorange error correction-based positioning with GA-ANFIS (compared with the conventional) is from about 44 m to 30 m, an improvement over 30%.

Further research is exploring the application of the methods proposed in dynamic applications in urban environments. With the training data obtained from the reference points offline, users will automatically obtain the rules online via communication links to nearby reference points for the accurate pseudorange error correction to improve positioning accuracy. This online data training mechanism, i.e., with frequently updated rules, will also be developed for users with real-time or post-processing applications, including robust ground vehicle or pedestrian trajectory estimation. We are also exploring the added value of various aspects of multi-constellation, multi-frequency, and multi-sensor/system solutions.

FIGURE 19 Location of three scenarios [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com and www.ion.org]

ACKNOWLEDGEMENT

This work was supported in part by the sponsorship of the National Natural Science Foundation of China (Grant No. 41974033, 41704022), Natural Science Foundation of Jiangsu Province (Grant No. BK20170780) and the Fundamental Research Funds for the Central Universities (Grant No. KFJJ20200703).

ORCID

Rui Sun https://orcid.org/0000-0003-2252-9944
Li-Ta Hsu https://orcid.org/0000-0002-0352-741X

REFERENCES

Baseri, H., & Belali-Owsia, M. (2017). A novel hybrid ICA-ANFIS model for prediction of manufacturing processes performance. *Proceedings of the Institution of Mechanical Engineers, Part E: Journal of Process Mechanical Engineering*, *231*(2), 181–190. https://doi.org/10.1177/0954408915585256

Biagi, L., & Caldera, S. (2013). An efficient leave one block out approach to identify outliers. *Journal of Applied Geodesy*, *7*(1), 11–19. https://doi.org/10.1515/jag-2012-0030

Chahnasir, E. S., Zandi, Y., Shariati, M., Dehghani, E., Toghroli, A., Mohamad, E. T., … Khorami, M. (2018). Application of support vector machine with firefly algorithm for investigation of the factors affecting the shear strength of angle shear connectors. *Smart Structures and Systems*, *22*(4), 413–424. https://doi.org/10.12989/sss.2018.22.4.413

Dabbagh, R., & Yousefi, S. (2019). A hybrid decision-making approach based on FCM and MOORA for occupational health and safety risk analysis. *Journal of Safety Research*, *71*, 111–123. https://doi.org/10.1016/j.jsr.2019.09.021

Deng, G. (2010). GPS satellite selection algorithm based on satellite elevation angle and GDOP. *Digit. Commun.*, *37*(2), 47–50.

Ghomsheh, V. S., Shoorehdeli, M. A., & Teshnehlab, M. (2007, June). Training ANFIS structure with modified PSO algorithm. *2007 Mediterranean Conference on Control & Automation*, Athens, Greece, 1–6. https://doi.org/10.1109/MED.2007.4433927

Goldberg, D. E., & Holland, J. H. (1988). Genetic algorithms and machine learning. *Machine Learning*, *3*, 95–99. https://doi.org/10.1023/A:1022602019183

Groves, P. D. (2011). Shadow matching: A new GNSS positioning technique for urban canyons. *The Journal of Navigation*, *64*(3), 417–430. https://doi.org/10.1017/S0373463311000087

Groves, P. D. (2013). *Principles of GNSS, inertial, and multisensor integrated navigation systems*. Artech House.

Groves, P. D., & Jiang, Z. (2013). Height aiding, C/N0 weighting and consistency checking for GNSS NLOS and multipath mitigation in urban areas. *The Journal of Navigation*, *66*(5), 653–669. https://doi.org/10.1017/S0373463313000350

Groves, P. D., Jiang, Z., Rudi, M., & Strode, P. (2013, September). A portfolio approach to NLOS and multipath mitigation in dense urban areas. *Proceedings of the 26th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS+ 2013)*, Nashville, TN, 3231–3247.

