A Novel Optimized Hybrid VMD-PCA-XGBoost Model for Forecasting Precipitation: Exemplified by the Beijing-Tianjin-Hebei Study Region in China

Qiaoli Kong, Qian Li, Qi Bai, Xiaolong Mi, Joseph Awange, Shi Wang, Yi Yang, and Guoli Bo

***Abstract*—Extreme precipitation events pose significant challenges to societal infrastructure and environmental stability, particularly in vulnerable regions like the Beijing-Tianjin-Hebei area of China. Traditional forecasting methods, such as numerical weather prediction, radar nowcasting, downscaling techniques, etc., frequently fail to capture the complex nonlinear dynamics of such events. In this study, we propose a novel hybrid model, variational mode decomposition-principal component analysis-extreme gradient boosting (VMD-PCA-XGBoost), which integrates VMD for effective signal processing, PCA for dimensionality reduction, and XGBoost for enhancing predictive accuracy. Utilizing the 2023 data from 13 global navigation satellite system and meteorological stations, our model, when rigorously compared with those of XGBoost and empirical mode decomposition-based models, achieves superior performance, with average critical success index, probability of detection, and false alarm rate of 52.14%, 73.07%, and 35.98%, respectively. These findings underscore the model's robustness and precision, offering a promising tool for improving precipitation forecasts. This study not only advances the methodological framework for atmospheric forecasting but also provides critical insights for enhancing disaster preparedness and mitigation strategies in climate-sensitive regions.**

***Index Terms*—Beijing-Tianjin-Hebei, empirical mode decomposition (EMD), extreme gradient boosting (XGBoost), precipitation prediction, variational mode decomposition (VMD).**

I. Introduction

EXTREME precipitation events, such as floods, landslides, and debris flows, have become increasingly frequent and severe worldwide due to the intensification of global climate change. These events result in significant casualties, infrastructure damage, and socioeconomic disruption [1], [2]. Recent examples include the 2021 floods in Tennessee, catastrophic flooding in Germany and Belgium, and heavy rainfall in Henan, China [3], [4], [5]. These incidents underscore the urgent need for improved precipitation forecasting systems to mitigate the impacts of such extreme weather events.

Traditional precipitation forecasting methods, such as numerical weather prediction (NWP) and remote sensing (RS) approaches, have been instrumental in weather prediction but often struggle with the nonlinear dynamics of extreme events [6], [7], [8]. The advent of global navigation satellite system (GNSS) data has opened new avenues for estimating atmospheric water vapor content, offering a promising tool for enhancing precipitation prediction [9], [10]. However, these methods face challenges in accurately capturing the complex interactions and nonlinear relationships inherent in extreme precipitation events, particularly in regions with diverse and varied topography like Beijing-Tianjin-Hebei. Research works proved that the forecasting capability of NWP is limited in complex terrain region due to the initial data uncertainties, computational errors, and atmospheric randomness, such as in the Sierras de Córdoba, mountains of western Canada and the north-western US [11], [12]. While, studies for precipitation forecasting in Ethiopia and other complex terrain demonstrated that remote sensing is significantly susceptible to weather conditions, resulting in poor performance in precipitation prediction in adverse weather [13], [14]. As for GNSS, experiments found that the observation processing process is quite complex, subject to tropospheric and ionospheric delays, multipath effects and other interfering factors, besides, its indirect correlation with precipitation presents challenges in capturing the involved complex interactions [15], [16], [17]. Therefore, it is difficult to overcome these challenges primarily and mainly because extreme precipitation events involve multiscale, multivariable nonlinear coupling mechanisms. The traditional precipitation forecasting methods are limited in their data fusion capabilities and model flexibility, making it difficult to fully represent and adapt to the complex atmospheric physical processes and abrupt features associated with precipitation formation.

Machine learning techniques, such as neural networks, random forest, and gradient boosting, have emerged as powerful tools for high-precision precipitation forecasting due to

Received 18 March 2025; revised 28 May 2025 and 23 June 2025; accepted 2 July 2025. Date of publication 7 July 2025; date of current version 18 July 2025. This work was supported in part by the National Natural Science Foundation of China under Grant 42374038, Grant 41704015, and Grant 42271436, 42404052, in part by the Shandong Natural Science Foundation of China under Grant ZR2023MD072, and Grant ZR2021MD030, and in part by the Talent introduction plan for Youth Innovation Team in universities of Shandong Province (innovation team of satellite positioning and navigation) and in part by the Research Grants Council of Hong Kong through the General Research under Grant 15212525. *(Corresponding author: Xiaolong Mi.)*

Qiaoli Kong, Qian Li, Shi Wang, Yi Yang, and Guoli Bo are with the College of Geodesy and Geomatics, Shandong University of Science and Technology, Qingdao 266590, China (e-mail: kqlabc3334@163.com; lqsea17@163.com; ws71712@163.com; yangyi6013@163.com; boguoli1113@163.com).

Qi Bai is with Shanghai Chart Center, Eastern Navigation Service Center, Maritime Safety Administration, Shanghai 200090, China (e-mail: bq974210@163.com).

Xiaolong Mi and Joseph Awange are with the Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hong Kong (e-mail: xiaolong.mi@polyu.edu.hk; joseph.awange@polyu.edu.hk).

Digital Object Identifier 10.1109/JSTARS.2025.3586351

their capabilities in analyzing and processing nonlinear and nonstationary signals (e.g., [18], [19], [20], [21], [22], [23]). These methods have demonstrated significant improvements in forecast accuracy across various contexts. For instance, an extended long short-term memory (ConvLSTM) network used for precipitation nowcasting achieved a critical success index (CSI) of 57.7%, a false alarm rate (FAR) of 19.5%, and a probability of detection (POD) of 66% [18]. Similarly, hybrid models combining wavelet analysis with artificial neural networks have outperformed traditional models in forecasting precipitation [19].

Recent advancements include the integration of machine learning with radar-based forecasting, where methods like random forest have reduced false alarm rates and increased the success rate of precipitation forecasts (e.g., [21]). The neural network MetNet-2, designed for forecasting rapidly changing precipitation variables up to 12 hours ahead, has shown superior accuracy and efficiency compared to traditional physics-based models [22]. Additionally, hybrid models such as XGBoost combined with numerical weather prediction (NWP) bias correction have achieved notable performance improvements [23].

Inspired by the advances in empirical mode decomposition (EMD) and neural networks, researchers have developed innovative models for precipitation forecasting. For example, a multivariate empirical mode decomposition-time dependent intrinsic cross-correlation-LSTM model effectively captures extreme events (e.g., [24]). Other studies have combined singular spectrum analysis and EMD with light gradient boosting machine and XGBoost to enhance long-term rainfall estimation accuracy (e.g., [25]).

In addition to machine learning and mode decomposition, principal component analysis (PCA) has been widely used in precipitation forecasting as it helps, on one hand, in predicting rainfall by explaining variance and distinguishing between rain and no-rain scenarios, and, on the other hand, in improving computational efficiency by reducing temporal dimensions [26], [27].

These studies collectively indicate that machine learning based methods such as EMD, VMD, PCA, and their combinations hold significant promise for improving precipitation forecasting. The multimodel ensemble approach is widely regarded as an effective strategy for leveraging individual model strengths and enhancing climate projections [28], [29]. However, there remains room to enhance forecasting accuracy and robustness, particularly in regions where the frequency and intensity of heavy precipitation events are rising such as Beijing-Tianjin-Hebei area [30]. Despite these advancements, few studies have specifically focused on such diverse and complex regions, highlighting a critical gap in the literature. Addressing this gap is essential, given the persistent challenges in meteorology posed by intensified climate instability from global warming [31], [32].

In response to these challenges, therefore, we propose a novel hybrid model, VMD-PCA-XGBoost, designed to exploit and access the advantages of the three models in order to achieve high-accuracy precipitation forecasts for these diverse and complex regions. It has been demonstrated that, compared with EMD, VMD's variational optimization framework avoids mode mixing—a critical weakness of EMD—by enforcing orthogonality and bandwidth constraints on decomposed modes [33], [44]. This ensures robust noise reduction while preserving signal integrity. It is particularly important for large datasets, as excessive dimensionality may lead to overfitting or increased computational complexity. These issues can be effectively mitigated by PCA. PCA based a dimensionality reduction mechanism can reduce redundant features, thereby enhancing the training efficiency and predictive accuracy of XGBoost [34]. By integrating these advanced techniques, the VMD-PCA-XGBoost model addresses the limitations of existing methods, offering a comprehensive approach to capturing the complex dynamics of extreme precipitation events.

