A Composite Uncertainty Forecasting Model for Unstable Time Series: Application of Wind Speed and Streamflow Forecasting

NA SUN [1], SHUAI ZHANG[2], TIAN PENG[1], JIANZHONG ZHOU [3], AND XINGUO SUN[4]

[1]College of Automation, Huaiyin Institute of Technology, Huai'an 223003, China
[2]Key Laboratory of Thermo-Fluid Science and Engineering, Ministry of Education, School of Energy and Power Engineering, Xi'an Jiaotong University, Xi'an 710049, China
[3]School of Hydropower and Information Engineering, Huazhong University of Science and Technology, Wuhan 430074, China
[4]Jiangsu Smart Factory Engineering Research Centre, College of Management and Engineering, Huaiyin Institute of Technology, Huai'an 223003, China

Corresponding author: Na Sun (sunna1347@126.com)

ABSTRACT Wind speed and streamflow series always are nonlinear and unstable because the effects of chaotic weather systems. These inherent features make them difficult to forecast, especially in a changing environment. To improve forecasting accuracy, an innovation uncertainty forecasting architecture is developed by coupling data decomposition method, feature selection, multiple artificial intelligence (AI) techniques and composite strategy to do unstable time series forecasting. In the designed architecture, the AVMD (adaptive variational mode decomposition) is first applied to excavate implicit information from the original time series. Then, the random forest is utilized to select the suitable inputs for each mode. After that, the GPR (Gaussian Process Regression), a very famous probabilistic AI technique, is driven by various neural networks (ELM (Extreme Learning Machine), BP (Back Propagation Neural Networks), GRNN(Generalized Regression Neural Networks) and RBF (Radial Basis Function Neural Networks)) to produce both deterministic and probabilistic forecasting results in a nonlinear manner to play strengths of each other. The effectiveness and applicability of the proposed approach is verified by unstable wind speed data and streamflow data, and also compared with eleven related models. Results indicate that the proposed model not only improves the forecasting accuracy for deterministic predictions, but also provides more probabilistic information for decision making. The proposed method achieves significantly better performance than the traditional forecasting models both on wind speed forecasting and streamflow forecasting with at least 50% average performance promotion over all the eleven competitors. Comprehensive comparisons demonstrate the superior performance of the proposed method than the involved models as a powerful tool for unstable series forecasting.

INDEX TERMS Unstable time series prediction, hybrid model, GPR, artificial intelligence, combination forecasting.

I. INTRODUCTION

Due to the deteriorating environment and the rapidly growing demand for energy, environmentally friendly energies, such as wind and water resources, have drawn worldwide attentions [1]. Predicting future changes of wind speed and streamflow series is an effective way to ensure rational

utilization on wind and water resources [2]. However, their inherent randomness and non-stationarity make a huge barrier to accurate forecasting.

Up to now, many approaches have been used to enhance the forecasting accuracy and reliability of energy series. The first one is using feature selection methods (FSMs). For examples, principle component analysis (PCA) [3], random forest (RF) [4], and cluster analysis (CA) [5]. These FSMs are used to handle multivariate data to reduce their

---

The associate editor coordinating the review of this manuscript and approving it for publication was Filbert Juwono [ID].

dimensions. The second way is applied the data decomposing techniques (DDTs) to divide the original single series into much subseries so as to extract more implicit information or reduce noise. Till now, there are many DDTs, such as empirical wavelet transform (EWT) [6], variational mode decomposition (VMD) [7], singular spectrum analysis (SSA)[8], empirical mode decomposition (EMD) [9], and many variants of the EMD (*e.g.* EEMD [10] and CEEMDAN [11]). These DDT have their respective pros and cons. EMD and its variant EEMD can extract waveforms of different frequencies from any signal step by step. EEMD can solve the mode mix problem of EMD by adding white noise [12]. The core idea of EWT is to construct a proper orthogonal wavelet filter bank to adaptively partition the Fourier spectrum of the signal, so as to extract the AM-FM component with tightly supported Fourier spectrum. Compared with EMD, EWT has the advantages of stronger theory, less computation, and less decomposed modes. VMD is an emerging DDT, its essential idea is the variational problem [13]. Different with the recursive filter subseries used by EMD, VMD turn the decomposing problem into a non-recursive and variational mode decomposition problem. In the VMD, the frequency center and bandwidth of each component are determined by searching the optimal solution of the variational model iteratively, so it can adaptively realize the division frequency domain of the signal and the effective separation of each component, highlight the local characteristics of the data. Recently, Sun *et al.* [14] proposed an adaptive variational mode decomposition (AVMD) by adding a decomposition quality factor (QF) as the termination condition to automatic determine the key parameter (the number of modes K) and then improve the performance of decomposition.

The above-mentioned FSMs and DDTs methods are used for data processing. The third way to improve the prediction accuracy is to improve the performance of the basic forecasting model. There are three typical forecasting models: physical models (PMs), data-driven models (DDMs), and hybrid models. With the rapid development of computer and information technology, and artificial intelligence (AI) theory, the DDMs are going famous than PMs in the aspects of easily modelling and less data requirement. AI techniques have many successful applications in energy series forecasting, showing their strong nonlinearity learning ability. Most famous AI models used in energy series forecasting are ELM (extreme learning machine)[15], BP (back-propagation neural network)[16], RBF (radial basis function) neural network [17], GRNN (generalized regression neural network) [18], SVM (support vector machine) [3], and ANFIS (adaptive neuro-fuzzy inference system)[19]. Nowadays, more studies have been done to explore the forecasting ability of deep learning techniques, including DBN(deep belief network)[20], CNN(Convolutional Neural Network) [21], LSTM (long short-term memory network) [22], and CNN-LSTM [23]. Although the above mentioned AI models can perform admirable effectiveness on short-term wind speed or streamflow prediction, while

neural networks based on gradient-based training, such as BP, often suffers from the disadvantages of lower training speed, local convergence and stopping criteria [15] as well as the increasing computation cost of SVM with the growing of data scale significantly [8]. For deep learning technique, such as LSTM, hyperparameters determination needs more experience and more computation cost. Due to the fact that the influence factors of energy series are various, the energy series often own several inherent features such as intermittent and stochastic merits. So, it is difficult to use a single AI model to achieve accurate and reliable prediction, especially in a changing condition, and a specific single model usually cannot obtain accurate results at various sites.

Generally, the forecasting effectiveness of single DDM is significantly affected by the corresponding inherent structure and parameters. For this purpose, hybrid model incorporating the advantages of multiple models or techniques have been rapidly developed. Liu *et al.* [18] combined a data pretreatment strategy, a modified multi-objective optimization algorithm, and several forecasting models together to construct a new model for short-term wind speed forecasting. Meanwhile, to improve effectiveness of single DDM, various swarm intelligence algorithms have been rapidly developed and applied. Yaseen *et al.* [19] proposed three bio-inspired adaptive neuro-fuzzy inference system (ANFIS) models for highly non-linear streamflow forecasting. Later, Lin *et al.* [20] developed a hybrid wind speed forecasting model by combining a deep belief network with genetic algorithm. Additionally, coupling DDTs, machine learning models, and/or optimization algorithms is also popular in improving the forecasting effectiveness. For examples, Liu *et al.* [24] developed a multi-step wind speed forecasting model based on decomposition, ensemble method and error correction algorithm. Zhang *et al.* [25] presented a hybrid model by combining VMD, the neural network, and Lorenz disturbance. Later, they adopted the Lorenz disturbance models to improve the EMD-BP model. Their method can eliminate the randomness of the wind speed sequences and significantly improve the forecasting accuracy [16]. The same year, they further developed a novel wind speed prediction model through decomposing wind speed into three parts, then analyzing each part and making predictions with different models according to their characteristics [26]. Based on these, another novel wind speed prediction method was proposed [17]. In their study, different models are built based on the characteristics of each part generated by the VMD, PCA-RBF, ARMA (Auto Regressive and Moving Average) model under the MCMC (Markov chain Monte Carlo) framework and probability distribution are built separately for the nonlinear part, linear and noise parts. Huang *et al.* [27] explored the combination of CNN and VMD for electricity price forecasting.

However, most of the previous researches have focused on deterministic forecasting with point result, which usually cannot meet the requirements of practical utilization. Whereas uncertainty analysis can provide supplementary

information to assistant relevant departments to do more safe and scientific decision making. Therefore, uncertainty analysis is particularly important for the topic of unsteady energy series forecasting. The most commonly used methods for probabilistic forecasting to construct the prediction intervals (PIs) are traditionally Bayesian, mean-variance, and Bootstrap [28]. These approaches have their own characteristics. Generally, the Bayesian method with high computational cost provides imprecise PIs. By contrast, the mean-variance has a negligible computational mass due to its calculation process is embedded in the model training process, but its PIs has a low empirical coverage probability. The Bootstrap is easy implementation and has a large computational requirement. Additionally, lower and upper bound estimation (LUBE) approach without a prior knowledge of point prediction result also is applied to construct PIs. But overtraining and high computational burden are primary reasons that are restraining its real applications. Recently, several statistical methods without any prior assumption have been suggested for probabilistic forecasting. The statistical methods mainly depend on the quantile analysis results of point prediction including kernel density estimation [29], quantile regression (QR) [22], and ensemble simulations [30] methods. Recently, based on VMD and QR, a probabilistic interval forecasting model was put forward and the eventual results quantified the potential risks of the wind power series [31]. To our knowledge, gaussian process regression (GPR) has gained more attentions in energy series forecasting, due to its seamless integration of hyperparameter estimation, model training, and uncertainty estimation. Zhang *et al.* [32] built a new probabilistic wind speed forecasting method based on autoregressive (AR) and GPR. From the above researches, it can be concluded that the current probabilistic prediction rarely involves in the analysis of how to generate good point predictions and multiple characteristics of data simultaneously. To provide both point and interval predictions for future energy series, a GPR based hybrid forecasting model is proposed in this study.

Based on the above literature review, this study focuses on merging the superiority of the dimension reduction, data decomposition, and multiple AI models to develop a composite uncertainty forecasting model for unstable energy series. This composite uncertainty forecasting model contains three major modules: a data pre-decomposing module, an input selection module, and a forecasting module. More specifically, AVMD is applied to decompose original energy series into several sub-series to fully dug the detail characteristic of original series. Then, RF is adopted to determine the best input vector. After that, several NNs, including BP, RBF, GRNN, and ELM are applied to forecast the prepared energy series. The results of these multiple individual forecasting models are weighted by the GPR to generate the final forecasting results. Finally, several famous indices are used to assess the performance of all relevant models. The proposed **AV**MD-**R**F-**M**NN-GPR, denoted as AVRM-GPR, model is compared fully with eleven forecasting models including individual and hybrid models.

The main contributions of this study are demonstrated as follows:

(1) The most effective data preprocessing method AVMD is adopted for extract more detailed information from original energy series. The obtained subseries is more predictable than the mother series.

(2) The RF as an input selection approach can get more nonlinear relationship between all input factors than the commonly used PACF (partial autocorrection function).