Gu, Y., Hsu, L. T., & Kamijo, S. (2015). GNSS/onboard inertial sensor integration with the aid of 3-D building map for lane-level vehicle self-localization in urban canyon. *IEEE Transactions on Vehicular Technology*, *65*(6), 4274–4287. https://doi.org/10.1109/TVT.2015.2497001

Guermah, B., El Ghazi, H., Sadiki, T., & Guermah, H. (2018, November). A robust GNSS LOS/multipath signal classifier based on the fusion of information and machine learning for intelligent transportation systems. *2018 IEEE International Conference on Technology Management, Operations and Decisions (ICTMOD)*, Marrakech, Morocco, 94–100. https://doi.org/10.1109/ITMC.2018.8691272

Hauke, J., & Kossowski, T. (2011). Comparison of values of Pearson's and Spearman's correlation coefficients on the same sets of data. *Quaestiones Geographicae*, *30*(2), 87–93. https://doi.org/10.2478/v10117-011-0021-1

Heinrichs, G., Lemke, N., Schmit, A., Neubauer, A., Kronberger, R., Rohmer, G., … Ehm, H. (2004, August). Galileo/GPS receiver architecture for high sensitivity acquisition. *Proceedings of the International Symposium on Signals, Systems, and Electronics (ISSSE'04)*, Linz, Austria.

Hsu, L. T. (2017, October). GNSS multipath detection using a machine learning approach. *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*, Yokohama, Japan, 1–6. https://doi.org/10.1109/ITSC.2017.8317700

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

Hsu, L. T., Tokura, H., Kubo, N., Gu, Y., & Kamijo, S. (2017). Multiple faulty GNSS measurement exclusion based on consistency check in urban canyons. *IEEE Sensors Journal*, *17*(6), 1909–1917. https://doi.org/10.1109/JSEN.2017.2654359

Jang, J. S. R., Sun, C. T., & Mizutani, E. (1997). Neuro-fuzzy and soft computing-a computational approach to learning and machine intelligence [Book Review]. *IEEE Transactions on Automatic Control*, *42*(10), 1482–1484.

Jiang, Z., & Groves, P. D. (2014). NLOS GPS signal detection using a dual-polarisation antenna. *GPS Solutions*, *18*(1), 15–26. https://doi.org/10.1007/s10291-012-0305-5

Lee, Y. C. (2002, June). Compatibility of the new military GPS signals with non-aviation receivers. *Proceedings of the 58th Annual Meeting of The Institute of Navigation and CIGTF 21st Guidance Test Symposium*, Albuquerque, NM, 581–597.

McGraw, G. A., & Braasch, M. S. (1999, January). GNSS multipath mitigation using gated and high resolution correlator concepts. *Proceedings of the 1999 National Technical Meeting of The Institute of Navigation*, San Diego, CA, 333–342.

Meguro, J. I., Murata, T., Takiguchi, J. I., Amano, Y., & Hashizume, T. (2009). GPS multipath mitigation for urban area using omnidirectional infrared camera. *IEEE Transactions on Intelligent Transportation Systems*, *10*(1), 22–30. https://doi.org/10.1109/TITS.2008.2011688

Miura, S., Hsu, L. T., Chen, F., & Kamijo, S. (2015). GPS error correction with pseudorange evaluation using three-dimensional maps. *IEEE Transactions on Intelligent Transportation Systems*, *16*(6), 3104–3115. https://doi.org/10.1109/TITS.2015.2432122

N'Diaye, A., Haile, J. K., Fowler, D. B., Ammar, K., & Pozniak, C. J. (2017). Effect of co-segregating markers on high-density genetic maps and prediction of map expansion using machine learning algorithms. *Frontiers in Plant Science*, *8*, 1434. https://doi.org/10.3389/fpls.2017.01434

Palamartchouk, K., Clarke, P. J., Edwards, S. J., & Tiwari, R. (2015, September). Dual-polarization GNSS observations for multipath mitigation and better high-precision positioning. *Proceedings of*

the 28th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS+ 2015), Tampa, FL, 2772–2779.

Phan, Q. H., Tan, S. L., McLoughlin, I., & Vu, D. L. (2013). A unified framework for GPS code and carrier-phase multipath mitigation using support vector regression. *Advances in Artificial Neural Systems*. https://doi.org/10.1155/2013/240564

Rong, Y., Zhang, G., Chang, Y., & Huang, Y. (2016). Integrated optimization model of laser brazing by extreme learning machine and genetic algorithm. *The International Journal of Advanced Manufacturing Technology*, *87*(9-12), 2943–2950. https://doi.org/10.1007/s00170-016-8649-6

Socharoentum, M., Karimi, H. A., & Deng, Y. (2016, October). A machine learning approach to detect non-line-of-sight GNSS signals in Nav2Nav. *23rd ITS World Congress*, Melbourne, Australia.