To validate the robustness, reliability, and accuracy of the proposed VMD-PCA-XGBoost model, we conducted comprehensive comparisons with existing models such as XGBoost, EMD-XGBoost, and VMD-XGBoost.

This study not only advances the methodological framework for atmospheric forecasting but also provides critical insights for enhancing disaster preparedness and mitigation strategies in topographically challenging areas. By focusing on the Beijing-Tianjin-Hebei area, this research contributes to a deeper understanding of regional climate dynamics and offers a scalable solution that can be adapted to other regions facing similar challenges.

II. Data and Methods

A. Study Region

The Beijing-Tianjin-Hebei region, a key economic zone in China, is particularly vulnerable to extreme weather events due to its location on the northern edge of the East Asian Summer Monsoon and its diverse topography [35]. In late July to early August 2023, the region experienced extreme heavy rainfall that triggered mountain floods, mudslides, and bridge collapses, resulting in approximately 165.8 billion RMB in economic losses and casualties [36]. Over the past two decades, this region has frequently experienced extreme heavy precipitation, highlighting the critical need for high-precision precipitation forecasting to aid governmental and public preparedness, implement effective measures to reduce disaster losses, and strengthen safety and social stability.

B. Data

The data used in this study include GNSS data, meteorological data, radiosonde data, and the fifth-generation European Center for Medium-Range Weather Forecasts Reanalysis (ERA5) data. In this experiment, the radiosonde data[1] were obtained from the Integrated Global Radiosonde Archive (IGRA), with a time resolution of 12 hours. The GNSS data were applied from the National Earthquake Data Center of China, with a sampling interval of 30 seconds. ERA5 reanalysis data[2] were released by the European Centre for Medium-Range Weather Forecasts (ECMWF), with a temporal resolution of 1 h and a spatial

[1] [Online]. Available: ftp://ftp.ncdc.noaa.gov/pub/data/igra
[2] [Online]. Available: https://cds.climate.copernicus.eu/datasets

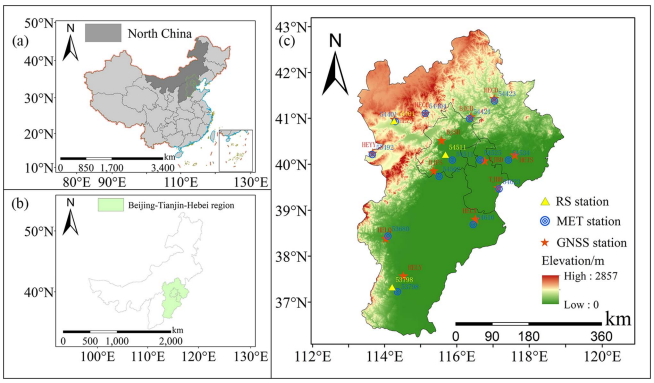

Fig. 1. Distribution map of showing monitoring stations in Beijing-Tianjin-Hebei region of China. (a) Location of North China, (b) location of the Beijing-Tianjin-Hebei region within North China, and (c), elevation and station distribution in the Beijing-Tianjin-Hebei region.

resolution of 0.25°×0.25°. Meteorological data[3] were collected from the QWeather platform, also with a time resolution of 1 hour.

The distribution of the GNSS, meteorological, and RS stations are shown in Fig. 1. The data time span ranges from 1 January 2023 to 27 October 2023, covering a total of 300 days.

In this study involving comprehensive analysis with radiosonde data and other data, due to missing entries in the radiosonde dataset, only the timestamps with available radiosonde data were retained, and other datasets were aligned accordingly to ensure temporal consistency. For other experiments without radiosonde data, all datasets shared a 1-h resolution and thus did not require temporal alignment. All values of missing meteorological data were filled using the mean values.

The meteorological data adjustment equations for pressure (P) and temperature (T) are based on the tropospheric model developed by Zhu et al. [37], as shown in the following equations. Relative humidity (RH) and hourly precipitation (R1h) were interpolated using the inverse distance weighting method

$$T = T_0 - \beta (h - h_0) \tag{1}$$

$$P = P_0 \left[1 - \tau (h - h_0)\right]^{\frac{g \cdot M}{R \cdot \beta}} \tag{2}$$

$$g = 9.8063 \cdot \left\{ 1 - 10^{-7} \frac{h + h_0}{2} [1 - 0.0026373 \cdot \cos(2\varphi) + 5.9 \cdot 10^{-6} \cdot \cos^2(2\varphi)] \right\} \tag{3}$$

where T is the temperature at the GNSS station (K), T_0 is the temperature at the meteorological station (K), P and P_0 are the air pressures at the GNSS and meteorological stations (hPa), respectively, h and h_0 are the elevations at the GNSS and meteorological stations (m), M is the molar mass of air (0.0289644 kg/mol), R is the ideal gas constant (8.31432 J/K·mol), g is the coefficient of gravity, φ is the latitude (rad), β is the temperature lapse rate, and τ is the ratio of temperature to lapse rate.

[3] [Online]. Available: https://q-weather.info/account/profile/

Radiosonde stations are generally not colocated with GNSS stations, however, if the horizontal distance between a radiosonde station and a GNSS station is less than 50 km and the elevation difference is less than 200 m [38], [39], the radiosonde station is considered to be colocated with the GNSS station.

C. Methods

1) Establishment of the Atmospheric Weighted Mean Temperature Model: In ground-based GNSS water vapor retrieval, obtaining real-time precipitable water vapor (PWV) is crucial, which also requires real-time atmospheric weighted mean temperature (Tm). Therefore, it is necessary to develop a localized Tm model. However, most localized Tm models are based on radiosonde data. Given the severe shortage of radiosonde data in the study region, studies have shown that Tm derived from the fifth-generation European Center for Medium-Range Weather Forecasts Reanalysis is relatively consistent with results obtained from radiosonde data (e.g., [40]).

Tm can be calculated based on the numerical integration method using ERA5 data and radiosonde data as follows [37]:

$$T_m = \frac{\int \frac{e_i}{T_i} dh}{\int \frac{e_i}{T_i^2} dh} \tag{4}$$

where h is the vertical height above the station, and $e_i$and T_i are the water vapor pressure and temperature of the *i*th pressure level, respectively.

Therefore, it is possible to construct a localized Tm model using those derived from ERA5 data along with ground meteorological factors. In this study, the linear Tm—Ts model proposed by Huang et al. [41] is employed to calculate the Tm from the following equation:

$$T_m\ (T_s, h, \varphi, \lambda, \mathrm{DOY}) = a_0 + a_1 T_s + a_2 h + a_3 \varphi + b_1 \cos\left(\frac{\mathrm{DOY}}{365.25} \cdot 2\pi\right) + b_2 \sin\left(\frac{\mathrm{DOY}}{365.25} \cdot 2\pi\right) + b_3 \cos\left(\frac{\mathrm{DOY}}{365.25} \cdot 4\pi\right) + b_4 \sin\left(\frac{\mathrm{DOY}}{365.25} \cdot 4\pi\right) \tag{5}$$

where T_s is the surface temperature (K), h is the station elevation (m), φ is the latitude (rad), DOY is the days of the year, and a_0, a_1, a_2, a_3, b_1, b_2, b_3, b_4 are the model coefficients.

Based on (5), the fitting coefficients are derived using ERA5 data from 2021 to 2023 and radiosonde data from 2016 to 2018, respectively. The Tm calculated from radiosonde data in 2019 is taken as the reference value as radiosonde data is severely lack. Tm values calculated using ERA5 data, the Bevis empirical formula, global pressure and temperature (GPT) 3_1, and GPT3_5 models are evaluated for accuracy by those derived from radiosonde data, respectively.

2) GNSS PWV Retrieval: The PRIDE Lab at GNSS Research Center of Wuhan University has developed an open-source software for GPS precise point positioning ambiguity resolution (PPP-AR) (i.e., PRIDE PPP-AR) [42]. In this study, Zenith Wet Delay (ZWD) is directly obtained from GNSS data using the PRIDE PPP-AR software. Tm values of 13 GNSS stations are then calculated based on the formula of the established Tm

model and surface temperature. The GNSS PWV is retrieved using the derived ZWD and Tm as follows:

$$\mathrm{PWV} = \Pi \cdot \mathrm{ZWD} = \frac{10^6}{\rho_w R_v \left(\frac{K_3}{T_m} + K_2'\right)} \cdot \mathrm{ZWD} \quad (6)$$

where Π is the water-gas conversion factor; ρ_w is the density of liquid water, which takes the value of 1×10^3Kg/m^3; R_v is the water vapor gas constant, which takes the value of 461.495 J/kg·K; and K_3 and K_2' are both gas constants, which take the values of 3.739×10^5, and 22.13 K/hPa, respectively.