(3) Different data characteristics can be portrayed effectively by multiple NNs, thus the weakness of individual models can be overcome through composite the forecasting results of multiple NNs.

(4) The weights of all these NNs is determined by a nonlinear way GPR, which can both provide the deterministic results and uncertainty bounds. This unique way can achieve many things such as nonlinear weighting, uncertainty analyzing at one stroke.

The remainder of this paper is arranged as follows. Section 2 presents the framework of the proposed model. Section 3 introduces the proposed forecasting framework and the methodologies used in the proposed model. Section 4 describes the construction and development of the models for energy series forecasting. Section 5 compares the performance of the proposed model with several other models on four different wind speed datasets and two streamflow datasets. Discussions are presented in Section 6. Finally, conclusions are drawn in Section 7.

II. THE FRAMEWORK AND PROCESS OF THE PROPOSED MODEL

The framework of the AVRM-GPR model is depicted in Figure 1, and the whole process is described in detail as follows:

Stage 1: (Data pre-Decomposing): In our previous paper [14], an AVMD (adaptive variational mode decomposition) was proposed by adding a decomposition quality factor (QF) as the termination condition to automatic determine the key parameter (the number of modes K) of VMD and thereby improve its decomposition performance. The AVMD is adopted to decompose the original series into several sub-series to make them easy to be predicted. The AVMD method is introduced in Section 3.1.

Stage 2 (Selecting the Input Vector for Each Subseries): The random forest (RF) technique is utilized to select the appropriate input vector of each sub-series. The introduction of the RF is presented in Section 3.2.

Stage 3 (Forecasting of Individual Engines): In this paper, BP, RBF, GRNN, and ELM which have strong nonlinear fitting ability, are selected as the individual forecasting engines to establish the developed AVRM-GPR model.

Stage 4 (Establishing the Proposed AVRM-GPR Model): All forecasting results of the above four engines are taken as inputs to drive the GPR model, a very famous probabilistic AI technique, to merge all advantages of these engines.
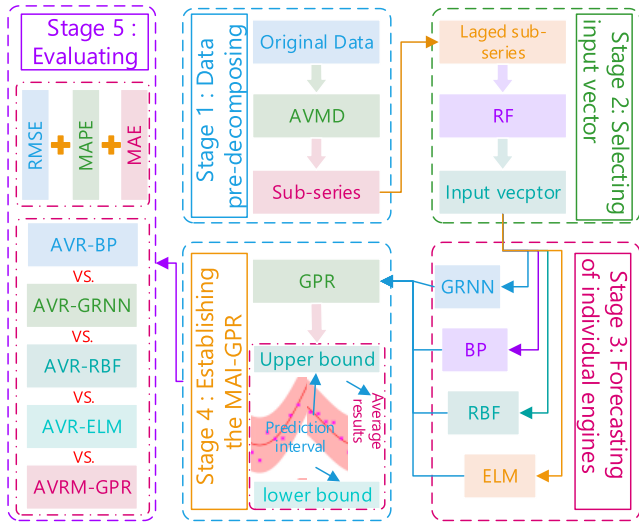
FIGURE 1. The flowchart of the proposed AVRM-GPR.

The AVRM-GPR model can provide both deterministic and probabilistic forecasting results.

Stage 5 (Evaluating): To illustrate the prediction performance of the proposed AVRM-GPR model, eleven benchmark models are used. These compared models can be divided into three types. The first one is the individual models, including GRNN, RBF, BP, and ELM. The second one is the hybrid models with PACF and different decomposition methods, including DWT-PACF-ELM, EEMD-PACF-ELM, and AVMD-PACF-ELM. The third one is coupling the individual model with AVMD and RF, which consists of **AV**MD-**R**F-**BP** (denoted by AVR-BP) model, AVR-GRNN, AVR-RBF, and AVR-ELM model.

III. METHODOLOGIES

A. ADAPTIVE VARIATIONAL MODE DECOMPOSITION

In the standard VMD method, the number of modes K is very important for the decomposition quality. As Dragomiretskiy and Zosso [13] mentioned, small K may cause that a few modes may be either shared by neighboring modes or mostly discarded, while large K may lead to excessive decomposition and produce some components with useless information. Therefore, a decomposition quality factor (QF) has been designed as the termination condition during the VMD decomposition process to determine the suitable K quickly and automatically.

For a given signal $f(t)$, the decomposition process of VMD is implemented by solving the following optimization problem:

$$\min_{\{u_k\},\{\omega_k\}} \left\{ \sum_k \left\| \partial_t \left[\left(\delta(t) + j/(\pi t) \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\},$$
$$k = 1, 2, \cdots, K$$
$$s.t. \sum_k u_k = f(t) \tag{1}$$

where u_k and ω_k represent all modes and their frequencies, respectively; $\delta(t)$ is the Dirac distribution and $*$ denotes the convolution operator.

The Eq. (1) can be translated into an unconstrained one by adding a quadratic penalty term and Lagrangian multipliers, as follows:

$$L(\{u_k\}, \{\omega_k\}, \lambda)$$
$$= \alpha \sum_k \left\| \partial_t \left[\left(\delta(t) + j/(\pi t) \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2$$
$$+ \left\| f(t) - \sum_k u_k \right\|_2^2 + \left\langle \lambda, f(t) - \sum_k u_k \right\rangle \tag{2}$$

where α denotes the balancing parameter of the data-fidelity constraint.

The alternate direction method of multipliers (ADMM) can be used to solve the unconstrained optimization problem Eq. (2). Through a series of iterative sub-optimizations by updating u_k^{n+1}, ω_k^{n+1}, and λ^{n+1}, the saddle point of the augmented Lagrangian can be found. The updating formulas are below:

$$\hat{u}_k^{n+1} = \frac{\hat{f}(\omega) - \sum_{i\neq k} \hat{u}_i(\omega) + \hat{\lambda}(\omega)/2}{1 + 2\alpha(\omega - \omega_k)^2}$$
$$\omega_k^{n+1} = \int_0^\infty \omega |\hat{u}_k(\omega)|^2 d\omega \bigg/ \int_0^\infty |\hat{u}_k(\omega)|^2 d\omega$$
$$\hat{\lambda}^{n+1}(\omega) = \hat{\lambda}^n(\omega) + \tau \left(\hat{f}(\omega) - \sum_{i\neq k} \hat{u}_k^{n+1}(\omega) \right) \tag{3}$$

where n is the total number of iterations; τ is the time-step of the dual ascent; $\hat{u}_k^{n+1}$, $\hat{u}_i(\omega)$, $\hat{f}(\omega)$, and $\hat{\lambda}(\omega)$ are the Fourier transform of u_k^{n+1}, $u_i(\omega)$, $f(\omega)$, and $\lambda(\omega)$, respectively.

Suppose the K modes $\{\text{Mode}_1, \text{Mode}_2, \ldots, \text{Mode}_K\}$ obtained by the standard VMD, the sample entropy values $\{S_i, i = 1, 2, \cdots, K\}$ of each mode can be calculated. The entire error between the original series $x(t)$ and the reconstructed series $\widehat{x}(t) = \text{Mode}_1 + \cdots + \text{Mode}_K$ is then calculated by:

$$Er = \sum_{i=1}^N |x(t) - \widehat{x}(t)| \tag{4}$$

The decomposition quality factor (QF) is defined as:

$$QF = C_v / Er \tag{5}$$

where C_v is the variation coefficient of sample entropy values for all modes.

The details of AVMD method can be found in our previous paper [14].

B. RANDOM FOREST

Random Forests (RF) algorithm was first developed in year 2001 to enhance the performance of classification and regression trees as well as reducing overfitting risk [33]. It is a powerful AI technique to extract features and/or map nonlinear relationship for high-dimensional data. So, it has been widely utilized in various realms since its birth, including rainfall forecasting [34], solar radiation forecasting [35], urban water consumption [36], and land cover classification [37]. It turned out to be a suitable tool for handling classification and regression tasks of high-dimensional data with small samples. In addition, it also can provide variable

importance measure (*VIM*) for each input feature variable by permuting each feature to measure its importance.

All data collected in this study are time series, so the rest of this subsection is limited to introduce the RF used in regression issues. A training set with n observations (denoted by Tr_n), which contains an input vector X with m features and its corresponding expected output vector Y, is written as:

$$Tr_n = \{(X_1, Y_1), (X_2, Y_2), \cdots, (X_n, Y_n)\},$$
$$X \in R^m, \quad Y \in R \tag{6}$$

The process of RF for solving a regression problem is briefly presented as follows [38]:

Phase 1: Generate N_{tr} random samples from the original training data according to bootstrap technique.

Phase 2: Grow unpruned regression trees to fit the bootstrap samples. In this stage, *mtry* variables are first randomly selected from all predictors at each node of the regression tree. Then, divide each node into a couple of sub-nodes according to the optimal split point (OSP) strategy. The OSP is a point that has ability to minimize the mean square error (MSE) using the selected *mtry* variables. The fission process is finished until achieving the default termination criterion, such as the smallest number in a child node.

Phase 3: Repeat phases 1~2 and aggregate N_{tr} trees as the final RF model.

Phase 4: Compute the average results of sub-trees as the final prediction.

One merit of the RF is embedding bootstrap technique into sample phase to enhance its ability in anti-noise interference. In the sample phase, many unrelated trees can be generated from different training samples obtained by bootstrap technique. The *VIM* value of each predictor can be calculated by the error of the out-of-bag sample (denoted by E_{OOB}). The out-of-bag samples are defined as the observations that are not in the bootstrap sample (denoted by S_n^B). In this study, MSE between expected output Y and the estimated output $\hat{Y}$ is adopted as the E_{OOB} to evaluate the generalization ability of each regression model. The *VIM* can be calculated through disturbing predictors one by one. The *VIM* value of a specific predictor is the mean value of the differences of E_{OOB} before and after perturbation of this predictor on all trees. Specifically, for a bootstrap sample S_n^B with total m predictors, its permuted sample denoted by OOB_B' is obtained by disturbing the predictor at the position j among the m predictors. The *VIM* value of the *j*-th variable is defined by the average difference between the original E_{OOB} and the E_{OOB}' for the disturbed OOB_B', and it is calculated by:

$$VIM\left(X^j\right) = \frac{1}{N_{tr}} \sum_{i=1}^{N_{tr}} \left(E_{OOB}^i - E_{OOB}^{'i}\right) \tag{7}$$

where N_{tr} represents the total number of the regression trees. The larger the *VIM* value, the more important the permutated predictor.

As mentioned above, the RF includes two key parameters: the number of regression trees N_{tr} and the maximum number

of variables *mtry* are used to grow a regression tree, *mtry* is specified as *m*/3 [39].

C. GAUSSIAN PROCESS REGRESSION (GPR)

Compared with NNs, the GPR, a non-parametric kernel-based probabilistic model [40], merges many machine learning tasks, including model training, hyper-parameters estimation and uncertainty quantification. So, GPR not only has the strong nonlinear learning capacity of NNs, but also has the inferential ability of Bayesian. Therefore, it can generate both deterministic and probabilistic results.

