**ORIGINAL ARTICLE**

# GNSS Site unmodeled error prediction based on machine learning

Nan Shen[1] · Liang Chen[2] · Lei Wang[2] · Ruizhi Chen[2]

## Abstract

A number of studies have attempted to reduce the effect of observation errors on Global Navigation Satellite Systems positioning through empirical error models. However, due to the complex spatiotemporal characteristics of observation errors, the effects of these errors cannot be eliminated, resulting in the unmodeled error in the positioning results. Although many studies have been carried out on unmodeled error mitigation, most of which only focus on positioning model optimization and fail to make use of historical observation data. We explore the relationship between unmodeled error and observation features and develop a new data-driven approach based on machine learning. Historical observations of a specific station are used to predict the unmodeled error of a positioning model. Time–frequency analysis is used to evaluate the prediction results. The feasibility of applying the method to the precise point positioning (PPP) kinematic positioning is verified by using IGS station data. It is clear from the findings that the data-driven model can effectively predict the unmodeled errors in GNSS positioning, especially in low-frequency components. In addition, the influencing factors of the method are explored in detail and the relevant settings are recommended.

**Keywords** GNSS · Unmodeled error · Machine learning · Convolutional neural network

## Introduction

GNSS positioning accuracy is affected by various observation errors, which seriously deteriorates its application in displacement monitoring such as settlement and landslides. Many studies have explored possible methods to mitigate the effects of these observation errors. To date, however, there has been no definitive model that can eliminate these errors caused by the complex spatiotemporal characteristics. These errors, which cannot be eliminated, are referred to as unmodeled errors. There are many unmodeled error mitigation strategies, a common strategy is to model observation errors. For example, Hoque and Jakowski (2008) investigated the higher-order ionospheric effects in precise GNSS positioning. Another well-known example is multipath mitigation, such as multipath ray-tracing (Lau and Cross 2007), multipath sidereal filtering (Choi et al. 2004; Ragheb et al. 2007; Shen et al. 2020), multipath hemispheric mapping (Dong et al. 2016), adaptive tracking of line-of-sight/non-line-of-sight signals (Chen et al. 2012; Chen et al. 2014; Chen et al. 2017) and so on. The effects of unmodeled error cannot be fully eliminated due to the complex spatiotemporal characteristics of observation errors.

The second strategy for error mitigation is to select the appropriate positioning strategy based on the characteristics of the observation residuals of the original positioning results. Most research on GNSS error characteristics has been carried out based on observation residuals. Zhang et al. (2017) compared and analyzed the unmodeled error in Global Positioning System (GPS) and BeiDou signals. The residual error analysis of the observation data shows that the unmodeled error is present in the positioning time series. The short-term correlation of unmodeled error is analyzed by correlation analysis, and two empirical models are used

✉ Liang Chen
    l.chen@whu.edu.cn

    Nan Shen
    nanshen@njtech.edu.cn

    Lei Wang
    lei.wang@whu.edu.cn

    Ruizhi Chen
    ruizhi.chen@whu.edu.cn

[1]  School of Geomatics Science and Technology, Nanjing Tech University, Nanjing 210000, PR China

[2]  State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, PR China

to estimate the correlation coefficient. Sequential adjustment is used in baseline resolution to take into account the short-term correlation of unmodeled error. Zhang et al. (2018) introduced a real-time adaptive weighting model which can mitigate the site-specific unmodeled error of undifferenced code observations. The model is a combination of the elevation-based model and the carrier-to-noise power density ratio (C/N0) based model, in which the parameters of the C/N0 model need to be determined through static observations of the receiver in a low multipath environment. Li et al. (2018a) proposed a procedure to test the significance of unmodeled errors and identify their components in GNSS observation. In this method, the unmodeled error is divided into three categories, and suggestions for processing each type of unmodeled error are also given. A recent study by Zhang and Li (2020) made use of this procedure to detect the significance of observation residuals. Then the unmodeled error mitigation method based on multi-epoch partial parameterization is adapted to process the observations with significant unmodeled error. In this kind of method, the choice of function model or stochastic model mainly depends on the hypothesis test of residuals (Wang et al. 2013), and other observation features are not fully utilized.

Another kind of strategy is the time–frequency analysis of geographic coordinate time series. Different trajectory models are defined to describe coordinate time series. Among them, the standard linear trajectory model (SLTM) is defined as the sum of three different types of displacements, including trend terms, jump terms, and periodic terms (Bevis et al. 2020). In addition, Bevis and Brown define the extended trajectory model (ETM), which adds one or more transients to the SLTM (Bevis and Brown 2014). Different parameter estimation methods are used to estimate the parameters involved in these models, including maximum likelihood estimation (Langbein 2017; Bos et al. 2020), Bayesian inference (Olivares-Pulido et al. 2020), and Kalman filter (Engels 2020). Usually, the error characteristics of coordinate time series are analyzed by the power spectrum, and the parameters of each error component can be estimated by the least square variance component (Teunissen and Amiri-Simkooei 2008). However, most of the above methods are for offline analysis of coordinate time series, which is not suitable for online applications.

The last strategy is based on regional filtering, which utilizes the regional correlation characteristics of unmodeled error and the reference station network to reduce the impact of the unmodeled error. In regional network analysis, regional-related errors caused by satellite orbit, earth orientation parameters, atmospheric effects, and other factors are called common-mode errors (CME) (Wdowinski et al. 1997; Dong et al. 2006). Many studies have attempted to mitigate the common model error through regional filtering techniques. The first detailed regional filtering study was

carried out to estimate coseismic and postseismic displacements (Wdowinski et al. 1997). This technique assumes that the CME is special uniform, which leads to a decrease of the calculated CME with the increase in regional network size. To handle this problem, the principal component analysis (Wold et al.1987) and the Karhunen–Loeve expansion (Kirby and Sirovich 1990) were introduced in the regional filtering, in which the nonuniform spatial response of the network stations to a CME source is considered (Dong et al. 2006). The entire network time series are taken into account and the time series is decomposed into various spatial and temporal coherent orthogonal modes in the regional filtering. Besides, the multiscale principal component analysis (PCA) techniques have been introduced into regional filtering (Li et al. 2017, 2018b). The wavelet denoising is applied to the coordinate time series before performing the PCA for common model error mitigation (Li et al. 2017). Similarly, the empirical mode decomposition (EMD) was adopted to denoise the coordinate time series before the common model error mitigation was performed using PCA (Li et al. 2018b). Besides, the multi-channel singular spectrum analysis was adopted to estimate common environmental impacts affecting GPS observations (Gruszczynska et al. 2018), and it is not suitable for site-related errors, such as multipath. All of these methods are based on time series analysis of coordinate residuals, so the features of observation are not fully utilized.

The observation features which are derived from the single epoch GNSS positioning results, such as satellite elevation angle, pseudorange residual, carrier-phase residual, signal strength, etc., are related to the unmodeled error. Some machine learning algorithms have been introduced for multipath detection (Hsu 2017). A convolutional neural network (CNN)-based carrier-phase multipath detection method has been proposed for static and kinematic GPS high-precision positioning (Quan et al. 2018). However, after detection, the weight loss or removal strategy needs further research (Lau and Cross 2006; Shen et al. 2020), and there is a lack of end-to-end unmodeled error removal research. So far, very little attention has been paid to the relationship between the unmodeled error and the observation features. Previous studies of unmodeled error mitigation are limited to correcting the observation error, positioning strategy, or coordinate time series. These studies have failed to demonstrate a link between the observation feature and the unmodeled error. Using historical observations, we seek to establish the relationship between the observation features and the unmodeled error. The purpose of this paper is to explore a data-driven unmodeled error prediction method using machine learning. The method is divided into offline training and online prediction, for which the machine learning algorithm is adopted. The research data in this paper is drawn from International GNSS Service (IGS) stations. This study offers

some important insights into the unmodeled error mitigation of a specific positioning model.

The remaining part of the paper proceeds as follows: the second section of this paper is concerned with the methodology used for this study. The experiments and results are presented next, followed by a discussion of the main issues of the proposed method and the conclusion.

## Methods

This part first provides the principle of an unmodeled error mitigation method based on machine learning. Then, the machine learning algorithm adopted in the proposed method is presented. Finally, the basic principle of the positioning model used in the experiment is introduced.

### Data-driven unmodeled error prediction

Similar to other applications based on machine learning, unmodeled error prediction is also divided into two processes: offline training and online prediction, which are introduced below.

### Online prediction

The unmodeled error mitigation in this work is achieved by training the unmodeled error prediction model of the positioning model through historical observation data. The trained unmodeled error prediction model is used to predict the unmodeled error in the positioning model to improve the accuracy of single epoch positioning. The unmodeled error mitigation process is shown in Fig. 1.