3) Empirical Mode Decomposition: Empirical mode decomposition is a signal processing technique proposed by Huang et al. [43] for analyzing nonlinear and non-stationary signals. It decomposes a complex signal into several simple intrinsic mode functions (IMFs) to extract local features of the signal. The main steps are as follows.

Step 1: Identifying extremes. All the local maxima and minima of the original signal are identified.
Step 2: Interpolation. Spline interpolation is performed on the identified maxima and minima to obtain the upper and lower envelope curves.
Step 3: Computing the mean. The mean envelope is computed as the average of the upper and lower envelope curves.
Step 4: Extracting the first IMF. The first IMF is obtained by subtracting the mean envelope from the original signal. The extracted IMF is then checked for the following conditions: 1) The number of extrema should differ by at most one from the number of zero crossings; 2) the number of extrema should be equal to the number of zero crossings.
Step 5: Iteration. The residual signal (original signal minus the extracted IMF) is treated as the new input signal. Steps 1 to 4 are repeated until the residual signal is either stationary or monotonic.

4) Variational Mode Decomposition: Variational mode decomposition (VMD) is a signal processing method proposed by Dragomiretskiy and Zosso [44], aimed at decomposing complex signals into multiple modes with specific frequency components. Compared to traditional time-frequency analysis methods such as EMD, ensemble empirical mode decomposition, and complete ensemble empirical mode decomposition with adaptive noise, VMD excels in noise reduction and feature extraction. The decomposition steps of VMD are outlined as follows.

Step 1: Parameter setting. The number of modes to be decomposed, denoted as K, and the penalty factor α, are determined. In this study, only K and α are set.
Step 2: Optimization problem formulation. An optimization problem is formulated to minimize the following objective function:

$$\min_{\{\mu_k\}} \left\{ \sum_{k=1}^{K} \left[\| \frac{\partial}{\partial t} \mu_k + \lambda_k \cdot \left(\mu_k * e^{j\omega_k t}\right) \|^2 + \alpha \cdot \|\hat{\mu}_k\|^2 \right] \right\} \quad (7)$$

where μ_k is the kth mode component, λ_k is the Lagrange multiplier, and $\hat{\mu}_k$ is the Fourier transform of the mode.

Step 3: Iterative solution. The optimization problem is solved iteratively using algorithms such as the alternating direction method of multipliers (ADMM), until convergence is achieved by Dragomiretskiy and Zosso [44]. In each iteration, the mode μ_k and the Lagrange multiplier λ_k are updated.
Step 4: Mode extraction. The modes μ_k are extracted from the optimization results. Each mode corresponds to a different frequency component, capturing distinct features of the signal.

5) Principal Component Analysis: Principal component analysis (PCA) is a statistical technique used for dimensionality reduction and feature extraction. It projects the original data onto a new coordinate system through linear transformation, such that the axes of the new system (the principal components) can explain the maximum variance in the data. The specific steps are as follows.

Step 1: Data standardization. The data is standardized using (8) so that each feature has a mean of 0 and a variance of 1. This step is essential because PCA is sensitive to the scale of the data

$$Z = \frac{X - \mu}{\sigma} \quad (8)$$

where Z is the standardized value, X is the original data, μ is the mean, and σ is the standard deviation.

Step 2: Computation of the covariance matrix. The covariance matrix of the standardized data is calculated to describe the relationships between different features.

$$C = \frac{1}{n-1} Z^T Z \quad (9)$$

where n is the number of samples.

Step 3: Computation of eigenvalues and eigenvectors. Singular value decomposition (SVD) is applied to gain the covariance matrix by the following equation, and the top *K* eigenvectors u(1), u(2),u(3), …,u(K) corresponding to the largest eigenvalues are selected:

$$[U,\ S,\ V] = \mathrm{svd}\,(C) \quad (10)$$

where U and V represent the left and right singular vectors of C, respectively, and S is a diagonal matrix containing the singular values of *C*.

Step 4: Selection of principal components (PC). The top r eigenvalues and their corresponding eigenvectors are selected based on their magnitude. Different values of *K* are chosen and iteratively computed. The smallest *K* value meeting the condition given by the below equation is selected

$$\frac{\frac{1}{m}\sum_{i=1}^{m} \|Z^{(i)} - Z_{\text{approx}}^{(i)}\|^2}{\frac{1}{m}\sum_{i=1}^{m} \|Z\|^2} \le t \quad (11)$$

where Z_{approx} is the approximation in PCA space after dimension reduction, t is the threshold value, and m is the number of samples. For instance, if $t = 0.05$, this means that the PCA algorithm retains 95% of the main information.

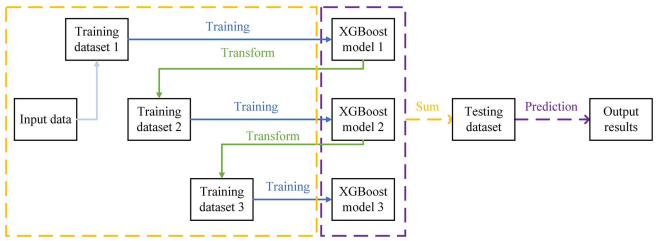

Fig. 2. XGBoost algorithm flowchart. The orange and purple boxes represent the model training process and results, respectively.

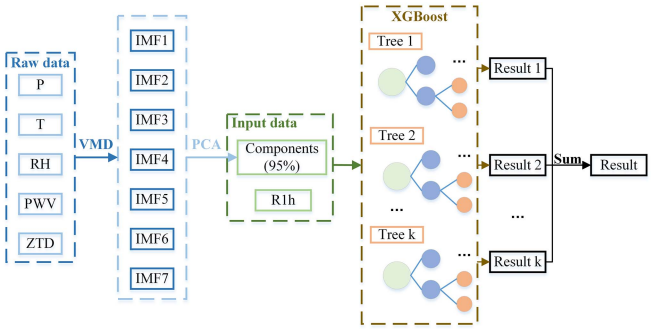

Fig. 3. VMD-PCA-XGBoost model flowchart.

6) XGBoost: XGBoost is an efficient gradient boosting algorithm widely used for classification and regression tasks. It works by constructing multiple decision trees and optimizing the loss function at each iteration, thereby enhancing the model's predictive performance. The algorithm's flowchart is shown in Fig. 2.

The objective function is defined as follows [45]:

$$\mathcal{L}(\emptyset) = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \tag{12}$$

where i is the ith sample, y_i is the observed value, $\hat{y}_i$ represents the predicted value, l is a loss function, and Ω is a regularization term used to penalize the model complexity in order to avoid overfitting. $\mathcal{L}(\emptyset)$ is the expression in linear space.

The tree ensemble model (13) is used for further model training. A new function f_t is introduced to minimize the following objective:

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} \omega_j^2 \tag{13}$$

$$\mathcal{L}^{(t)} = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)\right) + \sum_{k=1}^{K} \Omega(f_k) \tag{14}$$

where γ is the complexity of each leaf, T denotes the total number of leaves, λ is a tradeoff parameter, ω_j tabulates the score of the $j-$th leaf, and x_i is the $i-$th input data.

Equation (14) is approximated by the Taylor series expansion as follows:

$$\mathcal{L}^{(t)} \approx \sum_{i=1}^{n} \left[l\left(y_i, \hat{y}_i^{(t-1)}\right) + g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \sum_{k=1}^{K} \Omega(f_k) \tag{15}$$

$$g_i = \partial_{\hat{y}^{(t-1)}} l\left(y_i, \hat{y}^{(t-1)}\right) \tag{16}$$

$$h_i = \partial^2_{\hat{y}^{(t-1)}} l\left(y_i, \hat{y}^{(t-1)}\right) \tag{17}$$

where g_i and h_i are the first and second step statistics of the loss function, respectively, and are constant terms in the equation. Simplifying (15) yields the following:

$$\mathcal{L}^{(t)} = \sum_{i=1}^{n} \left[g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \Omega(f_t). \tag{18}$$

Substituting the parameters of the decision tree into the objective function, the sample set of the jth leaf is defined as $I_j = \{i | q(x_i = j)\}$. Expanding Ω, (18) can be rewritten as e (19). The optimal weight ω_j^* of the j leaf is given by the following:

$$\mathcal{L}^{(t)} = \sum_{j=1}^{T} \left[\sum_{i \in I_j} g_i + \frac{1}{2} \left(\sum_{i \in I_j} h_i + \lambda \right) \omega_j^2 \right] + \gamma T \tag{19}$$

$$\omega_j^* = -\frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \lambda}. \tag{20}$$

The optimal value $\mathcal{L}^{(t)}$ is obtained by calculation

$$\mathcal{L}^{(t)} = -\frac{1}{2} \sum_{j=1}^{T} \frac{\left(\sum_{i \in I_j} g_i\right)^2}{\sum_{i \in I_j} h_i + \lambda}. \tag{21}$$

7) Evaluation Metrics: In this article, the accuracy of the forecast model is validated using the standardized weather forecast evaluation league shown in Table I.

