IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received May 29, 2019, accepted June 9, 2019, date of publication June 13, 2019, date of current version July 3, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2922662

Short-Term Wind Speed Forecast With Low Loss of Information Based on Feature Generation of OSVD

NANTIAN HUANG[1], YINYIN WU[1], GUOWEI CAI[1], HEYAN ZHU[2], CHANGYONG YU[2], LI JIANG[2], YE ZHANG[3], JIANSEN ZHANG[4], AND ENKAI XING[1]

[1]Key Laboratory of Modern Power System Simulation and Control and Renewable Energy Technology, Ministry of Education, Northeast Electric Power University, Jilin 132012, China
[2]State Grid Liaoning Economic Research Institute, State Grid Liaoning Electric Power Co., Ltd., Shenyang 110000, China
[3]Liaoning Power Trading Center Company, State Grid Liaoning Power Co., Ltd., Shenyang 110000, China
[4]Shenyang Power Supply Company, State Grid Liaoning Electric Power Co., Ltd., Shenyang 110003, China

Corresponding author: Nantian Huang (huangnantian@126.com)

This work was supported in part by the National Key Research and Development Program of China under Grant 2016YFB0900104, and in part by the Research and Development of Industrial Technology in Jilin Province under Grant 2019C058-8, Specialized in Industrial Technology Development of Jilin Province.

ABSTRACT Improving the accuracy of wind speed forecast can reduce the randomness and uncertainty of the wind power output and effectively improve a system's wind power accommodation. However, the high-dimensional historical wind speed information should be taken into account in the wind speed forecast, which increases the complexity of the model and reduces the efficiency and accuracy of a forecast. Feature selection by the Filter method can effectively reduce the feature dimension, but losing all the information of low-importance features. Although the feature reduction can retain the partial information of all features, it causes the loss of the partial information of high-importance features. In order to reduce the information loss caused by traditional FS and FR, short-term wind speed forecast with low information loss based on OSVD feature generation is proposed. First, the original wind speed series is denoised by OVMD. Then, based on the 96-dimensional original wind speed feature set, the OSVD is used to generate features. Furthermore, the extended original feature set EFS is obtained by combining the initial feature set with the features generated by OSVD. Gini importance is used to measure the importance of all features in EFS, and the forward feature selection is combined with random forests to determine the optimal subset. Finally, the optimal model determined by the new method is compared with seven models to verify the advancement of the new method. The experiments show that it reduces the information loss. Thus, the model has a higher forecast accuracy than the traditional model.

INDEX TERMS Short-term wind speed forecast, feature selection, feature reduction, low information loss, variational mode decomposition, singular value decomposition.

I. INTRODUCTION

With the deterioration of the environment and the excessive use of nonrenewable energy, countries all over the world have begun to pay attention to the development of renewable energy [1]. As a relatively mature technology for renewable energy generation, wind power plays an irreplaceable role in the new energy market [2], [3]. However, the strong randomness, intermittence and uncontrollability of wind speed lead to considerable fluctuation of wind power outputs, which makes the safe and reliable operation of a power grid enormously challenging and restricts the development of wind power. Accurate and efficient wind speed forecast (WSF) can reduce the negative impact of wind power uncertainty [4], [5]. WSF methods can generally be divided into physical methods [6], [7], statistical methods [8]–[11], artificial intelligence (AI) methods [12]–[17], and others.

The physical method is based on a numerical weather prediction (NWP) model, which uses a series of meteorological data (wind direction, temperature, and humidity) and terrain information to establish WSFM [6], [7]. Unlike physical methods, statistical methods only use historical

The associate editor coordinating the review of this manuscript and approving it for publication was Shuaihu Li.

AI	artificial intelligience	LDA	linear discriminant analysis
AR	autoregressive model	MA	moving average
ARMA	autoregressive moving average	NWP	numerical weather prediction
ARIMA	autoregressive integrated moving average	OFS	original feature set
ANN	artificial neural networks	OFSS	optimal feature subset
BP-ANN	back propagation-artificial neural networks	OSVD	optimal singular value decomposition
CEEMD	complete ensemble empirical mode decomposition	OVMD	optimal variational mode decomposition
CEMD	complementary empirical mode decomposition	PCA	principal component analysis
EEMD	ensemble empirical mode decomposition	PSO	particle swarm optimization
EFS	extended feature set	RBFN	radial basis function network
ELM	extreme learning machines	RF	random forest
EMD	empirical mode decomposition	RFS	reducted feature set
FAHP	fuzzy analytic hierarchy process	SVD	singular value decomposition
FR	feature reduction	SVM	support vector machines
FS	feature selection	TSD	time series decomposition
GA-BP	Genetic algorithm-back propagation	VMD	variational mode decomposition
GSO	Gram-Schmidt orthogonalization	WT	wavelet transform
IMF	intrinsic mode function	WSF	wind speed forecast
IFS	Initial feature set	WSFM	wind speed forecast model
LA-SVM	least square support vector machine		

wind speed data to forecast. A large number of statistical methods have also been proposed to predict the wind, such as autoregressive (AR) [8], autoregressive moving average (ARMA) [9], and autoregressive integrated moving average (ARIMA) [10], Markov model [11], and Bayesian method [12]. In [8], the AR is combined with the average filter of the wind speed waveform to forecast wind speed forecast. In [9], using the ARMA to forecast lateral and longitudinal components which are the decomposition of the wind speed, and the results are combined to obtain the wind direction and speed forecasts. In [10], fractional-ARIMA models forecast wind speeds on the day-ahead and two-day-ahead horizons in North Dakota. Generally, artificial intelligence methods include artificial neural networks (ANN) [13], support vector machines (SVM) [14], [15], and extreme learning machines (ELM) [16], [17], random forest [18]. In [14], Using V-SVM which parameters are optimized using the Cuckoo search algorithm predict wind speed and the Grey correlation analysis is used to determine the input set. In [17], the stacked ELM is used to predict wind speed which is an advanced ELM algorithm under deep learning framework. In [18], the random forest is used as supervised forecasting model and a data driven dimension reduction procedure and a weighted voting method are utilized to optimize the random forest algorithm in the training process and the prediction process.

However, physical method performs poorly in short-term WSF with strong volatility and high accuracy requirements [15], [19]. The statistical methods have drawbacks in solving the problem of nonlinear and nonstationary wind speed [16]. The artificial intelligence method is suitable for nonlinear and nonstationary WSF and has high accuracy in short-term WSF. However, in the commonly used artificial intelligence method, ANN requires large amounts of training data and a long time to optimize parameters and can easily fall into local optimum [15]. SVM is sensitive to the selection of parameters and kernel functions, consumes huge time and space and is easy to over-fit [20]. ELM has high forecast efficiency, but its robustness and stability are poor [16]. With high accuracy and anti-noise ability, random forest (RF) is not prone to overfitting, is suitable for analyzing high-dimensional data, which do not need to be standardized datasets, and has fast training speed. It can also analyze the importance of feature during training [21]–[23].

Apart from these basic categories, preprocessing methods such as time-series decomposition (TSD), feature selection (FS) and feature reduction (FR), also play important roles in wind speed forecasting [24]. To reduce the randomness of wind speed time series, TSD methods, such as empirical mode decomposition (EMD) [21], wavelet transform (WT) [25]–[27], variational mode decomposition (VMD) [28], have been widely applied for WSF, before the data sets are constructed and input into the basic predicting model. However, WT is heavily influenced by the choice of wavelet basis functions and the level of decomposition [25]. EMD has modal aliasing and endpoint effects [16]. VMD effectively avoids the modal aliasing problem of EMD and its improved methods. It has better noise immunity, better performance and higher computational efficiency [28]. The decomposition modal number K and the update parameter of the VMD have considerable influence on the decomposition result. Considering the characteristics of the above methods, this study chooses OVMD to preprocess wind speed data, which optimizes two parameters of the VMD.The FS and FR methods are manifested to help in selecting the best inputs of the forecasting model, which would affect the model accuracy and efficiency greatly [24]. FS methods are generally divided into the wrapper method

and the filter method [29]. When there are many features, the wrapper method needs to rely on other learning algorithms. The number of calculations of the wrapper method is high and its practicability is low [30], [31]. The filter method is widely valued because it does not need to rely on other learning algorithms for feature selection and can avoid overfitting, for which the computational cost is small [29]. The FR methods mainly include principal component analysis (PCA) [31], linear discriminant analysis (LDA) [32] and singular value decomposition (SVD) [33]. Among these methods, the PCA have been influenced by artificial setting parameters, and the uniqueness of the orthogonal vector space of its eigenvalue matrix remains to be discussed [33]. The LDA method requires data to be tagged, which seeks to make the data points as easy to distinguish as possible after dimension reduction and is more suitable for data classification [34].

To attain better performance, researchers have concentrated on hybrid models such as PSO-ELM [15], PSO-SVM [19], EEMD-GA-BP [23], GSO-ELM [24], PCA-ANN [26], VMD-PACF-ELM [35], EMD-PACF-SVM [36]. In these hybrid models, the integration of TSD could improve the forecasting performance by overcoming the barriers of non-linearity and non-stationary of wind speed, the FS methods help to eliminate redundant information in feature sets and the FR methods help to reduce reduces the number of features, which can improve the efficiency and accuracy of model.

From the review of related works, we found that although complicated hybrid model have been studied, two problems still remain. Firstly, the FS and FR methods have different types of information loss. Specifically, the filter method in the FS determines an optimal subset by forward feature selection to construct a model. The optimal subset contains all features of high-importance and some features of low-importance, while discarding some low-importance features. This approach causes the input set to lose all of the information of the low-importance of the original feature set. However, the FR method maps high-dimensional data to a low-dimensional space to deducted the feature dimension of the input set, and obtains the reduced feature set. The deducted feature set contains part of the information of all the features in the original feature set, but also loses part of the information of the high-importance feature. Secondly, predictors with TSD and FS are all affected by their control parameters. For example, the ω of mode will aggregate or even overlap if K is too large, some modes will be divided into adjacent modes, or even discarded if K is too small [34], [37].

Inspired by these ideas, this paper proposes a short-term wind speed forecast method with low information loss, which reduces the information loss caused by traditional FR-FM, FS-FM or TSD-FS-FM, and improves the forecasting accuracy. At the same time, different optimization algorithms is used for the optimization of the essential parameters of those sub-models separately. First, the original wind speed sequence is decomposed into several IMFs using the OVMD method. The IMF with the smallest amplitude is eliminated and the remaining IMF is reconstructed into a new wind speed sequence. Second, the OSVD is used to reduce the dimension of the original feature set (OFS), and the number of features after dimension reduction is determined according to the best contribution rate. Thus, a reduction feature set (RFS) is obtained. Then, in order to further improve the forecast accuracy, RFS and OFS are combined to form an extended feature set (EFS). The importance of features in EFS is calculated and ranked by RF. Finally, the feature selection of forward Filter method is carried out on the EFS to determine the optimal subset according to the ranking of importance.

The structure of this paper is as follows: Chapter 2 introduces the theoretical background of OVMD, OSVD and RF. Chapter 3 describes the short-term WSFM proposed in this paper. Chapter 4 introduces the experiment and evaluates the proposed model. The last chapter summarizes this paper.

II. THERETICAL BACKGROUNDS

A. OVMD

1) BASIC PRINCIPLES OF VMD

VMD was first proposed by Dragomiretskiy and Zosso in 2014 [35]. The decomposition process of VMD consists of two parts: construction and solution. It involves three important concepts: classical Wiener filtering, Hilbert transform and frequency mixing.

(1) Construction of Variational Problems

The variational problem is how to decompose the original signal f into k modal functions $U_k(t)$ (subsequences). It is assumed that the finite bandwidth of each subsequence has a central frequency ω_k. The estimated bandwidth of each mode is minimized. The constraint condition is that the sum of the modal functions is equal to the original signal f.

(a) Analytical signals of each modal function $U_k(t)$ are obtained by Hilbert transform.

(b) The center frequency ω_k. is estimated by mixing the analytic signals of each mode, and the spectrum of each mode is moved to the fundamental frequency band.

(c) The bandwidth of each mode signal is estimated by H-Gauss smoothing of demodulated signal, which is the square of two norms of gradient.

Therefore, the constrained variational problem is as follows:

$$\begin{aligned} &\min_{U_{k,\omega k}} \quad \left\{ \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * U_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\} \\ &s.t. \quad \sum_k U_k = f \end{aligned} \tag{1}$$

Here, ∂_t represents partial derivative of t and $\delta(t)$ represents impulse function.