A gaussian process (GP) can be uniquely determined by its mean function $E(X)$ and covariance function matrix $K(X, X)$. The nonlinear function (denoted by $g(X)$) between the energy series y and its relevant predictors X can be considered as a GP:

$$g(X) \sim GP(E(X), K(X, X)) \tag{8}$$

In reality, due to the interference of noise, the energy series y be expressed as the superposition of mapping function $g(X)$ and noise ε:

$$y = g(X) + \varepsilon \tag{9}$$

Supposing that the noise ε conforms to the Gaussian distribution with the mean value of 0 and the variance of σ_n^2, denoted by $\varepsilon \sim N(0, \sigma_n^2 I_n)$, the prior distribution of the y can be written as:

$$y \sim N(0, K(X, X) + \sigma_n^2 I_n) \tag{10}$$

where I_n is the *n*-dimensional unit matrix; $K(X, X)$ represents the covariance function matrix which is a symmetric and positive semi-definite one.

The joint prior distributions of the observed y and the predicted y_p is:

$$\begin{bmatrix} y \\ y_p \end{bmatrix} \sim N\left(0, \begin{bmatrix} K & K_*^T \\ K_* & k(x_{test}, x_{test}) \end{bmatrix}\right) \tag{11}$$

where K and K_* are:

$$K = \begin{bmatrix} k(x_1, x_1), k(x_1, x_2), \cdots, k(x_1, x_n) \\ k(x_2, x_1), k(x_2, x_2), \cdots, k(x_2, x_n) \\ \cdots \quad\quad \cdots \quad\quad \cdots \\ k(x_n, x_1), k(x_n, x_2), \cdots, k(x_n, x_n) \end{bmatrix} \tag{12}$$

$$K_* = [k(x_{test}, x_1), k(x_{test}, x_2), \cdots, k(x_{test}, x_n)] \tag{13}$$

According to the properties of GP, the posterior distribution of the predicted value y_p for the new input x_{test} and output y under the given training input data X also conforms joint Gaussian distribution, so the output for the new input x_{test} is calculated:

$$p(y_p | X, y, x_{test}) \sim N(E(y_{test}), \text{cov}(y_{test}))$$
$$E(y_{test}) = K_* K^{-1} y$$
$$\text{cov}(y_{test}) = k(x_{test}, x_{test}) - K_* K^{-1} K_*^T \tag{14}$$

where $E(y_{test})$ and $\text{cov}(y_{test})$ are the expected value and posterior variance of y_P, respectively. The $\text{cov}(y_{test})$ can show the uncertainty of the predicted results.

The standard GPR can be determined by simplifying the mean function $E(X)$ of GP to be zero [41]. The covariance function $K(X, X)$ can be equivalent to the kernel function. Hence, the main task is determined the $K(X, X)$. Most commonly used kernel functions in the GPR are squared exponential kernel (SEK), rational quadratic kernel (RQK), periodic kernel (PK), and Matérn kernel (MaK) [42]. In this study, the SEK is adopted because its infinitely differentiable feature can make GPR smoother. The SEK is given by:

$$k\left(X, X^*\right) = \sigma_f^2 \exp\left(-\frac{1}{2}\left(X - X^*\right)^T P^{-1}\left(X - X^*\right)\right)$$
$$P = diag([\lambda_1, \lambda_2, \cdots, \lambda_m]^T) \tag{15}$$

where P is a diagonal matrix of smoothing coefficients λ; σ_f^2 is the signal variance linked to the general function variance. Generally, the parameter set $\theta = \{\sigma_f^2, \lambda_1, \lambda_2, \cdots, \lambda_m\}$ is called the hyper-parameter set of the kernel function in the GPR. The most commonly used method for solving hyper-parameters is maximum likelihood function (MLF).

IV. MODEL CONSTRUCTION AND DEVELOPMENT

In this section, the numerical modeling of the developed AVRM-GPR model in this study is performed to do multi-step ahead wind speed and streamflow forecasting.

A. DATA COLLECTION

In this study, two types of classic unstable time series: wind speed and streamflow are used. Four wind speed series (see Figure 2 (a)), collected from the famous wind farm, the Sotavento Galicia (SG) wind farm[1], are used. Each wind speed series includes 2700 data with 10-min time intervals. Moreover, two monthly streamflow series for the period year 1953 to 2015, collected from the Yangtze River Waterway Bureau, China, from different hydrological stations are obtained in this study as shown in Figure 2 (b). Figure 2 reveals the strong volatility of the original wind speed and the streamflow series. For each series, the first 75% of the whole dataset are selected as the training samples to establish the forecasting model, and the remaining 25% are used for validation.

The statistical information including average (Ave.), maximum (Max.), minimum (Min.), standard deviation (Std.), coefficient of variation (C_v) and the skewness coefficient (C_s) of these six series are presented in Table 1. The statistical information also verifies the strong volatility of the original wind speed and the streamflow series, indicating a hard forecasting task.

B. PARAMETER SETTINGS

Besides the proposed AVRM-GPR model, there are four **AV**MD-**R**F-based models as benchmark models including the **AV**MD-**R**F-**GRNN** (AVR-GRNN), the **AV**MD-**R**F-**RBF**
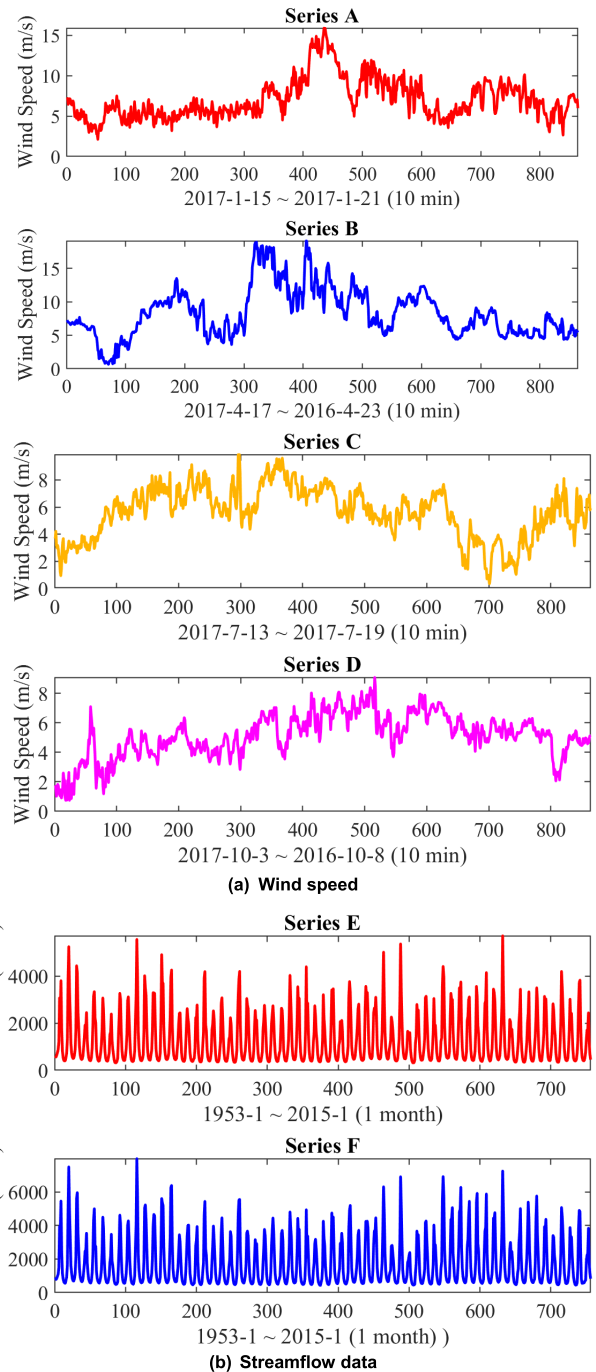
FIGURE 2. Four wind speed series and two streamflow series.

(AVR-RBF), the **AV**MD-**R**F-**BP** (AVR-BP), and the **AV**MD-**R**F-**ELM** (AVR-ELM). Similar to the proposed method, the number of inputs in GRNN, BP, ELM, and RBF is determined by RF. The other three hybrid models DWT-PACF-ELM, EEMD-PACF-ELM, and AVMD-PACF-ELM used PCAF to determine the inputs. Specifically, the Daubechies 10 is selected as the mother wavelet mother function for DWT according to Ref [43]. The EEMD method is employed with the ensemble number of 100 and the white noise amplitude

Series	Mean	Std.	Cv	Sk	Max	Min
Series A	8.64	4.53	0.53	0.87	29.82	0.35
Series B	7.50	3.30	0.44	1.51	31.88	2.91
Series C	6.14	2.06	0.34	1.47	17.49	2.13
Series D	6.06	2.09	0.34	1.39	15.42	1.89
Series E	1329	1071	0.81	1.26	5708	303
Series F	1811	1498	0.83	1.29	7990	426

of 0.2 times standard deviation [12]. For AVMD, the number of modes is searched in the range [2, 15] with increasing step of 1 [14]. For individual models, the number of hidden neurons of the ELM and BP neural networks are selected using grid search (GS) algorithm. The search range is set as $[m, 2n + 20]$, where n denotes the number of input neurons, m is set as $2n - 20$ when n is bigger than 10, otherwise m is set as 1 [44]. The Levenberg-Marquardt (trainlm) is selected as the training function for BP. The smoothing parameter of GRNN is determined by cross validation method (CV) in the range (0,1) with interval 0.01. For RBF model, the spread is searched by CV method in the range [0.5, 1.5] with 0.1 increments.

C. DECOMPOSITION RESULTS

The original series were processed by AVMD for the proposed model, and the decomposition result is shown in Figure 3. The initial sequence is broken into K modes. And the K is adaptively determined by the decomposition accuracy. The decomposition results reflect features of wind speed/streamflow sequences at different scale. The first mode in each graph represents the residual term, describing the trend information of wind speed and streamflow sequences. For the modes of the streamflow series E and F, mode 2 shows obvious periodicity.

D. PERFORMANCE EVALUATION

Three mainstream indexes *MAE* (mean absolute error), *MAPE* (mean absolute percentage error), and *RMSE* (root mean squared error) are utilized to evaluate the performance of all the involved models in this study. Equations of these indices are shown as follow [14]:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left(y_{obs,i} - y_{fore,i}\right)^2}, \quad RMSE > 0 \tag{16}$$

$$MAE = \frac{1}{N} \sum_{i=1}^{N} \left|y_{obs,i} - y_{fore,i}\right|, \quad MAE > 0 \tag{17}$$

$$MAPE = \frac{1}{N} \sum_{i=1}^{N} \left|\left(y_{obs,i} - y_{fore,i}\right)/y_{obs,i}\right|,$$
$$1 > MAPE > 0 \tag{18}$$

where $y_{obs,i}$ is the i-th measured wind speed or streamflow data, $y_{fore,i}$ is the i-th forecasted wind speed or streamflow data, and N is the total number of all sample data.