Soloviev, A., & van Graas, F. (2008, May). Utilizing multipath reflections in deeply integrated GPS/INS architecture for navigation in urban environments. *2008 IEEE/ION Position, Location and Navigation Symposium*, Monterey, CA, 383–393. https://doi.org/10.1109/PLANS.2008.4570094

Sun, R., Hsu, L. T., Xue, D., Zhang, G., & Ochieng, W. Y. (2019). GPS signal reception classification using adaptive neuro-fuzzy inference system. *The Journal of Navigation*, *72*(3), 685–701. https://doi.org/10.1017/S0373463318000899

Sun, R., Wang, G., Zhang, W., Hsu, L. T., & Ochieng, W. Y. (2020). A gradient boosting decision tree based GPS signal reception classification algorithm. *Applied Soft Computing*, *86*, 105942. https://doi.org/10.1016/j.asoc.2019.105942

Teunissen, P., & Montenbruck, O. (Eds.). (2017). *Springer handbook of global navigation satellite systems*. Springer.

Thornberg, D. B., Thornberg, D. S., DiBenedetto, M. F., Braasch, M. S., Van Graas, F., & Bartone, C. (2003). LAAS integrated multipath-limiting antenna. *NAVIGATION*, *50*(2), 117–130. https://doi.org/10.1002/j.2161-4296.2003.tb00323.x

Townsend, B., & Fenton, P. (1994, September). A practical approach to the reduction of pseudorange multipath errors in a L1 GPS receiver. *Proceedings of the 7th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GPS 1994)*, Salt Lake City, UT, 143–148.

Van Dierendonck, A. J., Fenton, P., & Ford, T. (1992). Theory and performance of narrow correlator spacing in a GPS receiver. *NAVIGATION*, *39*(3), 265–283. https://doi.org/10.1002/j.2161-4296.1992.tb02276.x

Van Nee, R. D., Siereveld, J., Fenton, P. C., & Townsend, B. R. (1994, April). The multipath estimating delay lock loop: Approaching theoretical accuracy limits. *Proceedings of 1994 IEEE Position, Location and Navigation Symposium-PLANS'94*, Las Vegas, NV, 246–251. https://doi.org/10.1109/PLANS.1994.303320

Wang, L., Groves, P. D., & Ziebart, M. K. (2013). GNSS shadow matching: Improving urban positioning accuracy using a 3D city model with optimized visibility scoring scheme. *NAVIGATION*, *60*(3), 195–207. https://doi.org/10.1002/navi.38

Wang, L., Groves, P. D., & Ziebart, M. K. (2015). Smartphone shadow matching for better cross-street GNSS positioning in urban environments. *The Journal of Navigation*, *68*(3), 411–433. https://doi.org/10.1017/S0373463314000836

Weill, L., & Fisher, B. (2002). *US Patent No. 6,370,207*. Washington, DC: U.S. Patent and Trademark Office.

Weir, I. (2016). *Spearman's correlation*. Statstutor, Mathematics Education Centre Loughborough University. http://www.statstutor.ac.uk/resources/uploaded/spearmans.pdf

Yozevitch, R., Moshe, B. B., & Levy, H. (2012, November). Breaking the 1 meter accuracy bound in commercial GNSS devices. *2012 IEEE 27th Convention of Electrical and Electronics Engineers in Israel*, Eilat, Israel, 1–5. https://doi.org/10.1109/EEEI.2012.6376966

Yozevitch, R., Moshe, B. B., & Weissman, A. (2016). A robust GNSS LOS/NLOS signal classifier. *NAVIGATION*, *63*(4), 429–442. https://doi.org/10.1002/navi.166

How to cite this article: Sun R, Fu L, Wang G, Cheng Q, Hsu L-T, Ochieng WY. Using dual-polarization GPS antenna with optimized adaptive neuro-fuzzy inference system to improve single point positioning accuracy in urban canyons. *NAVIGATION*. 2021;68:41–60. https://doi.org/10.1002/navi.408