(2) Solution of Variational Problems

By introducing Lagrange multiplier gamma $\gamma(t)$ and quadratic penalty factor α, the augmented Lagrange function

of the equation (1) is obtained.

$$L(\{Uk\},\{\omega k\},\gamma) = \alpha\sum_k \left\| \partial t\left[\left(\delta(t)+\frac{j}{\pi t}\right)\times Uk(t)\right]e^{j\omega kt}\right\|_2^2 + \left\|f-\sum_k Uk(t)\right\|_2^2 + \left[\gamma(t), f-\sum_k Uk(t)\right] \quad (2)$$

Alternate Direction Method of Multipliers (ADMM) based on dual decomposition and the Lagrange method is used to solve equation (2) for alternating iteration optimization of U_K, ω_k and γ.

$$\hat{U}_k^{n+1}(\omega) = \frac{\hat{f}(\omega)-\sum_{i\neq k}\hat{U}_i(\omega)+\frac{\hat{\gamma}(\omega)}{2}}{1+2\alpha(\omega-\omega_k)^2} \quad (3)$$

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega\left|\hat{U}_k(\omega)\right|^2 d\omega}{\int_0^\infty \left|\hat{U}_k(\omega)\right|^2 d\omega} \quad (4)$$

$$\hat{\gamma}^{n+1}(\omega) = \hat{\gamma}^{n}(\omega)+\tau\left(\hat{f}(\omega)+\hat{U}_k^{n+1}(\omega)\right) \quad (5)$$

Here, $\hat{U}_i(\omega)$, $\hat{U}_K^{n+1}$, $\hat{f}(\omega)$, $\hat{\gamma}(\omega)$ represents the Fourier transform of $U_i(\omega)$, U_K^n, $f(\omega)$, $\gamma(\omega)$, and n represents the number of iterations.

For a given solution accuracy, the iteration stops when the equation (6) is satisfied:

$$\sum_k \left\|U_k^{n+1}-U_k^n\right\|_2^2 < \varepsilon \quad (6)$$

Here, τ represents updated parameters and can be set to 0.

The specific implementation process of VMD is as follows:

(a) Initialization U_K^1, ω_K^1, γ^1 and maximum number of iterations N, $n=0$.

(b) For each mode U_K, it is updated according to equation (3) and (4).

(c) According to equation (5), update γ, $n=n+1$.

(d) Judging convergence according to equation (6): If it does not converge and n<N, repeat steps (b). Otherwise, the final modal function U_K and ω_k are obtained by stopping the iteration.

2) OVMD

The results show that the performance of VMD is mainly affected by the number of decomposed modal functions K and the updating step $\tau$of Lagrange multiplier when applied to wind speed sequence decomposition. In [34], if K is too large, the ω of mode will aggregate or even overlap. If K is too small, some modes will be divided into adjacent modes, or even discarded. Different τ will lead to different degrees of residual, which will affect the forecast accuracy. Therefore, a method based on central frequency observation to determine K, and a method based on minimization of residual error index (REI) to determine τ [20] are proposed.

First, the ω_k of decomposition modes under different K values are calculated and analyzed. Once a similar frequency occurs, the K at this time is determined as the best K for decomposition. Then, τ is optimized according to the root mean square error (RMSE) between the denoised time series and the original sequence, which can be simplified to the REI, which can be expressed as:

$$REI = \min\frac{1}{N}\sum_{i=1}^{N}\left|\sum_{k=1}^{k}U_k-f\right|_i \quad (7)$$

B. OSVD

SVD has ideal decorrelation. The method based on SVD can reconstruct features and separate useful information from data [38]. Using SVD to generate features can retain part of the information of all features in the original feature set and remove the correlation between features.

The known training matrix $A_{m\times c}$ represents m samples and c features. The rank of the matrix is$r$. Singular value decomposition of matrix A is performed.

$$A_{m\times c} = U_{m\times c}\Lambda_{c\times c}V_{mxc}^T \quad (8)$$

Here, U and V represent orthogonal matrices respectively, and $\mathbf{\Lambda}$ denotes the nonnegative diagonal matrices of $m\times c$:

$$\Lambda_{\text{mxc}}^T = \begin{bmatrix} S_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & S_c \end{bmatrix} \quad (9)$$

$S_1,\ldots,S_c$ is the singular value of matrix A, in which S_1>S_2>>S_c. According to the principal component theory, the larger the singular value, the more information it contains. Therefore, a new matrix A':

$$A'_{m\times h} = U(:,1:h)\times\Lambda_{h\times h} \quad (10)$$

where *U (:, 1: h)* represents the matrix corresponding to the front h column vectors in U, and $\mathbf{\Lambda}_{h\times h}$ represents the diagonal matrix corresponding to the first h larger singular values.

The selection of h affects the accuracy of WSF, so the corresponding model's MAPE is calculated based on the contribution rate of different singular values. The contribution rate under the minimum MAPE is selected as the best contribution rate, and the optimal singular value h is determined.

Contribution rate D:

$$D = \frac{\sum_j^h S_j}{\sum_i^n S_i} \quad (11)$$

Here, S_i represents the MAPE of the model corresponding to the i singular value, and S_j represents the MAPE of the model corresponding to the j singular value.

C. GINI IMPORTANCE

The Gini index is a measure of node impurity and is used as an evaluation index to measure the contribution of each feature to each tree in a random forest [39].

Assume that E is a dataset containing e samples. It can be divided into g categories, and e_i represents the number of samples contained class i (i =1, 2, …, g). Then, the Gini index of set E is:

$$Gini(S) = 1 - \sum_{i=1}^{n} P_i^2 \tag{12}$$

$P_i = e_i/e$, which represents the probability that any sample belongs to class i. When E contains only one class, its Gini index is 0. When all categories in E are evenly distributed, the Gini index takes the maximum value.

When a random forest uses a feature to partition nodes, E can be divided into v subsets ($E_{i,j} = 1, 2, \ldots, v$), and then the Gini index of E is:

$$Gini_{split}(S) = \sum_{j=1}^{m} \frac{s_j}{s} Gini\left(S_j\right) \tag{13}$$

where e_j represents the number of samples in the set E_j. According to the equation (13), the feature partition with the minimum Gini split value is the best.

In the process of node partitioning, first, RF calculates the Gini split value of each feature in the candidate feature subset after partitioning the node. The Gini Importance of the feature is obtained by subtracting the value from the Gini exponent of the front node of the partitioned node. Then, the feature with the greatest Gini importance is selected as the segmentation feature of the node. After the RF is constructed, the Gini importance of the same feature is linearly superposed and arranged in descending order. Then, the importance ranking of all features can be obtained.

D. RANDOM FOREST

RF is a machine learning algorithm that combines decision tree and Bagging idea [46]–[48]. RF uses multiple samples to construct different decision tree models, in which each decision tree is forecasted separately. The final forecast is obtained by voting.

RF is a set of predictors $\{p(x, \Theta_k), d = 1, 2, \ldots, d_{tree}\}$ composed of multiple CART [48] decision trees, where x represents the input vector and $\{\Theta_k\}$ represents the independent and identically distributed random vector which determines the growth process of a single tree. The d_{tree} is the number of decision trees. Predictor $\{p(x, \Theta_k)\}$ is a classified regression tree, which is completely grown by CART algorithm and does not prune.

The algorithm steps of RF are as follows:

(1) A new self-service sample set of group d is randomly extracted from the original dataset by resampling. The d regression trees were established using the CART algorithm. Each time the samples were not extracted, the d-group dataset was formed.

(2) The m_{try} features are randomly selected from the original set of samples with M features, and one of the m_{try} features with the best classification effect is selected for the splitting of the node.

(3) Every tree grows completely without pruning.

(4) A random forest was formed after the growth of d regression trees. Finally, the experimental data were forecasted.

III. THE PROPOSED METHOD

To improve the accuracy of WSF, a new wind speed forecast method, which has three main parts based on OVMD, OSVD and RF, is proposed. The structure is shown in Fig. 1. In the first part, OVMD is used to decompose and reconstruct the

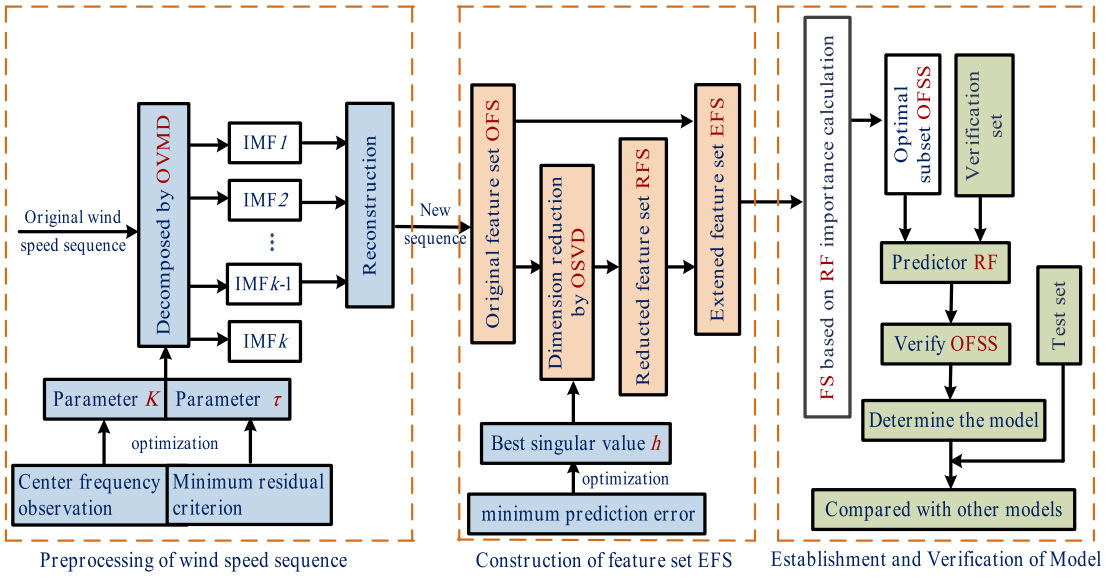


FIGURE 1. Model of the proposed method.

original wind speed series. The feature set OFS is established from the reconstructed wind speed sequence. In the second part, OSVD is used to generate features to get RFS, which is combined with OFS to obtain EFS. In the third part, RF is used to calculate the feature importance of EFS. The forward feature selection is carried out to determine the optimal subset of optimal feature subset (OFSS). In this method, parameters such as K and τ of OVMD and h of OSVD are optimized to improve the performance of the model, g is defined as a token of a feature in the feature set EFS.

The specific flow chart of this method is as follows:

Step 1: Data preparation: determine training set, test set, and verification set.

Step 2: The original wind speed sequence of the training set is processed by OVMD.

Step 2.1: Central frequency observation and residual index minimization criteria determine$K$ and τ of OVMD.

Step 2.2: By OVMD, the original wind speed sequence is decomposed into K modes and IMFs are obtained.

Step 2.3: Sort the IMF by magnitude and exclude the one with the smallest magnitude. The remaining *K-1* IMF was recombined to obtain a new wind speed sequence.

Step 3: Create feature sets.

Step 3.1: Establish a set of 96 historical wind speed features (F1, F2,. . . , F96) using new wind speed series.

Step 3.2: Use OSVD to generate features to get RFS (including h features, which are expressed as F97, F98,. . . , Fg). Combining OFS with RFS to obtain EFS (including g features, which are expressed as F1, F2,. . . , F96, F97,. . . , Fg.).

Step 4: Repeat steps 2 to 3 to process the verification set as well as the test set.

Step 5: Use RF to calculate and rank the importance of EFS features (assuming that descending rankings are: F1', F2', F3', F4', F5'. . .).

Step 6: RF as a predictor. Forward feature selection is performed according to importance ranking.

Step 6.1: The input set is {F1'} and the forecast model is established.

Step 6.2: The input set is {F1', F2'}, and the forecast model is established.

Step 6.3: Loop until the input set is {F1', F2', . . . , Fg'}, establish all models to forecast.

Step 6.4: Comparing the forecasted results of all models, the input set corresponding to the best results is determined as OFSS.

Step 7: Applying OFSS in the test set proves the performance of the model determined by the proposed method.

A. CONSTRUCTION OF FEATURE SET

Generally, the candidate feature set used in the existing research and references for feature selection is: (1) constructing a feature set using original wind speed data, temperature, atmospheric pressure, etc.; (2) decomposing the original wind speed data into an IMF using TSD, and constructing a feature set using the IMF. Because the FS method discards some low-importance features, the optimal subset loses all of the information of some low-importance features in the OFS. This loss of information leads to a reduction in the accuracy of the WSF. The traditional FR method maps all the data of the OFS from the high-dimensional space to the low-dimensional space, thereby reducing the data to obtain the RFS. The FS contains partial information about all features, but it also loses some of the information for all features. This loss of information increases the efficiency of the WSF, but reduces the accuracy.

Inspired by these ideas, this paper proposes a short-term WSFM with low information loss, which can reduce the information loss of high importance features while retaining some information of low importance features. The step of the construction of feature set is shown in the construction of feature set EFS of Fig. 1.First, OSVD method is used to reduce the dimension of OFS and obtain RFS. Given that this method maps high-dimensional data to low-dimensional space, each feature in RFS contains part of the information of all features of OFS. Then, the candidate feature set EFS is obtained by combining RFS and OFS, and the forward filter feature selection based on RF forecast accuracy is carried out for this feature set. The results of feature selection include some features of RFS and some features of OFS.