TABLE I
CONFUSION MATRIX

	Forecasting positive	Forecasting negative
Real positive	TP (True positive)	FN (False negative)
Real negative	FP (False positive)	TN (True negative)

In the context of precipitation, a positive instance represents the occurrence of rain, while a negative instance represents no rain. This study applies CSI, POD, and FAR as performance evaluation metrics for the model. To ensure accurate precipitation forecasting, an ideal model is expected to yield large CSI and POD values while minimizing the FAR. The calculation equations for these metrics are as follows:

$$\text{CSI} = \frac{\text{TP}}{\text{TP} + \text{FN} + \text{FP}} \tag{22}$$

$$\text{POD} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{23}$$

$$\text{FAR} = \frac{\text{FP}}{\text{TP} + \text{FP}}. \tag{24}$$

This study uses the root mean square error (RMSE), mean absolute error (MAE), and mean square error (MSE) to assess

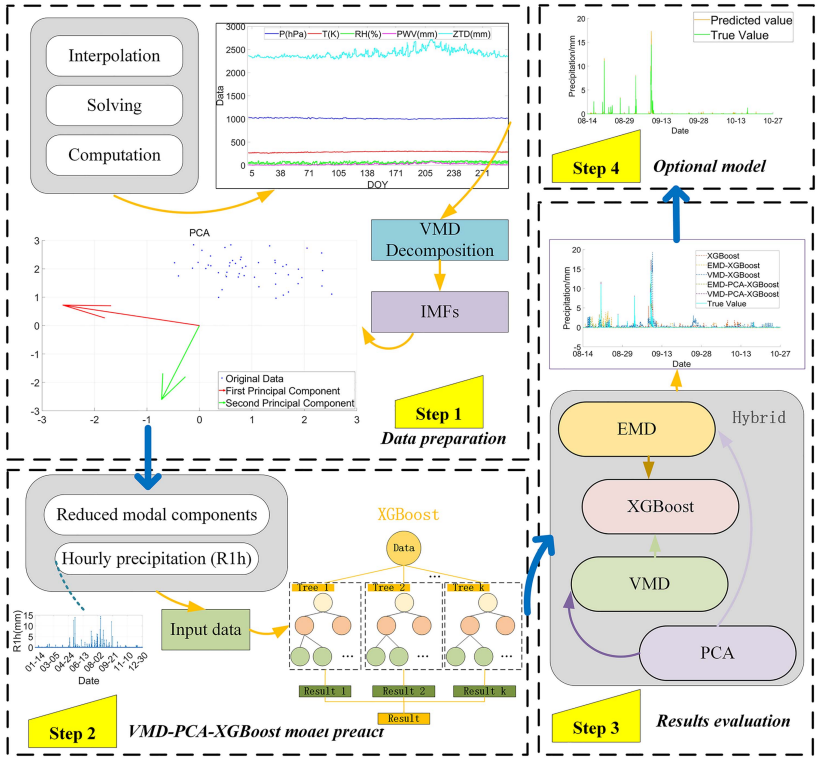

Fig. 4. Precipitation forecasting technical roadmap of data preparation (Step 1), VMD-PCA-XGBoost model (Step 2), results evaluation (including single model and other combined models) (Step 3), and display of forecasting result of the optional model.

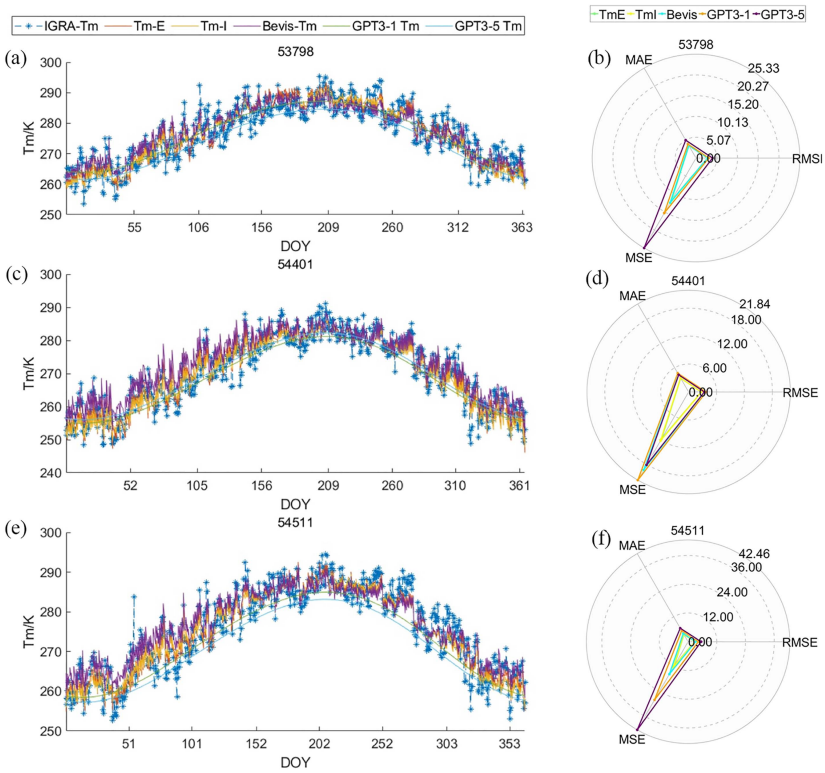

Fig. 5. Comparison of achieved Tm and their accuracy calculated by different models. (a), (b) Tm results for meteorological station 53798 and their accuracy. (c), (d) Tm results for meteorological station 54401 and their accuracy. (e), (f) Tm results for meteorological station 54511 and their accuracy. The locations of these meteorological stations are shown in the blue circles in subgraph (c) of Fig. 1. Tm-E and Tm-I are models fitted using ERA5 data and RS data, respectively, while Bevis and GPT3 are widely used models for estimating Tm.

the accuracy of the Tm model. The calculation equations are as follows:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(X_i - X_i')^2}{n}} \tag{25}$$

$$\mathrm{MAE} = \frac{\sum_{i=1}^{n}|X_i - X_i'|}{n} \tag{26}$$

$$\mathrm{MSE} = \frac{\sum_{i=1}^{n}(X_i - X_i')^2}{n} \tag{27}$$

where X_i represents the $i-$th value to be assessed, X_i' represents the $i-$th reference value, and n represents the number of samples.

D. Combined Forecasting Model

Based on the above calculated results and algorithms, this study constructs a novel hybrid model of VMD-PCA-XGBoost, as illustrated in the flowchart in Fig. 3. The technical roadmap is shown in Fig. 4.

III. Results

A. Establishment and Evaluation of the Tm Model Exemplified in the Beijing-Tianjin-Hebei Region

This study uses the built-in MATLAB function (named fitlm) to calculate and obtain the coefficients of the Tm fitting model. This function employs the least squares algorithm to determine the optimal parameters of the linear regression model by minimizing the squared error between the predicted values and the actual observed values, thereby achieving effective data fitting.

The Tm values calculated using the IGRA data from the three sounding stations in the Beijing-Tianjin-Hebei region in 2019 are taken as reference values to evaluate the accuracy of the Tm-E, Tm-I, Bevis, GPT3-1, and GPT3-5 models. Among these, Tm-E and Tm-I are models fitted using ERA5 data and RS data, respectively, while Bevis and GPT3 are widely used models for estimating Tm. The calculation results of each model and their corresponding MAE, RMSE, and MSE values are shown in Fig. 5.

In Fig. 5(b), (d), and (f), radar charts are adopted to present the accuracy evaluation metrics, such as MAE, RMSE and MSE, of different Tm estimation models at the radiosonde stations 53798, 54401, and 54511. In the radar charts, three axis radiating from the center represents MAE, RMSE, and MSE in clockwise order, respectively, and the magnitude of each metric is indicated by concentric circular grids. The closer the metric value is to the center of the circle, the higher the accuracy, and vice versa, the lower the accuracy.