As shown in Fig. 1, the output of single-epoch positioning is the basis for model training and unmodeled error prediction. The GNSS positioning in Fig. 1 is described as

$$Y = F(L) \tag{1}$$

where $L$ represents the positioning information such as observation data, navigation message, precise ephemeris, etc.; $F$ denotes the GNSS positioning model, which can be single-point positioning, precise point positioning, relative positioning, etc.; $Y$ is the unknown parameters to be estimated, and only the coordinates are considered. The coordinate and observation features extracted from single epoch positioning are used as the basis for model training and error prediction. The data for model training is extracted from historical observations, which will be described in more detail in the following subsection. The observation features in the model prediction are obtained from the single epoch positioning of the current epoch. The trained model predicts the
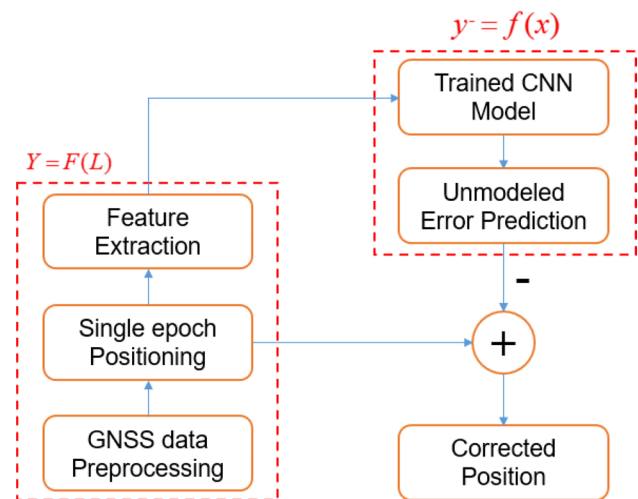


**Fig. 1** Unmodeled error mitigation process

unmodeled error, and the general definition of the prediction model in Fig. 1 is as follows

$$y^- = f(x) \tag{2}$$

where $x$ denotes observation features, including the original observation information and the intermediate processing results of single epoch GNSS positioning result; $f$ is the trained prediction model, which will be described in the following subsection; $y^-$ is the predicted unmodeled error in the current epoch. As shown in Fig. 1, after obtaining the original positioning value and the predicted unmodeled error, the final position estimation can be obtained by the following formula

$$\hat{Y} = Y - y^- \tag{3}$$

### Offline training

The above description is an online unmodeled error prediction process. However, how to train a reliable prediction model $f$ is the key to online prediction. The problem of unmodeled error prediction is essentially a regression problem in supervised learning. The machine learning algorithm is to learn the mapping relationship between input variables and output variables through training data. Training data is the basis of model training, and the expression of training data is given as follows

$$D = \{(x_1, y_1), (x_2, y_2) \cdots, (x_i, y_i) \cdots, (x_n, y_n)\}, \, y_i \in R \tag{4}$$

where $x_i$ represents the input of the $i$th sample, that is, the selected observation feature; $y_i$ denotes the output of the $i$th sample, that is, the unmodeled error.

The selected features can be divided into three categories: one is the features closely related to unmodeled error studied by previous researchers, such as satellite elevation, signal strength, observation residual, etc.; the other is GPS time which reflects the time characteristics of error; the last one is the quality features of observation, such as valid data flag, cycle-slip flag, cycle-slip count, etc. To eliminate the influence of unit and scale differences between features and treat all dimensional features equally, the features need to be normalized. Min–max normalization and z-score normalization are adopted to normalize these features. The details of observation features and corresponding normalization methods are shown in Table 1.

The features of all GPS satellites are organized in a matrix, forming a so-called feature matrix. Each column in the matrix represents the satellite, and each row in the matrix represents each feature. The size of the feature matrix is fixed, and the columns corresponding to the unavailable satellites will be reserved and filled with zeros. The output of the training data, the GNSS positioning unmodeled error, is usually obtained by the difference between the GNSS single-epoch positioning result and the relatively high-precision positioning result. More precise positioning results are used as reference positioning results, but these positioning results have a time delay. We used the difference between the positioning result of a single epoch and the positioning result of a whole day as the training data output. The unmodeled error is obtained as follows

$$y_i = Y_i - \overline{Y} \tag{5}$$

where $Y_i$ is the result of a single epoch solution and $\overline{Y}$ is the average value of a single epoch solution within a day. After preparing training samples, the key to the prediction of unmodeled error is to train a desirable prediction model with

these sample data. The machine learning algorithm used in this method is given below.

## Convolutional neural network

CNN has been widely used in the field of computer vision, and it has also performed well in other application fields (Goodfellow et al. 2016). Therefore, a deep convolutional neural network is designed to regress the unmodeled error in GNSS positioning.

### Network architecture

The network consists of an input layer, 4 convolutional layers, a pooling layer, a flattening layer, a fully connected layer, and an output layer. The unmodeled error CNN regression model is shown in Fig. 2.

The dimension of the input layer is 13*32, where 13 represents the number of features used and 32 represents the number of GPS satellites. The size of each convolution kernel in this network is 3 by 3, and the step size in both directions is set to 1. The convolution layer adopts the valid zero-padding strategy (Goodfellow et al. 2016), that is, to reject the zero-padding, resulting in a reduction of the length and width of the output data by 2. The number of convolution kernels in each convolution layer is set to 64, 64, 32, and 32, respectively, to extract multiple types of features.

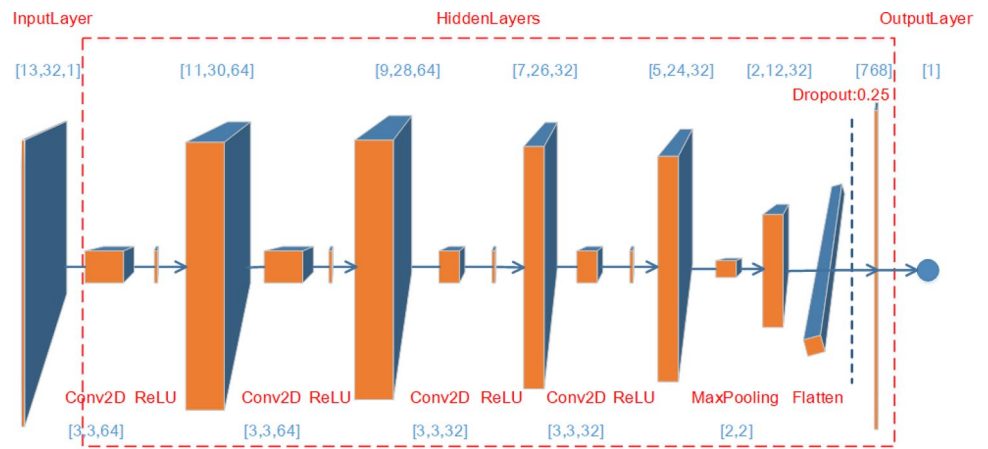The convolution operation of the convolution layer $l$ is described below (LeCun and Bengio 1998).

$$X_j^l = f_{\text{relu}}(\sum_{i=1}^{d} X_i^{l-1} \cdot K_{ij}^l + B_j^l) \tag{6}$$

where $X_i^{l-1}$ represents the feature matrix of the input of the convolution layer on the $i$th channel; $d$ denotes the number of input channels of the convolution layer; $K_{ij}^l$ is the $j$th con-

**Table 1** Observation features of the machine learning algorithm, $x$ denotes the original value of the feature, $\overline{x}$ is the mean of the original value, $\sigma$ represents the standard deviation of the original value, DoS is short for depends on samples

| Name | Type | Numerical range | Normalized method |
| --- | --- | --- | --- |
| GPS time | Float | DoS | $(x - \min(x))/(\max(x) - \min(x))$ |
| Satellite azimuth | Float | DoS | $(x - \min(x))/(\max(x) - \min(x))$ |
| Satellite elevation | Float | DoS | $(x - \min(x))/(\max(x) - \min(x))$ |
| Pseudo-range residual | Float | DoS | $(x - \overline{x})/\sigma$ |
| Carrier-phase residual | Float | DoS | $(x - \overline{x})/\sigma$ |
| Signal strength | Float | DoS | $(x - \overline{x})/\sigma$ |
| Valid data flag | Integer | [0, 1] | $x$ |
| Cycle-slip flag | Integer | [1, 2] | $x - 1$ |
| Ambiguity flag | Integer | [0, 1, 2, 3] | $x/3$ |
| Carrier-lock count | Integer | [0, DoS] | $x/\max(x)$ |
| Data outage count | Integer | [0, DoS] | $x/\max(x)$ |
| Cycle-slip count | Integer | [0, DoS] | $x/\max(x)$ |
| Data reject count | Integer | [0, DoS] | $x/\max(x)$ |

**Fig. 2** Unmodeled error CNN regression model



volution kernel of the convolution layer corresponding to the $i$th input channel; $\cdot$ denotes the convolution operator; $B_j^l$ represents the $j$th convolution kernel bias of the convolution layer; $X_j^l$ is the output feature matrix of the $j$th convolution kernel of the convolution layer, also known as feature map in the computer vision. After each convolutional layer, an activation layer is immediately followed. The activation layer essentially adds nonlinear elements to the model through the activation function, thereby making up for the expressive power of the linear model (Buduma and Locascio 2017). The activation layer does not change the dimensions of the input data. The rectified linear unit (ReLU) is generally selected as the activation function of the convolution layer. The definition of the activation function is as follows (Nair and Hinton 2010).

$$f_{\text{relu}}(x) = \max(0, x) \tag{7}$$

The activation function can reduce gradient vanishment and has a fast operation speed. There is a pooling layer after the last convolution layer, and the pooling size is set to 2 by 2. The size of the data passing through the pooling layer becomes half of that of the input, and the number of channels does not change. Convolution and pooling are regarded as infinitely strong priors (Goodfellow et al. 2016). The final flattening layer is responsible for converting multidimensional data into one-dimensional data.