IFS (Initial feature set) contains historical wind speed sequence features without any pre-treatment. IFS is used to: (1) Using IFS to reduce the dimension of the IFS, the resulting feature set is used as the input feature set of the model PCA-RF. (2) Using the OSVD to reduce the dimension of the IFS, the resulting feature set is used as the input feature set of the model OSVD-RF. (3) As an input feature set of the model OS-RF.

OFS (original feature set) contains 96-dimensional OVMD-processed historical wind speed sequence features. OFS is used to: (1) Using OSVD to perform feature reduction to generate features, thereby constructing a reduction feature set RFS. (2) As the input feature set of the OVMD-RF and OVMD-OFS$_{fs}$-RF.

RFS (reduced feature set) is a feature set obtained by the OSVD method for feature reduction of OFS. When the OSVD method is used in four data sets, the feature contribution rate of the OSVD is set to 95%. However, in different data sets, the number of features corresponding to 95% of the feature contribution rates is different. This causes the parameters determined by the OSVD method) to be different in the four data sets (the specific values are shown in Table 7). RFS is used to:(1) Combine with OFS to construct extended feature set EFS. (2) As an input feature set of the model OVMD-OSVD-RF in the type TSD-FR-FM.

EFS (extended feature set) is a feature set obtained by combining OFS and RFS, and is a candidate feature set for performing feature selection. The dimension of features of the OFS all are 96, and the Dimension of features of the RFS in the four data sets are different. Therefore, EFS has different feature numbers in the four data sets (111, 118, 112 and 110). It is used to:(1) Feature selection is performed as a candidate feature set to determine the OFSS.(2) As the input feature set

TABLE 1. Six index for evaluating model performance.

Index	Definition	Equation
MAE	Mean absolute error	$MAE=\left(\sum_{i=1}^{N}\left\|y_i-\hat{y}_i\right\|\right)/N$
RMSE	Root mean square error	$RMSE=\sqrt{\left(\sum_{i=1}^{N_t}\left[y_i-\hat{y}_i\right]^2\right)/N_t}$
MAPE	Mean absolute percentage error	$MAPE=\left(\sum_{i=1}^{N_t}\left\|\left(y_i-\hat{y}_i\right)/y_i\right\|\right)/N_t$
P_{MAE}	Promoting percentages of mean absolute error	$P_{MAE}=\left\|\left(MAE_1-MAE_2\right)/MAE_1\right\|$
P_{RMSE}	Promoting percentages of root mean square error	$P_{RMSE}=\left\|\left(RMSE_1-RMSE_2\right)/RMSE_1\right\|$
P_{MAPE}	Promoting percentages of mean absolute percentage error	$P_{MAPE}=\left\|\left(MAPE_1-MAPE_2\right)/MAPE_1\right\|$

TABLE 2. Division of test data sets.

Season	Division of dataset		
	Train set	Verification set	Test set
Spring	March 25th - April 9th	April 10th - April 13th	April 14th - April 17th
Summer	July 16th - July 31st	August 1st - August 4th	August 5th - August 8th
Autumn	August 30th - September 14th	September 15th - September 18th	September 19th - September 22nd
Winter	January 1st - January 16th J	January 17th - January 20th	January 21st - January 24th

of the model OVMD-OSVD-RF in the TSD-FR-FM, and the proposed model OVMD-EFS$_{fs}$-RF.

OFSS (the optimal feature subset) is the optimal subset determined by feature selection. The OFSS contains both features in the OFS as well as features in the RFS. If the feature name belongs to the range F1-F96, then this feature is a feature from the OFS. If the feature name belongs to the range F97-Fg (g takes 111, 18, 112, 110), then this feature is a feature from the RFS. For example, in Table 8, feature F106 is a feature in RFS. This feature is determined by feature selection as a feature of the optimal subset OFSS.

OFS is an original feature set containing a 96-dimensional historical wind speed sequence, which is used to:(1) Using OSVD to perform feature reduction to generate features, thereby constructing a reduction feature set RFS. (2) As the input feature set of the Basic Model and the TSD-FS. RFS is a reduction feature set (containing 15, 22, 16 and 14 features in each of the four data sets), which is used to:(1) Combine with OFS to construct an extended feature set EFS.(2) As an input feature set of the model OVMD-OSVD-RF in the type TSD-FR-FM, and the model OSVD-RF in the type FR-FM. EFS is a set of extended dimension features (including 111, 118, 112, and 110-dimensional features in each of the four data sets),which is used to:(1) Feature selection is performed as a candidate feature set to determine the OFSS.(2) As the input feature set of the model OVMD-OSVD-RF in the TSD-FR-FM, and the proposed model OVMD-EFS$_{fs}$-RF.OFSS is the optimal feature subset determined in this paper, and the specific feature composition is listed in Table 8-Table 11.

TABLE 3. Statistical information of the data.

Datasets	Statistic indices				
	Mean.(m/s)	Max.(m/s)	Min.(m/s)	Std.	Kurt.
Spring	4.91	10.67	1.48	2.142	1.95
Summer	3.93	11.26	0.62	1.83	5.67
Autumn	5.45	20.28	0.37	4.03	6.08
Winter	7.96	19.97	1.14	4.01	2.47

IV. EXPERIMENTS AND ANALYSIS

In this chapter, the experimental results of single-step and multistep forecast are compared to evaluate the performance of the proposed method. To evaluate the real performance of each method, all experiments were repeated 30 times. The RMSE, MAE and MAPE indices presented in the results were the average values of 30 trials. To evaluate the performance of each forecast method quantitatively, three error indices are used, mean absolute error (MAE), root mean square error (RMSE) and mean absolute percentage error (MAPE) (Table 1). Generally, the smaller these indicators, the better the forecast performance of the corresponding model. Here, N_t represents the sample size, yi and $\hat{y}_t$ represents the observed and forecasted values of time interval i, and v_A and v_B represent the index values of model A and model B, respectively.

TABLE 4. Description of several models.

Type	Model	Description
FR-FM	PCA-RF	PCA is used to generate features .The input feature set of RF based predictor is .a feature set obtained by PCA processing of a 96-dimensional feature set consisting of an unprocessed wind speed sequence.
	OSVD-RF	OSVD is used to generate features. The input feature set of RF based predictor is RFS.
TSD-FM	VMD-RF	VMD processes the original wind speed. The input feature set of RF based predictor is a 96-dimensional feature set consisting of a wind speed sequence processed by the VMD.
	OVMD-RF	OVMD processes the original wind speed. The input feature set of RF based predictor is OFS
TSD-FR-FM	OVMD-OSVD-RF	OVMD processes the original wind speed. OSVD generates features .The input feature set of RF based predictor is RFS.
	OVMD-EFS-RF	OVMD processes the original wind speed. OSVD generates features. The input feature set of RF based predictor is EFS.
TSD-FS-FM	OVMD-OFS_{fs}-RF	OVMD processes original wind speed. The input feature set of RF based predictor is OFS which is used to feature selection
Basic Model	OS-RF	The input feature set of RF based predictor is a 96-dimensional feature set consisting of an original wind speed sequence without any processing.
The proposed model	OVMD-EFS_{fs}-RF	OVMD processes original wind speed. OSVD generates features. The input feature set of RF based predictor is EFS which is used to feature selection

A. DATA

Experiments were conducted using data from 2009 provided by the National Renewable Energy Laboratory (NREL). The data is sampled at intervals of 15 minutes and contains 96 points per day. Based on the season, it is divided into four datasets [42]–[44] of spring, summer, autumn and winter, and 24 days are selected for experiments in four datasets (Table 2).The training set accounts for 60% of the dataset (16 days, 1536 points), the test set accounts for 20% (4 days, 384 points), and the verification set accounts for 20% (4 days, 384 points) [42], [45]–[47]. Table 3 gives statistical data (including average, maximum, minimum, standard deviation, and kurtosis).

The experiment used five types of models to verify the proposed method (Table 4). The first type is feature reduction-forecasting model (FR-FM), which is a WSFM based on feature reduction, including PCA-RF and OSVD-RF. The second type is time series decomposition-forecasting model (TSD-FM), which is a model based on time series decomposition, including VMD-RF, OVMD-RF and OVMD-OTSVD-RF. The third type is time series decomposition-feature reduction-forecasting model (TSD-FR-FM), including OVMD-OSVD-RF and OVMD-EFS-RF. The fourth type is time series decomposition-feature selection-forecasting model (TSD-FS-FM), including OVMD-OFS_{fs}-RF. The last type is the basic model in which the feature set was not processed. This paper compares these five types of models with the proposed model OVMD-EFS_{fs}-RF. The input feature set of all models are given in Table 4

B. EXPERIMENTAL RESULTS OF OVMD

The data was preprocessed using the OVMD method. As described in chapter 2, section A., the decomposed mode number K and the update parameter τ affect the decomposition effect of the OVMD. Tables 5 and 6 list the center frequencies of the IMFs at different K values. As the similar frequency starts from $K = 6$, K is determined to be 5. Because the REI of the original wind speed sequence needs to be as small as possible to ensure forecast accuracy, it is necessary to optimize τ before decomposition.

TABLE 5. Central frequencies under different K values of spring and summer.

Dataset	K	Central frequencies					
Spring	2	2.760×10^{-4}	0.336				
	3	2.327×10^{-4}	0.030	0.336			
	4	2.328×10^{-4}	0.021	0.129	0.340		
	5	2.357×10^{-4}	0.019	0.125	0.244	0.358	
	6	2.325×10^{-4}	0.019	0.125	0.253	0.323	0.364
Summer	2	4.70×10^{-4}	0.236				
	3	2.74×10^{-4}	0.035	0.240			
	4	2.30×10^{-4}	0.031	0.073	0.241		
	5	2.29×10^{-4}	0.031	0.073	0.241	0.322	
	6	2.29×10^{-4}	0.031	0.073	0.237	0.303	0.324

TABLE 6. Central frequencies under different K values of autumn and winter.

Dataset	K	Central frequencies					
Autumn	2	5.49×10^{-4}	0.256				
	3	4.17×10^{-4}	0.053	0.258			
	4	4.16×10^{-4}	0.053	0.236	0.324		
	5	2.20×10^{-4}	0.020	0.165	0.240	0.324	
	6	2.06×10^{-4}	0.019	0.163	0.222	0.306	0.337
Winter	2	3.96×10^{-4}	0.267				
	3	3.49×10^{-4}	0.020	0.267			
	4	3.43×10^{-4}	0.020	0.167	0.352		
	5	3.36×10^{-4}	0.019	0.163	0.235	0.355	
	6	3.33×10^{-4}	0.019	0.163	0.230	0.323	0.358

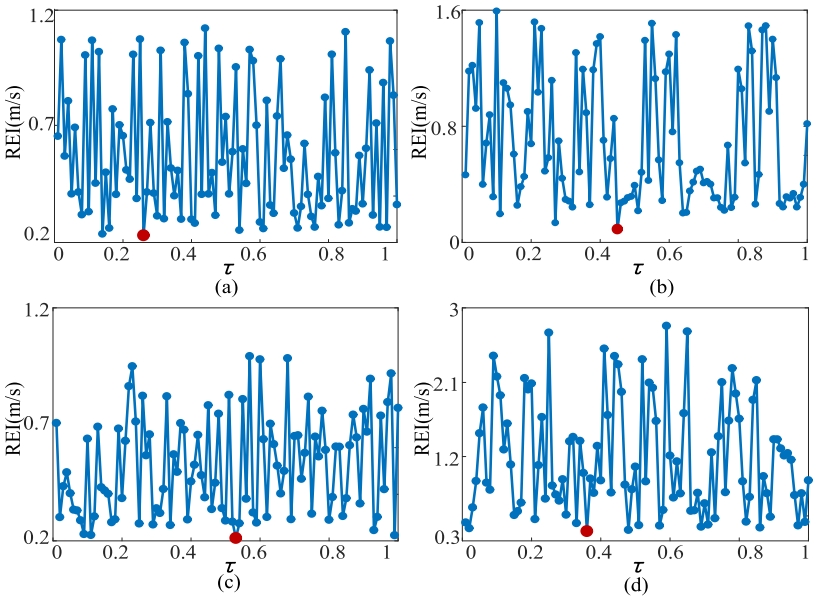


FIGURE 2. REI values for the data of the four study areas. (a) Spring. (b) Summer. (c) Autumn. (d) Winter.