In addition, Pearson's test is also applied to evaluate the agreement of the forecasting results with real data. The Pearson's test was developed by scientist Karl Pearson [24]. The Pearson's test correlation coefficient (CC) is given as:

$$CC = \frac{\sum_{i=1}^{N} \left(y_{obs,i} - \bar{y}_{obs}\right) \left(y_{fore,i} - \bar{y}_{fore}\right)}{\sqrt{\sum_{i=1}^{N} \left(y_{obs,i} - \bar{y}_{obs}\right)^2 \sum_{i=1}^{N} \left(y_{fore,i} - \bar{y}_{fore}\right)^2}},$$
$$1 > CC > 0 \tag{19}$$

where $\bar{y}_{obs}$ and $\bar{y}_{fore}$ are the average values of the measured and forecasted wind speed or streamflow data, respectively.

When the CC value is equal to 1, it indicates that there is a significant linear relationship between the actual data and the forecasting data, whereas, CC value is equal to 0, representing no relationship between these two datasets.

V. EXPERIMENTS

In this section, to illustrate the forecasting performance of the proposed AVRM-GPR model, six experimental tests are conducted for 1-step, 2-step, and 3-step predictions. Six datasets (Series A, B, C, D, E, and F) were collected to testify the effectiveness and efficiency of the proposed AVRM-GPR model for unstable energy series forecasting. The first four datasets are wind speed series among various seasons from the same wind farm, and the other two are streamflow series from different hydrological stations. In other word, the datasets of these experimental tests have different characteristics, because they are collected from different stations and have different time intervals. Therefore, these six datasets have obvious different complexities on non-stationarity, non-linearity, randomicity, and irregularity, which can provide sufficient evidence for the experimental conclusion. In each experimental test, the first 75% of the whole dataset are selected to establish the involved forecasting models, and the remaining 25% are used for testifying.

Based on the historical data, three comparisons were conducted to illustrate the applicability and efficiency of the developed combined model compared with eleven other benchmark models. These benchmark models are: 1) classic individual models: GRNN, RBF, BP, and ELM; 2) hybrid models employing diverse DDTs: EEMD-PACF-ELM, DWT-PACF-ELM, and AVMD-PACF-ELM; 3) AVR-based models: AVR-GRNN, AVR-RBF, AVR-BP, and AVR-ELM.
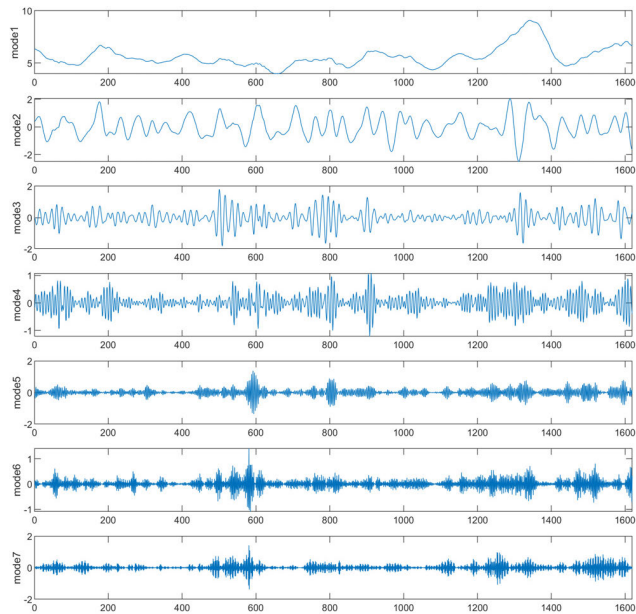
A. COMPARISON WITH CLASSIC INDIVIDUAL MODELS

This part reveals prediction abilities of the proposed model by contrasting the proposed model with several classic individual models, such as GRNN, RBF, BP, and ELM. The forecasting results of these models in 1-step ahead forecasting for Datasets A ~ F are listed in Table 2, where the bold part represents the best forecasting results. Clearly, the proposed model has a distinguished forecasting performance, whether for wind speed forecasting (Datasets A~D) or for streamflow forecasting (Datasets E and F). Specifically, for Dataset A, the smallest *RMSE*, *MAE*, and *MAPE* values are
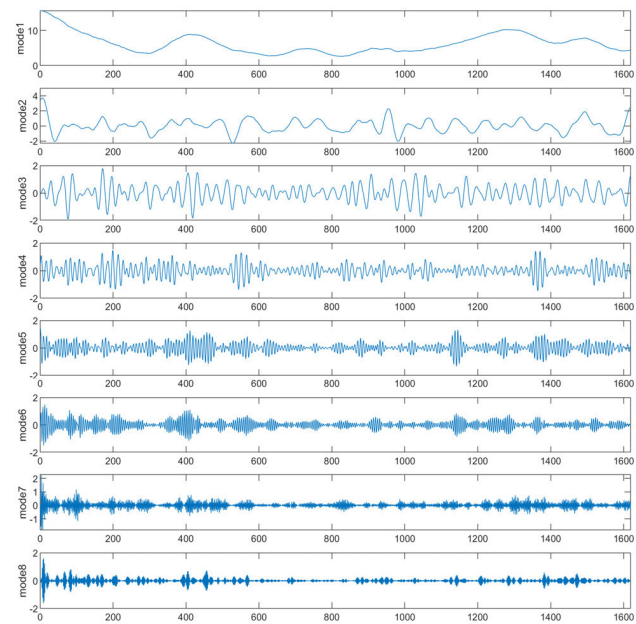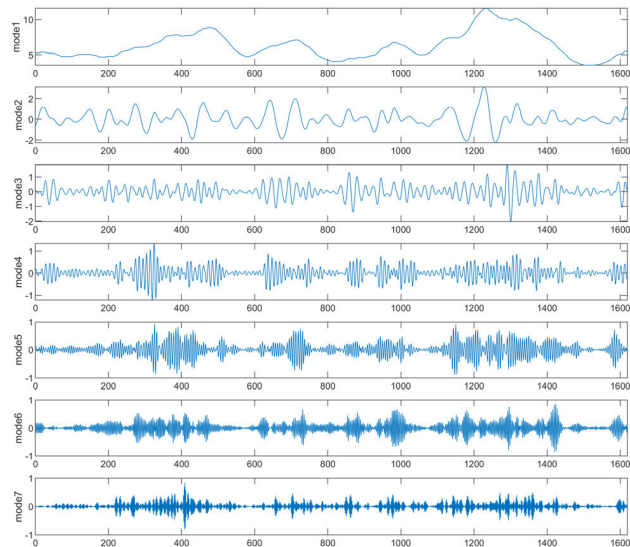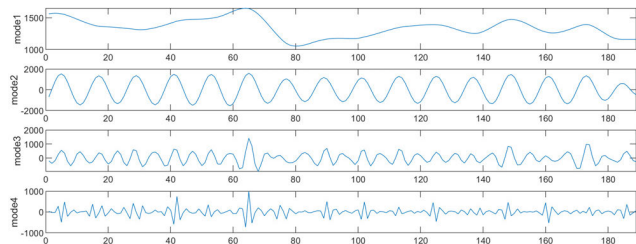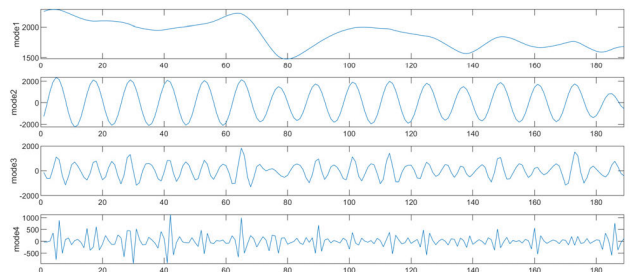
(a) for series A
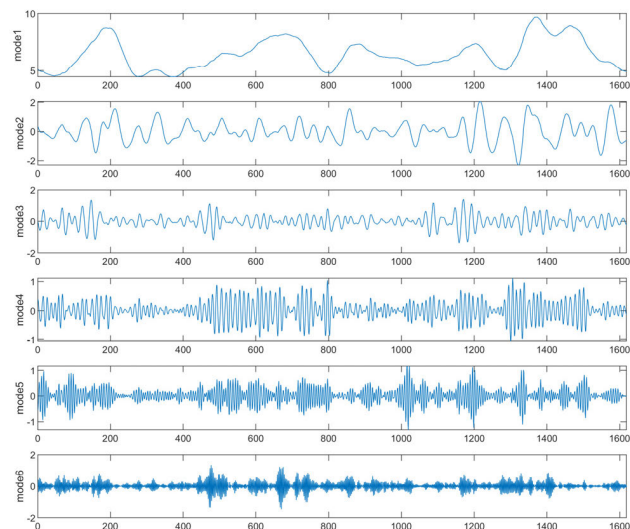
(b) for series B

(c) for series C

(d) for series D

(e) for series E

(f) for series F

FIGURE 3. Decomposition results in the testing period for all series using AVMD.

TABLE 2. Results for the proposed model and classic individual models in 1-step ahead forecasting.

Dataset	Indices	GRNN	RBF	BP	ELM	AVRM-GPR
A	RMSE	1.155	1.010	0.858	0.856	**0.238**
	MAE	0.876	0.766	0.626	0.627	**0.175**
	MAPE	0.195	0.199	0.132	0.141	**0.036**
B	RMSE	0.814	0.707	0.630	0.625	**0.238**
	MAE	0.662	0.531	0.457	0.457	**0.175**
	MAPE	0.114	0.085	0.073	0.073	**0.036**
C	RMSE	0.662	0.661	0.627	0.624	**0.270**
	MAE	0.494	0.479	0.441	0.443	**0.201**
	MAPE	0.096	0.090	0.082	0.083	**0.038**
D	RMSE	0.660	0.632	0.606	0.580	**0.21**
	MAE	0.462	0.456	0.416	0.404	**0.15**
	MAPE	0.073	0.074	0.064	0.063	**0.02**
E	RMSE	429	433	482	420	**50**
	MAE	228	268	337	233	**38**
	MAPE	0.135	0.204	0.333	0.150	**0.044**
F	RMSE	607	605	704	597	**69**
	MAE	343	364	485	339	**53**
	MAPE	0.160	0.189	0.306	0.155	**0.043**

all obtained from the developed AVRM-GPR model, with values of 0.238 m/s, 0.175 m/s, and 0.036, respectively. For Dataset E, the *RMSE*, *MAE*, and *MAPE* values are 50 m³/s, 38 m³/s, and 0.044, respectively, which are the best compared with the other four individual models.

Among all these four individual models, the priority of these forecasting model for the six Datasets is not identical, which indicate no perfect single model for energy series forecasting. Relatively, the *RMSE*, *MAE*, and *MAPE* values of ELM are nearly all smaller than those of the other individual models.

To further investigate the performance of the proposed model, these four neural networks have been also applied to do 2-3 step ahead forecasting for Datasets A∼F. The statistical error indices for 2-3 step ahead forecasting are presented in Figure 4 and Figure 5. The conclusions obtained from 2-3 step ahead are similar to those from 1-step ahead forecasting. It can be found that the AVRM-GPR model exhibit superior performance, while the best individual forecasting model for different Dataset is also different. While the ELM performs relative better than the other individual neural networks in most cases.

