Acta Geodyn. Geomater., Vol. 23, No. 2 (222), 191–206, 2026
DOI: 10.13168/AGG.2026.0012
journal homepage: https://www.irsm.cas.cz/acta

ORIGINAL PAPER

INTEGRATED SIGNAL DECOMPOSITION AND MACHINE LEARNING FOR GNSS HEIGHT TIME SERIES PREDICTION: A PERFORMANCE EVALUATION

Jia GUO [1], Liu ZHANG [2,*] and Zongheng LU [3]

[1] *School of Surveying and Land Information Engineering, Henan Polytechnic University, Jiaozuo 454003, China*
[2] *Department of Artificial Intelligence, School of Information Engineering, Jiangxi University of Science and Technology, Ganzhou, 341000, China*
[3] *School of Civil and Surveying & Mapping Engineering, Jiangxi University of Science and Technology, Ganzhou, China*

**Corresponding author's e-mail:* 13767377453@163.com

ARTICLE INFO

Article history:
Received 31 January 2026
Accepted 4 March 2026
Available online 16 March 2026

Keywords:
GNSS Height Time Series
Time Series Prediction
Machine Learning
Hybrid Model
Performance Evaluation

ABSTRACT

High-accuracy Global Navigation Satellite System (GNSS) height time series and predictions provide essential references for international terrestrial reference frame establish, crustal deformation monitoring, sea level change assessment and other geodynamic process. In this research, 11 machine learning models with 5 signal decomposition algorithms on GNSS time series prediction are comparatively investigated. For this purpose, we used 25-year GNSS height time series from 13 globally distributed IGS stations with less than 2 % data gaps in the dataset for prediction experiments. Firstly, Performance evaluation based on the evaluation error metrics was conducted on the selected 11 single-models, the Support Vector Machine (SVM) outperformance to the rest models, as an example of BGIS, it was found to be 2.45 mm, 0.05 mm, 3.12 mm, 0.74 and 1.23 for the MAE, MAPE, RMSE, R^2 and WQE respectively. Secondly, by integrating five signal decomposition algorithms—Singular Spectrum Analysis (SSA), Ensemble Empirical Mode Decomposition (EEMD), Variational Mode Decomposition (VMD), Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN), Empirical Mode Decomposition (EMD)—with the SVM model (the optimal standalone model), this study demonstrates that signal decomposition substantially enhances the prediction accuracy of GNSS height time series. Experimental validation identifies the SVM–SSA model as the top performer for long-term prediction; for instance, at the BGIS station, it achieves a WQE of 0.11 and an R^2 of 0.99, markedly outperforming the standalone SVM model (WQE = 1.23, R^2 = 0.74). Finally, this study supplements experiments on hybrid models combining signal decomposition algorithms with (Convolutional Neural Network) CNN and Transformer models, and comparative analysis across 13 globally distributed IGS stations shows that the SVM–SSA model attains the optimal performance, with a mean R^2 of 0.98 and a mean WQE of 0.20, compared with other hybrid models (e.g., EEMD–CNN, WQE = 0.46; EEMD–Transformer, WQE = 1.28). These results collectively indicate that the SVM exhibits more stable learning capability and stronger generalization within the decomposition–prediction framework, rendering it suitable for high-accuracy GNSS height forecasting across diverse global scenarios.

1. INTRODUCTION

Over the past three decades, the field of GNSS height time series prediction has seen remarkable advancements, largely driven by the enhanced capacity of machine learning to capture complex temporal features (Siemuri et al., 2022; Gao et al., 2022; Zhou et al., 2025; Jiao et al., 2025; Wang et al., 2026). As a critical data source for monitoring geophysical phenomena such as vertical crustal deformation (Richter et al., 2016), sea-level change (He et al., 2025), and surface mass redistribution (Wen et al., 2023), high-accuracy prediction of GNSS height time series is of significant importance for studying glacial isostatic adjustment and relative sea-level rise in coastal regions (Bogusz et al., 2019; Zhou et al., 2022; Huang et al., 2025). Numerous studies have indicated that GNSS time series exhibit pronounced seasonal variations, encompassing annual and semi-annual fluctuations (Chen et al., 2013; He et al., 2017; Klos et al., 2018; Savchuk et al., 2023). The primary drivers include seasonal mass loading from the atmosphere, oceans, and terrestrial water within the Earth system, along with long-period solid Earth tides, which induce periodic station displacements (Li et al., 2025a; Li et al., 2025b). These signals are further superimposed by seasonally correlated errors from the station environment (e.g., vegetation, temperature) and satellite orbit artifacts. This complex interplay of multiple geophysical and technical signals triggers corresponding variations, rendering traditional time series prediction methods—such as ARIMA and Kalman filtering—inadequate in both accuracy and adaptability to satisfy the multifaceted demands of GNSS vertical displacement prediction (Forootan et al., 2021; Gao et al., 2022).

To overcome the limitations of conventional approaches on GNSS Height Time Series Prediction, a range of machine learning and machine learning

Cite this article as: Guo J, Zhang L, Lu Z: Integrated signal decomposition and machine learning for GNSS height time series prediction: A performance evaluation. Acta Geodyn. Geomater., 23, No. 2 (222), 191–206, 2026. DOI: 10.13168/AGG.2026.0012

models have been progressively adopted and refined in the field of time series prediction. Among these, decision tree (DT) methods emerged in the 1980s the development of algorithms such as ID3 (Quinlan, 1986). In the 1990s, SVM overcame linear limitations through kernel functions (Boser et al., 1992), enabling their systematic application to time series prediction. By leveraging their ability to model nonlinear, non-stationary series with unknown prior distributions, SVM delivered reliable predictions for power load, weather, and environmental state monitoring. The results demonstrated that SVM could achieve more accurate forecasts for time series generated by complex systems without requiring a predefined system model (Sapankevych and Sankar, 2009). The generalized regression neural network (GRNN), in contrast, enables rapid predictions based on statistical learning theory. Following the year 2000, ensemble learning entered a period of rapid development: Random Forest (RF) enhanced model stability through bagging (Breiman, 2001), while the least-squares boosting (LSBoost) algorithm optimized performance by sequentially correcting residuals (Laurent and Massart, 2000). Against the backdrop of the machine learning revolution, CNN transformed image-processing tasks through convolutional operations and weight-sharing mechanisms (LeCun et al., 2002). CNN was introduced for energy time-series prediction, utilizing four sets of solar photovoltaic generation and power load data from three countries to predict next-day values. In comparative assessments with long short-term memory (LSTM) networks, CNN exhibited superior prediction accuracy and shorter training times, establishing them as an effective predictive model in this domain (Koprinska et al., 2018). Subsequently, LSTM networks, gated recurrent units (GRUs), and bidirectional long short-term memory (BiLSTM) networks were successively developed to address the vanishing gradient problem and enhance sequential dependency modeling in recurrent neural network (RNN). In 2017, the Transformer model redefined the approach to sequence modeling through its self-attention mechanism (Vaswani et al., 2017). In 2018, the temporal convolutional network (TCN) overcame the parallel-processing limitations of sequential data by employing dilated convolutions (Bai et al., 2018). A model based on the TCN was proposed for energy-related time-series prediction, with results on Spanish national electricity demand and electric vehicle charging station datasets showing that TCN achieved the lowest WAPE values of 0.0093 and 0.4228, respectively, establishing it as the optimal model (Lara-Benítez et al., 2020). However, single prediction models often face limitations in accuracy and robustness when dealing with the nonlinearity, nonstationarity, and noise interference inherent in GNSS time series. Beyond the optimization of individual models, hybrid strategies that integrate predictive models with data decomposition methods have increasingly emerged as a research focus, demonstrating superior performance across multiple domains. Meanwhile, GNSS-derived data has been effectively paired with advanced machine learning strategies to improve geoscientific prediction reliability. Su et al. (2026) demonstrated this by developing a GNSS-PWV-based rainfall forecast model, integrating a multi-algorithm fusion framework and rainfall event constraint to enhance accuracy. A hybrid model combining CEEMDAN and LSTM was proposed for financial time-series prediction. Evaluated on the daily closing prices of four major global stock indices, including the S&P500 and the Hang Seng Index from 2007 to 2017, the CEEMDAN-LSTM model achieved the highest accuracy in one-step-ahead prediction (Cao et al., 2019). An enhanced EMD-SVM model was developed for monthly runoff prediction. Applied to monthly runoff data from hydrological stations in the Wei River Basin, including Huaxian, Lintong, and Xianyang stations, the EMD-SVM model yielded the lowest RMSE, MAE, and MAPE values across all three stations, demonstrating its robustness and high consistency (Huang et al., 2014). A hydrological time-series analysis and prediction model based on SSA was proposed, using hydrological data from the Vouga River Basin in Portugal, covering three time series: annual precipitation, monthly runoff, and hourly water temperature. Since SSA overcomes the limitations of models like ARMA regarding data stationarity and normality, it accurately extracts precipitation trends, runoff trends and oscillatory components, as well as the complete water temperature signal, while demonstrating high predictive accuracy for the extracted components (Marques et al., 2006).

To address the limitations of these methods, we propose a novel framework that, we make a performance evaluation on different machine learning models on GNSS height time series prediction with various signal decomposition algorithms. A validation experiment was conducted by randomly selecting 13 global GNSS stations with long-term observations. The structure of this paper is organized as follows: Section 2 provides a brief description of the GNSS and environmental loading data, along with the relevant models used in our analysis. Section 3 presents the results and discussion, including the comparative analysis. Finally, Section 4 draws the conclusions of this study.