## Back propagation

The output $y$ is obtained through the above process, where the input $x$ passes through hidden layers such as a convolutional layer, a pooling layer, and a fully connected layer. This process is also called the forward propagation process. Through each epoch of the forward propagation, the value of the loss function can be obtained. The loss function is defined as follows

$$J(\theta) = \frac{1}{2n} \sum_{i=1}^{n} \left\| y_i - f(x_i;\theta) \right\|^2 \tag{8}$$

where $\theta$ is the optimization parameter, and $n$ is the number of samples. The mean square error cost is used here. The loss function value $J(\theta)$ can be obtained for each forward propagation, and the parameter information is adjusted by the gradient of the loss function. The process of spreading the loss function information back through the network is called backpropagation (Rumelhart et al. 1988). In this work, the stochastic gradient descent method is used to calculate the gradient.

$$\nabla_\theta J(\theta) = -\frac{1}{n} \sum_{i=1}^{n} (y_i - f(x_i;\theta)) \frac{\partial f(x_i;\theta)}{\partial \theta} \tag{9}$$

The minibatch stochastic gradient descent is adopted to improve computing efficiency, and the batch size is set to 32. After obtaining the gradient of the loss function, the parameters are updated by the following formula, and the network training process enters the next epoch.

$$\theta' = \theta - \varepsilon \nabla_\theta J(\theta) \tag{10}$$

where $\theta'$ is the updated parameter; $\varepsilon$ is the learning rate. In this work, an adaptive estimation algorithm of learning rate named adaptive moment estimation (Adam) is adopted. The Adam algorithm dynamically adjusts the parameter's learning rate based on the loss function's first and second-moment estimates of the parameters (Kingma and Ba 2015).

## Regularization

To reduce the generalization error of the model, two regularization methods were adopted which include early stopping, Dropout. Early stopping is to evaluate the performance of the model on the validation set during the training process, and stop training when the model

performance on the validation set starts to decline to avoid the problem of overfitting caused by continued training (Prechelt 1998). In this work, the early stop criterion is to stop training when the validation set error increases continuously for 50 epochs. Dropout is another method adopted in this work to prevent overfitting. The Dropout strategy can randomly inactivate neurons at a certain percentage and inactivated neurons do not participate in the forward propagation of the network (Srivastava et al. 2014), and backpropagation only updates the weights of activated neurons, that is, the weights of inactive neurons are not updated. In this work, Dropout is applied to the last hidden layer, and the inactivation ratio is set to 0.25, that is, within each minibatch epoch, 192 neurons in the last hidden layer are inactivated.

For the implementation of the CNN regression model mentioned above, we make full use of the open-source software TensorFlow to facilitate our research. Instead of implementing the model directly based on TensorFlow (Abadi et al. 2016), we adopted another open-source software: Keras, which is an encapsulation of TensorFlow to support rapid practice so that we can quickly convert ideas into results without paying too much attention to the underlying details (Gulli and Pal 2017).

## Evaluation of predicted results

To evaluate the unmodeled error of predictions, we compute the standard deviation (STD) of the difference between the original time series and the predicted time series. In addition, wavelet tools are utilized to analyze the original time series and the predicted time series. As an effective time–frequency analysis tool, wavelet is widely used in the field of image and audio, and also in the field of geodesy. As a time–frequency analysis method, the time–frequency resolution of the wavelet transform is variable. For example, in wavelet transform, the low-frequency part has a lower time resolution and higher frequency resolution, while the high-frequency part has a lower frequency resolution and higher time resolution (Ruch and Fleet 2009). The wavelet transform is widely used in signal analysis because of its adaptability to the signal. The definition of continuous wavelet transform (CWT) is as follows (Addison 2017)

$$\mathrm{CWT}_x^{\psi}(s, \tau) = \frac{1}{\sqrt{s}} \int_{-\infty}^{\infty} x(t)\psi\left(\frac{t - \tau}{s}\right) \mathrm{d}t \tag{11}$$

where $s$ is the scale factor, $t$ is the translation factor, $\psi$ is the mother wavelet, and $x$ is the signal to be analyzed. In the experiments, the CWT will be used to analyze the original time series and the predicted time series.

## Precise point positioning (PPP)

In the above, the prediction model of unmodeled error is proposed, and the machine learning algorithm is introduced. We will study the prediction of unmodeled error of the precise point positioning (PPP) kinematic positioning model. The basic principle of PPP is introduced here. PPP is a technology that uses a single receiver to achieve global high-precision positioning. This technology makes use of the precise satellite orbit and clock products provided by external organizations such as IGS and takes into account the fine modeling of various errors. The ionosphere-free (IF) combination of dual-frequency is the most widely used observable model in PPP, which can eliminate the first-order ionosphere influence. The ionosphere-free combination observation equations of PPP are demonstrated as follows (Kouba and H´eroux 2001; Malys and Jensen 1990),

$$\begin{cases} P_{r,\mathrm{IF}_{12}}^s = \frac{f_1^2}{f_1^2 - f_2^2} P_{r,1}^s - \frac{f_2^2}{f_1^2 - f_2^2} P_{r,2}^s = \rho_r^s + c(\mathrm{d}t_r - \mathrm{d}t^s) + T_r^s + e_{\mathrm{IF}_{12}} \\ \Phi_{r,\mathrm{IF}_{12}}^s = \frac{f_1^2}{f_1^2 - f_2^2} \Phi_{r,1}^s - \frac{f_2^2}{f_1^2 - f_2^2} \Phi_{r,2}^s = \rho_r^s + c(\mathrm{d}t_r - \mathrm{d}t^s) + T_r^s + B_{r,\mathrm{IF}_{12}}^s + \mathrm{d}\Phi_{r,\mathrm{IF}_{12}}^s + \varepsilon_{\mathrm{IF}_{12}} \end{cases} \tag{12}$$

where $f_i(i = 1, 2)$ denotes frequency; $P_{r,i}^s$ and $\Phi_{r,i}^s(i = 1, 2)$ are pseudorange observations and carrier phase observations, both in meters; $P_{r,\mathrm{IF}_{12}}^s$ is the ionosphere-free combination observation of pseudoranges; $\Phi_{r,IF_{12}}^s$ is the ionosphere-free combination observation of carrier phases; $\rho_r^s$ is the geometric range from the receiver to the satellite; $c$ is the speed of light in vacuum; $\mathrm{d}t_r$ is the receiver clock offset; $\mathrm{d}t_s$ is the satellite clock offset; $T_r^s$ is the tropospheric delay; $B_{r,\mathrm{IF}_{12}}^s$ is the carrier phase bias, including the ionosphere-free combination of integer ambiguity and phase delays; $\mathrm{d}\Phi_{r,\mathrm{IF}_{12}}^s$ is ionosphere-free carrier phase corrections, including receiver antenna phase center correction, satellite antenna phase center correction, earth tide correction, and phase wind-up correction; $e_{\mathrm{IF}}$ and $\varepsilon_{\mathrm{IF}}$ are measurement noises of pseudorange combined observation and carrier combined observation, respectively. For the latter method verification, only the floating resolution results of the ionospheric-free model are used. The satellite clock correction provided by IGS is based on the ionosphere-free combination observation, which absorbs the ionosphere-free combination of the satellite pseudorange hardware delay. The ionosphere-free combination of receiver pseudorange hardware delay is absorbed by the clock error of the receiver. The carrier phase hardware delay is absorbed by the carrier phase bias $B_{r,\mathrm{IF}_{12}}^s$ defined herein. Therefore, in (12), the description of pseudorange hardware delay and carrier phase hardware delay is ignored. To obtain high-precision centimeter-level positioning results using PPP, the error terms in (12) have to be carefully modeled or estimated together with the user position. In this

work, the open-source software RTKLIB (Takasu 2011) is used to realize PPP positioning, and kinematic positioning is tried.

## Experiments and results

The data used in this study comes from IGS stations. Since the model training is very time-consuming, 12 stations around the world are randomly selected for training, and the distribution of these stations is shown in Fig. 3. The observation data of these stations in October 2018 are used, and the sampling interval of these observation data is 30 s. Only the $L_1$ and $L_2$ observations of GPS in the observation data were used in the experiment. RTKLIB, an open-source GNSS data processing software, was used for the data processing. The positioning mode was set to PPP kinematic mode. The filter type was set to forward filter solution, and the elevation mask angle was set to 15°. Model correction items considered in model refinement include antenna phase center, phase wind-up, solid earth tide, ocean loading, polar tide, and others, as described by the RTKLIB manual

(Takasu 2011). Despite the addition of the various model corrections above, millimeter-scale level daily variations can occur due to imperfect ocean tides, other types of loading, and temperature effects on station monuments. Therefore, we assume that the actual displacement of the station is far less than the unmodeled error that we want to correct and take certain strategies to exclude the large displacement data from the training data.