APPENDIX A: The Performance Evaluations of ZYACF-L004 in Terms of the Axial Ratio, the Axial Ratio (dB), and the Cross-Polarization Discrimination (dB) against Various Frequencies

Frequency	1207 MHz	1227 MHz	1246 MHz	1561 MHz	1575 MHz	1602 MHz
AR	1.0823	1.0777	1.0727	1.052	1.1272	1.226
AR (dB)	0.687	0.65	0.61	0.44	1.04	1.77
XPD (dB)	28.06	28.54	29.09	31.93	24.46	19.84

Note: the Axial Ratio (AR) of the antenna is calculated by the ratio of major axis a and minor axis b, i.e., $AR = a/b$. Using dB as the unit, it can be written as:AR (dB) $= 20\log(a/b)$. Cross-Polarization Discrimination (XPD) is calculated by: XPD (dB) $= 24.8 - 20\log(AR(\text{dB}))$.

APPENDIX B: The Performance of ZYACF-L004 in Terms of RHCP Antenna Gain with Respect to Elevation Angles and Various Frequencies

Gain Frequency Elevation angle	1207 MHz	1227 MHz	1246 MHz	1561 MHz	1575 MHz	1602 MHz
0°	-5.56	-6.9	-7.68	-5.61	-5.64	-6.09
5°	-4.53	-5.54	-6.19	-4.04	-4.21	-4.92
10°	-3.24	-4.47	-5.61	-3.19	-3.54	-4.01
15°	-2.53	-3.99	-5.45	-2.83	-3.11	-3.62
20°	-2.04	-3.51	-4.41	-2.83	-2.26	-2.37
25°	-1.14	-2	-1.09	-2.01	-1.04	-1.33
30°	-0.17	-0.59	-1.09	-0.74	-0.18	-0.54
35°	0.66	0.04	-0.6	0.27	0.7	0.12
40°	1.13	0.33	-0.66	0.96	1.6	1.08
45°	1.64	0.80	-0.22	1.75	2.23	1.91
50°	2.07	1.23	0.5	2.62	2.9	2.51
55°	2.59	2.04	1.73	3.15	3.44	3.25
60°	3.09	2.75	2.56	3.94	4.24	3.77
65°	3.75	3.32	3	4.37	4.8	4.3
70°	4.15	3.80	3.41	4.96	5.19	4.67
75°	4.57	4.12	3.82	5.39	5.57	5.07
80°	4.87	4.46	4.04	5.71	6	5.48
85°	5.10	4.57	4.09	6.01	6.15	5.52
90°	5.12	4.75	4.28	6.12	6.36	5.75

APPENDIX C: The Performance of ZYACF-L004 in Terms of RHCP Antenna Gain with Respect to Elevation Angles and Various Frequencies

	Frequency Gain					
Elevation angle	1207 MHz	1227 MHz	1246 MHz	1561 MHz	1575 MHz	1602 MHz
0°	-7.46	-7.39	-8.25	-5.62	-6.55	-6.7
5°	-5.44	-5.48	-6.17	-4.29	-5.29	-5.18
10°	-4.08	-4.17	-5.07	-3.48	-4.45	-4.99
15°	-2.99	-3.41	-4.05	-3.01	-4	-4.11
20°	-2.14	-2.31	-2.99	-2.04	-3.04	-2.49
25°	-0.92	-1.01	-1.9	-0.86	-1.69	-1.6
30°	0.09	-0.11	-1.08	-0.1	-0.79	-0.76
35°	0.77	0.13	-0.6	0.63	-0.07	-0.17
40°	1.03	0.71	-0.09	1.2	0.48	0.93
45°	1.63	1.31	0.82	2.13	1.52	1.79
50°	2.26	1.93	1.42	2.94	2.38	2.43
55°	2.99	2.74	2.31	3.82	3.01	3.17
60°	3.51	3.21	3.94	4.21	3.69	3.81
65°	4	3.68	3.27	4.68	4.33	4.23
70°	4.36	3.89	3.52	5.15	4.82	4.65
75°	4.55	4.04	3.68	5.47	5.25	5.04
80°	4.56	4.08	3.79	5.73	4.63	5.3
85°	4.64	4.15	3.85	5.91	5.83	5.38
90°	4.77	4.36	4.01	6.08	6.11	5.55