As shown in Fig. 5, the Tm-E model consistently demonstrates smaller MAE, RMSE, and MSE values at all three stations compared to the other four Tm models (Tm-I, Bevis, GPT3-1, and GPT3-5). The accuracy evaluation metrics value lines of Tm-E are generally closer to the center of the radar charts compared with those of other models, indicating superior performance in all three accuracy evaluation metrics. Therefore, the Tm-E model is selected for GNSS PWV retrieval in this study. In contrast, the Bevis and GPT3-based models exhibit relatively larger errors at some stations, with particularly notable deviations along the MSE axis, suggesting comparatively lower estimation accuracy.

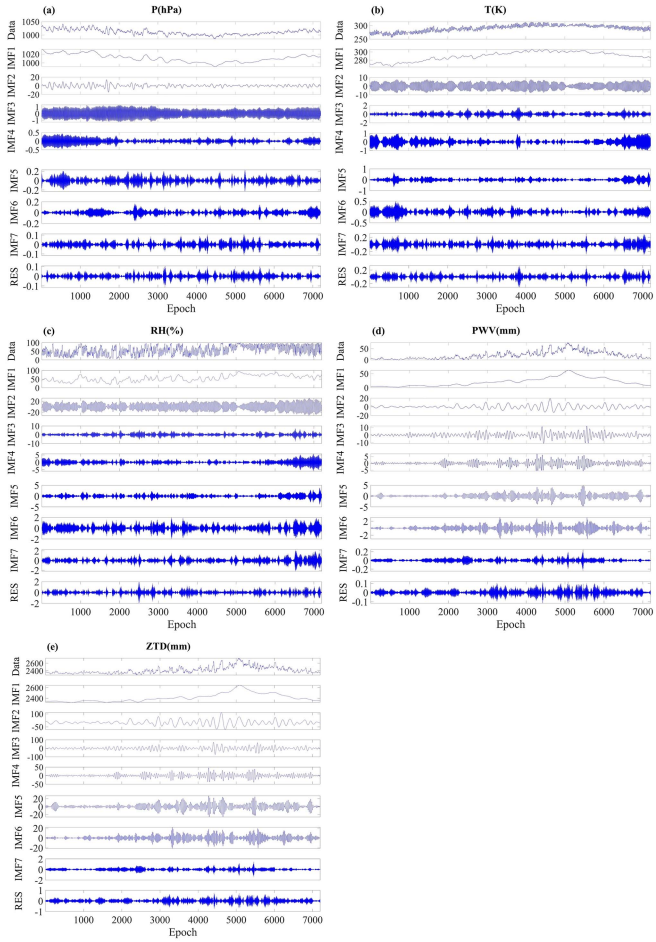

Fig. 6. VMD decomposition results of the GNSS station BJFS. (a)–(e) shows pressure, temperature, relative humidity, precipitable water vapor, and zenith tropospheric delay decomposition sequence, respectively.

B. VMD Decomposition Results

Taking the GNSS station BJFS as an example, the VMD decomposition results of pressure (*P*), temperature (*T*), RH, PWV, and zenith tropospheric delay (ZTD) are shown in Fig. 6.

C. Precipitation Forecasting Model Results and Analysis

This study applies the VMD method to decompose *P, T,* RH, PWV, and ZTD data from each GNSS station, followed by PCA dimensionality reduction (i.e., Step 1 in Fig. 4). The VMD dimensional-reduced data and hourly precipitation (R1h) data are used as input data for the VMD-PCA-XGBoost model (i.e., Step 2 in Fig. 4). Additionally, the original data (*P, T*, RH, PWV, ZTD, R1h) is used to train the XGBoost model. The EMD and VMD decomposed subsequences are separately used with R1h as input data for the EMD-XGBoost and VMD-XGBoost models. The EMD-reduced data combined with R1h is used

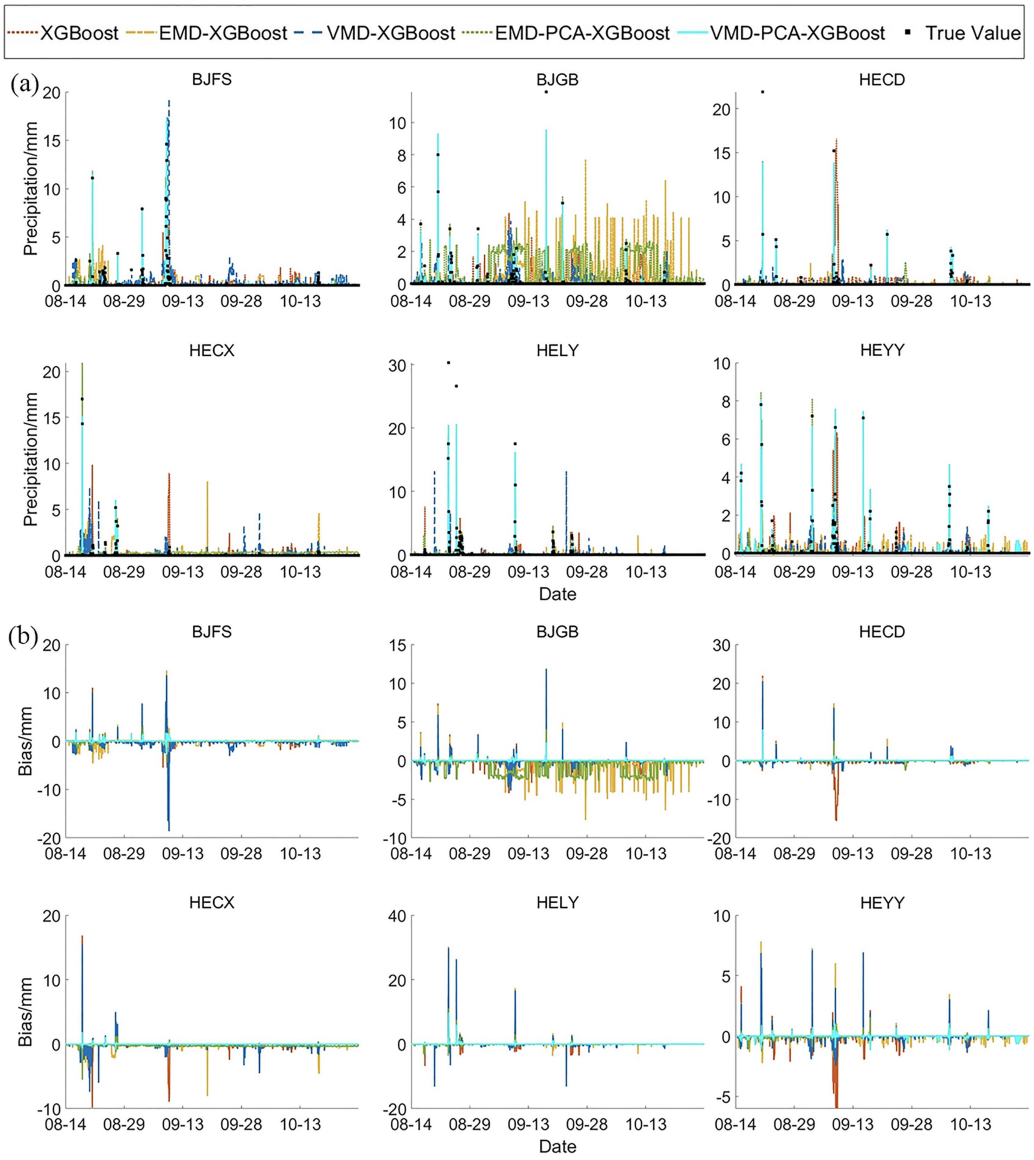

Fig. 7. Forecasting results of XGBoost, EMD-XGBoost, VMD-XGBoost, EMD-PCA-XGBoost, and VMD-PCA-XGBoost. (a) and (b) display the true precipitation values and the forecasting results on six GNSS stations with these five models and biases between them. The locations of these GNSS stations are shown with the red five-pointed star in subgraph (c) of Fig. 1.

as input data for the EMD-PCA-XGBoost model (i.e., Step 3 in Fig. 4). Finally, the predictive performances of these five models are compared and analyzed.

This study predicts precipitation at 13 GNSS stations in the Beijing-Tianjin-Hebei region using five models, calculating forecasting biases and evaluation metrics. For brevity, this study presents only the results of six GNSS stations in Fig. 7(a) and (b) with the rest provided in the supplementary material.