After obtaining the positioning results according to the above processing strategy, we extracted the features and labels according to the method described in the methods section. To reduce the influence of abnormal training data, we discarded the data of days with a large standard deviation in the positioning results. Data of days with label standard deviations greater than 10 cm were removed. Although the standard deviation of each day of kinematic positioning result of the HKWS station is greater than 10 cm, the processing results obtained from these data are also provided. We use the data on October 30 and October 31 as the test data and use the 8-day data and 18-day data that meet the conditions above as the training data, respectively. Besides, in the training data, all single epoch data are shuffled and a quarter of the data was used as the validation data set.
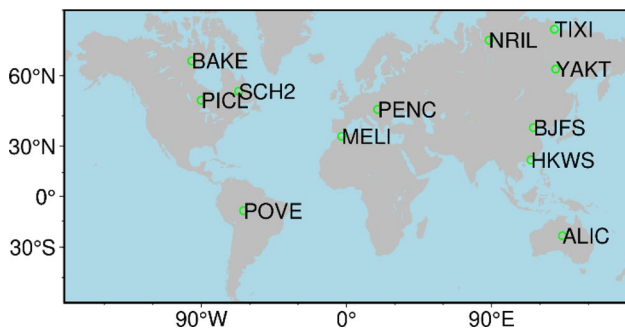
## Unmodeled error prediction of PPP kinematic positioning

The STD of the difference between the original time series and the predicted time series for the two-day test data is exhibited in Table 2. On the whole, the STD in the horizontal direction is smaller than that in the vertical direction, mainly because the measurement noise in the horizontal direction is smaller than that in the vertical direction. In addition, in the unmodeled error predictions for the next two days, there is no significant difference between the prediction results of day 1 and day 2. Taking

**Fig. 3** Geographical distribution of IGS stations

**Table 2** STD of prediction results of different stations

| station | First day | | | Second day | | |
|---|---|---|---|---|---|---|
| | North (mm) | East (mm) | Up (mm) | North (mm) | East (mm) | Up (mm) |
| ALIC | 20.61 | 20.15 | 54.51 | 19.16 | 16.02 | 55.70 |
| BAKE | 47.17 | 25.85 | 53.75 | 46.80 | 25.02 | 46.15 |
| BJFS | 23.96 | 15.40 | 44.16 | 22.20 | 14.73 | 41.93 |
| MELI | 45.82 | 27.06 | 28.35 | 43.84 | 25.98 | 23.35 |
| PENC | 25.74 | 22.78 | 50.55 | 23.23 | 20.38 | 51.38 |
| PICL | 49.93 | 19.53 | 28.76 | 47.87 | 22.77 | 26.51 |
| POVA | 28.30 | 30.91 | 57.25 | 30.24 | 20.74 | 62.47 |
| SCH2 | 45.83 | 19.18 | 27.32 | 40.54 | 19.55 | 28.77 |
| NRIL | 21.20 | 14.15 | 54.81 | 21.74 | 16.53 | 51.48 |
| TIXI | 19.09 | 15.11 | 60.28 | 19.02 | 16.70 | 53.27 |
| YAKT | 23.25 | 17.42 | 54.05 | 22.54 | 17.66 | 51.24 |
| HKWS | 38.96 | 21.07 | 86.92 | 39.29 | 19.93 | 79.81 |

the ALIC station as an example, the time–frequency analysis results of the predicted results of the north, east, and up coordinate components on the first day are shown in Figs. 4, 5, 6, respectively. In these figures, the top panel shows the original and the predicted time series, and the

bottom-left and bottom-right panels show the wavelet analysis results of the original and the predicted time series, respectively. It can be seen from Fig. 4 that both the original and the predicted time series of the north coordinate component have spectral components with a

**Fig. 4** Prediction results of the north coordinate component at the ALIC station
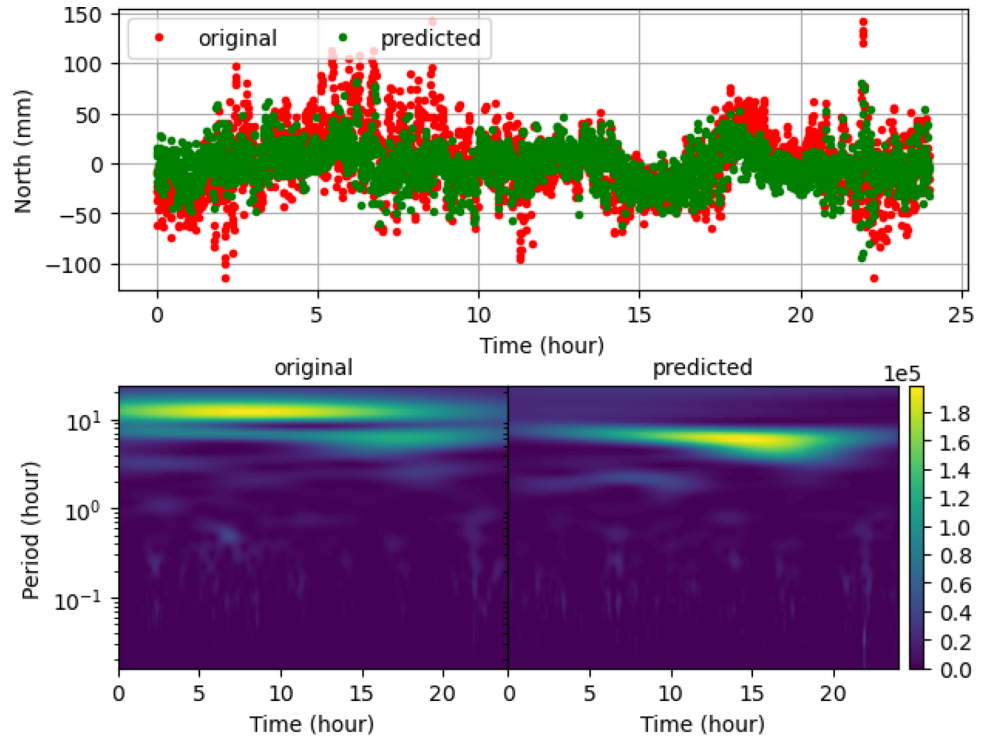


**Fig. 5** Prediction results of the east coordinate component at the ALIC station
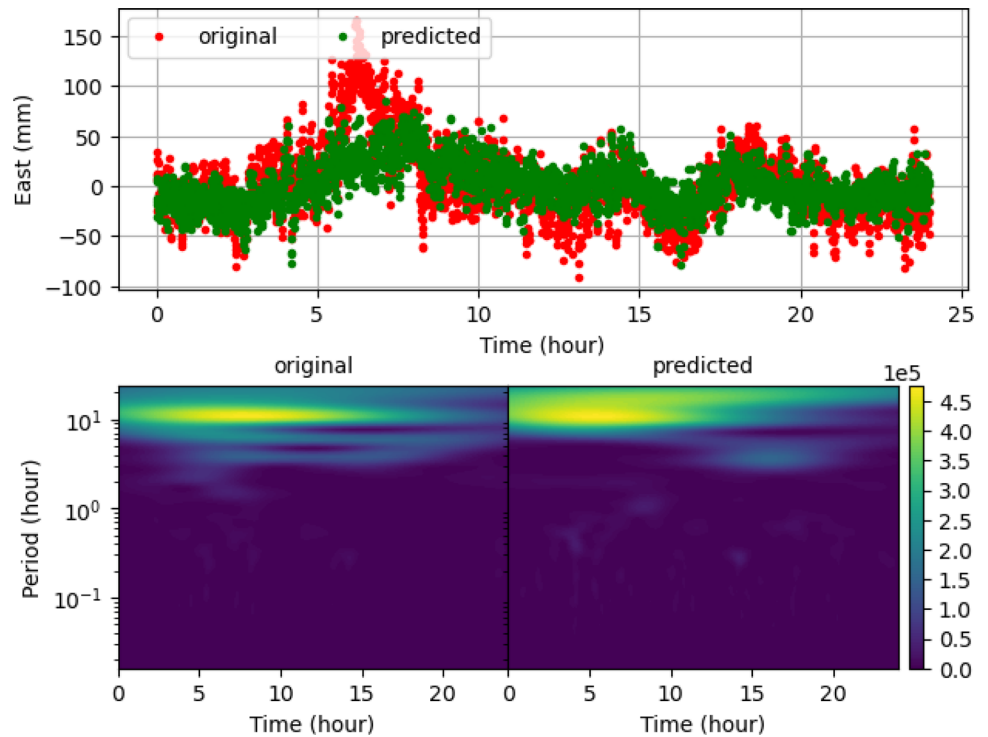
**Fig. 6** Prediction results of the up coordinate component at the ALIC station
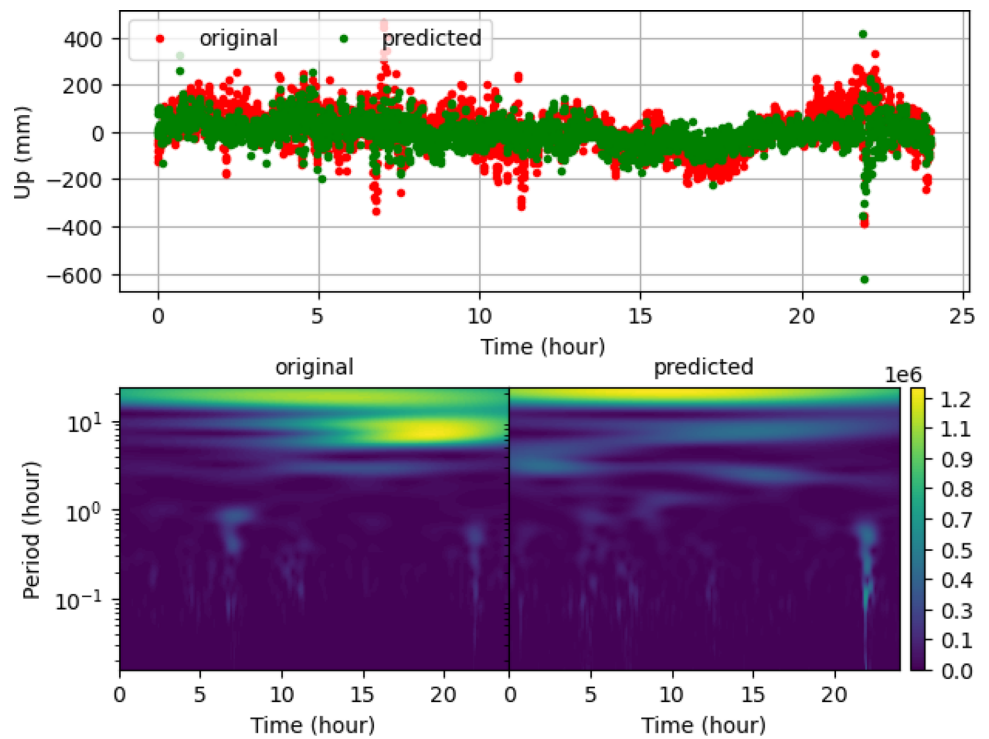