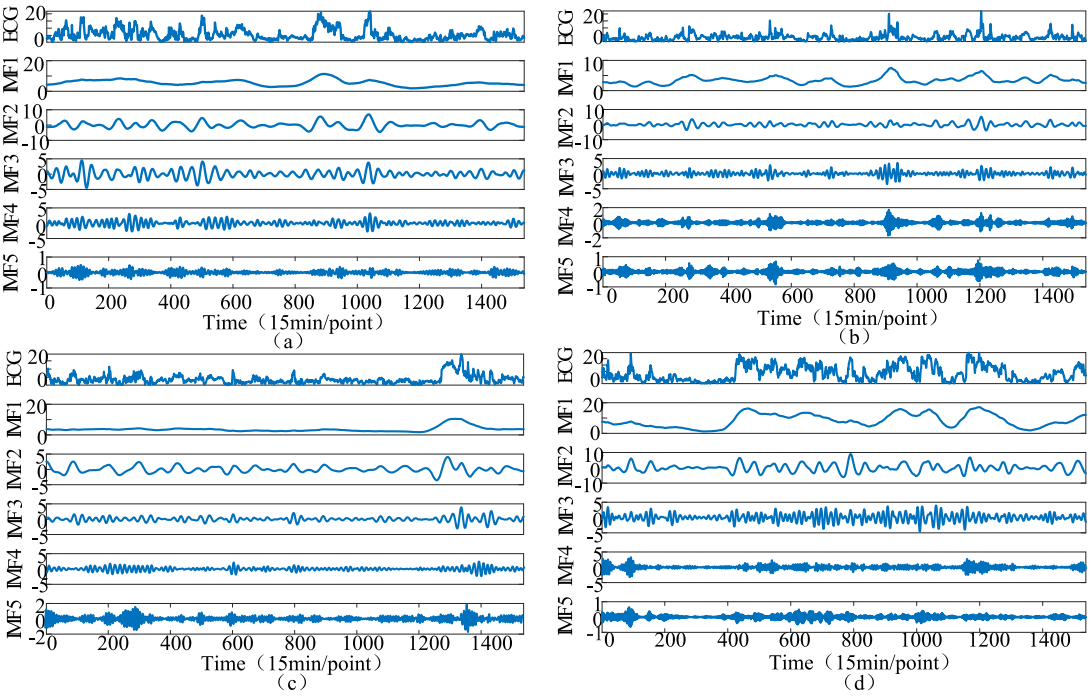


FIGURE 3. Result of OVMD. (a) Spring. (b) Summer. (c) Autumn. (d) Winter.

The REI value is shown in Fig. 2. In Fig. 2 (a), $\tau$is 0.26 when the REI value reaches the minimum value of 0.231. In Fig. 2 (b), τ is 0.46 when REI reaches a minimum value of 0.127. In Fig. 2(c), τ is 0.53 when the REI value reaches a minimum value of 0.212. In Fig. 2(d), τ is 0.37 when the REI reaches a minimum value of 0.316. The wind speed sequence was determined using OVMD after determining the best K and τ. Fig. 3 shows the original sequence and IMFS. The four original wind speed series are broken down into five subseries. IMF1 shows the general trend of the original wind speed series with the largest amplitude. IMF5 has the smallest amplitude and strong volatility. Other modes have significant periodicity. Fig. 4 shows the results of EMD decomposing the data of the spring dataset. It can be seen that the EMD has a significant end-effect which cause distortion and a large number of IMFs. Processing data using the OVMD method

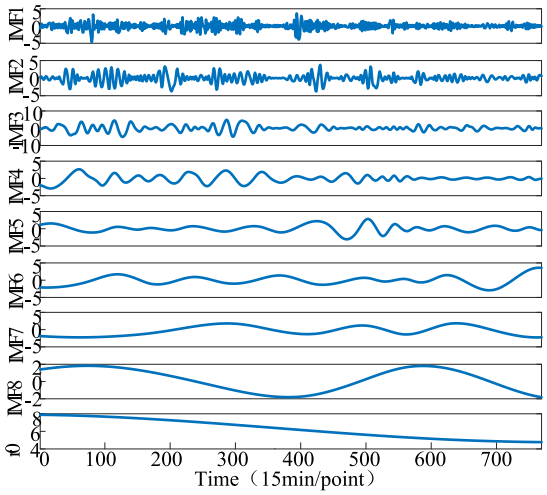


FIGURE 4. Result of EMD spring.

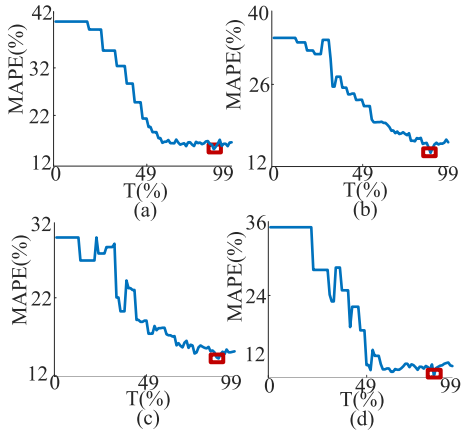


FIGURE 5. Determination of the best contribution rate. (a) Spring. (b) Summer. (c) Autumn. (d) Winter.

avoids endpoint effects effectively and reduces the number of IMFs compared to EMD. Therefore, the new method uses OVMD to decompose and reconstruct the wind speed series. The one with the smallest amplitude in the IMF is removed, and then the other IMFs are recombined to obtain a new wind speed series, which is used for later experiments.

C. GENERATION OF FEATURES AND CONSTRUCTION OF EFS

To reduce the loss of information caused by filter method and improve the accuracy of forecast, OSVD is used to generate features to construct a new set of candidate features. As shown in chapter 2, section B., OSVD needs to determine the optimal number of singular values h. Fig. 5 shows the forecast accuracy of WSF using only the feature set generated by OSVD under different contribution rates. The optimal singular value h is shown in Table 7. After determining the parameters of the OSVD, the method is used to reduce the dimensions of the OFS (including 96-dimensional features) to obtain the RFS (RFS of different data sets contains different number of features). This feature set reduces several

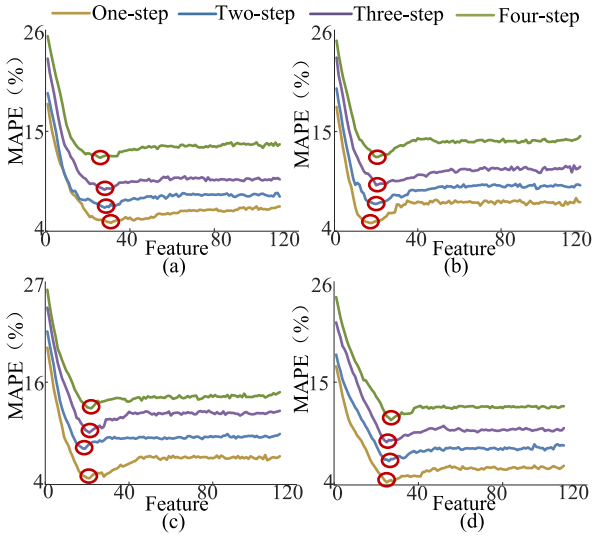


FIGURE 6. Process of feature selection (a) Spring. (b) Summer. (c) Autumn. (d) Winter.

TABLE 7. 7 Value of h in different dataset.

Dataset	Value of h with different forecast scales.			
	One-step	Two-step	Three-Step	Four-step
Spring	15	15	15	15
Summer	22	22	22	22
Autumn	16	16	16	16
Winter	14	14	14	14

features compared with the OFS. However, it is noteworthy that the features in the RFS are not any of the features in the OFS, because the OSVD maps raw data from high-dimensional space to low-dimensional space and new feature retains some important information of all features in the OFS.

D. FEATURE SELECTION

A multistep WSFM based on a recursive strategy is proposed in this paper. First, RF is used to calculate the importance of all features of the EFS. Then, the forward feature selection method is used, and the EFS is used to select features to determine the optimal subset according to the importance of features. The d_{tree} is set to the default value of 500 and m_{try} also takes the default experience value of t/3 [50]. Table 8-11 lists the results of feature selection, and Fig. 6 shows the process of feature selection. The following points deserve attention:

(1) Under different forecast targets, the importance of the same feature is different. Taking the dataset of spring in Table 8 as an example, F4 ranks fourth in one-step and two-step models, but fifth in three-step and four-step models. Therefore, different forecast targets should be modeled separately.

(2) The optimal subset contains features in the OFS as well as the RFS. The results of the optimal subset of different forecast scales contain several features in the OFS and in the RFS. For example, the optimal subset contains 10 features of

TABLE 8. Results of feature selection of spring.

Model	OFSS
One-step	F1,F2,F3,F4,F100,F5,F102,F97,F99,F6,F13,F71,F19,F15,F96,F98,F24,F23,F109,F82,F89,F41,F26,F77,F105,F43,F62,F65,F94,F110,F106
Two-step	F1,F2,F3,F4,F14,F97,F15,F98,F16,F105,F109,F12,F89,F5,F102,F6,F13,F96,F100,F71,F94,F82,F68,F25,F92,F85,F63,F106
Three-step	F1,F2,F3,F98,F4,F97,F100,F94,F96,F102,F95,F110,F20,F6,F5,F91,F24,F70,F31,F106,F78,F101,F90,F76,F77,F23
Four-step	F1,F2,F3,F97,F4,F100,F98,F33,F106,F77,F58,F109,F67,F18,F30,F61,F13,F68,F108,F94,F96,F92,F31,F20,F102,F76

TABLE 9. Results of feature selection of summer.

Model	OFSS
One-step	F1,F2,F3,F114,F109,F5,F97,F51,F110,F4,F50,F90,F91,F36,F49,F101,F104
Two-step	F1,F2,F3,F4,F97,F114,F5,F51,F103,F90,F25,F91,F53,F6,F24,F107,F18,F61,F13,F101
Three-step	F1,F2,F3,F109,F4,F97,F114,F5,F95,F39,F89,F6,F20,F29,F44,F84,F69,F26,F100,F104
Four-step	F1,F2,F3,F118,F109,F4,F103,F100,F101,F88,F5,F108,F104,F97,F115,F6,F29,F44,F110,F114

TABLE 10. Results of feature selection of autumn.

Model	OFSS
One-step	F1,F2,F3,F4,F111,F99,F5,F16,F103,F105,F102,F100,F110,F108,F18,F96,F65,F19,F85,F75,F6
Two-step	F1,F2,F3,F111,F99,F4,F110,F103,F105,F104,F14,F5,F108,F100,F26,F95,F10,F23,F102
Three-step	F1,F2,F3,F111,F100,F4,F106,F13,F103,F102,F108,F14,F95,F107,F96,F24,F31,F101,F94,F38,F105
Four-step	F1,F2,F111,F3,F100,F38,F4,F103,F106,F96,F13,F12,F107,F102,F98,F110,F29,F93,F94,F63,F23,F101

TABLE 11. Results of feature selection of winter.

Model	OFSS
One-step	F1,F2,F3,F96,F4,F18,F95,F5,F109,F89,F43,F29,F91,F101,F79,F102,F24,F69,F25,F27,F74,F103,F45,F16,F108
Two-step	F1,F2,F3,F109,F104,F102,F26,F96,F4,F103,F94,F75,F17,F51,F18,F93,F105,F45,F100,F61,F79,F62,F38,F67,F97,F107
Three-step	F1,F96,F2,F95,F3,F93,F91,F105,F101,F110,F44,F4,F60,F86,F94,F25,F43,F71,F98,F73,F92,F85,F74,F106,F15
Four-step	F1,F2,F96,F95,F3,F109,F104,F47,F100,F97,F37,F102,F71,F13,F107,F105,F4,F38,F42,F70,F84,F90,F14,F27, F108,F52,F57

TABLE 12. Number of features in different feature sets in OFSS.

Dataset	Feature dimension of OFSS							
	One-step		two-step		three-step		four-step	
	OFS	RFS	OFS	RFS	OFS	RFS	OFS	RFS
Spring	22	9	21	7	19	7	19	7
Summer	11	6	15	5	15	5	10	10
Autumn	1	8	10	9	12	9	13	9
Winter	20	5	18	8	20	5	19	8

the OFS and the RFS, each accounting for 50% in the four-step model in the dataset of summer.

This result shows that the method of construction of feature set proposed in this paper is effective in WSF. The main information of all high-importance features is retained by preserving the features in the OFS. By retaining the features in the RFS, the information of some low-importance features can be preserved. Thus, the new method reduces the loss of information effectively caused by traditional feature reduction or feature selection methods.

E. EXPERIMENT RESULT OF SINGLE-STEP FORECAST

Table 13 lists the RMSE, MAE and MAPE for the single-step WSF of the eight methods in four data sets. The following points are discussed:

(1) In FS-FM, the MAPE, MAE and RMSE of OSVD-RF are lower than PCA-RF, which shows that OSVD performs better in single-step WSF. In TSD-FM, the MAP, RMSE and MAE of OVMD-RF are lower than VMD-RF, indicating the effectiveness of OVMD in single-step WSF.