Overall, the forecasting ability of these four individual models for all Datasets are not identical, which emphasizes the individual model is not suitable in every situation. To resolve this issue, a new hybrid model which learns from strong points of many individual models should be proposed.

B. COMPARISON WITH MODELS EMPLOYING DIVERSE DDTS

The aim of this part is to testify the superiority of the combination of AVMD and RF as data pretreatment for
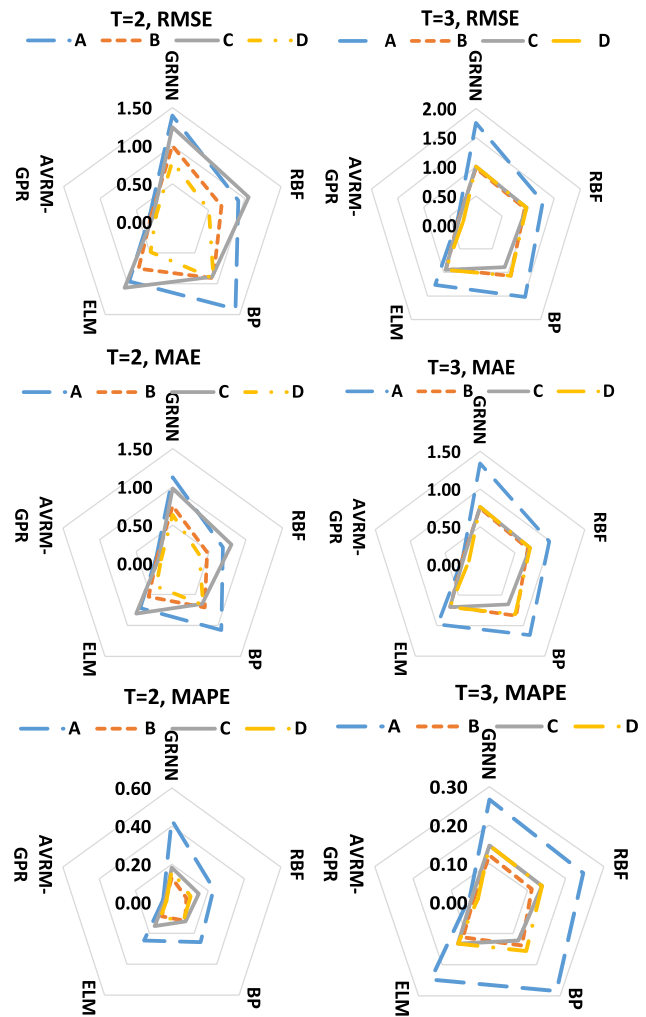

FIGURE 4. Forecasting performance of GRNN, RBF, BP, ELM, and AVRM-GPR in 2-3 step forecasting for Datasets A∼ D.

the proposed model AVRM-GPR. The relatively excellent single model ELM is used as the basic prediction module. The **AV**MD-**R**F-ELM (AVR-ELM) is compared with other three hybrid models with different DDTs and using PACF input determination method. Diverse DDTs includes EEMD, DWT, and AVMD. So, the three hybrid benchmark models are denoted as EEMD-PACF-ELM, DWT-PACF-ELM, and AVMD-PACF-ELM. Tables 3∼5 display the calculated statistical error values for 1-3 step forecasting results, respectively. Among these three benchmark models, AVMD-PACF-ELM always performs the best for all datasets and all lead times, followed by the EEMD-PACF-ELM, the DWT-PACF-ELM produces the worst results. For example, for Dataset A, the *RMSE* of AVMD-PACF-ELM, EEMD-PACF-ELM, and DWT-PACF-ELM for 1-step forecasting are 0.242 m/s, 0.363 m/s, and 0.493 m/s. For 2-step forecasting, the *RMSE* values are 0.317 m/s, 0.489 m/s, and 0.520 m/s. For 3-step forecasting, the *RMSE* values are 0.361 m/s, 0.550 m/s, and 0.633 m/s. These results indicate the AVMD is more suitable for preprocessing the original energy series than EEMD
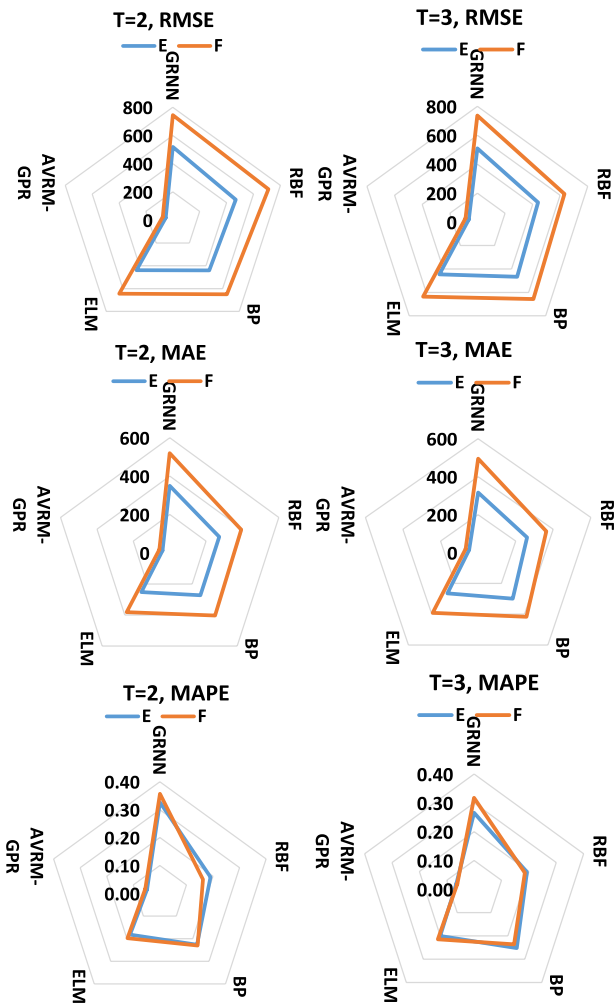
FIGURE 5. Forecasting performance of GRNN, RBF, BP, ELM, and AVRM-GPR in 2-3 step forecasting for Datasets E and F.

TABLE 3. Results for the combined models with diverse DDTs in 1-step ahead forecasting.

Dataset	indices	DWT-PACF-ELM	EEMD-PACF-ELM	AVMD-PACF-ELM	AVR-ELM
A	RMSE	0.493	0.363	0.242	**0.240**
	MAE	0.350	0.254	0.177	**0.176**
	MAPE	0.078	0.049	0.036	**0.036**
B	RMSE	0.342	0.330	0.314	**0.304**
	MAE	0.256	0.242	0.234	**0.227**
	MAPE	0.041	0.039	0.038	**0.037**
C	RMSE	0.353	0.305	0.284	**0.273**
	MAE	0.243	0.204	0.207	**0.201**
	MAPE	0.045	0.038	0.039	**0.038**
D	RMSE	0.348	0.328	0.313	**0.306**
	MAE	0.239	0.224	0.235	**0.225**
	MAPE	0.038	0.039	0.039	**0.036**
E	RMSE	86	81	53	**51**
	MAE	66	62	41	**40**
	MAPE	0.084	0.073	0.043	**0.044**
F	RMSE	128	128	74	**71**
	MAE	99	96	56	**55**
	MAPE	0.089	0.086	0.043	**0.043**

TABLE 4. Results for the combined models with diverse DDTs in 2-step ahead forecasting.

Dataset	indices	DWT-PACF-ELM	EEMD-PACF-ELM	AVMD-PACF-ELM	AVR-ELM
A	RMSE	0.520	0.489	0.317	**0.312**
	MAE	0.373	0.343	0.233	**0.229**
	MAPE	0.078	0.068	0.048	**0.047**
B	RMSE	0.342	0.364	0.311	**0.301**
	MAE	0.255	0.262	0.232	**0.224**
	MAPE	0.041	0.042	0.038	**0.036**
C	RMSE	0.393	0.373	0.275	**0.272**
	MAE	0.280	0.257	0.200	**0.198**
	MAPE	0.052	0.048	0.038	**0.037**
D	RMSE	0.453	0.439	0.386	**0.377**
	MAE	0.324	0.319	0.281	**0.279**
	MAPE	0.052	0.053	0.054	**0.045**
E	RMSE	223	216	67	**59**
	MAE	152	148	52	**44**
	MAPE	0.142	0.143	0.065	**0.053**
F	RMSE	300	286	94	**86**
	MAE	213	191	72	**67**
	MAPE	0.155	0.127	0.066	**0.061**

and DWT. Additionally, the AVR-ELM is always superior to the best hybrid benchmark model AVMD-PACF-ELM, with lower statistical error values. For Dataset A, the *RMSE* of AVR-ELM for 1-3 step forecasting are 0.24 m/s, 0.312 m/s, and 0.358 m/s. This indicates the combination of AVMD and RF can dig more information from the original series than the coupling of AVMD and PACF. Therefore, the combination of AVMD and RF is suitable as the preprocessing module for developing the energy series forecasting model.

C. COMPARISON WITH AVR-BASED MODELS
To further prove the effectiveness of our proposed combined model AVRM-GPR, it is compared with the other four AVR-based models. The statistical error indices including *RMSE*, *MAPE*, and *MAE* values of all five involved forecasting models for these six experimental tests are presented in Tables 6∼8, where the best value of each index is highlighted in boldface. It can be found that (1) for Dataset A, among the four AVR-based models, the performance of the AVR-ELM model is the best, followed by the AVR-BP

model, later the AVR-GRNN, that of the AVR-RBF is the worst for all three steps ahead wind speed forecasting. The *RMSE*, *MAE*, and *MAE* values of the AVR-ELM model are all the smallest. For instance, the *RMSE* of the AVR-ELM in 1-step ahead wind speed forecasting is 0.24 m/s while those of the AVR-GRNN, AVR-RBF, and AVR-BP are 0.51 m/s,

TABLE 5. Results for the combined models with diverse DDTs in 3-step ahead forecasting.

Dataset	indices	DWT-PACF-ELM	EEMD-PACF-ELM	AVMD-PACF-ELM	AVR-ELM
A	RMSE	0.633	0.550	0.361	**0.358**
	MAE	0.460	0.401	0.264	**0.263**
	MAPE	0.103	0.082	0.054	**0.053**
B	RMSE	0.379	0.406	0.359	**0.343**
	MAE	0.280	0.297	0.271	**0.258**
	MAPE	0.045	0.048	0.044	**0.042**
C	RMSE	0.439	0.431	0.324	**0.322**
	MAE	0.312	0.307	0.237	**0.237**
	MAPE	0.058	0.057	0.045	**0.044**
D	RMSE	0.581	0.592	0.520	**0.465**
	MAE	0.420	0.427	0.398	**0.346**
	MAPE	0.067	0.076	0.068	**0.055**
E	RMSE	252	266	78	**74**
	MAE	174	180	61	**57**
	MAPE	0.160	0.163	0.077	**0.073**
F	RMSE	356	354	108	**104**
	MAE	253	246	82	**79**
	MAPE	0.164	0.149	0.077	**0.074**

TABLE 6. Results for the proposed model and AVR-based models in 1-step ahead forecasting.