2. DATA AND METHODS

2.1. GNSS HEIGHT TIME SERIES AND PREPROCESSING

To evaluate the prediction performance of integrated signal decomposition and machine learning, we selected long-period GNSS time series for experiments. The Combined JPL and GIPSY GNSS daily height time series from Extended Solid Earth Science ESDR System (ES3, available at http://garner.ucsd.edu/pub/measuresESESES products/Timeseries/, Bock et al., 2022; Huang et al., 2025), which recorded between 2000.0 and 2025.0 were used

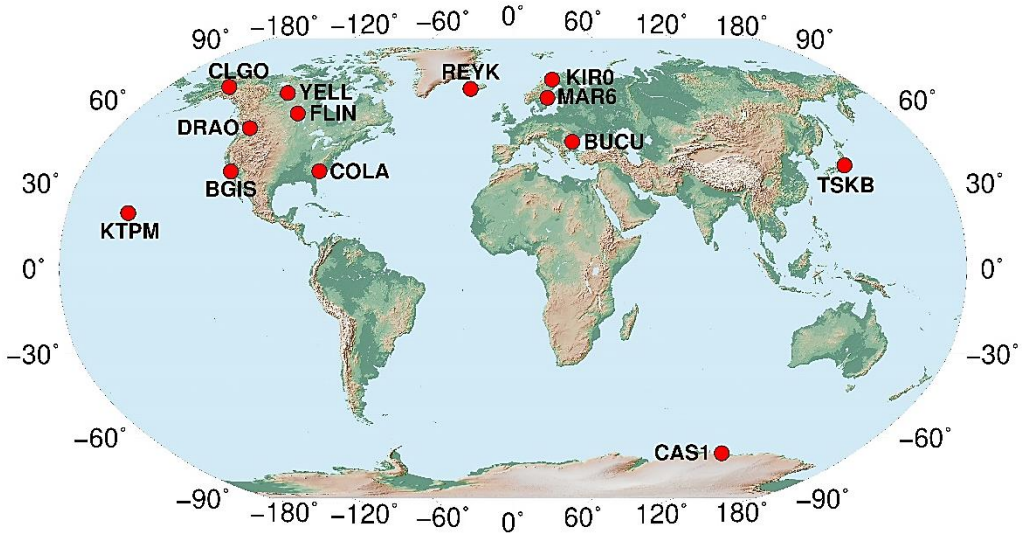

Fig. 1 Distribution of the selected 13 GNSS stations analyzed in this work.

Table 1 Basic Information of the 13 Selected GNSS Stations.

Station	Lat (°).	Lon (°).	Date Gap
BGIS	33.97	33.97	1.03 %
BUCU	44.46	26.13	1.54 %
CAS1	-66.28	110.52	1.56 %
CLGO	64.87	212.14	0.09 %
COLA	34.08	278.88	1.90 %
DRAO	49.32	240.38	0.30 %
FLIN	54.73	258.02	0.59 %
KIR0	67.88	21.06	0.42 %
KTPM	19.34	204.84	1.74 %
MAR6	60.60	17.26	0.11 %
REYK	64.14	338.04	1.05 %
TSKB	36.11	140.09	0.63 %
YELL	62.48	245.52	0.06 %

to evaluate performance comparison of different machine learning algorithms. The locations of the selected 13 global distributed GNSS stations are shown in Figure 1 and such a selection is based on the following considerations. Firstly, each GNSS station should have at least 25 years of continuous time series data between 2000.0 and 2025.0. Secondly, GNSS height time series must exhibit low data gaps of less than 2 % to eliminate the effect of data gapping on prediction accuracy, and the average rate of data gapping is about 0.85 % for the 13 GNSS sites (see Table 1). Thirdly, to validate the generalizability of different prediction algorithms, we randomly selected GNSS stations worldwide, ensuring that the chosen sites are approximately globally distributed.

In addition, the combined JPL and GIPSY GNSS daily height time series were processed with gross error elimination with 3IQR (He et al., 2020) and offset correction, which mitigated the impacts of certain local effects. The corrected GNSS height series are presented in Figure 2.

2.2. MACHINE LEARNING ALGORITHMS FOR TIME SERIES PREDICTION

2.2.1. LSTM

As an improved variant of RNN, LSTM networks are designed to address the problems of vanishing gradients, exploding gradients, and insufficient long-term memory retention that conventional RNN suffer from when modeling long-term dependencies (Hochreiter and Schmidhuber, 1997). Generally, an LSTM network is composed of three gating units: the forget gate, the input gate, and the output gate. Specifically, the forget gate

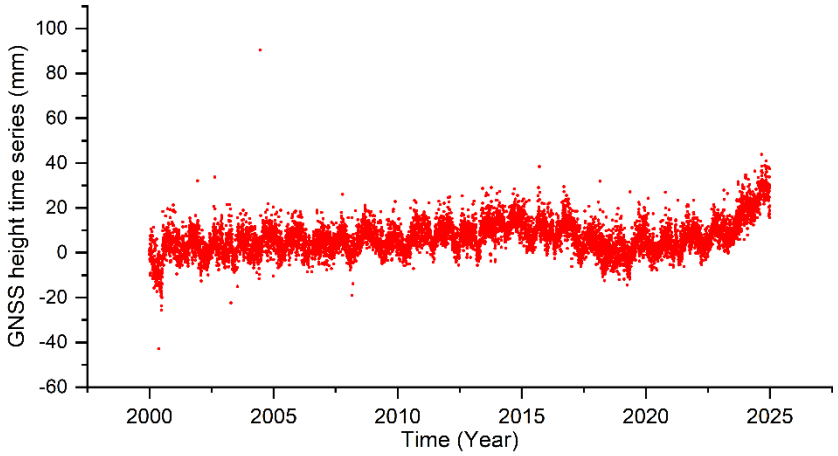

Fig. 2 Height Time Series from REYK.

determines whether to retain or discard historical information; the input gate regulates the extent to which new information is written into the cell state; and the output gate governs how the current cell state contributes to the final output (Liu et al., 2022; Landi et al., 2021).

2.2.2. BILSTM

The deep-bidirectional LSTMs (Schuster and Paliwal, 1997) are an extension of the described LSTM models. The BiLSTM covers two parallel LSTM layers in both the forward and backward propagation directions. One layer performs is the operations following the same direction of the data sequence and the other layer applies its operations on in reverse direction of the data sequence. It has been reported that using BiLSTM models outperforms regular LSTMs (Baldi et al., 1999).

2.2.3. GRNN

The GRNN, first proposed by Specht (1991), is a neural network architecture designed to model the nonlinear functional relationships between a response variable and a set of independent explanatory variables. The training process of the GRNN is designed in such a way to allow for the free parameters to be chosen automatically. The automatic optimization of the free parameters is achieved by iteratively training and testing the network on a wide range of potential values for each free parameter and selecting the value that gives the smaller accuracy error.

2.2.4. DT

The DT method is a widely used data mining technique, which can be applied to develop classification systems based on multiple covariates or construct predictive algorithms for target variables (Song and Lu, 2015) Similar to other analytical methodologies, the DT approach is not without inherent limitations. A primary drawback lies in its susceptibility to overfitting and underfitting, a concern that becomes particularly pronounced when dealing with small-scale datasets. Moreover, this method simplifies intricate relationships between input variables and the target variable by partitioning the original input variables into distinct, meaningful subgroups, which may potentially overlook nuanced or non-linear interdependencies inherent in the data. Loh (2014) and Strobl (2014) provided a comprehensive review of the statistical literature of classification tree methods.

2.2.5. LSBOOST

LSBoost (Pragati et al., 2024), is a member of the boosting algorithm family in machine learning. It enhances the overall performance of the model by iteratively fitting weak learners (typically DTs) to the residuals generated from the previous iteration. This algorithm minimizes the least squares loss function, which quantifies the squared deviation between the predicted values and the actual values. Mathematically, it can be represented as the minimization of the sum of squared deviations between the observed values and the predicted values.

$$min\ i\ \text{mise} \sum_{i=1}^{N}(y_i - F(x_i))^2 \quad (4)$$

2.2.6. TCN

The TCN approach was initially developed to examine long-range patterns via a hierarchy of temporal convolutional filters (Lea et al., 2017). As a specialized CNN variant, TCN features convolutional layers that maintain consistent input-output dimensionality. Using one-dimensional convolution for sequential data, it integrates dilated and causal convolution to enable residual connections and capture long-range dependencies—designs that enhance network depth while mitigating gradient vanishing. Notably, causal convolution constrains outputs to depend only on current and historical input features, eliminating look-ahead bias and information leakage in predictive tasks. Extended convolution, by contrast, expands causal convolution's scope, improving performance on long-period time series and facilitating effective learning.

2.2.7. RF

RF, or random decision forest, proposed by Breiman (2001), is an ensemble learning method for both of classification and regression problems. The RF forms the so-called "forest" by ensembling multiple DTs, and there are several widely used ensemble strategies, such as Bagging (Breiman, 1996), Boosting (Freund and Schapire, 1997), and random subspace technique (Ho, 1998). Unlike global machine learning models such as artificial neural networks (ANNs) and support vector regression (SVR)—which strive to construct a single global model directly from data—ensemble learning models (with RF as a representative) operate by building a set of base models and then integrating them. This paradigm enables ensemble learning models to often exhibit superior performance, especially when addressing complex systems.

2.2.8. CNN

A CNN is a biologically-inspired deep neural network (DNN). It is widely used in fields including image and video recognition, image classification, medical image analysis, and natural language processing, and also exhibits significant advantages in processing time-based streaming data (Sainath et al., 2013; Schmidhuber, 2012; Brocardo et al., 2017). Structurally, a CNN consists of sequential convolutional layers, where each layer's output connects only to local regions of the previous layer's input. A filter (or weight matrix) is slid over the input data, with the dot product (i.e., convolution operation) calculated at each position, enabling the model to learn filters that recognize specific input patterns.