(2) MAPE, MAE and RMSE of model OVMD-OSVD-RF are lower than OSVD-RF and OVMD-RF. This shows that the combination of these two methods improves the accuracy of short-term WSF.

(3) In TSD-FR-FM, the MAPE, RMSE and MAE of the model OVMD-EFS-RF are significantly lower than OVMD-OSVD-RF, indicating that EFS improves the accuracy of WSF.

(4) The MAPE, RMSE and MAE of the model OVMD-EFS_{fs}-RF in TSD-FR-FM are significantly lower than OVMD-OFS_{fs}-RF in TSD-FS-FM, indicating that the

TABLE 13. Three evaluation indicators (single step forecast).

Model	Spring			Summer			Autumn			Winter		
	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE
OS-RF	0.79	0.73	18.66	0.64	0.54	17.76	0.88	0.80	19.65	0.82	0.72	15.72
PCA-RF	0.75	0.68	17.99	0.49	0.47	15.39	0.84	0.72	16.94	0.78	0.71	13.02
OSVD-RF	0.65	0.58	15.30	0.48	0.43	14.46	0.66	0.56	14.48	0.73	0.67	12.29
VMD-RF	0.70	0.57	14.78	0.49	0.46	14.85	0.73	0.67	16.75	0.74	0.69	12.55
OVMD-RF	0.64	0.51	12.02	0.48	0.39	12.69	0.57	0.46	12.70	0.67	0.63	10.09
OVMD-OFS$_{fs}$-RF	0.61	0.47	11.01	0.42	0.35	11.54	0.54	0.44	12.12	0.63	0.59	9.77
OVMD-OSVD-RF	0.46	0.41	9.97	0.34	0.33	10.77	0.50	0.41	11.46	0.59	0.56	9.26
OVMD-EFS-RF	0.38	0.27	6.70	0.27	0.24	7.92	0.36	0.31	8.66	0.47	0.46	6.48
OVMD-EFSfs-RF	0.31	0.20	4.91	0.17	0.15	4.89	0.25	0.21	5.67	0.28	0.25	4.28

TABLE 14. P_{INDEX} of the proposed model and the comparison model.

Model	Spring			Summer			Autumn			Winter		
	P_{RMSE}	P_{MAE}	P_{MAPE}	P_{RMSE}	P_{MAE}	P_{MAPE}	P_{RMSE}	P_{MAE}	P_{MAPE}	P_{RMSE}	P_{MAE}	P_{MAPE}
OS-RF	60.62	73.68	72.16	73.91	71.63	72.44	74.87	73.9	71.15	65.26	64.68	72.79
PCA-RF	58.61	72.70	70.07	66.14	67.15	68.19	7.93	71.04	66.53	63.79	64.50	67.14
OSVD-RF	51.84	67.90	64.76	64.67	64.26	66.15	62.73	62.65	60.85	60.99	61.84	65.23
VMD-RF	55.61	66.78	64.15	65.57	66.59	67.05	66.29	69.14	66.15	61.66	63.45	65.92
OVMD-RF	51.41	59.15	60.07	64.76	60.06	61.42	56.72	54.83	55.37	57.53	59.82	57.60
OVMD-OFS$_{fs}$-RF	49.18	57.74	55.40	59.52	57.14	57.63	53.70	52.27	53.22	55.56	57.63	56.81
OVMD-OSVD-RF	31.65	50.75	50.28	50.60	53.24	54.57	51.09	49.71	50.51	52.61	54.18	53.79
OVMD-EFS-RF	17.45	26.74	25.28	38.08	36.02	38.24	32.34	33.27	34.60	39.75	44.84	33.95

proposed new method of low information loss effectively improves the accuracy of short-term WSF.

In different data sets, the indicators of the model OVMD-EFS$_{fs}$-RF are the lowest of the eight models, which shows that it has the highest accuracy in single-step WSF.

To evaluate the performance improvement of the proposed method based on other methods, the P_{INDEX} value is listed in Table 14. The model determined by the proposed method is considerably improved on the basis of the other seven models; especially for all models in the basic model, FR-FM and TSD-FM, the performance improvement is higher than 50%. To more directly demonstrate the superiority of the proposed model, select the best model in the FR-FM and TSD-FM types and all the models in the TSD-FR-FM and TSD-FS-FM type for the section E of chapter 4 test according to the experimental results in Tables 13-14. OSVD-RF, OVMD-RF, OSVD-OVMD-RF, OVMD-OFS$_{fs}$-RF and OSVD-EFS-RF were set as comparative experimental groups and compared with the proposed model.

Fig. 7-10 shows the forecasted results of the four data sets, respectively. At the same time, Fig. 7-10 shows the local analysis of the interval length of 20 points with the peak (valley) value as the center point. The left subgraph is a partial map of the valley value, and the right subgraph is a partial map of the peak. Specifically:

(1) In the partial map of peak (valley) values, the model OVMD-EFS$_{fs}$-RF has higher forecast accuracy when the wind speed is higher (lower). The model also got better forecasts when the wind speed suddenly changed. The trend of the forecasted sequence is consistent with the original sequence and is the best among the eight models.

(2) The model OVMD-EFS-RF performance is second only to OVMD-EFS$_{fs}$-RF, which is significantly better than other models. On the one hand, the effectiveness of the optimal subset is explained. On the other hand, EFS improves the accuracy of forecast.

(3) Obviously, the prediction results of the model OVMD-OFS$_{fs}$-RF are not as good as the model OVMD-EFS$_{fs}$-RF, and even the prediction results opposite to the original data changes. This shows that the model determined by the method of low information loss performs better than the traditional model OVMD-OFS$_{fs}$-RF.

(4) In the peak-to-valley partial map, OSVD-RF obtains better performance than PCA-RF. The OVMD-RF achieved better performance than VMD-RF. This explains the effectiveness of the OSVD and OVMD methods.

Obviously, the forecasted sequence obtained by the proposed model best matches the original sequence. It better tracks the original wind speed sequence and does not show a completely opposite trend to the original wind

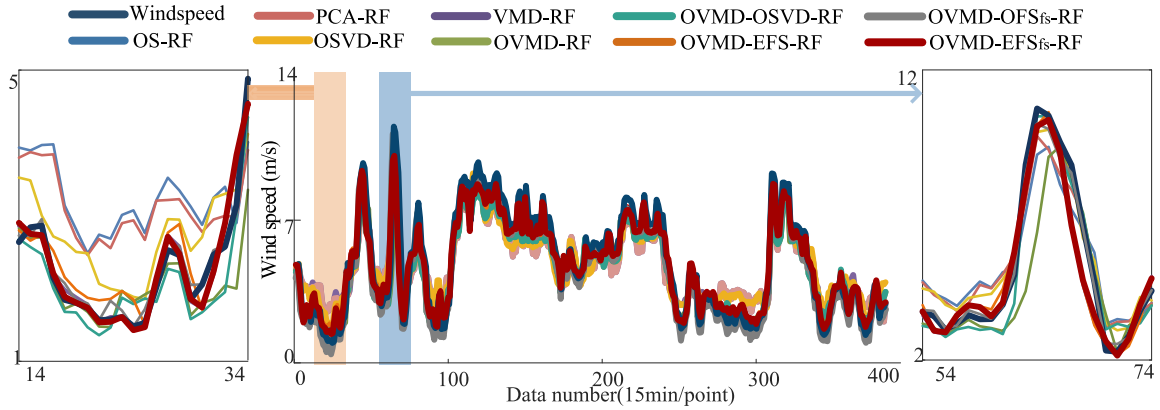


FIGURE 7. Result of wind speed forecast in spring.

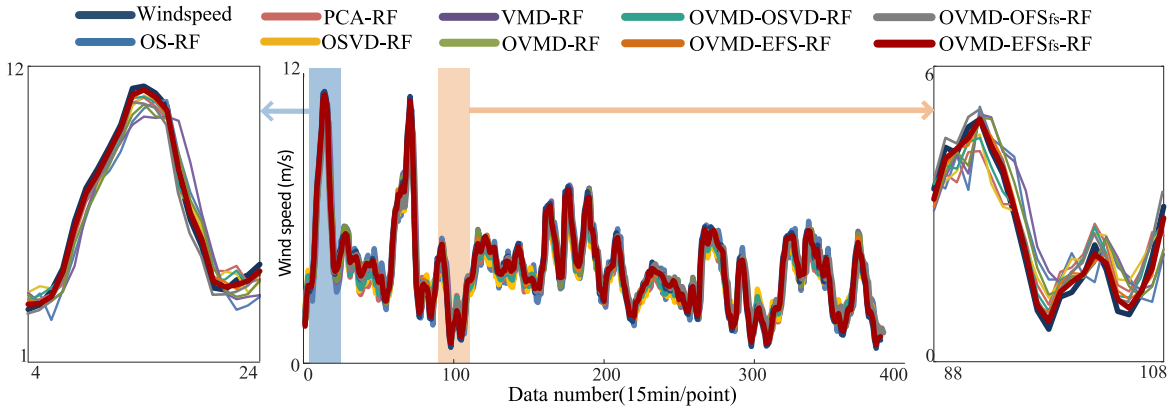


FIGURE 8. Result of wind speed forecast in summer.

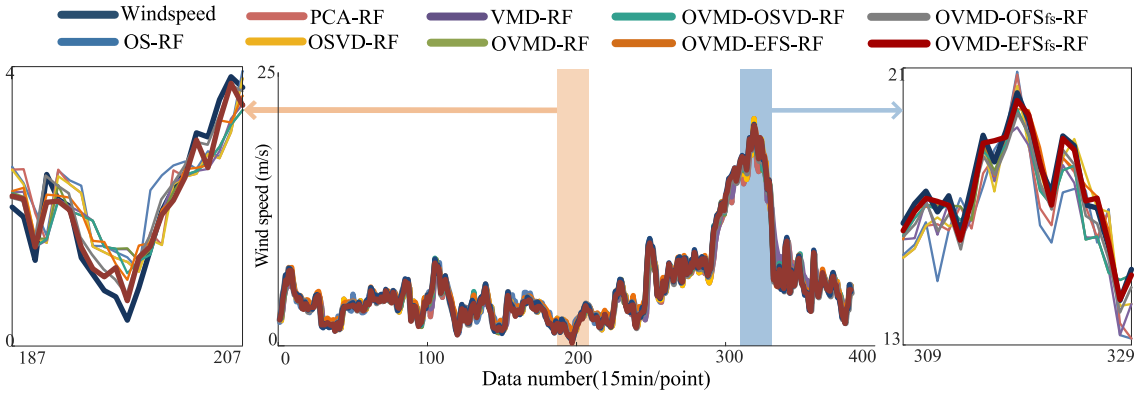


FIGURE 9. Result of wind speed forecast in autumn.

speed sequence. Other models, when the wind speed fluctuates greatly, will get some forecast points with the exact opposite of the actual sequence.

To more visually demonstrate the predictive effect of the model, the RMSE, MSE and MAPE of the selected models in the FR-FM, TSD-FM, TSD-FS-FM and TSD-FR-FM types were compared (Fig. 11). The trend of the index changes is clearly consistent: the accuracy of the model is improved according to the order of OSVD-RF, OVMD-RF, OVMD-OFS$_{fs}$-RF, OVMD-OSVD-RF, OVMD-EFS-RF and the proposed model. This trend is reflected in all evaluation indicators and in all data sets. This result illustrates the traditional TSD-FS-FS model does not improve the performance of the model as much as the proposed model, and the contribution of FR, TSD and EFS in improving model performance.

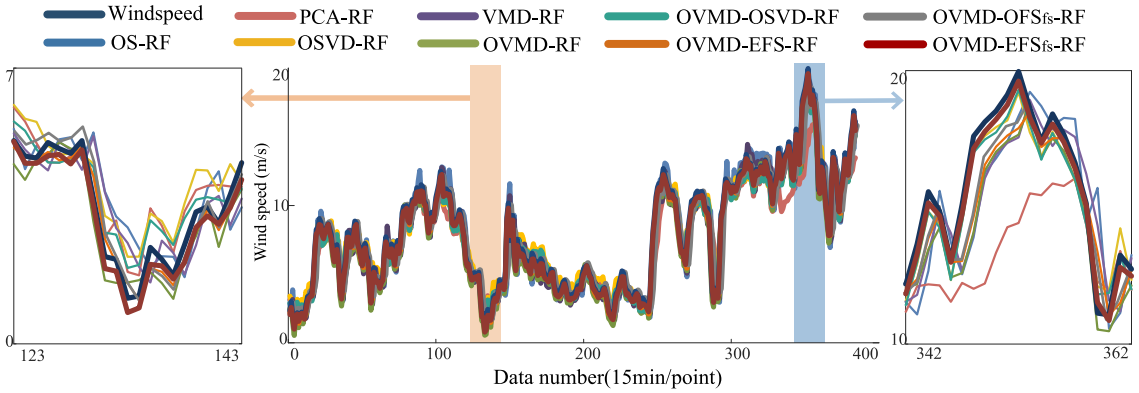


FIGURE 10. Result of wind speed forecast in winter.