Dataset	indices	AVR-GRNN	AVR-RBF	AVR-BP	AVR-ELM	AVRM-GPR
A	RMSE	0.51	0.749	0.517	0.240	**0.238**
	MAE	0.397	0.583	0.394	0.176	**0.175**
	MAPE	0.085	0.167	0.082	0.036	**0.036**
B	RMSE	0.595	0.622	0.383	0.304	**0.291**
	MAE	0.473	0.495	0.291	0.227	**0.216**
	MAPE	0.073	0.082	0.047	0.037	**0.035**
C	RMSE	0.453	0.518	0.346	0.273	**0.27**
	MAE	0.357	0.392	0.258	0.201	**0.201**
	MAPE	0.071	0.074	0.049	0.038	**0.038**
D	RMSE	0.393	0.340	0.311	0.306	**0.212**
	MAE	0.316	0.257	0.227	0.225	**0.148**
	MAPE	0.057	0.044	0.036	0.036	**0.024**
E	RMSE	184	140	147	51	**50**
	MAE	140	112	114	40	**38**
	MAPE	0.159	0.135	0.136	0.044	**0.044**
F	RMSE	263	289	223	71	**69**
	MAE	193	204	172	55	**53**
	MAPE	0.152	0.141	0.139	0.043	**0.043**

TABLE 7. Results for the proposed model and AVR-based models in 2-step ahead forecasting.

Dataset	indices	AVR-GRNN	AVR-RBF	AVR-BP	AVR-ELM	AVRM-GPR
A	RMSE	0.704	0.842	0.506	0.312	**0.31**
	MAE	0.569	0.664	0.382	0.229	**0.227**
	MAPE	0.108	0.195	0.087	0.047	**0.047**
B	RMSE	0.603	0.631	0.418	0.301	**0.29**
	MAE	0.472	0.502	0.325	0.224	**0.217**
	MAPE	0.073	0.083	0.053	0.036	**0.035**
C	RMSE	0.51	0.518	0.46	0.272	**0.263**
	MAE	0.404	0.394	0.347	0.198	**0.193**
	MAPE	0.081	0.074	0.065	0.037	**0.037**
D	RMSE	0.493	0.405	0.380	0.377	**0.225**
	MAE	0.403	0.310	0.281	0.279	**0.160**
	MAPE	0.074	0.052	0.045	0.045	**0.026**
E	RMSE	213	150	177	59.195	**53**
	MAE	162	120	137	44.487	**39**
	MAPE	0.178	0.153	0.16	0.053	**0.048**
F	RMSE	281	279	250	85.620	**75**
	MAE	202	204	192	67.121	**57**
	MAPE	0.154	0.15	0.164	0.061	**0.055**

TABLE 8. Results for the proposed model and AVR-based models in 3-step ahead forecasting.

Dataset	indices	AVR-GRNN	AVR-RBF	AVR-BP	AVR-ELM	AVRM-GPR
A	RMSE	0.648	0.797	0.537	0.358	**0.347**
	MAE	0.501	0.628	0.415	0.263	**0.252**
	MAPE	0.134	0.171	0.086	0.053	**0.051**
B	RMSE	0.534	0.651	0.451	0.343	**0.333**
	MAE	0.419	0.521	0.348	0.258	**0.251**
	MAPE	0.067	0.085	0.057	0.042	**0.04**
C	RMSE	0.531	0.547	0.497	0.322	**0.321**
	MAE	0.402	0.415	0.381	0.237	**0.238**
	MAPE	0.073	0.078	0.072	0.044	**0.045**
D	RMSE	0.540	0.488	0.467	0.465	**0.241**
	MAE	0.429	0.374	0.348	0.346	**0.162**
	MAPE	0.075	0.062	0.055	0.055	**0.025**
E	RMSE	202	202	163	74	**62**
	MAE	152	152	130	57	**48**
	MAPE	0.158	0.17	0.158	0.073	**0.064**
F	RMSE	290	262	259	104	**86**
	MAE	204	200	194	79	**66**
	MAPE	0.135	0.165	0.164	0.074	**0.061**

0.749 m/s, and 0.517 m/s, respectively. The same conclusion can be found in the 2- and 3-step ahead wind speed forecasting results and streamflow forecasting results. (2) from the comparison of the AVR-ELM and the proposed model AVRM-GPR, it can be seen that the *RMSE*, *MAE*, and *MAE* of the AVRM-GPR model are nearly all less than those of the AVR-ELM model for all three horizons. (3) the proposed model is more valid than the other AVR-based model for all forecasting steps. To be more specific, the *MAPE* values of the developed model for Dataset A in three steps forecasting are 0.036, 0.047, and 0.051, respectively, which are lower than the corresponding values obtained from the other AVR-based combined models. These results indicate that the idea of series decomposition, dimensionality reduction and
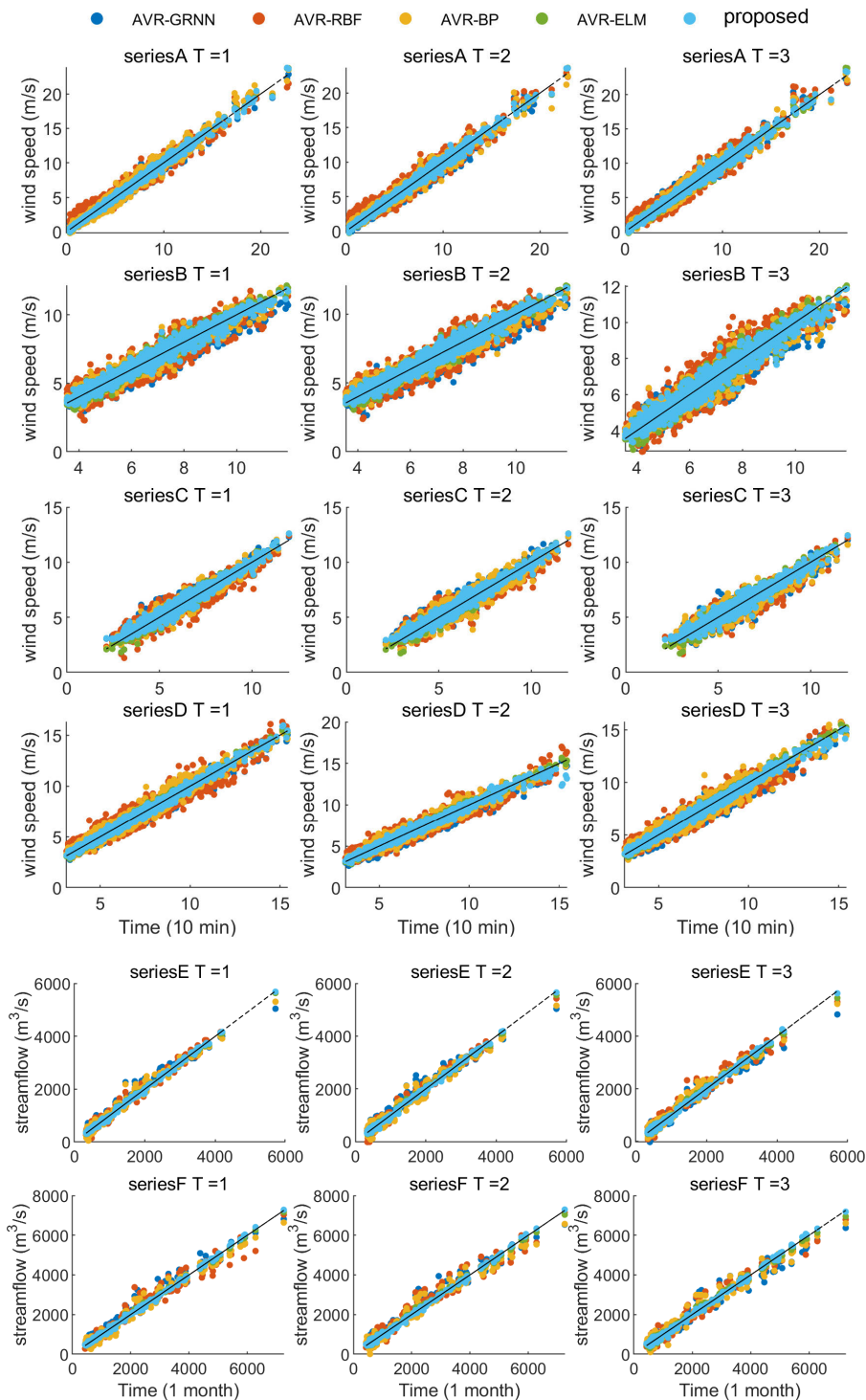
FIGURE 6. Scatters of the proposed AVRM-GPR and four AVR-based model for series A~F.

multiple AI techniques driven GPR can effectively promote the prediction performance of the forecasting model.

Additionally, the scatters of the predicated and the measured data for the proposed AVRM-GPR and four AVR-based models are plotted in Figure 6. These scatters can reflect the forecasting performance of the prediction model in high, middle, and low regions. Among the four benchmark models, the AVR-RBF always provides the worst results, especially in high and middle regions, the AVR-GRNN produce worse results for streamflow forecasting, the AVR-BP provides

over-forecasted results in middle region and under-predicted results in high region, while AVR-ELM always show the best performance. Compared with the best AVR-based model AVR-ELM, the scatters of the predicated data and the measured data for the AVRM-GPR are more close to the 45° line with a narrower coverage and evenly distributed on both sides of the 45° line, for all six experimental tests and all three lead times. Though the forecasting performance deteriorates with the extension of lead time, the deterioration of the proposed model is less than these of the other models. This indicates the superior merits of the AVRM-GPR.

The results of Series B (wind speed) and Series E (streamflow) are used to further illustrate the effective of the proposed model. Figures 7 and 8 exhibit the results of the benchmark models and the proposed AVRM-GPR models using Series B and Series E, where the blue line denotes the observed data, the green line is the worst benchmark model, the cyan line represents the best benchmark model, the magenta line means the deterministic results and the grey bounds are the uncertainty bounds of the AVRM-GPR. It can be seen from Figure 7 that the predicted data provided by the AVRM-GPR are closer to the measured data than the benchmark models, especially for relative high wind speed. Meanwhile, the observed data are entirely in the 50% predicted interval (PI) provided by the AVRM-GPR for all lead times. The same conclusion can be found from Figure 8.
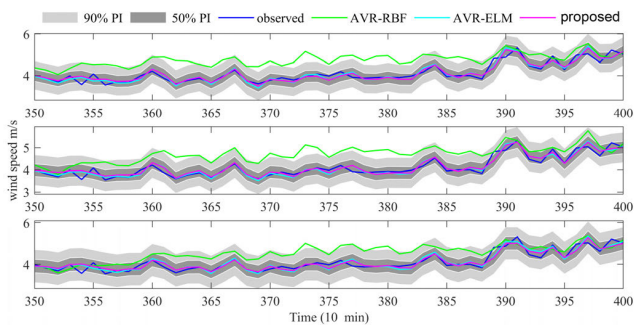

FIGURE 7. Forecasting results of Series B. (blue line: the observed data; green line: the worst results among all benchmark models; cyan line: the best results among all benchmark models; magenta line: the deterministic results of the AVRM-GPR, and the grey bounds are the uncertainty bounds provided by the AVRM-GPR).