2.2.9. GRU

The GRU networks fall into the category of RNN, i.e., neural networks whose underlying topology of inter-neuronal connections contains at least one cycle. GRU model was proposed by Cho et al. (2014). on statistical machine translation. It combines with the forget gate and the input gate into a single update gate. It is also mixed with cellular state and hidden state. GRU can thus keep knowledge and delete it as needed. This method performs quite well in time series. The equations controlling the GRU's updating mechanism consist as follows (Jia et al., 2025):

$$\begin{aligned} r_t &= \sigma(W_r x_t + U_r h_{t-1} + b_r) \\ z_t &= \sigma(W_z x_t + U_z h_{t-1} + b_z) \\ \tilde{h}_t &= tanh(W_h x_t + U_h(r_t \times h_{t-1}) + b_h) \\ h_t &= (1 - z_t) \times h_{t-1} + z_t \times \tilde{h}_t \end{aligned} \quad (5)$$

x_t represents input vector; h_{t-1} is previous time step memory state; h_t is current time step memory state; r_t represents reset gate state; z_t is update gate state; $\tilde{h}$ represents current candidate state; W_r represents weight matrix for reset gate; W_z is weight matrix for update gate; W_h represents weights matrix for the candidate state; σ represents sigmoid activation function

2.10. TRANSFORMER

The Transformer architecture, originally proposed by Vaswani et al. (2017), Devlin et al. (2019) and Han et al. (2021, 2022) is a class of machine learning models that take self-attention or scaled dot-product operation as the primary learning mechanism. It has revolutionized time series prediction by utilizing self-attention to capture long-range dependencies without relying on recurrent or convolutional structures. A key advantage of the Transformer in these fields is its ability to enhance prediction power by increasing model size, where model capacity is typically controlled by the number of layers, commonly set between 12 and 128.

The canonical self-attention in (Vaswani et al., 2017) is defined based on the tuple inputs, query, key and value, which performs the scaled dot-product as

$$Attention(Q, K, V) = Soft\,max(\frac{QK^T}{\sqrt{d_k}})V \quad (6)$$

Where Q, K, V are query, key, and value matrices, $Q \in R^{L_Q \times d}$, $K \in R^{L_K \times d}$, $V \in R^{L_V \times d}$ and d is the input dimension d_k is the dimension of the key vectors, QK^T computes similarity scores between queries and keys, the scaling factor $\sqrt{d_k}$ prevents gradient vanishing in Soft max.

2.2.11. SVM

SVM and SVR are based on statistical learning theory, or VC theory (VC-Vapnik, Chervonenkis), developed over the last several decades (Sapankevych and Sankar, 2009). Unlike most of the traditional neural network models which implement the empirical risk minimization principle, SVMs implement the structural risk minimization principle which seeks to minimize an upper bound of the generalization error rather than minimize the training error.

SVM is linear learning machines which means that a linear function ($y = f(x) = wx + b$) is always used to solve the regression problem. The best line is defined to be that line which minimize the following cost (Thissen et al.,2003):

$$\begin{aligned} Q &= \frac{1}{2} \geq \|w\|^2 + C \sum_{i=1}^{N} L^{\varepsilon}(x_i, y_i, f) \\ \text{subject to} &\begin{cases} y_i - wx_i - b \leq \varepsilon + \xi_i \\ wx_i + b - y_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases} \end{aligned} \quad (7)$$

The first term $C\frac{1}{n}\sum_{i=1}^{n} L_{\varepsilon}(d_i, y_i) + \frac{1}{2}$ is the empirical error (risk). They are measured by the ε −insensitive loss function. The second term $\sum_{i=1}^{n} L_{\varepsilon}(d_i, y_i) + \frac{1}{2}\|w\|$ on the other hand, C is the regularization term. C is called the regularization constant and it determines the trade-off between the empirical risk and the regularization term. Increasing the value of C leads to an increase in the relative importance of the empirical risk with respect to the regularization term.

2.3. TIME SERIES DECOMPOSITION METHODS

2.3.1. EMD

EMD is a signal preprocessing algorithm that was introduced by Huang et al. (1998). The EMD is an adaptive data processing technique tailored for decomposing nonlinear and nonstationary time series, with extensive applications across diverse research domains. It is based on the simple assumption that any signal consists of different subtasks. Non-linear and non-stationary time series can be decomposed into a group of zero mean and quasi-periodic signals, where each component is called IMF and a residue component, as follows:

$$\boldsymbol{x(t)} = \sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} \boldsymbol{h_i(t)} + \boldsymbol{r(t)} \tag{8}$$

where $x(t)$ represents the IMF components, and $h_i(t)$ is a residual component. The residual $r(t)$ could be a constant, or a function that contains only a single extremum from which no additional IMFs can be extracted.

2.3.2. EEMD

EEMD is an extended form of the EMD method., in which EMD is performed on an ensemble of initial signals, each perturbed by low-amplitude white noise (Wu and Huang, 2009). EEMD was improved from EMD to overcome modal aliasing problems by adding white noise (Wu and Huang, 2004), and it has been widely used for the decomposition of nonlinear and nonstationary signals. EEMD has the advantages of robust self-adaptability and local variation. This method adds white noise of finite amplitude to the signal $x_i(t) = x(t) + w_i(t)$.
where $x(t)$ represents the original pure target signal function as a function of time, $x_i(t)$ represents the $i-$ th noise added signal f unction varying with time, and $w_i(t)$ represents $i-$ th Gaussian white noise function that changes over time.

2.3.3. VMD

Dragomiretskiy and Zosso (2013) proposed the VMD method VMD makes full use of variational models to produce a given number of discrete subcomponents. The VMD algorithm for evaluating the bandwidth of a one-dimensional signal proceeds as follows: firstly, the Hilbert transform is applied to each mode to derive the associated analytic signal, thereby obtaining a unilateral frequency spectrum; secondly, the frequency spectrum of each mode is mixed with an exponential function tuned to its corresponding estimated center frequency, which shifts the spectrum to the baseband; finally, the bandwidth is estimated via Gaussian smoothness (e.g., the squared L2-norm of the gradient). Then, the constrained variational problem is given by Dragomiretskiy and Zosso:

$$\min_{u_k, w_k} = \left\{ \sum_{k=1}^{K} \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-jw_k t} \right\|_2^2 \right\}, \quad s.t. \sum_{k=1}^{K} u_k(t) = f(t) \tag{9}$$

where K is the number of models. t represents the time script, $f(t)$ is the signal needed to be decomposed, δ implicates the Dirac distribution respectively. j is complex square root of -1, and $*$ denotes the convolution operator.

2.3.4. CEEMDAN

CEEMDAN is a new noise-aided data analysis method proposed by Torres et al. (2011). As an improved version of EEMD, it addresses the issue where the EEMD algorithm cannot completely remove Gaussian white noise after signal reconstruction, which in turn leads to reconstruction errors. This method can more effectively eliminate mode mixing, with reconstruction errors close to zero and significantly reduced computational costs. Define the operator $E_j(\cdot)$ that produces the jth mode obtained by EMD, The final residue can be expressed as:

$$R_M(t) = S(t) - \sum_{j=1}^{M} \overline{IMF_j} \tag{10}$$

In this case, M denotes the total number of IMFs. The IMFs together represent the characteristics of the original signal at different time scales. The residue clearly shows the trend of the original sequence; it exhibits greater smoothness and effectively reduces the prediction error.

2.3.5. SSA

SSA is a non-parametric time series signal analysis method used to extract trends, periodic components, and noise from signals (Hassani, 2007). Combining the ideas of Principal Component Analysis (PCA) and time series analysis, SSA serves as a powerful signal decomposition tool.

The basic process of SSA decomposition consists of the following four steps:

The first step called the embedding step, Convert the original time series $x(t)$ into a Trajectory Matrix. First, select a window length L, then construct a $K * L$ trajectory matrix X, where $K = N - L + 1$ and N is the length of the time series.

$$X = \begin{pmatrix} x_1 & x_2 & \cdots & x_L \\ x_2 & x_3 & \cdots & x_{L+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_K & x_{K+1} & \cdots & x_N \end{pmatrix} \tag{11}$$

Among them, each column of the trajectory matrix corresponds to a segment of the original time series, referred to as a delay vector.

The next step is the Singular Value Decomposition (SVD) of the trajectory matrix into a sum of rank-one bi-orthogonal elementary matrices

$$X = X_1 + \cdots + X_L \tag{12}$$

The elementary matrices $X_i$are given by $X_i = s_i U_i V_i^T$. The collection (s_i, U_i, V_i) is called the i − th eigentriple of the SVD.

In the third step, the grouping step. It is made a partition of the indices set into m disjoint subsets $I_1, \ldots, I_m$, corresponding to split the elementary matrices into m groups. Let $I = \{i_1, \ldots, i_p\}$, then the resultant matrix $X_I$is defined as $X_I = X_{i_1} + \cdots + X_{i_p}$. The resultant matrices are computed for $I = I_1, \ldots, I_m$ and substituting in Equation (13) one obtains the new expansion

$$X = X_{I_1} + \cdots + X_{I_m} \tag{13}$$

where the trajectory matrix is represented as a sum of m resultant matrices. The choice of the sets $I_1, \ldots, I_m$ is called the eigentriple grouping.