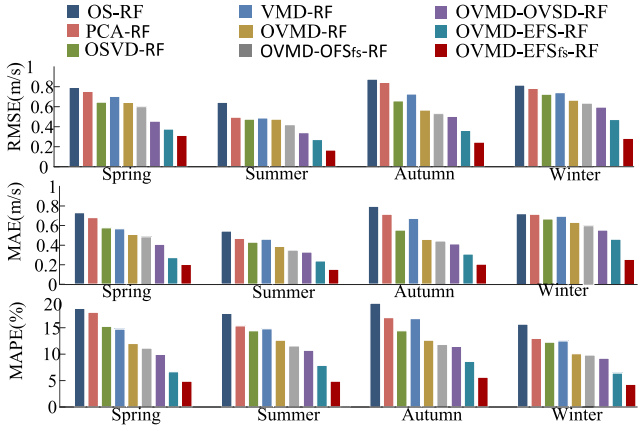


FIGURE 11. RMSE, MAE, and MAPE for each model.

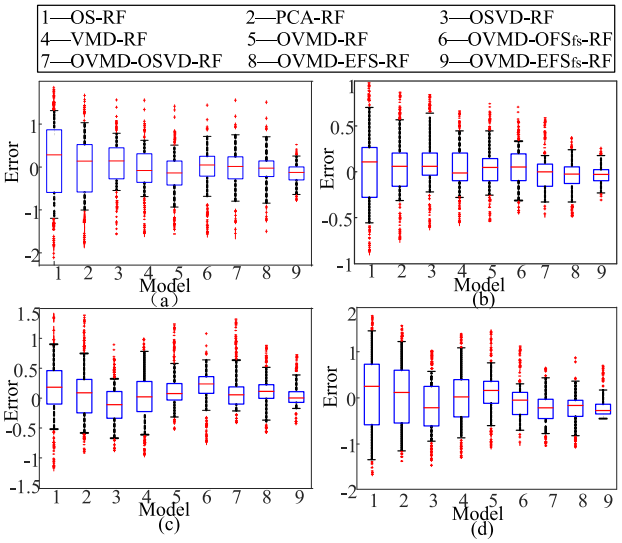


FIGURE 12. Box plot of the distribution of errors. (a) Spring. (b) Summer. (c) Autumn. (d) Winter.

The distribution of forecasted errors for each model is shown in Fig. 12. The box plot clearly shows that the forecast error of OSVD-EFS$_{fs}$-RF is distributed at approximately 0, which is more concentrated and closer to zero than the distribution of forecast errors of the other eight models. This means that the proposed model can obtain highly accurate forecasts at almost all points. However, the forecast errors of other models in some data points are large, and the original wind speed sequence cannot be well tracked.

F. RESULTS OF MULTISTEP FORECAST

This section evaluates the performance of the multistep forecast of the proposed model through extensive experiments. At the same time, according to the experimental results given in subsection chapter 4, section E, the models OSVD-RF, OVMD-RF, OVMD-OSVD-RF, OVMD-EFS- RF, OVMD-OFS$_{fs}$-RF and OVMD-EFS$_{fs}$-RF. Tables 15-18 list the RMSE, MAE and MAPE for the different models. It can be observed that the order of performance of each model in the comparative experimental group is consistent with that in chapter 4, section E.

In general, the indicators of the proposed model OVMD-EFS$_{fs}$-RF are much lower than other models. In single-step and multistep forecast, OVMD-RF performed better than OSVD-RF. Fig. 13-16 shows a comparison of MAPE for all models in different forecast scales. Specifically:

(1) The MAPE values of the model OSVD-RF are smaller than the PCA-RF in different forecast scales, and the MAPE values of the model OVMD-RF are smaller than the VMD-RF. This shows that the methods OSVD and OVMD can also improve the accuracy of forecast in multistep forecast.

(2) The MAPE value of the model OVMD-OSVD-RF is smaller than OSVD-RF, OVMD-RF. It is stated that the combination of the two methods is also effective in multistep forecast.

(3) The effect of the model OVMD-EFS-RF is better than that of OVMD-OSVD-RF, which verifies the effectiveness of EFS in multistep forecast.

(4) The RMSE, MAE and MAPE values of the model OVMD-OFS$_{fs}$ are higher than the model OVMD-EFS$_{fs}$, indicating that the model determined by the proposed new method of low information loss improves the prediction effect of the traditional TSD-FS-FM.

TABLE 15. Multi-step forecast results in spring.

Model	One-step			two-step			three-step			four-step		
	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE
OS-RF	0.79	0.73	18.66	0.91	0.80	20.52	1.28	0.89	23.42	1.58	1.00	25.00
PCA-RF	0.73	0.68	17.99	0.83	0.74	19.26	0.94	0.81	21.71	1.36	0.90	22.91
OSVD-RF	0.66	0.58	15.30	0.73	0.63	18.00	0.81	0.75	20.12	0.90	0.85	21.98
VMD-RF	0.70	0.59	14.78	0.81	0.72	16.21	0.90	0.80	18.62	1.13	0.89	20.30
OVMD-RF	0.62	0.51	12.02	0.70	0.60	15.13	0.79	0.70	17.00	0.85	0.83	19.02
OVMD-OFS$_{fs}$-RF	0.61	0.49	11.01	0.68	0.57	13.50	0.74	0.67	15.22	0.82	0.79	17.14
OVMD-OSVD-RF	0.48	0.41	9.97	0.62	0.52	11.57	0.71	0.63	13.91	0.80	0.76	15.98
OVMD-EFS-RF	0.38	0.28	6.70	0.46	0.37	8.17	0.51	0.48	9.77	0.63	0.57	13.80
OVMD-EFS$_{fs}$-RF	0.32	0.25	4.91	0.38	0.33	6.56	0.47	0.41	8.53	0.53	0.50	12.02

TABLE 16. Multi-step forecast results in summer.

Model	One-step			two-step			three-step			four-step		
	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE
OS-RF	0.64	0.54	17.76	0.74	0.63	19.59	1.02	0.89	22.86	1.41	1.01	24.56
PCA-RF	0.50	0.48	15.39	0.60	0.59	17.44	0.84	0.68	19.77	1.11	0.80	22.36
OSVD-RF	0.48	0.43	14.46	0.55	0.55	16.81	0.72	0.60	17.39	0.85	0.69	20.40
VMD-RF	0.49	0.46	14.85	0.54	0.52	16.06	0.81	0.64	18.90	1.06	0.76	21.59
OVMD-RF	0.48	0.39	12.69	0.57	0.44	14.64	0.64	0.55	16.20	0.75	0.65	19.42
OVMD-OFS$_{fs}$-RF	0.42	0.35	11.54	0.52	0.42	13.39	0.59	0.50	15.38	0.69	0.60	18.06
OVMD-OSVD-RF	0.34	0.33	10.77	0.38	0.40	12.13	0.48	0.47	14.17	0.59	0.56	17.35
OVMD-EFS-RF	0.27	0.24	7.18	0.32	0.28	9.01	0.39	0.38	11.05	0.47	0.48	14.40
OVMD-EFS$_{fs}$-RF	0.17	0.15	4.89	0.24	0.26	6.97	0.29	0.31	9.03	0.36	0.40	12.10

TABLE 17. Multi-step forecast results in autumn.

Model	One-step			two-step			three-step			four-step		
	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE
OS-RF	0.88	0.80	19.65	0.96	0.82	21.41	1.26	1.05	23.98	1.60	1.11	25.87
PCA-RF	0.84	0.72	16.94	0.89	0.75	19.17	0.95	0.90	21.79	1.36	0.95	23.41
OSVD-RF	0.65	0.56	14.15	0.69	0.66	16.73	0.73	0.73	18.69	0.98	0.83	21.05
VMD-RF	0.73	0.67	16.75	0.78	0.72	18.45	0.85	0.82	20.93	1.26	0.88	22.85
OVMD-RF	0.57	0.46	12.70	0.65	0.49	15.09	0.76	0.62	17.97	0.89	0.70	20.66
OVMD-OFS$_{fs}$-RF	0.54	0.44	12.12	0.60	0.47	14.17	0.71	0.59	16.84	0.82	0.66	19.03
OVMD-OSVD-RF	0.50	0.41	11.46	0.57	0.45	13.69	0.68	0.56	15.73	0.77	0.63	18.34
OVMD-EFS-RF	0.36	0.31	8.07	0.44	0.36	10.49	0.52	0.45	13.00	0.65	0.50	15.03
OVMD-EFS$_{fs}$-RF	0.25	0.21	5.70	0.28	0.27	8.89	0.31	0.35	10.64	0.39	0.44	13.24

(5) The MAPE value of the model OVMD-EFS$_{fs}$-RF is the smallest, indicating that the model has the highest forecast accuracy, which reflected that the method based on EFS for feature selection to obtain the optimal subset can also improve the forecast accuracy in multistep WSF.

Table 19 and Fig. 13-16 were analyzed with the focus on models OSVD-RF, OVMD-RF, OVMD-OSVD-RF, OVMD-EFS-RF, OVMD-OFS$_{fs}$ and OVMD-EFS$_{fs}$-RF. Thus, the contribution of the OSVD and OVMD methods towards the improvement of forecast accuracy is evaluated. Table 19 lists the P_{INDEX} of the model. The maximum index of the model OSVD-RF indicates that the proposed model has the highest improvement. Fig. 15-16 show the trend of the models metrics as the range of forecasts increases. Specifically:

(1) Compared with OVMD-OFS$_{fs}$-RF, the proposed model has less error growth at different time scales. This shows that the prediction effect of the traditional TSD-FS-FM at different time scales is not as good as that of the proposed low information loss method.

TABLE 18. Multi-step forecast results in winter.

Model	One-step			two-step			three-step			four-step		
	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE
OS-RF	0.82	0.72	15.72	0.89	0.77	17.42	0.99	0.91	20.10	1.26	1.00	23.19
PCA-RF	0.75	0.72	13.02	0.81	0.74	15.39	0.90	0.87	18.42	1.17	0.90	20.66
OSVD-RF	0.71	0.67	12.30	0.75	0.72	13.29	0.82	0.79	16.15	0.89	0.83	18.46
VMD-RF	0.73	0.68	12.55	0.81	0.71	15.06	0.85	0.84	17.53	0.90	0.88	19.78
OVMD-RF	0.66	0.63	10.09	0.71	0.69	13.05	0.76	0.73	15.27	0.86	0.77	17.50
OVMD-OFS$_{fs}$-RF	0.63	0.59	9.97	0.68	0.64	12.00	0.72	0.70	14.15	0.81	0.75	16.33
OVMD-OSVD-RF	0.58	0.56	9.26	0.64	0.60	11.18	0.68	0.65	13.21	0.76	0.73	15.35
OVMD-EFS-RF	0.46	0.47	6.08	0.51	0.49	8.38	0.57	0.53	10.15	0.63	0.63	12.52
OVMD-EFSfs-RF	0.24	0.25	4.28	0.27	0.28	6.59	0.29	0.30	8.62	0.41	0.40	11.02

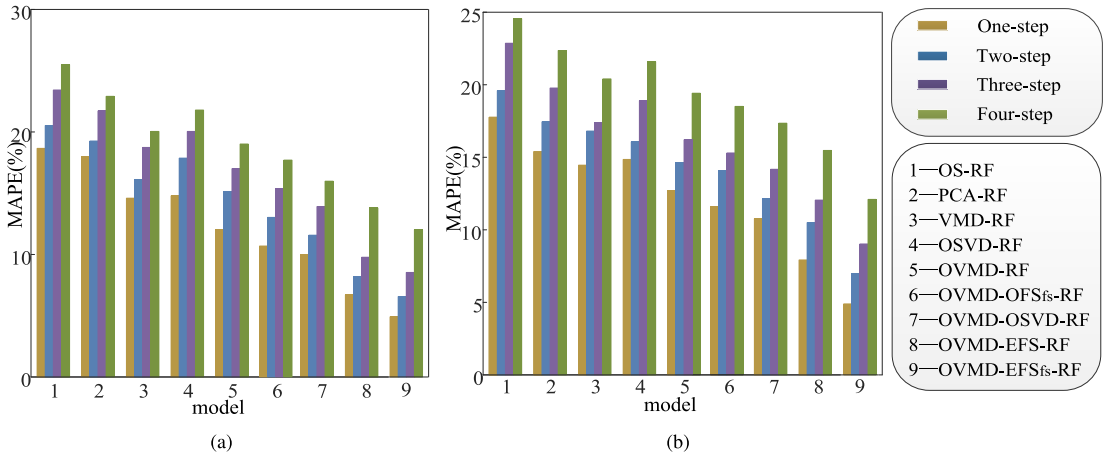


FIGURE 13. Comparison of MAPE with different forecast scales. (a) Spring. (b) Summer.