As shown in Fig. 7(a), the predicted values of the VMD-PCA-XGBoost model owns the best agreement with the true values, while the forecasting results of other models show greater discrepancies from the true values. From Fig. 6(b), it is evident that the models with the larger biases are the XGBoost model, followed by the EMD-XGBoost model and the VMD-XGBoost model.

The model evaluation metrics for all stations are presented in Fig. 8. As shown in Fig. 8(a)–(c), for an individual station, the VMD-PCA-XGBoost model achieves a CSI of over 70%, POD reaching 100%, and FAR as low as 20% . For the HEZJ and TJBD stations, the performance of the EMD-PCA-XGBoost model outperforms that of the VMD-PCA-XGBoost model, which may be due to the fact that the data from these two stations are more suitable for EMD decomposition. The specific data for the average evaluation metrics of all models are shown in Fig. 8(d).

Comprehensively, Fig. 8 illustrates that the VMD-PCA-XGBoost model outperforms the other models in terms of performance. Both CSI and POD show improvements or significant improvements compared to the other models, while the FAR generally decreases across the board. Specifically, compared to the XGBoost model, CSI of VMD-PCA-XGBoost increases

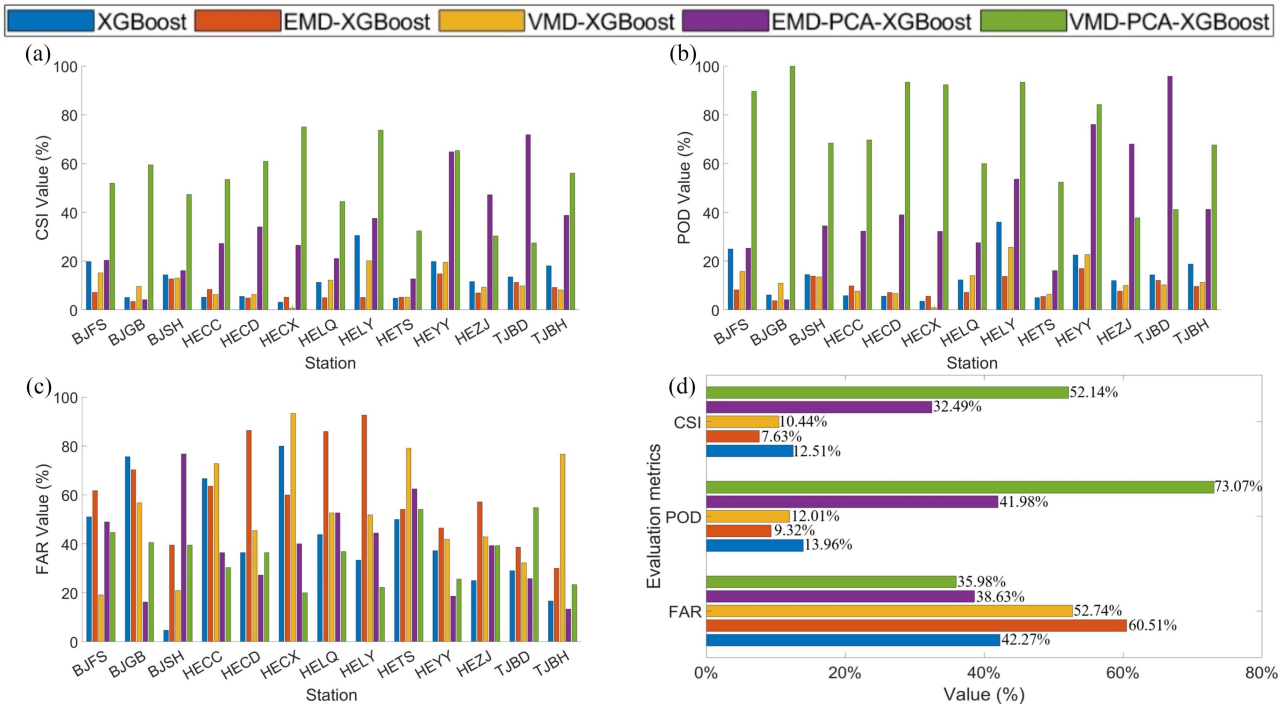

Fig. 8. Distribution of model evaluation metrics of CSI, POD (b) and FAR (c) for all stations, and the average values of the forecasting evaluation metrics of XGBoost, EMD-XGBoost, VMD-XGBoost, EMD-PCA-XGBoost, and VMD-PCA-XGBoost models.

by 39.63%, POD by 59.11%, and FAR decreases by 6.29% . Compared to the EMD-XGBoost model, CSI of VMD-PCA-XGBoost increases by 44.51%, POD by 63.75%, and FAR decreases by 24.53% . Compared to the VMD-XGBoost model, CSI of VMD-PCA-XGBoost improves by 41.70%, POD by 61.06%, and FAR reduces by 16.76% . Even when compared to the EMD-PCA-XGBoost model, the VMD-PCA-XGBoost still demonstrates a 19.65% increase in CSI, 31.09% in POD, and 2.65% decrease in FAR, indicating overall best performance of VMD-PCA-XGBoost model.

The above analysis proved that, VMD-PCA-XGBoost model exhibits the best predictive performance among all the models evaluated. Compared to the EMD algorithm, VMD effectively avoids mode mixing and is more stable when processing signal edges.

In summary, the XGBoost combined model with VMD and PCA outperforms the single XGBoost model, EMD-XGBoost model, VMD-XGBoost model, and EMD-PCA-XGBoost model in forecasting performance.

The results and analysis demonstrate the effectiveness of the proposed model in the study region, while its applicability to other geographic regions and under diverse climatic conditions, including extreme weather events, still requires further investigation. Besides, current limitations of VMD-PCA-XGBoost include reliance on dense GNSS networks, which may hinder applicability in data-sparse regions, therefore, the method is expected to be generalizable to other regions provided sufficient GNSS and meteorological data are available. Additionally, the model's performance under sub-hourly precipitation extremes warrants further testing. To improve the model's adaptability in extreme weather scenarios not covered by the current dataset, future work integrating multisource data to enhance the reliability and robustness of the prediction will be carried out thoroughly.

IV. Conclusion

Given the frequent occurrence of extreme heavy precipitation in complex terrain regions with the limitations of traditional forecasting methods, this study introduces a novel hybrid model, VMD-PCA-XGBoost, for high-precision precipitation forecasting using GNSS water vapor inversion. This approach effectively integrates multiple data sources and advanced techniques, significantly enhancing the accuracy and reliability of precipitation predictions.

The key contributions of this study include the development of a practical regional Tm model and the establishment of an optimized precipitation forecasting model. The process involves decomposing multiple meteorological factors, such as pressure, temperature, relative humidity, precipitable water vapor, and zenith total delay, into intrinsic mode components using VMD. PCA is then applied to reduce the dimensionality of these components, which are combined with precipitation data to serve as input for the model. Exemplified by the Beijing-Tianjin-Hebei region of China, specifically, the proposed VMD-PCA-XGBoost model demonstrates superior performance compared to XGBoost, EMD-XGBoost, VMD-XGBoost, and EMD-PCA-XGBoost models.

In the Beijing-Tianjin-Hebei region, the GPT model and Bevis empirical formula show limited applicability for calculating Tm, particularly the GPT model, which exhibits the poorest

performance in terms of MAE, RMSE, and MSE at the three sounding stations. Conversely, precision evaluations of the Tm model using ERA5 and sounding data indicate that ERA5 data can effectively construct a regional Tm model in the absence of sounding data, facilitating real-time GNSS PWV estimation.

The precipitation forecasting results using the EMD and VMD-based XGBoost combined models reveal that EMD may introduce significant errors at the time series boundaries, known as end-point effects, which can degrade the precision of the overall results. The meteorological data's instability due to climatic fluctuations introduces noise, to which EMD is highly sensitive, potentially failing to decompose signals effectively under high-noise conditions. In contrast, VMD's frequency-domain analysis method is more stable when processing edges signal and exhibits greater resistance to noise in high-noise environments.

The application of VMD denoising and PCA dimensionality reduction enhances the performance of the XGBoost model. Experiment results show that, compared with the XGBoost, EMD-XGBoost, VMD-XGBoost, and EMD-PCA-XGBoost models, the VMD-PCA-XGBoost model provides the highest reliability and accuracy in precipitation forecasting (except for the HEZJ and TJBD stations), followed by the EMD-PCA-XGBoost model. At the HECX station, the CSI achieved by the VMD-PCA-XGBoost model reached 75%, demonstrating greater reliability compared to similar models used in previous studies [18], [21]. For instance, based on ConvLSTM networks, the CSI was only 57.7% [18], and, based on machine learning method of Random Forest, the CSI of all stations only reached 49% [21].