The last step in Basic SSA transforms each resultant matrix of the grouped decomposition (13) into a new one- dimensional series of length N, and is called diagonal averaging. Diagonal averaging transfers matrix Y to a series $g_0, \ldots, g_{N-1}$ by the formula:

$$g_k = \begin{cases} \dfrac{1}{k+1}\displaystyle\sum_{m-1}^{k+1} y^*_{m,k-m+2} & 0 \le k < L^* - 1 \\ \dfrac{1}{L^*}\displaystyle\sum_{m=1}^{L^*} y^*_{m,k-m+2} & L^* - 1 \le k < K^* \\ \dfrac{1}{N-k}\displaystyle\sum_{m=k-K^*+2}^{N-k+1} y^*_{m,k-m+2} & K^* \le k < N \end{cases} \tag{14}$$

For proper choices of L and the sets $I_1, \ldots, I_m$, it is possible to associate the components F_k with the trend, oscillations or noise of the original time series F. For a sufficiently vast class of series, its continuation can be successfully accomplished if a number of conditions are gathered. Namely, if the time series has a structure, an algorithm identifying this structure is achieved, a method for the time series continuation is available, and the structure of the time series is preserved for the time period that is to be continued.

2.3.6. PREDICTION ERROR METRICS

We employed the Root Mean Square Error (RMSE), Mean Absolute Error (MAE), We adopted Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Coefficient of Determination (R^2) evaluation metrics to assess the performance of the proposed SVM-SSA prediction model. RMSE, MAE, and MAPE were utilized to quantify the discrepancy between predicted values and actual values. R^2 was applied to assess the fitting quality and predictive capability of the model. The following are the equation expressions of these evaluation metrics (Shen et al., 2025):

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \tag{15}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i| \tag{16}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\hat{y}_i - y_i}{y_i}\right| \times 100\% \tag{17}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{n}(\bar{y}_i - y_i)^2} \tag{18}$$

Where, n denotes the number of data points, y_i stands for the observed value, $\hat{y}_i$ represents the predicted value, and $\bar{y}_i$ is also the average of the observed values.

To provide a more thorough assessment of the model, we adopt the weighted quality evaluation (WQE), a recent and comprehensive evaluation index proposed by Zhou et al. (2025). The WQE is formulated as follows:

$$WQE = 2\left(\frac{X_{RMSE}}{X_{max} - X_{min}}\right) + 3\left(\frac{X_{MAE}}{X_{max} - X_{min}}\right) + MAPE + 4(1 - R^2) \tag{19}$$

Where, X_{max} and X_{min} represent the maximum and minimum values of the dataset, respectively. A lower WQE value signifies superior performance, and WQE is dimensionless.

3. RESULT AND DISCUSSION

In this section, we systematically compare and analyze the prediction accuracy of different machine learning single models and ensemble models with 13 globally distributed IGS time series.

3.1. GENERAL PARAMETERS SETTING FOR DIFFERENT PREDICTION MODELS

The models related in machine learning for GNSS height time series prediction was developed and trained on the MATLAB 2023b platform. To ensure an objective evaluation, all models were assessed using chronologically partitioned training (80 %), validation (10 %), and test (10 %) sets. To eliminate the impact of varying parameter settings, we adopted a consistent parameter configuration based on empirical values, allowing for a more systematic comparative analysis of the accuracy of different prediction models. Its architecture consisted of a fully connected neural network with two hidden layers, each containing 16 neurons (Charco et al., 2020). The model was optimized using the Adam optimizer with an initial learning rate of 0.001 (Tabar and Sisman, 2025; Chen et al., 2024). Training was conducted for 70 (Gao et al., 2024) epochs with a batch size of 90, and a Dropout rate of 0.2 (Chen et al., 2024) was applied as presented in Table 2.

Table 2 Parameter Settings.

Type	Parameter settings
Optimizer	Adam
Learning Rate	0.001
Training Epochs	70
Batch Size	90
Hidden Layers	2
Hidden Units	[16, 16]
Dropout Rate	0.2

Table 3 Comparative Analysis of Single-Model Results at the BGIS Station.

Single model	Example of BGIS				
	MAE	**MAPE**	**RMSE**	**R^2**	**WQE**
SVM	**2.45**	**0.05**	**3.12**	**0.74**	**1.23**
RF	2.54	0.05	3.20	0.73	1.29
CNN	2.73	0.05	3.43	0.69	1.40
TCN	2.66	0.05	3.37	0.70	1.41
BiLSTM	2.96	0.06	3.71	0.64	1.69
LSTM	3.05	0.06	3.81	0.62	1.78
GRU	3.06	0.06	3.81	0.62	1.78
Transformer	3.06	0.06	3.82	0.62	1.78
GRNN	3.25	0.06	4.04	0.57	1.99
DT	3.33	0.07	4.19	0.54	2.13
LSBoost	3.32	0.07	4.20	0.53	2.14

Table 4 Comparative Analysis of Single-Model Results at the FLIN Station.

Single model	FLIN				
	MAE	**MAPE**	**RMSE**	**R^2**	**WQE**
SVM	**4.33**	**0.08**	**6.27**	**0.83**	**1.00**
TCN	4.86	0.09	6.70	0.80	1.13
Transformer	5.19	0.09	7.06	0.78	1.23
RF	5.18	0.09	7.14	0.78	1.25
BiLSTM	5.62	0.10	7.55	0.75	1.38
CNN	5.38	0.10	7.62	0.75	1.40
DT	6.06	0.11	7.93	0.73	1.51
GRU	6.44	0.11	8.33	0.70	1.64
LSBoost	6.31	0.11	8.48	0.69	1.69
GRNN	6.92	0.12	8.97	0.65	1.87
LSTM	7.17	0.12	9.21	0.63	1.96

3.2. COMPARATIVE ANALYSIS OF SINGLE-MODEL PREDICTIVE PERFORMANCE

To evaluate the performance of different machine learning models in predicting GNSS height time series, this study constructed eleven distinct models—namely SVM, GRU, LSTM, BiLSTM, RF, LSBoost, TCN, CNN, DT, GRNN, and Transformer—based on GNSS time-series listed in previous Figure 1. The evaluation of eleven single-model methods revealed that for all stations, the SVM model yielded the smallest WQE values. This indicates that, among the 13 long-term GNSS height time series analyzed, SVM achieved the highest prediction accuracy. Table 3 and Table 4 present the comparative analysis prediction evaluation error metrics under different models for BGIS and FLIN stations by random.

As shown in Tables 3 and 4, the SVM model achieved the best predictive performance at both stations. For BGIS the SVM model attained an R^2 of 0.74 and a WQE of 1.23, followed by the RF model (R^2 = 0.73, WQE = 1.29). The relatively strong performance of the RF model can be attributed to its ensemble learning mechanism, which leverages multiple DTs to capture multi-scale information within GNSS height time series in parallel, thereby offering certain advantages in robustly fitting complex temporal patterns. However, the inherent piecewise linear fitting characteristic of DTs limits their capacity to accurately capture continuous and highly nonlinear deformation processes, resulting in slightly lower accuracy compared to the SVM model. In contrast, the LSBoost model demonstrated relatively poor predictive performance (R^2 = 0.53, WQE = 2.14).

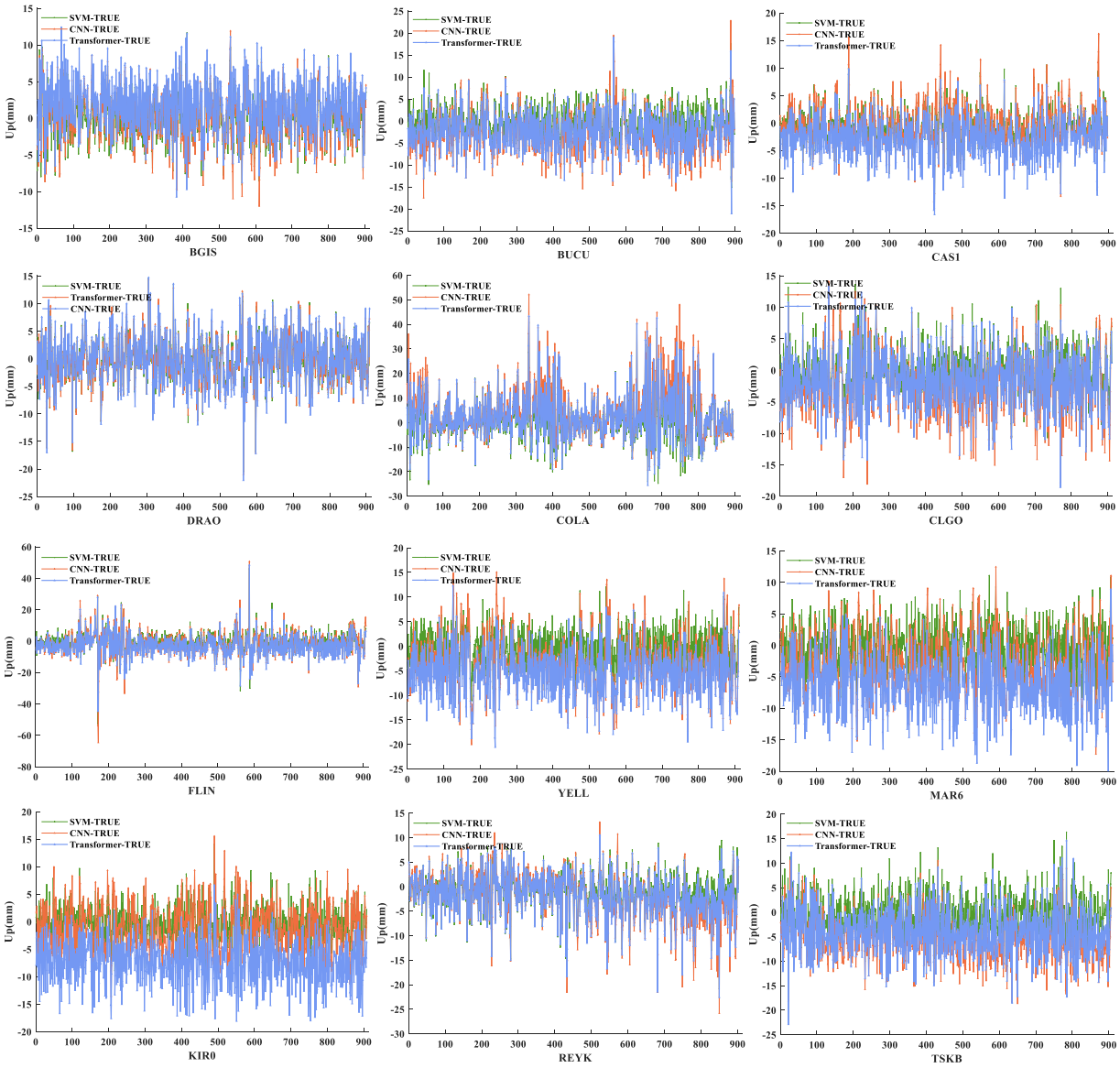

Fig. 3 Comparative Analysis of Predicted Curves for SVM, CNN, and Transformer Models ($TS\_{Prediction}$-TS_{True}).