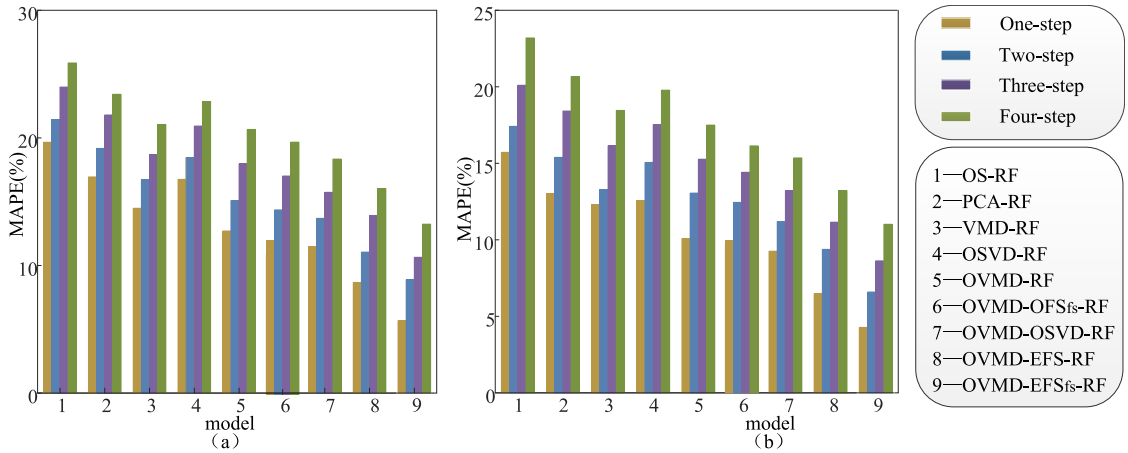


FIGURE 14. Comparison of MAPE with different forecast scales. (a) Autumn. (b) Winter.

(2) The RMSE, MAE and MAPE of the model OVMD-EFS$_{fs}$-RF have the minimum and minimum growth ratios, which indicated that its performance is better than that of other models.

(3) As the range of forecast increases, the index of the model becomes larger, which indicates that the longer the forecast time scale is, the lower the forecast accuracy of the model is.

(4) The performance of the model OVMD-RF is always better than that of the model OSVD-RF. To a certain extent, this indicates that the OVMD method contributes more to the improvement of accuracy than the OSVD method.

TABLE 19. Percentage of performance improvement of the proposed model.

	Model	One-step			two-step			three-step			four-step		
		P_{RMSE}	P_{MAE}	P_{MAPE}	P_{RMSE}	P_{MAE}	P_{MAPE}	P_{RMSE}	P_{MAE}	P_{MAPE}	P_{RMSE}	P_{MAE}	P_{MAPE}
spring	OSVD-RF	51.05	56.11	67.90	47.31	47.64	63.53	42.18	45.36	57.58	40.86	40.91	45.31
	OVMD-RF	51.41	59.15	60.07	45.20	45.20	56.62	41.18	41.03	49.81	37.50	39.74	36.80
	OVMD-OFS_{fs}-RF	45.18	47.74	55.40	44.12	42.11	51.41	36.49	38.81	43.96	35.37	36.71	29.87
	OVMD-OSVD-RF	32.42	38.08	50.75	38.08	36.50	43.28	34.64	34.75	38.66	33.69	33.62	24.79
	OVMD-EFS-RF	14.81	10.23	26.74	15.57	10.89	19.63	9.37	14.41	12.68	15.96	11.19	12.87
summer	OSVD-RF	64.67	64.26	66.15	59.87	59.45	46.84	42.18	45.36	57.58	40.28	28.10	30.27
	OVMD-RF	64.76	60.06	61.42	57.69	45.99	41.06	41.18	41.03	49.81	35.69	16.48	21.81
	OVMD-OFS_{fs}-RF	59.52	57.14	57.63	53.85	44.10	39.95	38.85	38.00	41.29	30.83	33.33	33.00
	OVMD-OSVD-RF	50.60	53.24	54.57	51.38	42.73	35.06	34.64	34.75	38.66	29.19	30.06	30.41
	OVMD-EFS-RF	38.08	36.02	38.24	37.00	25.22	19.52	9.37	14.41	12.68	23.4	16.67	15.97
autumn	OSVD-RF	62.73	62.65	60.85	56.83	53.40	58.52	59.28	48.58	48.10	35.62	36.96	16.64
	OVMD-RF	56.72	54.83	55.37	58.58	41.06	52.37	53.93	43.55	44.30	52.45	48.13	37.03
	OVMD-OFS_{fs}-RF	53.70	52.27	53.22	53.33	42.56	47.26	46.33	40.68	36.82	52.04	46.33	30.43
	OVMD-OSVD-RF	51.09	49.71	50.51	37.04	36.04	42.54	34.15	34.18	36.32	46.54	45.22	28.23
	OVMD-EFS-RF	32.34	33.27	34.60	26.88	9.72	33.59	25.21	19.18	25.10	40.00	12.00	11.84
winter	OSVD-RF	60.99	61.84	65.23	64.04	61.43	50.44	57.73	51.87	43.08	54.39	52.05	40.31
	OVMD-RF	57.53	59.82	57.60	62.00	59.81	49.54	59.23	43.64	40.81	52.45	48.13	37.03
	OVMD-OFS_{fs}-RF	55.56	57.63	56.81	60.29	56.25	45.08	58.72	41.14	39.08	49.38	46.67	32.52
	OVMD-OSVD-RF	52.61	54.18	53.79	57.51	54.16	41.07	54.19	37.68	32.37	46.54	45.22	28.23
	OVMD-EFS-RF	39.75	44.84	33.95	46.65	43.44	29.75	40.43	22.39	23.56	35.62	36.96	16.64

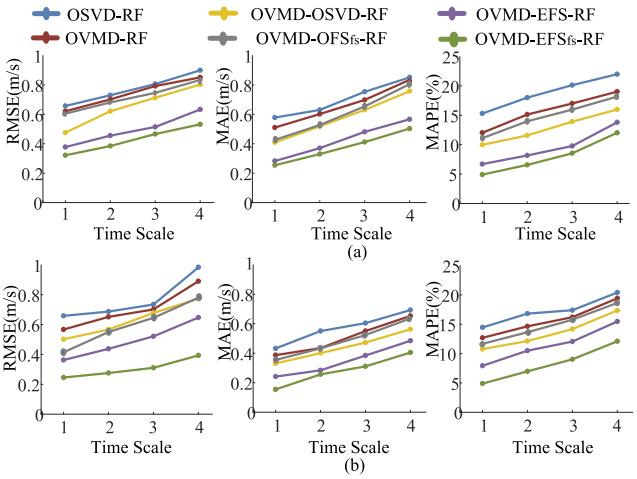


FIGURE 15. Trends in multistep prediction results for different models. (a) Spring. (b) Summer.

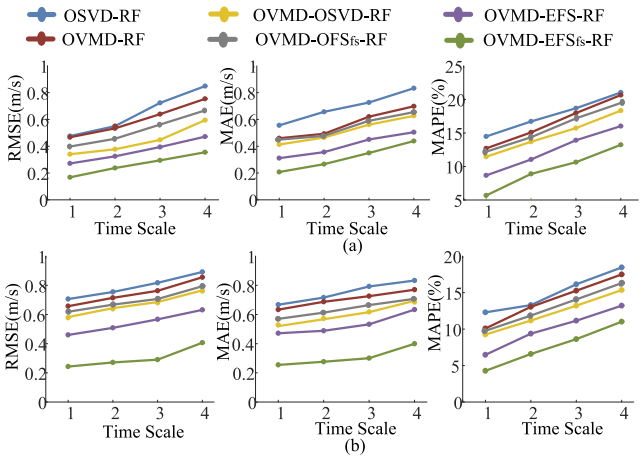


FIGURE 16. Trends in multistep prediction results for different models. (a) Autumn. (b) Winter.

G. PERFORMANCE EVALUATION

The accuracy and efficiency of the model need to be considered in practical applications. However, the paradox is that high-efficiency models have lower accuracy or higher precision models have lower efficiency. Therefore, a good WSFM should find a balance between accuracy and efficiency. In Fig. 17, the model that takes the longest time is OVMD-EFS-RF (which includes the largest number of features) and that which takes the shortest time is OSVD-RF (which includes the fewest number of features). In terms of accuracy, the proposed model is much better than several other models, and it ranks fourth in terms of efficiency. However, the efficiency of the model OVMD-OFS_{fs}-RF is ranked third and the accuracy is ranked fourth, which overall performance is far less than the proposed model. Therefore, the model achieves a good balance between accuracy and efficiency, achieving both high precision and good efficiency. Thus, it is applicable in the real world.

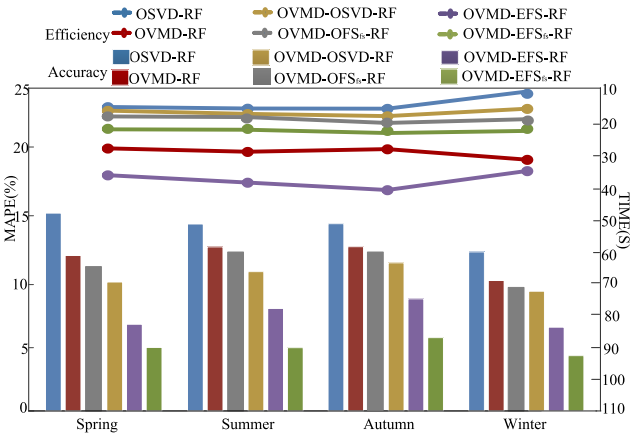


FIGURE 17. Comparison of accuracy and efficiency of the model.

V. CONCLUSION

A short-term wind speed forecast model for OVMD-EFS$_{fs}$-RF with low information loss is proposed in this paper. The effectiveness and advancement of the new method was verified by comparing the data of four data sets provided by NREL with single-step and multi-step forecast. Specifically:

(1) The original wind speed sequence is preprocessed by OVMD to reduce the influence of noise, outliers and abnormal points on forecast accuracy. The parameters of the OVMD were optimized using a center frequency-based observation method and a minimum residual criterion.

(2) The feature is generated by OSVD and combined with the original feature set to obtain EFS. The forward feature selection was then carried out for EFS. The partial generation feature is retained to reduce the information loss of the high importance feature caused by the FR method. Retaining the high importance features in the original feature set reduces the overall loss of information for the low importance features of the FS-FM and TSD-FS-FM methods.

(3) Using RF to perform forward feature selection based on EFS, determine the optimal subset, overcome the prediction difficulty caused by the excessive feature set dimension, and eliminate the negative impact of redundant information on prediction accuracy.

In the future work, a method to evaluate the information loss will be worked hard to study. The method proposed in this paper will also be more comprehensively verified.

REFERENCES

[1] J. Wang, W. Zhang, J. Wang, T. Han, and L. Kong, "A novel hybrid approach for wind speed prediction," *Inf. Sci.*, vol. 273, pp. 304–318, Jul. 2014.

[2] P. S. Georgilakis, "Technical challenges associated with the integration of wind power into power systems," *Renew. Sustain. Energy Rev.*, vol. 12, no. 3, pp. 852–863, 2008.

[3] H. Z. Wang, G. B. Wang, G. Q. Li, J. C. Peng, and Y. T. Liu, "Deep belief network based deterministic and probabilistic wind speed forecasting approach," *Appl. Energy*, vol. 182, pp. 80–93, Nov. 2016.

[4] Z.-S. Zhang, Y.-Z. Sun, and L. Cheng, "Potential of trading wind power as regulation services in the California short-term electricity market," *Energy Policy*, vol. 59, pp. 885–897, Aug. 2013.

[5] J. Wang and J. Hu, "A robust combination approach for short-term wind speed forecasting and analysis—Combination of the ARIMA (autoregressive integrated moving average), ELM (extreme learning machine), SVM (support vector machine) and LSSVM (least square SVM) forecasts using a GPR (Gaussian process regression) model," *Energy*, vol. 93, pp. 41–56, Dec. 2015.

[6] W. Zhang, Z. Qu, K. Zhang, W. Mao, Y. Ma, and X. Fan, "A combined model based on CEEMDAN and modified flower pollination algorithm for wind speed forecasting," *Energy Convers. Manage.*, vol. 136, pp. 439–451, Mar. 2017.

[7] J. Zhao, Z.-H. Guo, Z.-Y. Su, Z.-Y. Zhao, X. Xiao, and F. Liu, "An improved multi-step forecasting model based on WRF ensembles and creative fuzzy systems for wind speed," *Appl. Energy*, vol. 162, pp. 808–826, Jan. 2016.