V. Discussion

The study confirms the VMD-PCA-XGBoost model's superiority in robustness, reliability, and precision for precipitation forecasting. The proposed hybrid model shows significant application potential and can be applied to other extreme weather events, such as typhoon and sandstorm, etc. For typhoon forecasting, the model can be integrated with satellite-based cloud imagery and sea surface temperature (e.g., from the Fengyun satellite series), along with surface observations of pressure and wind speed. By capturing the nonlinear interactions between typhoon paths, intensity, and surrounding atmospheric variables, the model facilitates short term and imminent precipitation prediction of affected areas. For sandstorm forecasting, VMD-PCA-XGBoost model can incorporate surface features from Gaofen satellites, GNSS PWV, and ground-based data about wind, pressure, and particulate matter, enabling effective tracking and prediction of sandstorm development.

The model maybe also transferable to other regions besides the Beijing-Tianjin-Hebei region, offering a scalable solution for other regions facing similar climate challenges. In regions with complex terrain and variable precipitation patterns, the model can be trained with local meteorological data to adapt and provide accurate forecasts. Immediate next steps include real-time testing using VMD-PCA-XGBoost with China's Fengyun-4A satellite data and operational the weather research and forecasting outputs to improve the forecasting accuracy and robustness, and validate other hybrid physics-machine learning forecasting. Future research may also explore the model's adaptability to different climatic conditions and its integration with other forecasting systems to further enhance predictive capabilities.

Moreover, independent component analysis (ICA), as a higher statistical version of PCA, can be combined with VMD and XGBoost to construct VMD-ICA-XGBoost model. ICA, as a powerful statistical method, can effectively separate independent components from mixed signals. Therefore, future studies will explore the feasibility of a VMD-ICA-XGBoost combination for high precision precipitation forecasting.

Acknowledgment

The authors thank the National Oceanic and Atmospheric Administration for providing the radiosonde data, CMONOC for providing the GNSS data, ECMWF for providing ERA5 data, Wuhan University for providing PRIDE PPP-AR software.

References

[1] Y. Hirabayashi et al., "Global flood risk under climate change," *Nature Climate Change*, vol. 3, no. 9, pp. 816–821, Sep. 2013, doi: 10.1038/nclimate1911.

[2] S. Westra et al., "Future changes to the intensity and frequency of short-duration extreme rainfall," *Rev. Geophys.*, vol. 52, no. 3, pp. 522–555, Sep. 2014, doi: 10.1002/2014RG000464.

[3] E. Koks, K. Van Ginkel, M. Van Marle, and A. Lemnitzer, "Brief communication: Critical infrastructure impacts of the 2021 mid-July western European flood event," *Natural Hazards Earth Syst. Sci.*, vol. 22, no. 12, pp. 3831–3838, Nov. 2022, doi: 10.5194/nhess-22-3831-2022.

[4] L. Deng et al., "The remote effect of binary Typhoon Infa and Cempaka on the '21.7' heavy rainfall in Henan Province, China," *J. Geophysical Res., Atmospheres*, vol. 127, no. 16, Aug. 2022, Art. no. e2021JD036260, doi: 10.1029/2021JD036260.

[5] S. Gangrade et al., "Unraveling the 2021 Central Tennessee flood event using a hierarchical multi-model inundation modeling framework," *J. Hydrol.*, vol. 625, Oct. 2023, Art. no. 130157, doi: 10.1016/j.jhydrol.2023.130157.

[6] F. Giorgi and L. O. Mearns, "Introduction to special section: Regional climate modeling revisited," *J. Geophysical Res., Atmospheres*, vol. 104, no. D6, pp. 6335–6352, Mar. 1999, doi: 10.1029/98JD02072.

[7] A. Hou et al., "The global precipitation measurement mission," *Bull. Amer. Meteorological Soc.*, vol. 95, no. 5, pp. 701–722, Jan. 2014, doi: 10.1175/BAMS-D-13-00164.1.

[8] C. Kidd and G. Huffman, "Global precipitation measurement," *Meteorol. Appl.*, vol. 18, no. 3, pp. 334–353, Sep. 2011, doi: 10.1002/met.284.

[9] M. Bevis et al., "GPS meteorology: Remote sensing of atmospheric water vapor using the Global Positioning System," *J. Geophysical Res., Atmospheres*, vol. 97, pp. 15787–15801, Oct. 1992, doi: 10.1029/92JD01517.

[10] P. Benevides, J. Catalão, and P. Miranda, "On the inclusion of GPS precipitable water vapour in the nowcasting of rainfall," *Natural Hazards Earth Syst. Sci.*, vol. 15, no. 12, pp. 2605–2616, Dec. 2015, doi: 10.5194/nhess-15-2605-2015.

[11] M. Schirmer and B. Jamieson, "Verification of analysed and forecasted winter precipitation in complex terrain," *Cryosphere*, vol. 9, no. 2, pp. 587–601, Mar. 2015, doi: 10.5194/tc-9-587-2015.

[12] G. Casaretto et al., "High-resolution NWP forecast precipitation comparison over complex terrain of the Sierras de Córdoba during RELAMPAGO-CACTI," *Weather Forecasting*, vol. 37, no. 2, pp. 241–266, Feb. 2022, doi: 10.1175/WAF-D-21-0006.1.

[13] F. A. Hirpa, M. Gebremichael, and T. Hopson, "Evaluation of high - resolution satellite precipitation products over very complex terrain in Ethiopia," *J. Appl. Meteorol. Climatol.*, vol. 49, no. 5, pp. 1044–1051, May 2010, doi: 10.1175/2009JAMC2298.1.

[14] Y. Derin et al., "Multiregional satellite precipitation products evaluation over complex terrain," *J. Hydrometeorol.*, vol. 17, no. 6, pp. 1817–1836, Jun. 2016, doi: 10.1175/JHM-D-15-0197.1.

[15] S. D. P. Williams and F. G. Nievinski, "Tropospheric delays in ground-based GNSS multipath reflectometry—Experimental evidence from