Compared to LSBoost, the SVM model achieved a 39.62 % improvement in R^2 and a reduction of 0.91 in WQE. At the FLIN station, the SVM model attained an R^2 of 0.83 and a WQE of 1.00. The TCN achieved the second-best performance ($R^2 = 0.80$, WQE = 1.13). While its temporal convolutional architecture enables efficient capture of local temporal patterns, the fixed receptive field of the convolutional layers restricts its ability to adequately represent long-term trends spanning multiple years in the sequence, ultimately resulting in slightly lower overall accuracy compared to the SVM model. In contrast, the LSTM model exhibited relatively poor predictive performance ($R^2 = 0.63$, WQE = 1.96). Compared to the LSTM model, the SVM model achieved a 31.75 % improvement in R^2 and a reduction of 0.96 in WQE.

These results demonstrate that the RBF kernel of the SVM holds a significant advantage in distinguishing between valid signals and noise. It not only adapts effectively to the non-stationary and strongly nonlinear regime observed at the BGIS station, but also maintains robust fitting under the conditions of periodic noise and stable long-term trends present at the FLIN station. Although the RF, CNN, TCN, and Transformer models exhibit slight advantages in their respective strengths—such as feature extraction or dependency modeling—their overall performance in predicting GNSS height time series is generally inferior to that of the SVM model, due to limitations in their adaptability to specific scenarios inherent in their respective architectures.

In addition, to further validate the model performance in practical time-series fitting, Figure 3 visualizes the residual time series (predicted values minus observed values) for the SVM, CNN, and Transformer models across twelve GNSS stations. The results indicate that the SVM model more accurately captures both the overall trends and local

Table 5 Comparative Analysis of Hybrid Model Results at the BGIS Station.

Hybrid model	BGIS Station				
	MAE	MAPE	RMSE	R^2	WQE
SSA-SVM	**0.59**	**0.01**	**0.74**	**0.99**	**0.11**
EEMD-SVM	1.22	0.02	1.55	0.94	0.35
VMD-SVM	1.39	0.03	1.77	0.92	0.44
CEEMDAN-SVM	1.72	0.03	2.08	0.89	0.59
EMD-SVM	1.75	0.03	2.23	0.87	0.67
SVM	2.45	0.05	3.12	0.74	1.23

Table 6 Comparative Analysis of Hybrid Model Results at the FLIN Station.

Hybrid model	FLIN Station				
	MAE	MAPE	RMSE	R^2	WQE
SSA-SVM	**0.93**	**0.02**	**1.27**	**0.99**	**0.09**
VMD-SVM	2.34	0.04	3.16	0.96	0.34
EEMD-SVM	2.07	0.04	3.62	0.94	0.39
CEEMDAN-SVM	3.31	0.06	4.57	0.91	0.59
EMD-SVM	3.38	0.06	4.88	0.90	0.66
SVM	4.33	0.08	6.27	0.83	1.00

fluctuations of the series at most stations, demonstrating superior stability and generalization capability.

3.3. COMPARATIVE ANALYSIS OF SVM HYBRID MODELS INTEGRATED WITH FIVE SIGNAL DECOMPOSITION ALGORITHMS

Recent studies have shown that preprocessing the original series using signal decomposition methods can effectively enhance the predictive performance of GNSS height time series. For example, adaptive SSA has been used to accurately extract periodic signals from GNSS height time series, while a VMD-LSTM model demonstrated strong performance in handling GNSS sequences with complex noise (Li et al., 2024; Chen et al., 2023). Given that the SVM demonstrated the best predictive performance in the single-model comparison (Section 3.2), this study further developed a hybrid modeling framework based on "signal decomposition–SVM prediction." Five signal decomposition methods—SSA, VMD, EEMD, CEEMDAN, and EMD—were employed to isolate the effective components (trend and periodic terms) from the noise in GNSS height series, thereby reducing the interference of random noise on the SVM fitting process and enabling an investigation into the predictive performance of different decomposition-SVM hybrid frameworks. Based on data from the BGIS and FLIN stations, a quantitative evaluation of the SVM model and various hybrid models was conducted using five metrics: MAE, MAPE, RMSE, R^2, and WQE. The results are presented in Tables 5 and 6.

Combined with the data in Tables 5 and 6, the analysis demonstrates that incorporating signal-decomposition methods significantly improve the predictive accuracy of the standalone SVM model. To visually illustrate the advantage of the "signal decomposition–prediction" hybrid framework over the single-model approach, Figures 4 and 5 present the prediction residuals (predicted values minus observed values) for the standalone SVM and various hybrid models (SSA-SVM, VMD-SVM, EEMD-SVM, CEEMDAN-SVM, and EMD-SVM) at the BGIS and FLIN stations, respectively.

Moreover, the results presented in Tables 5 and 6 indicate that the SSA-SVM hybrid model achieved the best predictive performance among all hybrid models. At the BGIS station, the EMD-SVM hybrid model demonstrated relatively poor predictive performance ($R^2 = 0.87$, WQE = 0.67). In contrast, EEMD-SVM, which introduces Gaussian white noise to alleviate the mode-mixing problem inherent in EMD, achieves a preliminary separation of useful components from noise, leading to improved prediction results ($R^2 = 0.94$, WQE = 0.35). However, the noise-added process introduces residual noise, resulting in slightly lower accuracy compared to the SSA-SVM model. By comparison, the SSA-SVM model delivered the best performance, attaining an R^2 of 0.99 and a WQE of only 0.11. Relative to the EMD-SVM model, this represents an improvement of 13.79 % in R^2 and a reduction of 0.56 in WQE. At the FLIN station, the EMD-SVM hybrid model showed relatively poor predictive performance ($R^2 = 0.90$, WQE = 0.66). In comparison, the VMD-SVM model, benefiting from its adaptive mode decomposition capability, performed more effectively in extracting periodic components such as annual and seasonal signals ($R^2 = 0.96$, WQE = 0.34). However, it still exhibits limited coverage of long-period trends spanning multiple years, resulting in slightly lower accuracy than the SSA-SVM model. By comparison, the SSA-SVM model achieved the best predictive

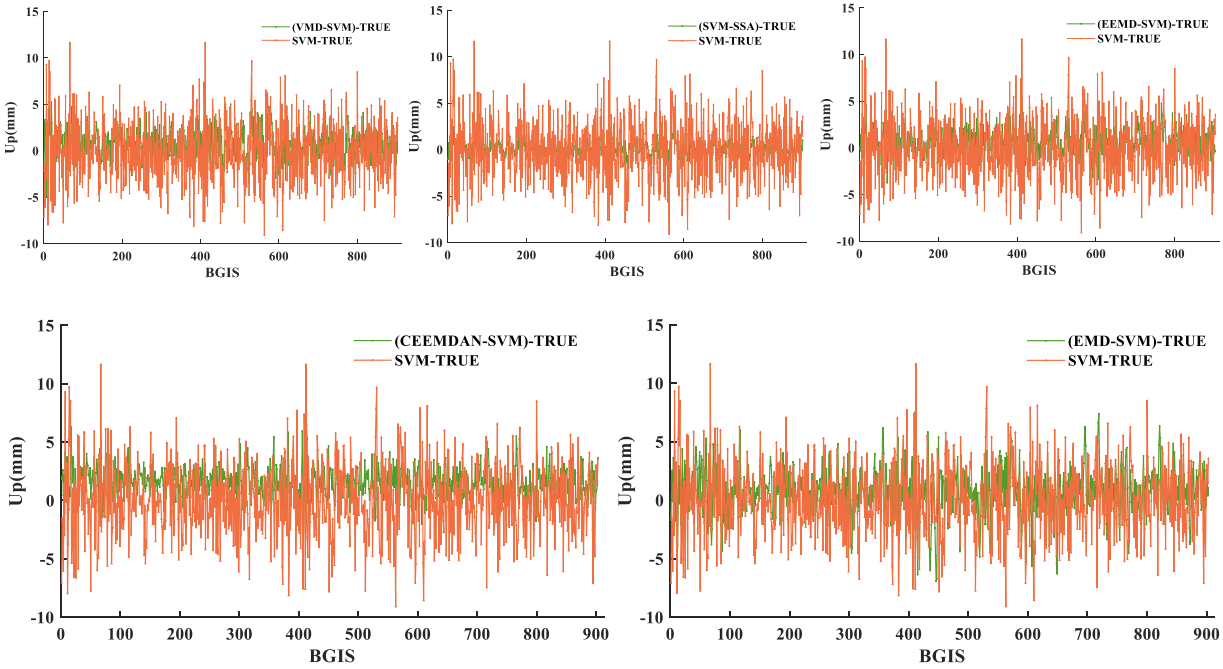

Fig. 4 Comparison of Predictive Models at the BGIS Station.