**Fig. 7** Prediction results of the north coordinate component at the SCH2 station
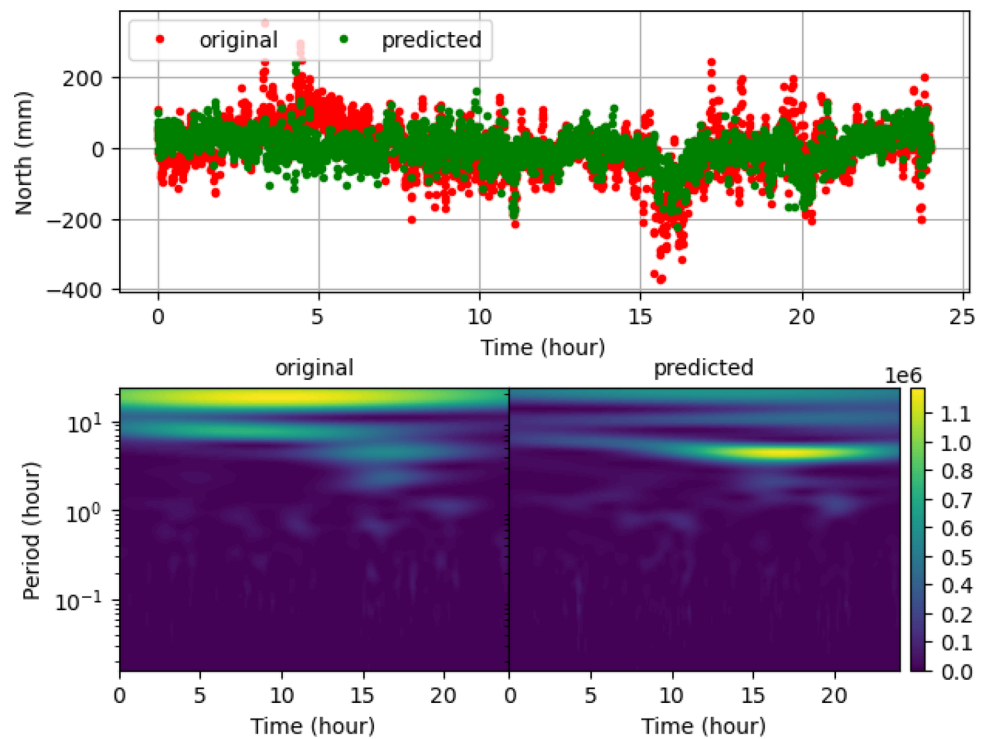


period of about 8 h, especially between 10 and 20 h. The spectral components with a period of about 10 h in the unmodeled error of the east coordinate component have been effectively predicted, and it can be seen from Fig. 5 that it is mainly concentrated between 0 and 15 h. It can

be seen from Fig. 6 that the unmodeled error of the up coordinate component is noisier than in the other two directions, and its spectral composition is more complex. Nonetheless, the time–frequency analysis of the original and the predicted time series also show a certain degree of

similarity. Time–frequency analysis results of the SCH2 station are also given, as shown in Figs. 7, 8, 9. Unlike the previous station, the STD of the north component of this station is larger than that of the other two components. As can be seen from Figs. 7, 8, 9, the original time series of the north component is noisier than that of the other two components.
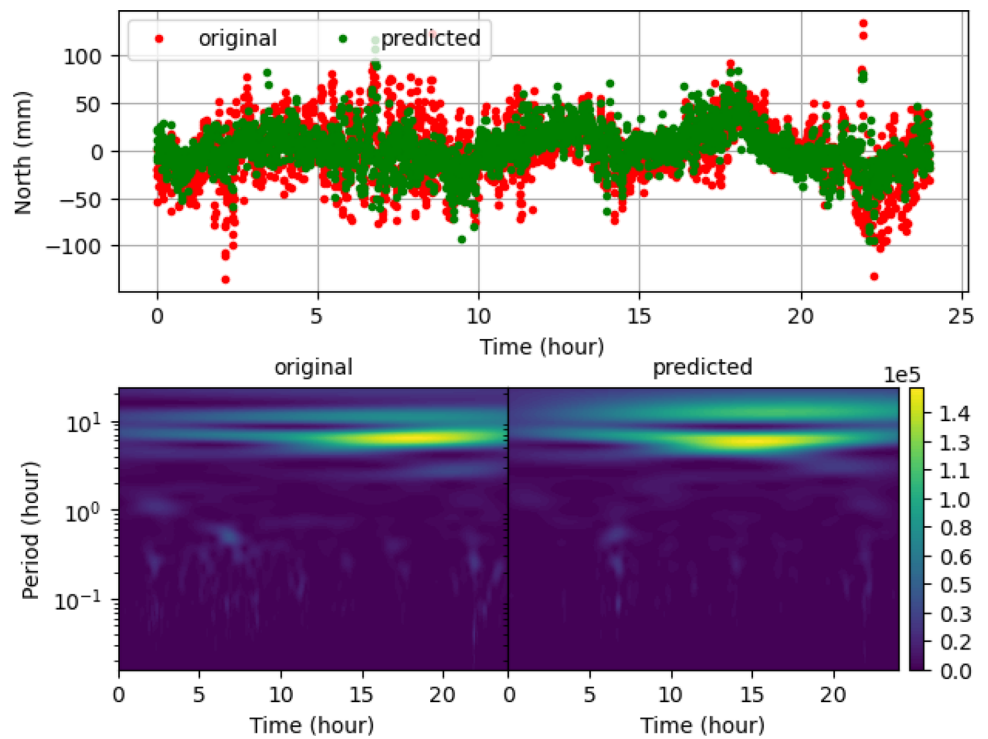
**Fig. 8** Prediction results of the east coordinate component at the SCH2 station
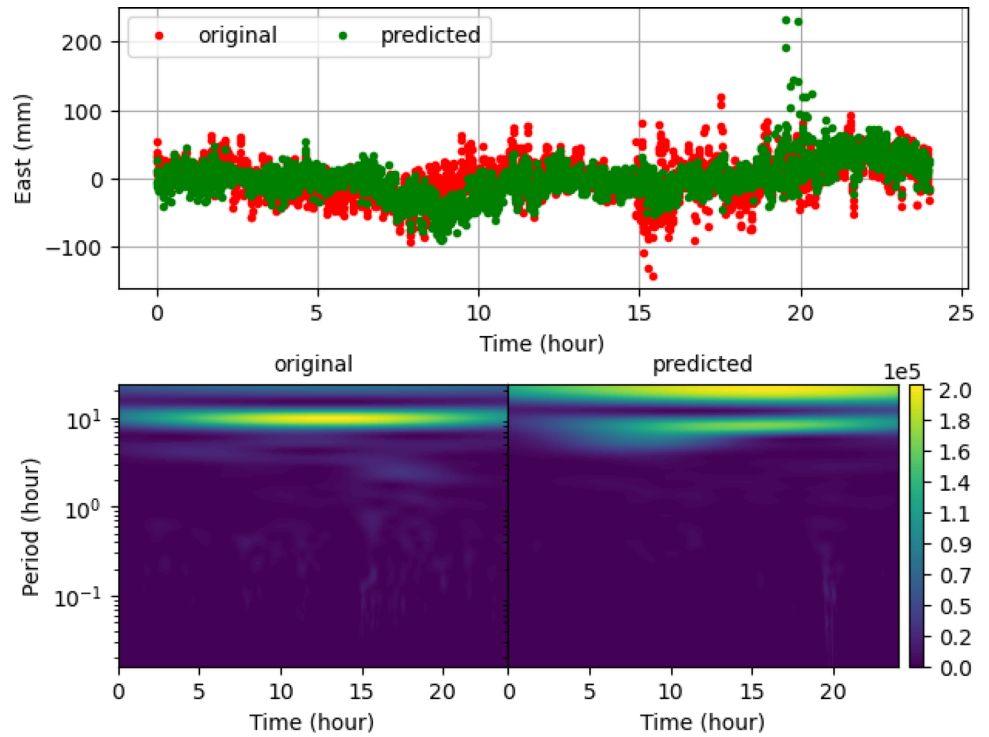