VI. DISCUSSION
A. ASSOCIATION STRENGTH

Pearson's test can reveal the association strength between the actual data and the forecasting result. In this section, the discussion of association strength based on Pearson's correlation coefficients are calculated to further justify the excellent ability of the proposed AVRM-GPR compared with the other elven competitors. Generally, the correlation coefficient (CC) is equal to 1, indicating a significant linear relationship between the actual data and the forecasting data, whereas CC is equal to 0, representing no relationship between these two datasets. The average CC values of different lead times are
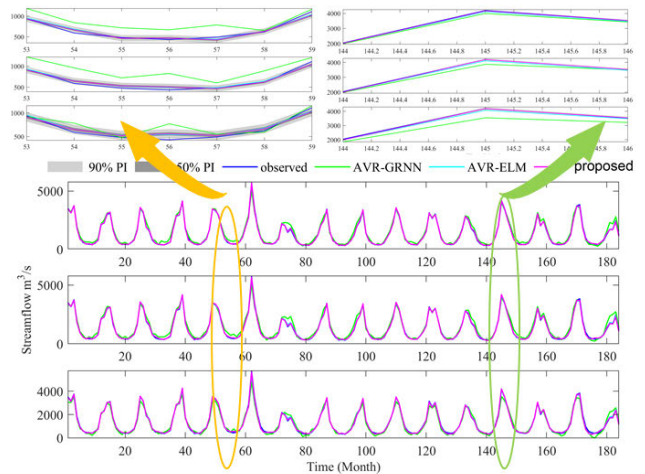

FIGURE 8. Forecasting results of Series E. (blue line: the observed data; green line: the worst results among all benchmark models; cyan line: the best results among all benchmark models; magenta line: the deterministic results of the AVRM-GPR, and the grey bounds are the uncertainty bounds provided by the AVRM-GPR).
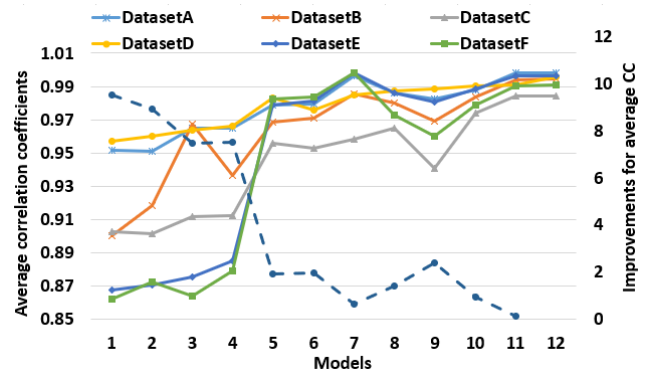

FIGURE 9. Comparison of correlation coefficient.

presented in Figure 9, where Model1 to Model12 represent GRNN, RBF, BP, ELM, DWT-PACF-ELM, EEMD-PACF-ELM, AVMD-PACF-ELM, AVR-GRNN, AVR-RBF, AVR-BP, AVR-ELM, and the proposed AVRM-GPR model. It can be found that the CC values of the proposed AVRM-GPR are significantly higher than these of individual models and hybrid models with different DDTs. The overall forecasting accuracy of the proposed AVRM-GPR model is no less than these of the AVR-based models and AVMD-ELM model. The average improvement percentages of average CC value from the GRNN, RBF, BP, ELM, DWT-PACF-ELM, EEMD-PACF-ELM, AVMD-PACF-ELM, AVR-GRNN, AVR-RBF, AVR-BP, and AVR-ELM on all cases of the testing data are 9.53%, 8.91%, 7.46%, 7.51%, 1.94%, 1.98%, 0.65%, 1.41%, 2.38%, 0.95%, and 0.11%, respectively.

B. IMPROVEMENTS

In this subsection, three indicators (P_{RMSE}, P_{MAE}, and P_{MAPE}) are adopted to discuss the accuracy and effectiveness of our developed model in detail. P_{RMSE}, P_{MAE}, and

TABLE 9. Improved percentages between compared models and the proposed model on all six datasets for all steps forecasting.

models	1-Step			2-Step		
	P_{RMSE}	P_{MAE}	P_{MAPE}	P_{RMSE}	P_{MAE}	P_{MAPE}
GRNN	76	75	71	80	81	83
RBF	74	73	71	72	71	71
BP	73	72	70	78	78	76
ELM	72	70	65	72	72	73
DWT-ELM	58	60	61	48	47	44
EEMD-ELM	59	61	62	47	45	41
AVMD-ELM	47	48	47	16	17	17
AVR-GRNN	39	38	38	60	62	61
AVR-RBF	32	29	29	58	60	61
AVR-BP	13	13	9	49	50	50
AVR-ELM	10	11	8	12	13	11

models	3-Step			Ave.		
	P_{RMSE}	P_{MAE}	P_{MAPE}	P_{RMSE}	P_{MAE}	P_{MAPE}
GRNN	78	78	76	78	78	76
RBF	75	74	71	74	73	71
BP	77	77	75	76	75	73
ELM	75	74	72	73	72	70
DWT-ELM	49	48	45	52	52	50
EEMD-ELM	49	47	44	52	51	49
AVMD-ELM	18	19	19	27	28	28
AVR-GRNN	53	55	53	51	52	51
AVR-RBF	56	58	58	49	49	49
AVR-BP	46	48	47	36	37	36
AVR-ELM	15	15	15	12	13	11

P_{MAPE} represent the improvement percentages of the indices *RMSE*, *MAE*, and *MAPE*, respectively. Table 9 presents the mean percentages of these three error improvements values compared with eleven models for three steps forecasting in all observed series. It is found that remarkable performance improvement of indices *RMSE*, *MAE*, and *MAPE* on all three horizons. The average improvement percentages of *MAPE* from GRNN, RBF, BP, ELM, DWT-PACF-ELM, EEMD-PACF-ELM, AVMD-PACF-ELM, AVR-GRNN, AVR-RBF, AVR-BP, and AVR-ELM on all cases of the testing data are 76%, 71%, 73%, 70%, 50%, 49%, 28%, 51%, 49%, 36%, and 11%, respectively.

To sum up, we can safely and reasonably conclude that: (a) the proposed forecasting model is insensitive to the collected data, time interval, and the number of the training data due to its excellent forecasting effectiveness; (b) the proposed AVRM-GPR achieves extremely competitive results and provide more information on uncertainty, so it is a valuable alternative for stochastic, nonlinear, and non-stationary energy series forecasting problems.

C. LIMITATIONS

Although the proposed AVRM-GPR shows a good performance in the wind speed and streamflow forecasting, there remain aspects of this model that need to be improved, which can be summarized as follows.

(1) The proposed model is just verified in the field of energy systems, including wind farms and water resources system, other unstable series like load and price have not been tested.

(2) The proposed model is focused mainly on short-term wind speed forecasting. More effort can be invested in long-term wind speed prediction to further improve the efficiency of operation and scheduling in a wind power system.

(3) In the proposed model, the GPR was adopted to merge advantages of many neural networks and quantify the forecasting uncertainty. However, thus far no perfect theory exists that can help effectively determine the kernel function when using GPR. In the current study, the kernel function in GPR was determined by empirical studies. Accordingly, the effective determination of the appropriate kernel function when developing energy series forecasting models should be investigated in future.

(4) Additionally, further work also can consider several other impact factors, such as humidity, air pressure, as inputs to enhance the energy series forecasting accuracy.

VII. CONCLUSION

As the serious environmental pollution, climate change and energy crisis, clean energies, such as wind and water resources, have attracted more attentions. One of the means to promote the efficient and rational utilization of such energy is forecasting. However, the energy series always is non-stationary and highly non-linear, these intrinsic properties make their accurate forecasting very challengeable. Merging merits of data decomposition method, feature selection, and multiple AI techniques, we develop a new composite uncertainty model. In the developed model, the AVMD (adaptive variational mode decomposition) is first applied to excavate implicit information from the original time series. Then, the RF (random forest) is utilized to select the suitable inputs for each mode. After that, the forecasting results of multiple neural networks are taken as inputs to drive the GPR model (called AVRM-GPR) for unstable energy time series forecasting, such as wind speed and streamflow. Eleven models in three categories, *i.e.*, the individual models, hybrid models with different DDTs, and AVR-based models, are compared with the proposed model. Experimental results on 1-3 step forecasting indicate the AVRM-GPR can provide more accurate results than any involved forecasting benchmark models. Additionally, the proposed AVRM-GPR can provide both deterministic and probabilistic forecasting results. The obtained uncertainty information can benefit many relevant management departments in the allocation of wind/water energy, the integration of wind energy into the power system, and the safe of the power system. By comprehensively comparing the results of one to three horizons,

the average P_{MAPE} of the ELM, AVMD-PACF-ELM, and AVR-ELM on all tested cases are 70%, 28%, and 11%, respectively. The results also indicate the positive influence of time series decomposition method, and feature selection imposing on the developed model. Therefore, by reasonably integrating time series decomposition method, feature selection, and designing suitable composite strategy, the performance of the unstable energy series forecasting model could be improved.

REFERENCES

[1] C. Tian, Y. Hao, and J. Hu, "A novel wind speed forecasting system based on hybrid data preprocessing and multi-objective optimization," *Appl. Energy*, vol. 231, pp. 301–319, Dec. 2018.

[2] N. Sun, J. Zhou, G. Liu, and Z. He, "A hybrid wind speed forecasting model based on a decomposition method and an improved regularized extreme learning machine," *Energy Procedia*, vol. 158, pp. 217–222, Feb. 2019.

[3] R. Noori, A. R. Karbassi, A. Moghaddamnia, D. Han, M. H. Zokaei-Ashtiani, A. Farokhnia, and M. G. Gousheh, "Assessment of input variables determination on the SVM model performance using PCA, gamma test, and forward selection techniques for monthly stream flow prediction," *J. Hydrol.*, vol. 401, nos. 3–4, pp. 177–189, May 2011.

[4] H. Tyralis, G. Papacharalampous, and A. Langousis, "A brief review of random forests for water scientists and practitioners and their recent history in water resources," *Water*, vol. 11, no. 5, p. 910, Apr. 2019.

[5] Y. R. Fan, W. Huang, G. H. Huang, Z. Li, Y. P. Li, X. Q. Wang, G. H. Cheng, and L. Jin, "A stepwise-cluster forecasting approach for monthly streamflows based on climate teleconnections," *Stochastic Environ. Res. Risk Assessment*, vol. 29, no. 6, pp. 1557–1569, Mar. 2015.

[6] H. Liu, H. Wu, and Y. Li, "Smart wind speed forecasting using EWT decomposition, GWO evolutionary optimization, RELM learning and IEWT reconstruction," *Energy Convers. Manage.*, vol. 161, pp. 266–283, Apr. 2018.