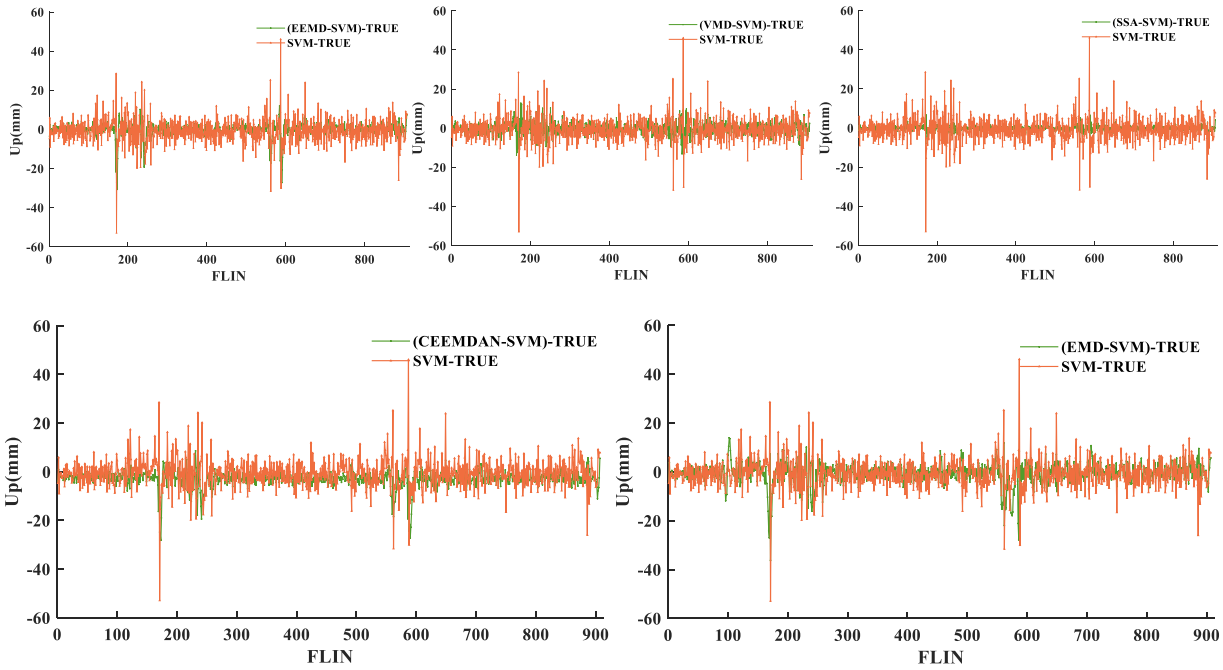

Fig. 5 Comparison of Predictive Models at the FLIN Station.

performance, with an R^2 of 0.99 and a WQE of 0.09. Relative to the EMD-SVM model, this corresponds to an improvement of 10.00 % in R^2 and a reduction of 0.57 in WQE.

These results demonstrate the superior predictive performance of the SSA-SVM model for GNSS height series. By leveraging the singular value decomposition mechanism, SSA accurately separates the effective components (trend and periodic terms) from the noise in the GNSS height time series. This provides the SVM model with feature inputs that are characterized by low noise interference and high integrity of the signal components, thereby enabling the RBF kernel of the SVM to fully exploit its capability to learn complex temporal dependencies.

To systematically evaluate the global applicability of hybrid models, this study further calculated the average performance of different hybrid models across thirteen GNSS stations, with the results presented in Table 7.

As indicated in Table 7, the SSA-SVM hybrid model achieved the best average performance across all stations ($R^2 = 0.98$, WQE = 0.20), with all of its error metrics consistently lower than those of the other

Table 7 Comparison of Average Hybrid-Model Performance Across 13 Stations .

Hybrid model	MAE	MAPE	RMSE	R^2	WQE
SSA-SVM	1.25	0.09	1.50	0.98	0.20
EEMD-SVM	2.08	0.13	2.80	0.93	0.50
VMD-SVM	2.56	0.18	3.09	0.92	0.57
EMD-SVM	3.47	0.24	4.33	0.83	1.05
CEEMDAN-SVM	4.89	0.26	5.46	0.70	1.60

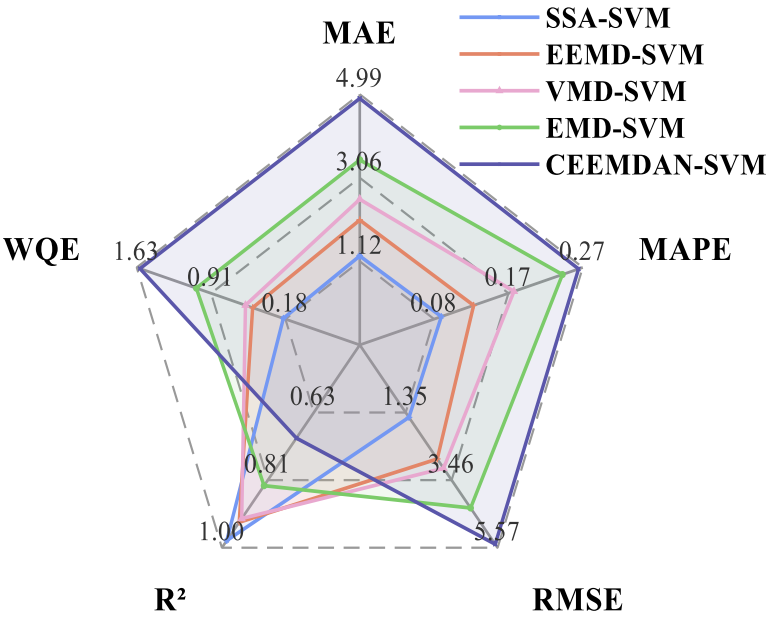

Fig. 6 Radar Chart Comparison of Hybrid Models Across Multiple Metrics.

hybrid models. The EEMD-SVM ($R^2 = 0.93$, WQE = 0.50) and VMD-SVM ($R^2 = 0.92$, WQE = 0.57) models performed second best, while the EMD-SVM ($R^2 = 0.83$, WQE = 1.05) and CEEMDAN-SVM ($R^2 = 0.70$, WQE = 1.60) models exhibited relatively poor average performance. Among them, the SSA-SVM model showed an improvement of 40.00 % in R^2 and a reduction of 1.40 in WQE relative to the CEEMDAN-SVM model.

To provide a more intuitive visualization of the comprehensive performance of different hybrid models across multiple evaluation metrics, Figure 6 employs a radar chart format to compare the multi-metric values derived from the average performance of each model across all stations.

The above comparative results convincingly demonstrate that signal decomposition methods significantly enhance the predictive accuracy of the standalone SVM model, with the SSA-SVM hybrid framework exhibiting superior generalizability—it maintains stable predictive performance across globally distributed stations covering diverse geographic and geophysical characteristics, thereby offering a reliable technical solution for long-term GNSS height prediction.

3.4. COMPARISON OF SVM, CNN, TRANSFORMER WITHIN A DECOMPOSITION-ENSEMBLE FRAMEWORK

Besides, we conducted a supplementary experiment using hybrid models that combine signal decomposition methods with other machine learning models. Due to space constraints, only the results of CNN and Transformer are presented, and the average performance metrics of each hybrid model across all 13 GNSS stations are summarized in Table 8. The experimental results show that SVM-SSA still achieves the smallest WQE, further confirming our conclusion.

As shown in Table 8, SSA-SVM outperformed all other hybrid models. For example, the SSA-SVM model showed an improvement of 4.26 % in R^2 and a reduction of 0.26 in WQE compared to the EEMD-CNN model; compared to the EEMD-Transformer model, it achieved an improvement of 27.27% in R^2 and a reduction of 1.08 in WQE. Synthesizing the analyses from Sections 3.2 to 3.4 reveals that although machine learning models such as CNN and Transformer exhibit relatively strong performance at specific stations, the SVM-based hybrid framework—when integrated into a “decomposition-prediction” structure—learns the

Table 8 **C**omparison of Average Hybrid Model Performance.

Hybrid model	MAE	MAPE	RMSE	R^2	WQE
SSA-SVM	1.25	0.09	1.50	0.98	0.20
EEMD-CNN	1.88	0.12	2.60	0.94	0.46
VMD-CNN	2.28	0.14	2.91	0.92	0.57
SSA-CNN	2.41	0.12	3.07	0.90	0.61
CEEMDAN-CNN	2.70	0.17	3.33	0.90	0.66
EMD-CNN	2.60	0.18	3.40	0.88	0.79
SSA-Transformer	3.71	0.17	4.24	0.79	1.13
CEEMDAN-Transformer	3.54	0.18	4.20	0.79	1.15
VMD-Transformer	3.79	0.21	4.55	0.79	1.18
EEMD-Transformer	3.85	0.22	4.70	0.77	1.28
EMD-Transformer	4.16	0.19	5.01	0.63	1.81

more well-defined temporal components (trend and dominant periodic terms) extracted via SSA decomposition in a more stable and efficient manner, thereby achieving optimal global robustness. In contrast, machine learning models with more complex parameterization are prone to overfitting the residual fluctuations in the decomposed subsequences of GNSS height time series, thereby compromising their generalization ability. Thus, this study confirms the distinct advantage of SVM as the predictor in hybrid models for processing GNSS height time series.