**Fig. 9** Prediction results of the up coordinate component at the SCH2 station
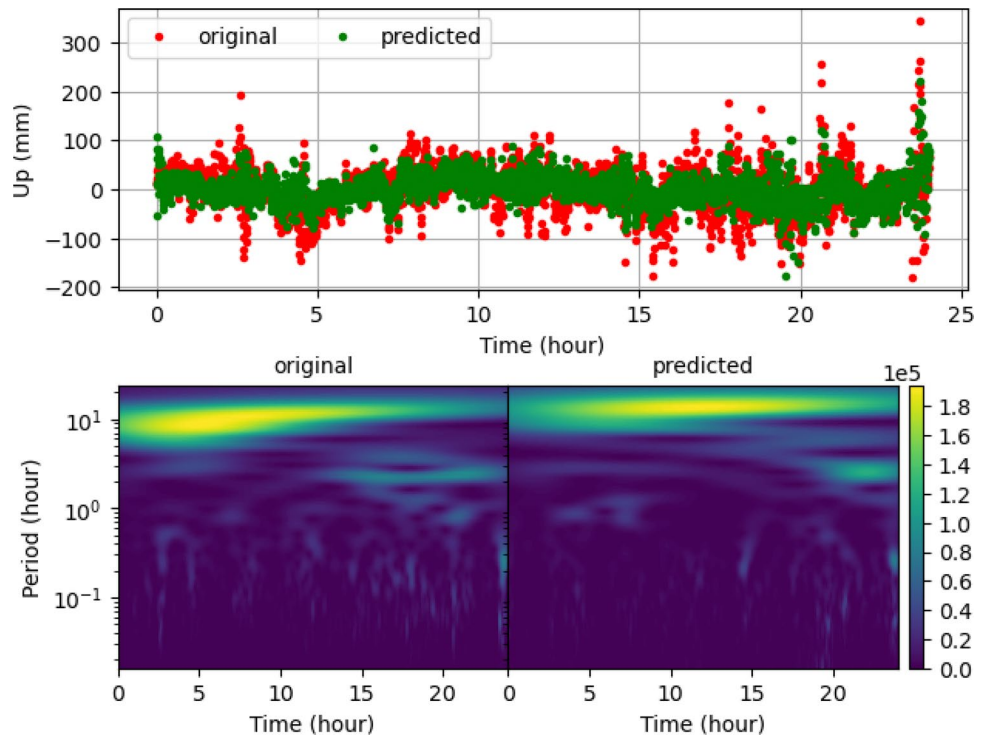
**Fig. 10** Prediction results of the north coordinate component using the combined filtering mode at the ALIC station



**Fig. 11** Prediction results of the east coordinate component using the combined filtering mode at the ALIC station
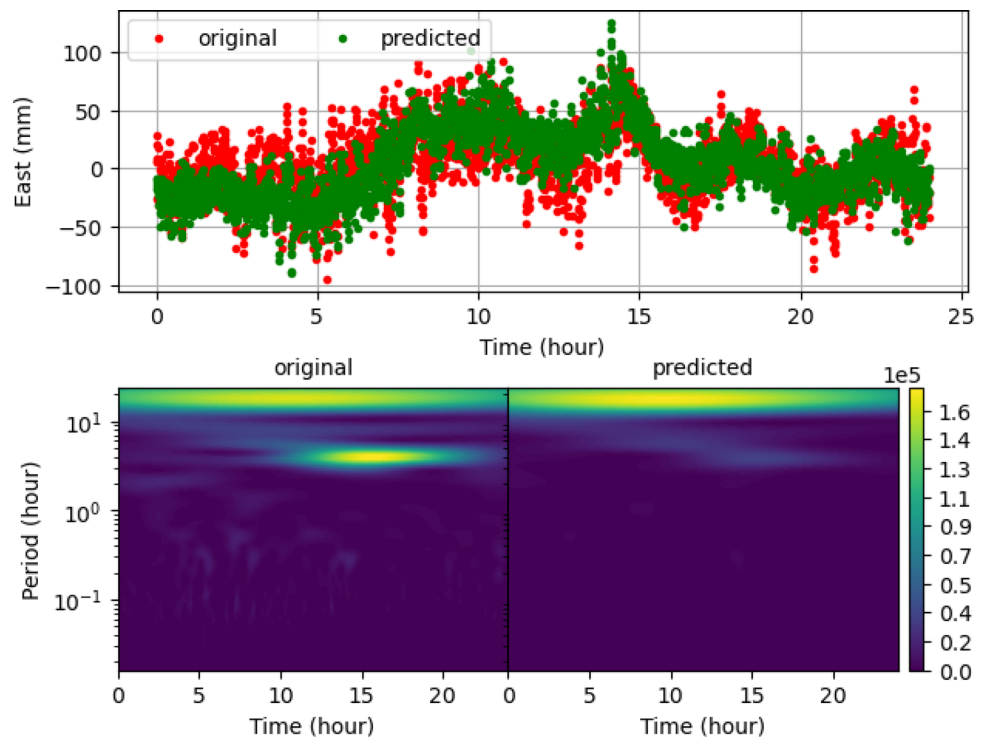


## Unmodeled error prediction of the combined filtering mode

In the above experiments, the forward filtering mode is used, and the prediction results corresponding to the combined filtering mode are given below. Using 8 days' data from the ALIC station, the STD of the difference between the original time series and the predicted time series on the first day are 18.12 mm, 16.59 mm, 52.58 mm, and the STD on the second day are 17.41 mm, 15.55 mm, 48.56 mm. The

STD of combined filtering is smaller than that of the forward filtering, which is mainly because the noise is further suppressed in combined filtering. Wavelet analysis results of the predicted and original time series of the ALIC station are also given, as shown in Figs. 10, 11, 12. Compared with the forward filtering mode, the wavelet analysis results of the combined filtering mode are more similar to those of the original time series. It can be seen from the above experiments that the unmodeled error of the low-frequency component is effectively predicted.

## Discussions

In the previous section, experiments of data-driven unmodeled error prediction, the number of days of training data was set to 8. The discussion of the number of days of training data on the prediction result is given here. Besides, the influence of different features on prediction is explored, and the corresponding STD is given when a different feature is excluded.

### Influence of more training data on unmodeled error prediction

To further verify the influence of training data of different days on the prediction results, the prediction results of different training data of the ALIC station are provided in Figs. 13, 14. In these figures, the X-axis represents the
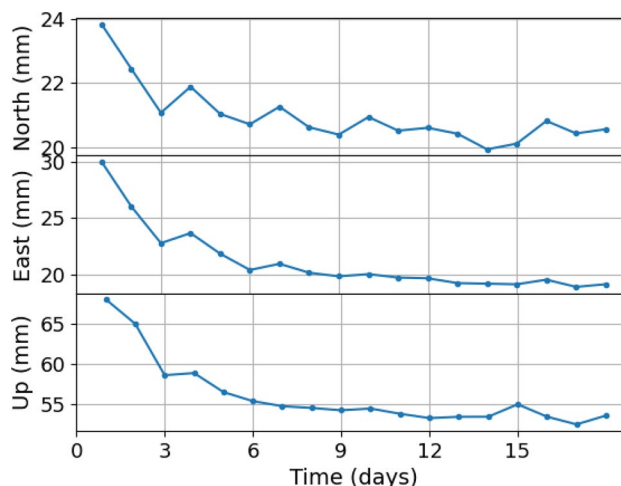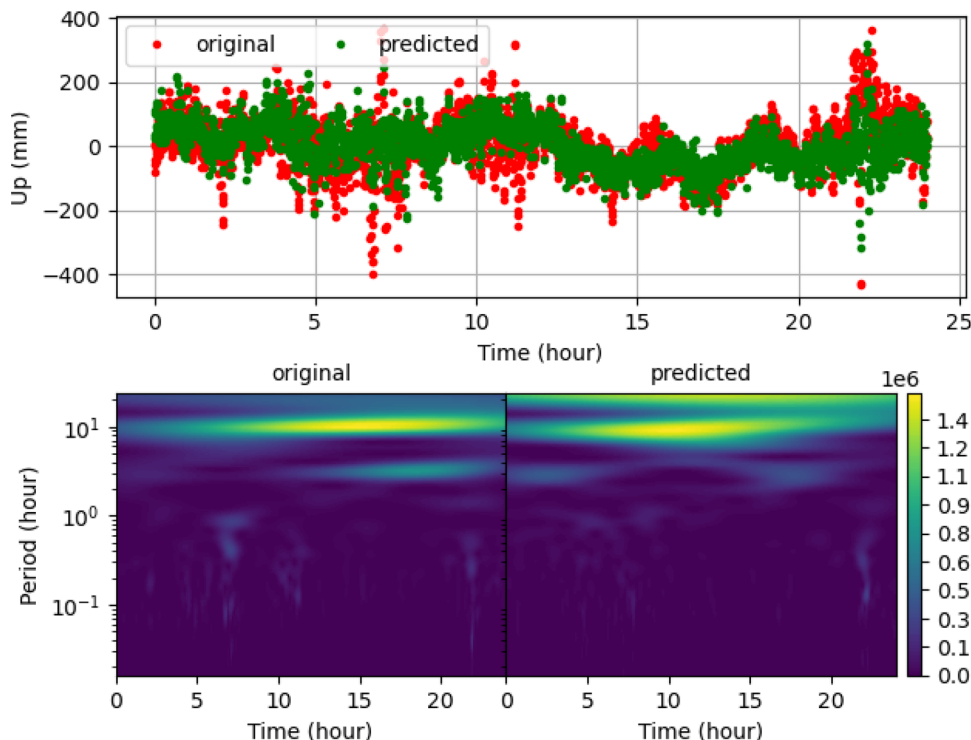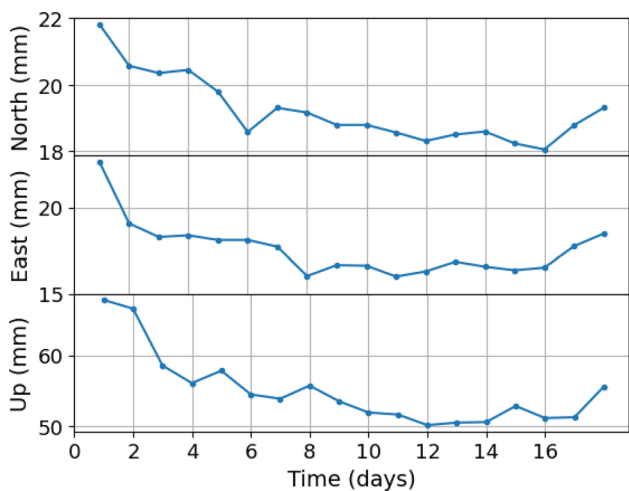