[7] J. Zhou, N. Sun, B. Jia, and P. Tian, "A novel decomposition-optimization model for short-term wind speed forecasting," *Energies*, vol. 11, no. 7, pp. 1752–1779, Jul. 2018.

[8] W. Fu, K. Wang, J. Tan, and K. Zhang, "A composite framework coupling multiple feature selection, compound prediction models and novel hybrid swarm optimizer-based synchronization optimization strategy for multi-step ahead short-term wind speed forecasting," *Energy Convers. Manage.*, vol. 205, Feb. 2020, Art. no. 112461.

[9] X. Zhao, X. Chen, Y. Xu, D. Xi, Y. Zhang, and X. Zheng, "An EMD-based chaotic least squares support vector machine hybrid model for annual runoff forecasting," *Water*, vol. 9, no. 3, p. 153, Feb. 2017.

[10] Q. F. Tan, X. H. Lei, X. Wang, H. Wang, X. Wen, Y. Ji, and A. Q. Kang, "An adaptive middle and long-term runoff forecast model using EEMD-ANN hybrid approach," *J. Hydrol.*, vol. 567, pp. 776–780, Dec. 2018.

[11] W. Fu, K. Zhang, K. Wang, B. Wen, P. Fang, and F. Zou, "A hybrid approach for multi-step wind speed forecasting based on two-layer decomposition, improved hybrid DE-HHO optimization and KELM," *Renew. Energy*, vol. 164, pp. 211–229, Feb. 2021.

[12] Z. Wu and N. E. Huang, "Ensemble empirical mode decomposition: A noise-assisted data analysis method," *Adv. Adapt. Data Anal.*, vol. 1, no. 1, pp. 1–41, Jan. 2009.

[13] K. Dragomiretskiy and D. Zosso, "Variational mode decomposition," *IEEE Trans. Signal Process.*, vol. 62, no. 3, pp. 531–544, Feb. 2014.

[14] N. Sun, J. Zhou, L. Chen, B. Jia, M. Tayyab, and T. Peng, "An adaptive dynamic short-term wind speed forecasting model using secondary decomposition and an improved regularized extreme learning machine," *Energy*, vol. 165, pp. 939–957, Dec. 2018.

[15] L. Chen, N. Sun, C. Zhou, J. Zhou, Y. Zhou, J. Zhang, and Q. Zhou, "Flood forecasting based on an improved extreme learning machine model combined with the backtracking search optimization algorithm," *Water*, vol. 10, no. 10, pp. 1362–1379, Sep. 2018.

[16] Y. Zhang, G. Pan, C. Zhang, and Y. Zhao, "Wind speed prediction research with EMD-BP based on lorenz disturbance," *J. Electr. Eng.*, vol. 70, no. 3, pp. 198–207, Jun. 2019.

[17] Y. Zhang, Y. Zhao, C. Kong, and B. Chen, "A new prediction method based on VMD-PRBF-ARMA-E model considering wind speed characteristic," *Energy Convers. Manage.*, vol. 203, Jan. 2020, Art. no. 112254.

[18] Z. Liu, P. Jiang, L. Zhang, and X. Niu, "A combined forecasting model for time series: Application to short-term wind speed forecasting," *Appl. Energy*, vol. 259, Feb. 2020, Art. no. 114137.

[19] Z. M. Yaseen, W. H. M. W. Mohtar, A. M. S. Ameen, I. Ebtehaj, S. F. M. Razali, H. Bonakdari, S. Q. Salih, N. Al-Ansari, and S. Shahid, "Implementation of univariate paradigm for streamflow simulation using hybrid data-driven model: Case study in tropical region," *IEEE Access*, vol. 7, pp. 74471–74481, 2019.

[20] K.-P. Lin, P.-F. Pai, and Y.-J. Ting, "Deep belief networks with genetic algorithms in forecasting wind speed," *IEEE Access*, vol. 7, pp. 99244–99253, 2019.

[21] C.-J. Huang and P.-H. Kuo, "Multiple-input deep convolutional neural network model for short-term photovoltaic power forecasting," *IEEE Access*, vol. 7, pp. 74822–74834, 2019.

[22] K. Wang, W. Fu, T. Chen, B. Zhang, D. Xiong, and P. Fang, "A compound framework for wind speed forecasting based on comprehensive feature selection, quantile regression incorporated into convolutional simplified long short-term memory network and residual error correction," *Energy Convers. Manage.*, vol. 222, Oct. 2020, Art. no. 113234.

[23] C.-J. Huang and P.-H. Kuo, "A deep CNN-LSTM model for particulate matter (PM2.5) forecasting in smart cities," *Sensors*, vol. 18, no. 7, p. 2220, Jul. 2018.

[24] H. Liu, Z. Duan, F.-Z. Han, and Y.-F. Li, "Big multi-step wind speed forecasting model based on secondary decomposition, ensemble method and error correction algorithm," *Energy Convers. Manage.*, vol. 156, pp. 525–541, Jan. 2018.

[25] Y. Zhang, Y. Zhao, and S. Gao, "A novel hybrid model for wind speed prediction based on VMD and neural network considering atmospheric uncertainties," *IEEE Access*, vol. 7, pp. 60322–60332, 2019.

[26] Y. Zhang, S. Gao, J. Han, and M. Ban, "Wind speed prediction research considering wind speed ramp and residual distribution," *IEEE Access*, vol. 7, pp. 131873–131887, 2019.

[27] C. Huang, Y. Shen, Y. Chen, and H. Chen, "A novel hybrid deep neural network model for short-term electricity price forecasting," *Int. J. Energy Res.*, vol. 44, pp. 1–15, Sep. 2020, doi: 10.1002/er.5945.

[28] J. Naik, P. K. Dash, and S. Dhar, "A multi-objective wind speed and wind power prediction interval forecasting using variational modes decomposition based multi-kernel robust ridge regression," *Renew. Energy*, vol. 136, pp. 701–731, Jun. 2019.

[29] Y. He and H. Li, "Probability density forecasting of wind power using quantile regression neural network and kernel density estimation," *Energy Convers. Manage.*, vol. 164, pp. 374–384, May 2018.

[30] L. Ye, J. Zhou, X. Zeng, J. Guo, and X. Zhang, "Multi-objective optimization for construction of prediction interval of hydrological models based on ensemble simulations," *J. Hydrol.*, vol. 519, pp. 925–933, Nov. 2014.

[31] Y. Zhang, K. Liu, L. Qin, and X. An, "Deterministic and probabilistic interval prediction for short-term wind power generation based on variational mode decomposition and machine learning methods," *Energy Convers. Manage.*, vol. 112, pp. 208–219, Mar. 2016.

[32] C. Zhang, H. Wei, X. Zhao, T. Liu, and K. Zhang, "A Gaussian process regression based hybrid approach for short-term wind speed prediction," *Energy Convers. Manage.*, vol. 126, pp. 1084–1092, Oct. 2016.

[33] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[34] P.-S. Yu, T.-C. Yang, S.-Y. Chen, C.-M. Kuo, and H.-W. Tseng, "Comparison of random forests and support vector machine for real-time radar-derived rainfall forecasting," *J. Hydrol.*, vol. 552, pp. 92–104, Sep. 2017.

[35] C. Voyant, G. Notton, S. Kalogirou, M.-L. Nivet, C. Paoli, F. Motte, and A. Fouilloy, "Machine learning methods for solar radiation forecasting: A review," *Renew. Energy*, vol. 105, pp. 569–582, May 2017.

[36] G. Chen, T. Long, J. Xiong, and Y. Bai, "Multiple random forests modelling for urban water consumption forecasting," *Water Resour. Manag.*, vol. 31, no. 15, pp. 1–15, Sep. 2017.

[37] P. Thanh Noi and M. Kappas, "Comparison of random forest, k-nearest neighbor, and support vector machine classifiers for land cover classification using sentinel-2 imagery," *Sensors*, vol. 18, no. 1, pp. 18–38, Dec. 2017.

[38] W. Fang, S. Huang, Q. Huang, G. Huang, E. Meng, and J. Luan, "Reference evapotranspiration forecasting based on local meteorological and global climate information screened by partial mutual information," *J. Hydrol.*, vol. 561, pp. 764–779, Jun. 2018.

[39] A. Lahouar and J. Ben Hadj Slama, "Hour-ahead wind power forecast based on random forests," *Renew. Energy*, vol. 109, pp. 529–541, Aug. 2017.

[40] J. Vanhatalo, J. Riihimki, J. Hartikainen, P. Jylnki, and A. Vehtari, "GPstuff: Bayesian modeling with Gaussian processes," *J. Mach. Learn. Res.*, vol. 14, no. 1, pp. 1175–1179, Apr. 2013.

[41] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*, vol. 2, no. 3. Cambridge, MA, USA: MIT Press, 2006.

[42] S. Zhu, X. Yuan, Z. Xu, X. Luo, and H. Zhang, "Gaussian mixture model coupled recurrent neural networks for wind speed interval forecast," *Energy Convers. Manage.*, vol. 198, Oct. 2019, Art. no. 111772.

[43] Y. Su, G. Huang, and Y.-L. Xu, "Derivation of time-varying mean for non-stationary downburst winds," *J. Wind Eng. Ind. Aerodynamics*, vol. 141, pp. 39–48, Jun. 2015.

[44] T. Peng, J. Zhou, C. Zhang, and Y. Zheng, "Multi-step ahead wind speed forecasting using a hybrid model based on two-stage decomposition technique and AdaBoost-extreme learning machine," *Energy Convers. Manage.*, vol. 153, pp. 589–602, Dec. 2017.

TIAN PENG received the Ph.D. degree in hydropower and information engineering from the Huazhong University of Science and Technology, Wuhan, China, in 2019. She is currently an Associate Professor with the College of Automation, Huaiyin Institute of Technology. Her research interests include deep learning, artificial intelligence, and their application in intelligent prediction of clean energies.

NA SUN received the Ph.D. degree in hydropower and information engineering from the Huazhong University of Science and Technology, Wuhan, China, in 2019. She is currently an Associate Professor with the College of Automation, Huaiyin Institute of Technology. Her research interests include deep learning, artificial intelligence, and their application in intelligent prediction of clean energies.

JIANZHONG ZHOU was born in Wuhan, China, in December 1959. He received the B.S. degree in automatic control from the Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 1982. He is currently a Professor with the School of Hydropower and Information Engineering, Huazhong University of Science and Technology (HUST), Wuhan. His research interests include optimal operation and control of hydropower energy systems.

SHUAI ZHANG received the master's degree from the Xi'an University of Technology, Shaanxi, China, in 2014. He is currently pursuing the Ph.D. degree with the School of Energy and Power Engineering, Xi'an Jiaotong University, Shaanxi, China. His research interests include research and development of general CFD software and application of machine learning in saving energy.

XINGUO SUN received the Ph.D. degree from the Faculty of Infrastructure Engineering, Dalian University of Technology, Dalian, China, in 2016. He is currently an Associate Professor with the Department of Economics and Management, Huaiyin Institute of Technology. His research interests include water resources management and hydrological modeling.

. . .