4. CONCLUSION

Based on height time series data from 13 globally distributed GNSS stations, this study systematically compared the performance of eleven machine learning models and "decomposition-prediction" hybrid models for long-term GNSS height prediction. First, the robustness and generalization capability of each standalone model were evaluated across different stations. Subsequently, a hybrid framework based on "signal decomposition–SVM prediction" was constructed to verify the accuracy improvement by signal decomposition. Finally, by comparing the performance of SVM, CNN, and Transformer within the same decomposition framework, the SVM coupled with SSA decomposition was found to have stable predictive capability and strong generalization. Based on the above analysis, the following conclusions are drawn:

1. Among the eleven machine learning models, the SVM exhibited the best robustness and generalizability for GNSS height time series prediction, with a superior average WQE of 1.31 across 13 stations, reflecting its strong ability to distinguish valid signals from noise and adapt to site-specific temporal characteristics.
2. The "signal decomposition–SVM prediction" hybrid framework significantly improved prediction accuracy. The SSA-SVM hybrid model performed best, achieving an R^2 of 0.99 at both the BGIS and FLIN stations, with WQE reduced to 0.11 and 0.09, respectively, demonstrating excellent fitting capability and effective noise suppression.
3. By comparing SSA-SVM with CNN- and Transformer-based hybrid models, the SVM was confirmed to have distinct advantages as a predictor in the decomposition-prediction framework. The SVM stably learns the trend and periodic components extracted by SSA, while the more parametrically complex CNN and Transformer are prone to overfitting high-frequency fluctuations in decomposed subsequences, weakening their generalization.

This study demonstrates that the SSA-SVM hybrid model effectively captures the trend and periodic components in GNSS height time series, providing a high-accuracy and reliable technical solution for long-term GNSS height prediction. Future research will incorporate additional signal decomposition methods, investigate the influence of multivariate inputs, and advance model light weighting and real-time prediction capability to further enhance its practical value in dynamic crustal deformation monitoring and disaster early warning.

ACKNOWLEDGMENTS

This work was supported by 2024 National College Students' Innovation and Entrepreneurship Training Program Project (202410407026). GNSS time series data were obtained from http://garner.ucsd.edu/pub/measuresESESES_products/Timeseries/.

REFERENCES

Bai, S.: 2018, An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. arXiv preprint arXiv:1803.01271. DOI: 10.48550/arXiv.1803.01271

Baldi, P., Brunak, S., Frasconi, P., Soda, G. and Pollastri, G.: 1999, Exploiting the past and the future in protein secondary structure prediction. Bioinformatics, 15, 11, 937–946. DOI: 10.1093/bioinformatics/15.11.937

Bock, Y., Moore, A., Argus, D., Fang, P., Jiang, S. and Kedar, S.: 2022, Extended solid Earth science ESDR system (ES3): Algorithm theoretical basis document, Sept. 19. NASA MEaSUREs project.

Bogusz, J., Klos, A. and Pokonieczny, K.: 2019, Optimal strategy of a GPS position time series analysis for post-glacial rebound investigation in Europe. Remote Sens., 11, 10, 1209. DOI: 10.3390/rs11101209

Boser, B.E., Guyon, I.M. and Vapnik, V.N.: 1992, A training algorithm for optimal margin classifiers. In: Proceedings of the Fifth Annual Workshop on Computational Learning Theory, 144–152. DOI: 10.1145/130385.130401

Breiman, L.: 1996, Bagging predictors. Mach. Learn., 24, 2, 123–140. DOI: 10.1007/BF00058655

Breiman, L.: 2001, Random forests. Mach. Learn., 45, 1, 5–32. DOI: 10.1023/A:1010933404324

Brocardo, M.L., Traore, I., Woungang, I. and Obaidat, M.S.: 2017, Authorship verification using deep belief network systems. Int. J. Commun. Syst., 30, 12, e3259. DOI: 10.1002/dac.3259

Cao, J., Li, Z. and Li, J.: 2019, Financial time series forecasting model based on CEEMDAN and LSTM. Physica A: Stat. Mech. Appl., 519, 127–139. DOI: 10.1016/j.physa.2018.11.061

Charco, J.L., Roque-Colt, T., Egas-Arizala, K., Pérez-Espinoza, C.M. and Cruz-Chóez, A.: 2020, Using multivariate time series data via long-short term memory network for temperature forecasting. In: International Conference on Systems and Information Sciences, Cham, Springer International Publishing, 38–47. DOI: 10.1007/978-3-030-59194-6_4

Chen, H., Lu, T., Huang, J., He, X., Yu, K., Sun, X., ... and Huang, Z.: 2023, An improved VMD-LSTM model for time-varying GNSS time series prediction with temporally correlated noise. Remote Sens., 15, 14, 3694. DOI: 10.3390/rs15143694

Chen, Q., van Dam, T., Sneeuw, N., Collilieux, X., Weigelt, M. and Rebischung, P.: 2013, Singular spectrum analysis for modeling seasonal signals from GPS time series. J. Geodyn., 72, 25–35. DOI: 10.1016/j.jog.2013.05.005

Chen, W., Luo, H., Li, J. and Chi, J.: 2024, Long-term trend prediction of pandemic combining the compartmental and deep learning models. Sci. Rep., 14, 1, 21068. DOI: 10.1038/s41598-024-72005-x

Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H. and Bengio, Y.: 2014, Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078. DOI: 10.48550/arXiv.1406.1078

Devlin, J., Chang, M.W., Lee, K. and Toutanova, K.: 2019, Bert: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 1, 4171–4186. DOI: 10.18653/v1/N19-1423

Dragomiretskiy, K. and Zosso, D.: 2013, Variational mode decomposition. IEEE Trans. Signal Process., 62, 3, 531–544. DOI: 10.1109/TSP.2013.2288675

Forootan, E. et al.: 2021, Analyzing GNSS measurements to detect and predict bridge movements using the Kalman filter (KF) and neural network (NN) techniques. Geomatics, 1, 1, 65–80. DOI: 10.3390/geomatics1010006

Freund, Y. and Schapire, R.E.: 1997, A decision-theoretic generalization of on-line learning and an application to boosting. J. Comput. Syst. Sci., 55, 1, 119–139. DOI: 10.1006/jcss.1997.1504

Gao, R., Huo, Y., Bao, S., Tang, Y., Antic, S.L., Epstein, E., ... and Landman, B.A.: 2019, Distanced LSTM: time-distanced gates in long short-term memory models for lung cancer detection. In: International Workshop on Machine Learning in Medical Imaging, Cham, Springer International Publishing, 310–318. DOI: 10.1007/978-3-030-32692-0_36

Gao, W., Li, Z., Chen, Q., Jiang, W. and Feng, Y.: 2022, Modelling and prediction of GNSS time series using GBDT, LSTM and SVM machine learning approaches. J. Geod., 96, 10, 71. DOI: 10.1007/s00190-022-01662-5

Han, K., Wang, Y., Chen, H., Chen, X., Guo, J., Liu, Z., ... and Tao, D.: 2022, A survey on vision transformer. IEEE Trans. Pattern Anal. Mach. Intell., 45, 1, 87–110. DOI: 10.1109/TPAMI.2022.3152247

Han, X., Zhang, Z., Ding, N., Gu, Y., Liu, X., Huo, Y., ... and Zhu, J.: 2021, Pre-trained models: Past, present and future. Ai Open, 2, 225–250. DOI: 10.1016/j.aiopen.2021.08.002

Hassani, H.: 2007, Singular spectrum analysis: methodology and comparison. J. Data Sci., 5, 2, 239–257. DOI: 10.6339/JDS.2007.05(2).396

He, X., Huang, J., Montillet, J.P., Wang, S., Kermarrec, G., Shum, C.K., Hu, S. and Wang, F.: 2025, A noise reduction approach for improve North American regional sea level change from satellite and in situ observations. Surv. Geophys., 1–32. DOI: 10.1007/s10712-025-09894-8

He, X., Montillet, J.P., Fernandes, R., Bos, M., Yu, K., Hua, X. and Jiang, W.: 2017, Review of current GPS methodologies for producing accurate time series and their error sources. J. Geodyn., 106, 12–29. DOI: 10.1016/j.jog.2017.01.004

He, X., Yu, K., Montillet, J.P., Xiong, C., Lu, T., Zhou, S., Ma, X., Cui, H. and Ming, F.: 2020, GNSS-TS-NRS: An Open-source MATLAB-Based GNSS time series noise reduction software. Remote Sens., 12, 21, 3532. DOI: 10.3390/rs12213532

Ho, T.K.: 1998, The random subspace method for constructing decision forests. IEEE Trans. Pattern Anal. Mach. Intell., 20, 8, 832–844. DOI: 10.1109/34.709601

Hochreiter, S. and Schmidhuber, J.: 1997, Long short-term memory. Neural Comput., 9, 8, 1735–1780. DOI: 10.1007/978-3-642-24797-2_4

Huang, J., He, X., Hu, S. and Ming, F.: 2025, Impact of offsets on GNSS time series stochastic noise properties and velocity estimation. Adv. Space Res., 75, 4, 3397–3413. DOI: 10.1016/j.asr.2024.12.016

Huang, N.E., Shen, Z., Long, S.R., Wu, M.C., Shih, H.H., Zheng, Q., ... and Liu, H.H.: 1998, The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. Proc. R. Soc. Lond. A, Math. Phys. Eng. Sci., 454, 1971, 903–995. DOI: 10.1098/rspa.1998.0193

Huang, S., Chang, J., Huang, Q. and Chen, Y.: 2014, Monthly streamflow prediction using modified EMD-based support vector machine. J. Hydrol., 511, 764–775. DOI: 10.1016/j.jhydrol.2014.01.062

Jia, N., Jia, T. and Zhang, Z.: 2025, A residual GRU method with deep cross fusion for Alzheimer's disease progression prediction using missing variable-length time series data. Biomed. Signal Process. Control, 102, 107253. DOI: 10.1016/j.bspc.2024.107253

Jiao, J., Wang, H., Dang, Y., Ren, Y., Yue, C., Wu, X., Cui, H. and Wang, X.: 2025, Noise-resilient GNSS coordinate time series prediction using AVMD-sLSTM-transformer hybrid model. Adv. Space Res., 76, 11. DOI: 10.1016/j.asr.2025.09.040