**Fig. 13** First-day prediction results of different days of training data for PPP kinematic mode at the ALIC station

number of training days, the Y-axis represents the STD of the difference between the original time series and the predicted time series. From the results, it can be seen that as the number of days increases, the STD decreases and then stabilizes. After using more than about 14 days of data, the STD stabilized. For all stations, the predicted results of 8-day training data and 18-day training data are compared as shown in Figs. 15, 16. It can be seen from these figures that the prediction results of most stations have been improved after using more training data. The main reason is that PPP



**Fig. 12** Prediction results of the up coordinate component using the combined filtering mode at the ALIC station
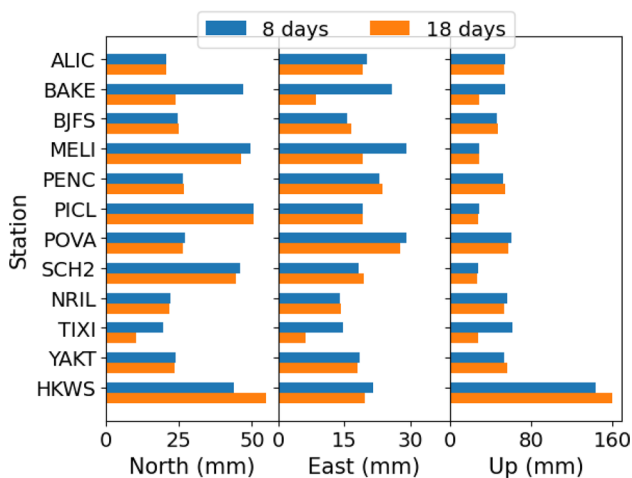
**Fig. 14** Second-day prediction results of different days training data for PPP kinematic mode at the ALIC station
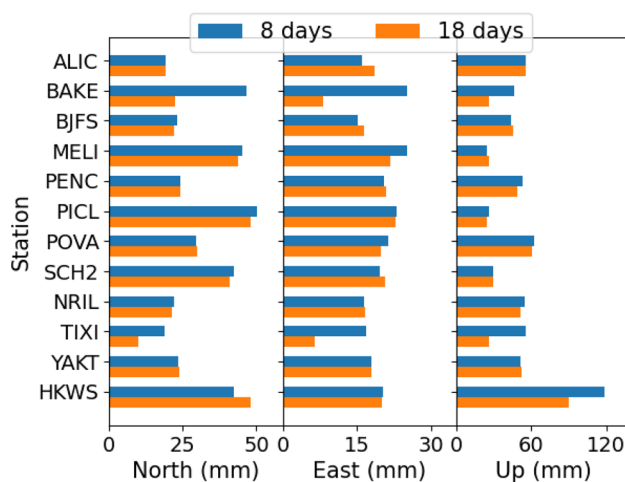


**Fig. 16** Comparison of the second-day prediction results of 8-day training data and 18-day training data



**Fig. 15** Comparison of the first-day prediction results of 8-day training data and 18-day training data

positioning results show low-frequency characteristics as well as high-frequency characteristics, and the 8-day training data is not enough to meet the training requirements of the model, that is, the model is under-fitting. Therefore, in PPP kinematic positioning mode, more training data is more conducive to the prediction of unmodeled error.

### Influence of different features on prediction results

To further explore the influence of different features on the predicted results, we try to exclude features one by one. Taking two stations ALIC and BAKE as examples, the STD of the difference between the original time series and the predicted time series obtained when the different feature is excluded are shown in Tables 3 and 4. When GPS time is

excluded from the feature, the STD of the difference between the original time series and the predicted time series of the ALIC station changes obviously, and some components of the BAKE station also increase. Sidereal filtering is one of the proposed methods in multipath mitigation of sites, where time is a crucial parameter. When the satellite altitude, satellite azimuth, and signal-to-noise ratio are excluded, there is also an increase in the STD of some components. The multipath hemispheric mapping model is a multipath mitigation method based on the satellite elevation and satellite azimuth of the station, and many stochastic models for positioning depend on the satellite elevation and the signal-to-noise ratio. Observation residuals have a significant impact on the prediction results because the residuals reflect the suitability of the stochastic model. Besides, the STD is also influenced by the features of the flag type such as the valid data flag, which reflect the quality of the current positioning.

Different features have different effects on the prediction results, and the same feature has different effects on different stations. The influence of these features on the prediction results is nonlinear, and its regularity cannot be obtained intuitively. Therefore, in the feature design, it is recommended to include as many features as possible that may have an impact on the unmodeled error, and the determination of the weight of each feature is handed over to a large amount of training data.

### Conclusions

This study set out to develop a data-driven method for unmodeled error prediction of a specific positioning model. An unmodeled error prediction method based on machine learning has been proposed and the convolution neural

**Table 3** STD of the ALIC station when the different feature is excluded

| Excluded feature | First day | | | Second day | | |
|---|---|---|---|---|---|---|
| | North (mm) | East (mm) | Up (mm) | North (mm) | East (mm) | Up (mm) |
| None | 20.61 | 20.15 | 54.51 | 19.16 | 16.02 | 55.70 |
| GPS time | 20.10 | 20.08 | 64.22 | 18.67 | 17.17 | 62.53 |
| Satellite azimuth | 20.37 | 19.97 | 54.92 | 18.82 | 15.99 | 53.49 |
| Satellite elevation | 20.52 | 20.55 | 53.95 | 18.53 | 16.61 | 53.35 |
| Pseudo-range residual | 19.48 | 20.17 | 57.14 | 18.09 | 15.46 | 57.34 |
| Carrier-phase residual | 21.57 | 23.93 | 62.08 | 19.80 | 18.15 | 61.77 |
| Signal strength | 20.26 | 19.96 | 53.22 | 18.78 | 15.92 | 51.74 |
| Valid data flag | 21.65 | 20.10 | 55.41 | 18.84 | 16.10 | 55.05 |
| Cycle-slip flag | 23.72 | 22.72 | 63.91 | 21.00 | 16.17 | 65.07 |
| Ambiguity flag | 21.46 | 21.41 | 55.21 | 19.68 | 15.75 | 52.00 |
| Carrier-lock count | 20.52 | 20.30 | 53.43 | 18.50 | 15.93 | 53.28 |
| Data outage count | 20.25 | 20.16 | 54.30 | 18.55 | 16.77 | 54.47 |
| Cycle-slip count | 20.83 | 21.85 | 52.96 | 19.75 | 15.87 | 51.89 |
| Data reject count | 20.47 | 20.63 | 59.62 | 18.20 | 16.22 | 59.28 |

**Table 4** STD of the BAKE station when the different feature is excluded

| Excluded Feature | First day | | | Second day | | |
|---|---|---|---|---|---|---|
| | North (mm) | East (mm) | Up (mm) | North (mm) | East (mm) | Up (mm) |
| None | 47.17 | 25.85 | 53.75 | 46.80 | 25.02 | 46.15 |
| GPS time | 48.11 | 23.55 | 52.33 | 40.64 | 23.44 | 43.89 |
| Satellite azimuth | 45.26 | 24.75 | 51.09 | 44.03 | 23.65 | 43.33 |
| Satellite elevation | 50.95 | 26.14 | 53.31 | 45.84 | 26.53 | 46.18 |
| Pseudo-range residual | 45.87 | 24.20 | 51.99 | 42.67 | 25.28 | 44.66 |
| Carrier-phase residual | 45.55 | 31.52 | 54.24 | 44.63 | 32.40 | 50.88 |
| Signal strength | 72.73 | 37.53 | 61.04 | 66.03 | 32.51 | 51.99 |
| Valid data flag | 47.34 | 24.82 | 51.89 | 44.63 | 24.16 | 44.47 |
| Cycle-slip flag | 89.83 | 44.05 | 120.66 | 92.61 | 49.31 | 113.86 |
| Ambiguity flag | 48.73 | 24.06 | 53.96 | 44.29 | 26.75 | 49.48 |
| Carrier-lock count | 75.08 | 40.13 | 71.18 | 81.05 | 37.72 | 58.30 |
| Data outage count | 47.11 | 24.28 | 53.87 | 42.96 | 25.58 | 46.15 |
| Cycle-slip count | 46.05 | 40.68 | 51.29 | 39.17 | 36.88 | 44.83 |
| Data reject count | 45.98 | 24.07 | 52.43 | 41.86 | 27.56 | 45.12 |

network is adopted as the regression model. The effectiveness of this method applying to the PPP kinematic positioning is verified by the historical observation data of IGS stations distributed all over the world. Wavelet analysis was employed to evaluate the prediction results. The most obvious finding from this study is that the data-driven model can effectively predict the unmodeled error in the GNSS positioning, especially in the low-frequency component. More training data to complete the training is more conducive to obtaining better prediction results. However, the prediction results do not keep improving with increasing training data but started to stabilize after reaching about 14 days of training data. In addition, we have investigated 13 training features and found that for various stations, different features have different importance in the neural network time series

prediction. The impact of these features on the prediction results is nonlinear and complex, and it is recommended to include as many features as possible that may have an impact on unmodeled error. This is the first study that has evaluated the effectiveness of the GNSS unmodeled error prediction method based on historical data. Although only the PPP kinematic positioning was used in the experiment, this method should apply to other positioning strategies, which will be carried out in future research.