coastal sites," *J. Geophys. Res. Solid Earth*, vol. 122, no. 3, pp. 2310–2327, Mar. 2017, doi: 10.1002/2016JB013612.
[16] S. Xia, S. Jin, and X. Jin, "Estimation and evaluation of zenith tropospheric delay from single and multiple GNSS observations," *Remote Sens.*, vol. 15, no. 23, Nov. 2023, Art. no. 5457, doi: 10.3390/rs15235457.
[17] Y. Xu, Z. Yang, H. Zhou, and F. Zhang, "An initial investigation of the non-isotropic feature of GNSS tropospheric delay," *Satell. Navigation*, vol. 5, Jan. 2024, Art. no. 2, doi: 10.1186/s43020-023-00122-5.
[18] X. Shi et al., "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," in *Proc. Adv. Neural Inf. Process. Syst.*, Sep. 2015, vol. 28, pp. 802–810, doi: 10.48550/arXiv.1506.04214.
[19] A. Solgi, V. Nourani, and M. Marzouni, "Evaluation of nonlinear models for precipitation forecasting," *Hydrological Sci. J.*, vol. 62, no. 16, pp. 2695–2704, Nov. 2017, doi: 10.1080/02626667.2017.1392529.
[20] A. Asanjan et al., "Short-term precipitation forecast based on the PERSIANN system and LSTM recurrent neural networks," *J. Geophysical Res., Atmospheres*, vol. 123, no. 22, pp. 12–543, Oct. 2018, doi: 10.1029/2018JD028375.
[21] Y. Mao and A. Sorteberg, "Improving radar-based precipitation nowcasts with machine learning using an approach based on random forest," *Weather Forecasting*, vol. 35, no. 6, pp. 2461–2478, Dec. 2020, doi: 10.1175/WAF-D-20-0080.1.
[22] L. Espeholt et al., "Deep learning for twelve hour precipitation forecasts," *Nature Commun.*, vol. 13, no. 1, pp. 1–10, Sep. 2022, doi: 10.1038/s41467-022-32483-x.
[23] J. Dong et al., "Enhancing short-term forecasting of daily precipitation using numerical weather prediction bias correcting with XGBoost in different regions of China," *Eng. Appl. Artif. Intell.*, vol. 117, Jan. 2023, Art. no. 105579, doi: 10.1016/j.engappai.2022.105579.
[24] K. Johny, M. Pai, and S. Adarsh, "A multivariate EMD-LSTM model aided with Time Dependent Intrinsic Cross-correlation for monthly rainfall prediction," *Appl. Soft Comput.*, vol. 123, Jul. 2022, Art. no. 108941, doi: 10.1016/j.asoc.2022.108941.
[25] R. Narimani, C. Jun, A. Saedi, S. Bateni, and J. Oh, "A multivariate decomposition–ensemble model for estimating long-term rainfall dynamics," *Climate Dyn.*, vol. 61, no. 3, pp. 1625–1641, Dec. 2022, doi: 10.1007/s00382-022-06646-x.
[26] S. Manandhar, S. Dev, Y. Lee, S. Winkler, and Y. Meng, "Systematic study of weather variables for rainfall detection," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2018, pp. 3027–3030, doi: 10.1109/IGARSS.2018.8517667.
[27] K. Rajkumar and K. Subrahmanyam, "Analysis of principal components and classification enhancement for rainfall prediction," *Indian J. Comput. Sci. Eng.*, vol. 12, no. 2, pp. 348–361, Apr. 2021, doi: 10.21817/indjcse/2021/v12i2/211202047.
[28] C. Tebaldi and R. Knutti, "The use of the multi-model ensemble in probabilistic climate projections," *Philos. Trans.. Ser. A, Math., Phys., Eng. Sci.*, vol. 365, no. 1857, pp. 2053–2075, Aug. 2007, doi: 10.1098/rsta.2007.2076.
[29] T. Li et al., "Machine learning to optimize climate projection over China with multi-model ensemble simulations," *Environ. Res. Lett.*, vol. 16, no. 9, Sep. 2021, Art. no. 094028, doi: 10.1088/1748-9326/ac1d0c.
[30] X. Wen et al., "Changes of temperature and precipitation extremes in China: Past and future," *Theor. Appl. Climatol.*, vol. 126, pp. 369–383, Oct. 2016, doi: 10.1007/s00704-015-1584-x.
[31] K. Trenberth, "Changes in precipitation with climate change," *Climate Res.*, vol. 47, pp. 123–138, Mar. 2011, doi: 10.3354/cr00953.
[32] B. Clarke, F. Otto, R. Stuart-Smith, and L. Harrington, "Extreme weather impacts of climate change: An attribution perspective," *Environ. Res., Climate*, vol. 1, no. 1, Jun. 2022, Art. no. 012001, doi: 10.1088/2752-5295/ac6e7d.
[33] P. Seyrek, B. Şener, A. M. Özbayoğlu, and H. Ö. Ünver, "An evaluation study of EMD, EEMD, and VMD for chatter detection in milling," *Procedia Comput. Sci.*, vol. 200, pp. 160–174, Mar. 2022, doi: 10.1016/j.procs.2022.01.215.
[34] S. Bhattacharya et al., "A novel PCA-firefly based XGBoost classification model for intrusion detection in networks using GPU," *Electronics*, vol. 9, no. 2, Jan. 2020, Art. no. 219, doi: 10.3390/electronics9020219.
[35] Y. Wang, H. Wang, and L. Song, "Amplified precipitation extremes since 21st century in the Beijing-Tianjin-Hebei urban agglomeration, China," *Atmospheric Res.*, vol. 311, Dec. 2024, Art. no. 107695, doi: 10.1016/j.atmosres.2024.107695.
[36] Y. Yu et al., "Tropical cyclone over the western Pacific triggers the record-breaking '21/7' extreme rainfall in Henan, central-eastern China," *Environ. Res. Lett.*, vol. 17, no. 12, Nov. 2022, Art. no. 124003, doi: 10.1088/1748-9326/aca2c4.
[37] G. Zhu et al., "A new approach for the development of grid models calculating tropospheric key parameters over China," *Remote Sens.*, vol. 13, no. 17, Sep. 2021, Art. no. 3546, doi: 10.3390/rs13173546.
[38] W. Zhang et al., "The use of ground-based GPS precipitable water measurements over China to assess radiosonde and ERA-interim moisture trends and errors from 1999 to 2015," *J. Climate*, vol. 30, no. 19, pp. 7643–7667, Oct. 2017, doi: 10.1175/JCLI-D-16-0591.1.
[39] J. Awange, "The Global Positioning System," in *GNSS Environmental Sensing: Revolutionizing Environmental Monitoring*, 2nd ed., Berlin, Germany: Springer, 2018, ch. 2, pp. 25–42, doi: 10.1007/978-3-319-58418-8.
[40] H. Hersbach et al., "The ERA5 global reanalysis," *Quart. J. Roy. Meteorological Soc.*, vol. 146, no. 730, pp. 1999–2049, Jul. 2020, doi: 10.5194/gmd-13-1999-2020.
[41] L. Huang et al., "High-precision GNSS PWV retrieval using dense GNSS sites and in-situ meteorological observations for the evaluation of MERRA-2 and ERA5 reanalysis products over China," *Atmospheric Res.*, vol. 276, Oct. 2022, Art. no. 106247, doi: 10.1016/j.atmosres.2022.106247.
[42] J. Geng, X. Chen, Y. Pan, S. Mao, and K. Zhang, "PRIDE PPP-AR: An open-source software for GPS PPP ambiguity resolution," *GPS Solutions*, vol. 23, no. 91, pp. 1080–5370, Jul. 2019, doi: 10.1007/s10291-019-0888-1.
[43] N. Huang et al., "The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis," *Proc. Roy. Soc. A. Math., Phys. Eng. Sci.*, vol. 454, no. 1971, pp. 903–995, Mar. 1998, doi: 10.1098/rspa.1998.0193.
[44] K. Dragomiretskiy and D. Zosso, "Variational mode decomposition," *IEEE Trans. Signal Process.*, vol. 62, no. 3, pp. 531–544, Feb. 2014, doi: 10.1109/TSP.2013.2288675.
[45] M. Ma et al., "XGBoost-based method for flash flood risk assessment," *J. Hydrol.*, vol. 598, Jul. 2021, Art. no. 126382, doi: 10.1016/j.jhydrol.2021.126382.

Qiaoli Kong received the Ph.D. in space geodesy degree from the College of Geodesy and Geomatics degree, Shandong University of Science and Technology, Qingdao, China, in 2015.

Her research interests include precise orbit determination of low Earth orbit satellites, satellite positioning and navigation systems, Earth rotation parameter estimation, and time series analysis in geodesy.

Qian Li received the B.E. degree in surveying and mapping engineering from the School of Surveying and Urban Spatial Information, Henan University of Urban Construction, Pingdingshan, China, in 2023. She is currently working toward the M.E. degree in space geodesy at Shandong University of Science and Technology, Qingdao, China.

Her research interests include ground-based GNSS water vapor inversion to forecast precipitation.

Qi Bai received the M.E. degree in space geodesy from the College of Geodesy and Geomatics degree, Shandong University of Science and Technology, Qingdao, China, in 2024.

He is currently working with the Shanghai Chart Center of Maritime Safety Administration, Shanghai, China. His research interests include on GNSS meteorology and its applications.

Xiaolong Mi received the Ph.D. degree in GNSS and space geodesy from the University of Chinese Academy of Sciences, Beijing, China, in 2022, and the Ph.D. degree in spatial sciences from Curtin University, Perth, WA, Australia, in 2023.

His primary research interests include GNSS and low Earth orbit technologies for positioning, navigation, and timing along with the application of artificial intelligence in earth and space sciences.

Joseph Awange received the Ph.D. degree in Geodesy from Stuttgart University, Stuttgart, Germany, in 2001.

In 2004, he did his postdoctoral research in Kyoto University, Japan. He is currently a Professor of Environmental Geoinformatics with the Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hong Kong, having previously worked at Curtin University, Australia.

Shi Wang received the B.E. degree in surveying and mapping engineering from the Shandong University of Technology, Zibo, China, in 2022. He is currently working toward the M.E. degree in surveying and mapping science and technology in the Shandong University of Science and Technology, Qingdao, China.

His research interests include the impact of high-order ionosphere on satellite orbit determination.

Yi Yang received the B.E. degree in surveying and mapping engineering from the School of Environmental Arts and Architectural Engineering, Heilongjiang University of Technology, Heilongjiang, China, in 2023. He is currently working toward the M.E. degree in space geodesy at Shandong University of Science and Technology, Qingdao, China.

His research interests include utilizing GNSS-derived precipitable water vapor to investigate the dynamics and impacts of typhoons.

Guoli Bo received the B.E. in surveying and mapping engineering degree from Jining University, Qufu, China, in 2023. He is currently working toward the M.E. degree in space geodesy at Shandong University of Science and Technology, Qingdao, China.

His research interests include PM prediction.