Klos, A., Bos, M.S. and Bogusz, J.: 2018, Detecting time-varying seasonal signal in GPS position time series with different noise levels. GPS Solut., 22, 1, 21. DOI: 10.1007/s10291-017-0686-6

Koprinska, I., Wu, D. and Wang, Z.: 2018, Convolutional neural networks for energy time series forecasting. In: 2018 International Joint Conference on Neural Networks, 1–8. IEEE. DOI: 10.1109/IJCNN.2018.8489399

Landi, F., Baraldi, L., Cornia, M. and Cucchiara, R.: 2021, Working memory connections for LSTM. Neural Netw., 144, 334–341. DOI: 10.1016/j.neunet.2021.08.030

Lara-Benítez, P., Carranza-García, M., Luna-Romera, J.M. and Riquelme, J.C.: 2020, Temporal convolutional networks applied to energy-related time series forecasting. Appl. Sci., 10, 7, 2322. DOI: 10.3390/app10072322

Laurent, B. and Massart, P.: 2000, Adaptive estimation of a quadratic functional by model selection. Ann. Stat., 1302–1338. DOI: 10.1214/aos/1015957395

Lea, C., Flynn, M.D., Vidal, R., Reiter, A. and Hager, G.D.: 2017, Temporal convolutional networks for action segmentation and detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 156–165. DOI: 10.48550/arXiv.1611.05267

LeCun, Y., Bottou, L., Bengio, Y. and Haffner, P.: 2002, Gradient-based learning applied to document recognition. Proc. IEEE, 86, 11, 2278–2324. DOI: 10.1109/5.726791

Li, C., Yang, P., Zhang, T. and Guo, J.: 2024, Periodic signal extraction of GNSS height time series based on adaptive singular spectrum analysis. Geod. Geodyn., 15, 1, 50–60. DOI: 10.1016/j.geog.2023.04.003

Li, Z., Yang, K., Jiang, W., Deng, L. and Zou, Y.: 2025a, Impacts of Period Offset and Period Variation on modeling seasonal signals in GNSS coordinate time series from both theoretical and practical perspectives. Geo-Spat. Inf. Sci., 1–16. DOI: 10.1080/10095020.2025.2583793

Li, Z., Jiang, W., van Dam, T., Zou, X., Chen, Q. and Chen, H.: 2025b, A review on modeling environmental loading effects and their contributions to nonlinear variations of global navigation satellite system coordinate time series. Eng. J., 47, 4, 29–41. DOI: 10.1016/j.eng.2024.09.001

Liu, Y., Hao, X., Zhang, B. and Zhang, Y.: 2020, Simplified long short-term memory model for robust and fast prediction. Pattern Recognit. Lett., 136, 81–86. DOI: 10.1016/j.patrec.2020.05.033

Loh, W.Y.: 2014, Fifty years of classification and regression trees. Int. Stat. Rev., 82, 3, 329–348. DOI: 10.1111/insr.12016

Marques, C.A.F., Ferreira, J.A., Rocha, A., Castanheira, J.M., Melo-Gonçalves, P., Vaz, N. and Dias, J.M.: 2006, Singular spectrum analysis and forecasting of hydrological time series. Phys. Chem. Earth Pt. A/B/C, 31, 18, 1172–1179. DOI: 10.1016/j.pce.2006.02.061

Pragati, A., Mishra, M., Panigrahi, R.R. and Gadanayak, D.A.: 2024, Bayesian Optimized LSBoost Model for the DC fault location in a MT-VSC-HVDC Network. In: 2024 IEEE Third International Conference on Power Electronics, Intelligent Control and Energy Systems, 1101–1106. IEEE. DOI: 10.1109/ICPEICES62430.2024.10719101

Quinlan, J.R.: 1986, Induction of decision trees. Mach. Learn., 1, 1, 81–106. DOI: 10.1007/BF00116251

Richter, A., Ivins, E., Lange, H., Mendoza, L., Schröder, L., Hormaechea, J.L., Casassa, G., Marderwald, E., Fritsche, M., Perdomo, R. and Horwath, M.: 2016, Crustal deformation across the Southern Patagonian Icefield observed by GNSS. Earth Planet. Sci. Lett., 452, 206–215. DOI: 10.1016/j.epsl.2016.07.042

Sainath, T.N., Kingsbury, B., Mohamed, A.R., Dahl, G.E., Saon, G., Soltau, H., ... and Ramabhadran, B.: 2013, Improvements to deep convolutional neural networks for LVCSR. In: 2013 IEEE Workshop on Automatic Speech Recognition and Understanding, 315–320. IEEE. DOI: 10.1109/ASRU.2013.6707749

Sapankevych, N.I. and Sankar, R.: 2009, Time series prediction using support vector machines: a survey. IEEE Comput. Intell. Mag., 4, 2, 24–38. DOI: 10.1109/MCI.2009.932254

Savchuk, S., Doskich, S., Gołda, P. and Rurak, A.: 2023, The seasonal variations analysis of permanent GNSS station time series in the central-east of Europe. Remote Sens., 15, 15, 3858. DOI: 10.3390/rs15153858

Schmidhuber, J.: 2012, Multi-column deep neural networks for image classification. In: Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition, 3642–3649. DOI: 10.1109/CVPR.2012.6248110

Schuster, M. and Paliwal, K.K.: 1997, Bidirectional recurrent neural networks. IEEE Trans. Signal Process., 45, 11, 2673–2681. DOI: 10.1109/78.650093

Shen, W., Wang, C., Huang, M.M., Li, B. and Chen, Z.: 2025, Multiple complex construction noise signal decomposition and prediction methods: based on WOA-VMD and stacked LSTM. Measurement, 118201. DOI: 10.1016/j.measurement.2025.118201

Siemuri, A., Selvan, K., Kuusniemi, H., Valisuo, P. and Elmusrati, M.S.: 2022, A systematic review of machine learning techniques for GNSS use cases. IEEE Trans. Aerosp. Electron. Syst., 58, 6, 5043–5077. DOI: 10.1109/TAES.2022.3219366

Song, Y.Y. and Lu, Y.: 2015, Decision tree methods: applications for classification and prediction. Shanghai Arch. Psychiatry, 27, 2, 130–135. DOI: 10.11919/j.issn.1002-0829.215044

Specht, D.F.: 1991, A general regression neural network. IEEE Trans. Neural Netw., 2, 6, 568–576. DOI: 10.1109/72.97934

Strobl, C.: 2014, Discussion on fifty years of classification and regression trees. Int. Stat. Rev., *82*, 3, 349–352. DOI: 10.1111/insr.12059

Su, M., Chen, C., Li, Z., Jiang, W., Gao, Y., Shang, J. and Zhou, X.: 2026, Rainfall amount forecast using GNSS-PWV based on machine learning fusion strategy and the constraint of rainfall event. IEEE Trans. Geosci. Remote Sens., 20. DOI: 10.1109/TGRS.2026.3659172

Tabar, M.E. and Sisman, Y.: 2025, Deep learning-based modeling and prediction of GNSS time series: A comparative analysis of adaptive optimization algorithms. Adv. Space Res., 76, 4, 2086–2103. DOI: 10.1016/j.asr.2025.06.018

Thissen, U.V.B.R., Van Brakel, R., De Weijer, A.P., Melssen, W.J. and Buydens, L.M.C.: 2003, Using support vector machines for time series prediction. Chemom. Intell. Lab. Syst., 69, 1-2, 35–49. DOI: 10.1016/S0169-7439(03)00111-4

Torres, M.E., Colominas, M.A., Schlotthauer, G. and Flandrin, P.: 2011, A complete ensemble empirical mode decomposition with adaptive noise. In: 2011 IEEE International Conference on Acoustics, Speech and Signal Processing, 4144–4147. IEEE. DOI: 10.1109/ICASSP.2011.5947265

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., ... and Polosukhin, I.: 2017, Attention is all you need. Adv. Neural Inf. Process. Syst., 30. DOI: 10.48550/arXiv.1706.03762

Wang, J., Liu, Q., Ma, D., Zhang, Y. and Yu, X.: 2026, An integrated VKA-LSTM model for GNSS height time series prediction. GPS Solut., 30, 1, 29. DOI: 10.1007/s10291-025-01994-7

Wen, Z., Rao, W. and Sun, W.: 2023, Contribution of loading deformation to the GNSS vertical velocity field in the Chinese mainland. Geophys. J. Int., 233, 3, 1655–1670. DOI: 10.1007/s10712-025-09894-8

Wu, Z. and Huang, N.E.: 2004, A study of the characteristics of white noise using the empirical mode decomposition method. Proc. R. Soc. Lond. A: Math., Phys. Eng. Sci., 460, 2046, 1597–1611. DOI: 10.1098/rspa.2003.1221

Wu, Z. and Huang, N.E.: 2009, Ensemble empirical mode decomposition: a noise-assisted data analysis method. Adv. Adapt. Data Anal., 1, 1, 1–41. DOI: 10.1142/S1793536909000047

Zhou, D., Liu, Y., Feng, Y., Zhang, H., Fu, Y., Liu, Y. and Tang, Q.: 2022, Absolute sea level changes along the coast of China from tide gauges, GNSS, and satellite altimetry. J. Geophys. Res., Oceans, 127, 9, e2022JC018994. DOI: 10.1029/2022JC018994

Zhou, Y., He, X., Montillet, J. P., Wang, S., Hu, S., Sun, X., ... and Ma, X.: 2025, An improved ICEEMDAN-MPA-GRU model for GNSS height time series prediction with weighted quality evaluation index. GPS Solut., 29, 3, 113. DOI: 10.1007/s10291-025-01867-z