# References

Abadi M et al (2016) Tensorflow: a system for large-scale machine learning. In: 12th symposium on operating systems design and implementation, Savannah, USA, 2–4 Nov 2016, pp 265–283

Addison PS (2017) The illustrated wavelet transform handbook: introductory theory and applications in science, engineering, medicine, and finance. CRC Press

Bevis M, Brown A (2014) Trajectory models and reference frames for crustal motion geodesy. J Geod 88:283–311. https://doi.org/10.1007/s00190-013-0685-5

Bevis M, Bedford J, Caccamise DJ II (2020) The art and science of trajectory modelling. In: Montillet JP, Bos M (eds) Geodetic time series analysis in earth sciences. Springer Geophysics. Springer, Cham. https://doi.org/10.1007/978-3-030-21718-1_1

Bos MS, Montillet JP, Williams SDP, Fernandes RMS (2020) Introduction to geodetic time series analysis. In: Montillet JP, Bos M (eds) Geodetic time series analysis in earth sciences. Springer Geophysics. Springer, Cham. https://doi.org/10.1007/978-3-030-21718-1_2

Buduma N, Locascio N (2017) Fundamentals of deep learning: designing next-generation machine intelligence algorithms. O'Reilly Media, Inc.

Chen L, Ali-LöyttyPiche´ SR, Wu L (2012) Mobile tracking in mixed line-of-sight/non-line-of-sight conditions: algorithm and theoretical lower bound. Wirel Pers Commun 65(4):753–771

Chen L, Piché R, Kuusniemi H, Chen R (2014) Adaptive mobile tracking in unknown non-line-of-sight conditions with application to digital TV networks EURASIP. J Adv Signal Process 1:22

Chen L et al (2017) Robustness, security and privacy in location-based services for future IoT: a survey. IEEE Access 5:8956–8977

Choi K, Bilich A, Larson KM, Axelrad P (2004) Modified sidereal filtering: implications for high-rate GPS positioning. Geophys Res Lett. https://doi.org/10.1029/2004GL021621

Dong D, Fang P, Bock Y, Webb F, Prawirodirdjo L, Kedar S, Jamason P (2006) Spatiotemporal filtering using principal component analysis and Karhunen–Loeve expansion approaches for regional GPS network analysis. J Geophys Res. https://doi.org/10.1029/2005JB003806C

Dong D, Wang M, Chen W, Zeng Z, Song L, Zhang Q, Cai M, Cheng Y, Lv J (2016) Mitigation of multipath effect in GNSS short baseline positioning by the multipath hemispherical map. J Geodesy 90(3):255–262

Engels O (2020) Stochastic modelling of geophysical signal constituents within a Kalman filter framework. In: Montillet JP, Bos M (eds) Geodetic time series analysis in earth sciences. Springer Geophysics. Springer, Cham. https://doi.org/10.1007/978-3-030-21718-1_8

Goodfellow I, Bengio Y, Courville A (2016) Deep learning. MIT Press, Cambridge

Gruszczynska M, Rosat S, Klos A et al (2018) Multichannel singular spectrum analysis in the estimates of common environmental effects affecting GPS observations. Pure Appl Geophys 175:1805–1822. https://doi.org/10.1007/s00024-018-1814-0

Gulli A, Pal S (2017) Deep learning with Keras. Packt Publishing Ltd, Birmingham

Hoque M, Jakowski N (2008) Mitigation of higher order ionospheric effects on GNSS users in Europe. GPS Solut 12(2):87–97

Hsu L-T (2017) GNSS multipath detection using a machine learning approach. In: 2017 IEEE 20th international conference on intelligent transportation systems (ITSC), Yokohama, Japan, 16–19 Aug 2017. IEEE

Kingma DP, Ba JL (2015) Adam: a method for stochastic optimization. In: ICLR2015: international conference on learning representations 2015, Venue San Diego, CA, 7–9 May 2015

Kirby M, Sirovich L (1990) Application of the Karhunen-Loeve procedure for the characterization of human faces. IEEE Trans Pattern Anal Mach Intell 12(1):103–108

Kouba J, Héroux P (2001) Precise point positioning using IGS orbit and clock products. GPS Solut 5(2):12–28

Langbein J (2017) Improved efficiency of maximum likelihood analysis of time series with temporally correlated errors. J Geod 91:985–994. https://doi.org/10.1007/s00190-017-1002-5

Lau L, Cross P (2006) A new signal-to-noise-ratio based stochastic model for GNSS high-precision carrier phase data processing algorithms in the presence of multipath errors. In: Proceedings of ION GNSS 2006, Institute of navigation, Fort worth, TX, 26–29 Sep 2006, pp 276–285

Lau L, Cross P (2007) Development and testing of a new ray-tracing approach to GNSS carrier-phase multipath modeling. J Geodesy 81(11):713–732

LeCun Y, Bengio Y (1998) Convolutional networks for images, speech, and time series. In: Arbib MA (ed) The handbook of brain theory and neural networks. MIT Press, pp 255–258

Li Y, Xu C, Yi L (2017) Denoising effect of multiscale multiway analysis on high-rate GPS observations. GPS Solut 21(1):31–41

Li B, Zhang Z, Shen Y, Yang L (2018a) A procedure for the significance testing of unmodeled errors in GNSS observations. J Geodesy 92(10):1171–1186

Li Y, Xu C, Yi L, Fang R (2018b) A data-driven approach for denoising GNSS position time series. J Geodesy 92(8):905–922

Malys S, Jensen PA (1990) Geodetic point positioning with GPS carrier beat phase data from the CASA UNO Experiment. Geophys Res Lett 17(5):651–654

Nair V, Hinton GE (2010) Rectified linear units improve restricted Boltzmann machines. In: Proceedings of the 27th international conference on machine learning, Haifa, Israel, 21–24 June 2010, pp 807–814

Olivares-Pulido G, Teferle FN, Hunegnaw A (2020) Markov chain monte carlo and the application to geodetic time series analysis. In: Montillet JP, Bos M (eds) Geodetic time series analysis in earth sciences. Springer Geophysics. Springer, Cham. https://doi.org/10.1007/978-3-030-21718-1_3

Prechelt L (1998) Early stopping-but when? In: Montavon G, Orr GB, Müller KR (eds) Neural Networks: tricks of the trade. Lecture notes in computer science, vol 7700. Springer, Berlin, Heidelberg, pp 55–69

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(12):2052

Ragheb AE, Clarke PJ, Edwards SJ (2007) GPS sidereal filtering: coordinate and carrier-phase-level strategies. J Geodesy 81(5):325–335

Ruch DK, Van Fleet PJ (2009) Wavelet theory: an elementary approach with applications. John Wiley & Sons

Rumelhart DE, Hinton GE, Williams RJ (1988) Learning representations by back-propagating errors. Nature 323(6088):533–536

Shen N, Chen L, Wang L, Lu X, Tao T, Yan J, Chen R (2020) Site-specific real-time GPS multipath mitigation based on coordinate time series window matching. GPS Solut 24(3):82

Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R (2014) Dropout: a simple way to prevent neural networks from overfitting. J Mach Learn Res 15(1):1929–1958

Takasu T (2011) Rtklib: An open source program package for GNSS positioning. Tech Rep, 2013 Software and documentation

Teunissen PJG, Amiri-Simkooei AR (2008) Least-squares variance component estimation. J Geod 82:65–82. https://doi.org/10.1007/s00190-007-0157-x

Wang L, Feng Y, Wang C (2013) Real-time assessment of GNSS observation noise with single receivers. J Glob Position Sys 12(1):73–82

Wdowinski S, Bock Y, Zhang J, Fang P, Genrich J (1997) Southern California permanent GPS geodetic array: spatial filtering of daily positions for estimating coseismic and postseismic displacements induced by the 1992 landers earthquake. J Geophys Res 102(8):18057–18070

Wold S, Esbensen K, Geladi P (1987) Principal component analysis. Chemom Intell Lab Syst 2(1–3):37–52

Zhang Z, Li B (2020) Unmodeled error mitigation for single-frequency multi-GNSS precise positioning based on multi-epoch partial parameterization. Meas Sci Technol 31(2):25008

Zhang Z, Li B, Shen Y (2017) Comparison and analysis of unmodelled errors in GPS and BeiDou signals. Geod Geodyn 8(1):41–48

Zhang Z, Li B, Shen Y, Gao Y, Wang M (2018) Site-specific unmodeled error mitigation for GNSS positioning in urban environments using a real-time adaptive weighting model. Remote Sens 10(7):1157

**Liang Chen** is a professor at State Key Laboratory of Surveying, Mapping and Remote Sensing Information Engineering, Wuhan University. His research interests include navigation new signal theory and method, ubiquitous smartphone positioning, and indoor and outdoor seamless positioning and navigation.

**Lei Wang** is currently an associate research fellow at the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing at Wuhan University. His research interests include GNSS precise positioning and LEO navigation augmentation system.

**Nan Shen** is currently a researcher at Nanjing Tech University. His research interests focus on precise GNSS data processing and its applications.

**Ruizhi Chen** is a professor and director of the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing in Wuhan University, and an academician of the Finnish Academy of Science and Letters. He is an international scholar working in the field of navigation and positioning. He is committed to the theoretical research and core technology development for seamless indoor/outdoor positioning using smartphones.