[8] M. Lydia, S. S. Kumar, A. I. Selvakumar, and G. E. P. Kumar, "Linear and non-linear autoregressive models for short-term wind speed forecasting," *Energy Convers. Manage.*, vol. 112, pp. 115–124, Mar. 2016.

[9] E. Erdem and J. Shi, "ARMA based approaches for forecasting the tuple of wind speed and direction," *Appl. Energy*, vol. 88, no. 4, pp. 1405–1414, 2011.

[10] R. G. Kavasseri and K. Seetharaman, "Day-ahead wind speed forecasting using f-ARIMA models," *Renew. Energy*, vol. 34, no. 5, pp. 1388–1393, 2009.

[11] Z. Song, Y. Jiang, and Z. Zhang, "Short-term wind speed forecasting with Markov-switching model," *Appl. Energy*, vol. 130, no. 3, pp. 103–112, 2014.

[12] S. Baran, "Probabilistic wind speed forecasting using Bayesian model averaging with truncated normal components," *Comput. Statist. Data Anal.*, vol. 75, no. 3, pp. 227–238, 2014.

[13] H. Mori and S. Okura, "An ANN-based method for wind speed forecasting with S-transform," in *Proc. IEEE Region 10 Conf.*, Nov. 2016, pp. 642–645.

[14] P. Jiang, Y. Wang, and J. Wang, "Short-term wind speed forecasting using a hybrid model," *Energy*, vol. 119, pp. 561–577, Jan. 2017.

[15] D. Bai, J. He, and X. Wang, "Combination model for forecasting wind speed forecasting wind speed based on adaptive PSO-ELM," *Acta Energiae Solaris Sinica*, vol. 36, no. 3, pp. 792–797, 2015.

[16] Z. Yihui *et al.*, "A hybrid short-term wind speed forecasting model based on ensemble empirical mode decomposition and improved extreme learning machine," *Power Syst. Protection Control*, vol. 10, no. 42, pp. 29–34, Dec. 2017.

[17] X. Luo, J. Sun, L. Wang, W. Wang, W. Zhao, J. Wu, J.-H. Wang, and Z. Zhang, "Short-term wind speed forecasting via stacked extreme learning machine with generalized correntropy," *IEEE Trans. Ind. Informat.*, vol. 14, no. 11, pp. 4963–4971, Nov. 2018.

[18] Z. Sun, H. Sun, and J. Zhang, "Multistep wind speed and wind power prediction based on a predictive deep belief network and an optimized random forest," *Math. Problems Eng.*, vol. 7, no. 10, p. 115, Jul. 2018.

[19] L. Ran, K. Yong-qin, and Z. Xiao-Qian, "Forecasting of wind speed with least squares support vector machine based on genetic algorithm," in *Proc. Int. Conf. Consum. Electron., Commun. Netw.*, Apr. 2011, pp. 358–361.

[20] Y. U. Ao, L. Chen, and J. T. Peng, "Short-term wind speed forecasting by using hysteretic ELM model," *Comput. Technol. Develop.*, vol. 27, no. 6, pp. 131–141, Jun. 2017.

[21] Y. Ren, P. N. Suganthan, and N. Srikanth, "A comparative study of empirical mode decomposition-based short-term wind speed forecasting methods," *IEEE Trans. Sustain. Energy*, vol. 6, no. 1, pp. 236–244, Dec. 2017.

[22] H. Liu, X. Mi, and Y. Li, "An experimental investigation of three new hybrid wind speed forecasting models using multi-decomposing strategy and ELM algorithm," *Renew. Energy*, vol. 123, pp. 694–705, Aug. 2018.

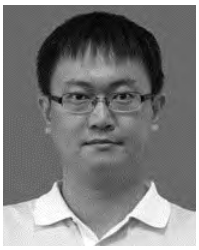
[23] S. Wang, N. Zhang, L. Wu, and Y. Wang, "Wind speed forecasting based on the hybrid ensemble empirical mode decomposition and GA-BP neural network method," *Renew. Energy*, vol. 94, pp. 629–636, Aug. 2016.

[24] C. Li, Z. Xiao, X. Xia, W. Zou, and C. Zhang, "A hybrid model based on synchronous optimisation for multi-step short-term wind speed forecasting," *Appl. Energy*, vol. 215, pp. 131–144, Apr. 2018.

[25] H. Liu, H.-Q. Tian, D.-F. Pan, and Y.-F. Li, "Forecasting models for wind speed using wavelet, wavelet packet, time series and artificial neural networks," *Appl. Energy*, vol. 107, pp. 191–208, Jul. 2013.

[26] C. Skittides and W. G. Früh, "Wind forecasting using principal component analysis," *Renew. Energy*, vol. 69, no. 3, pp. 365–374, Sep. 2014.

[27] M. Yousefi, D. Hooshyar, M. Yousefi, W. Khaksar, K. S. M. Sahari, and F. B. I. Alnaimi, "An artificial neural network hybrid with wavelet transform for short-term wind speed forecasting: A preliminary case study," in *Proc. Int. Conf. Sci. Inf. Technol.*, Oct. 2015, pp. 95–99.
[28] M. Ali, A. Khan, and N. U. Rehman, "Hybrid multiscale wind speed forecasting based on variational mode decomposition," *Int. Trans. Elect. Energy Syst.*, vol. 28, no. 1, 2018, Art. no. e2466.
[29] S. Qiu, J. Wang, C. Tang, and D. Du, "Comparison of ELM, RF, and SVM on E-nose and E-tongue to trace the quality status of mandarin (*Citrus unshiu* Marc.)," *J. Food Eng.*, vol. 166, pp. 193–203, Dec. 2015.
[30] G. Chandrashekar and F. Sahin, *A Survey on Feature Selection Methods*. New York, NY, USA: Pergamon, 2014.
[31] R. Kohavi and G. H. John, "Wrappers for feature subset selection," *Artif. Intell.*, vol. 97, nos. 1–2, pp. 273–324, 1997.
[32] D. E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*, vol. 8, no. 7. 1989, pp. 2104–2116.
[33] S. Wan, X. Zhang, B. Nan, and L. Zhang, "Fault diagnosis of rolling bearing based on PPCA and 1.5-dimensional energy spectrum," *Electr. Power Automat. Equip.*, vol. 38, no. 6, p. 25, 2018.
[34] G. Chen and Q. Zeng, "Face recognition method based on LLE algorithm," *J. Comput. Appl.*, vol. 10, no. 24, pp. 176–177, Jun. 2007.
[35] C. Zhang, J. Zhou, C. Li, W. Fu, and T. Peng, "A compound structure of ELM based on feature selection and parameter optimization using hybrid backtracking search algorithm for wind speed forecasting," *Energy Convers. Manage.*, vol. 143, pp. 360–376, Jul. 2017.
[36] C. Zhang, H. Wei, J. Zhao, T. Liu, T. Zhu, and K. Zhang, "Short-term wind speed forecasting using empirical mode decomposition and feature selection," *Renew. Energy*, vol. 96, pp. 727–737, Oct. 2016.
[37] J. Chorowski, J. Wang, and J. M. Zurada, "Review and performance comparison of SVM-and ELM-based classifiers," *Neurocomputing*, vol. 128, pp. 507–516, Mar. 2014.
[38] H. Zhou, J. Q. He, and Z. B. Chen, "Regional wind energy resource forecasting based on SVD and support vector machine," *Adv. Mater. Res.*, vols. 1070–1072, pp. 247–252, Dec. 2014.
[39] A. A. Abdoos, "A new intelligent method based on combination of VMD and ELM for short term wind power forecasting," *Neurocomputing*, vol. 203, pp. 111–120, Aug. 2016.
[40] J. Wang, W. Yang, P. Du, and T. Niu, "A novel hybrid forecasting system of wind speed based on a newly developed multi-objective sine cosine algorithm," *Energy Convers. Manage.*, vol. 163, pp. 134–150, May 2018.
[41] P. Jiang and C. Li, "Research and application of an innovative combined model based on a modified optimization algorithm for wind speed forecasting," *Measurement*, vol. 124, pp. 395–412, Aug. 2018.
[42] D. Wang, H. Luo, O. Grunder, and Y. Lin, "Multi-step ahead wind speed forecasting using an improved wavelet neural network combining variational mode decomposition and phase space reconstruction," *Renew. Energy*, vol. 113, pp. 1345–1358, Dec. 2017.
[43] X. Kong, X. Liu, R. Shi, and K. Y. Lee, "Wind speed prediction using reduced support vector machines with feature selection," *Neurocomputing*, vol. 169, pp. 449–456, Dec. 2015.
[44] N. Huang, E. Xing, G. Cai, Z. Yu, B. Qi, and L. Lin, "Short-term wind speed forecasting based on low redundancy feature selection," *Energies*, vol. 11, no. 7, p. 1638, 2018.
[45] T. Niu, J. Wang, K. Zhang, and P. Du, "Multi-step-ahead wind speed forecasting based on optimal feature selection and a modified bat algorithm with the cognition strategy," *Renew. Energy*, vol. 118, pp. 213–229, Apr. 2018.
[46] K. Shi, Y. Qiao, W. Zhao, Q. Wang, M. Liu, and Z. Lu, "An improved random forest model of short-term wind-power forecasting to enhance accuracy, efficiency, and robustness," *Wind Energy*, vol. 21, no. 12, pp. 1383–1394, 2018.
[47] A. Lahouar and J. B. H. Slama, "Hour-ahead wind power forecast based on random forests," *Renew. Energy*, vol. 109, pp. 529–541, Aug. 2017.
[48] E. J. Natenberg, J. Zack, J. Manobianco, G. Van Knowe, T. Melino, and D. J. Gagne, "Application of a random forest approach to model output statistics for use in day ahead wind power forecasts," in *Proc. Symp. Role Stat. Methods Weather Climate Predict.*, vol. 5, 2013, pp. 235–241.
[49] H. Nantian, W. Da, and L. Zaiming, "Characteristics of power quality complex perturbation in complex noise environment," *Chin. J. Sci. Instrum.*, vol. 39, no. 4, pp. 83–90, Apr. 2018.
[50] N. Huang, G. Lu, and D. Xu, "A permutation importance-based feature selection method for short-term electricity load forecasting using random forest," *Energies*, vol. 9, no. 10, p. 767, 2016.



NANTIAN HUANG was born in 1980. He received the B.S. degree from the Harbin Institute of Technology, Harbin, China, in 2003, the M.S. degree from the Dalian University of Technology, Dalian, China, in 2006, and the Ph.D. degree from the Harbin Institute of Technology, Harbin, China, in 2012. He is currently an Associate Professor with the School of Electrical Engineering, Northeast Electric Power University.

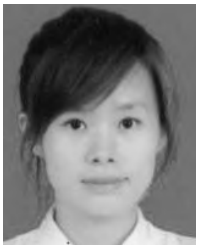
YINYIN WU was born in 1994. She received the B.S. degree in electronic information engineering from Northeast Electric Power University, China, in 2017, where she is currently pursuing the M.S. degree with the Department of Electrical Engineering.

GUOWEI CAI was born in 1968. He received the B.S. and M.S. degrees from Northeast Electric Power University, China, in 1990 and 1993, respectively, and the Ph.D. degree from the Harbin Institute of Technology, Harbin, China, in 1999. Since 2004, he has been a Professor with the School of Electrical Engineering, Northeast Electric Power University, Jilin, China.

HEYAN ZHU was born in 1987. He received the master's degree from the University of Bath, in 2012. He is currently an Engineer with a middle-level title. He is also with the Economic Technology Research Institute, State Grid Liaoning Electric Power Company, Ltd. He is a professional in power systems.

CHANGYONG YU was born in 1975. He received the bachelor's degree in relay protection and automatic telecontrol from Northeastern Electric Power University, in 1998. He is currently a Senior Engineer with a senior title. He is also with the Economic Technology Research Institute, State Grid Liaoning Electric Power Company, Ltd. His main research interest includes the power systems.

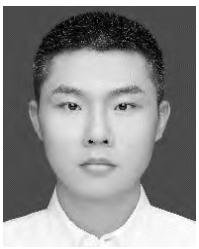
LI JIANG was born in 1983. He received the master's degree in electrical engineering and automation from the North China Electric Power University, Baoding, in 2013. He is currently a Senior Engineer with a senior title. He is also with the Economic Technology Research Institute, State Grid Liaoning Electric Power Company, Ltd. His main research interest includes the power systems.



YE ZHANG was born in 1989. She received the master's degree in electrical engineering and automation from Shenyang Engineering College, in 2011.

She is currently a master engineer with a mid-level title. She is also with the Liaoning Electric Power Trading Center Company, Ltd. Her research interest includes the power systems.

JIANSEN ZHANG was born in 1982. He received the master's degree in agricultural engineering from the Shenyang University of Technology, in 2011. He is currently an Engineer with a middle-level title. He was with the Shenyang Power Supply Company, State Grid Liaoning Power Company, Ltd. His research interest includes agricultural electrification and automation.



ENKAI XING was born in 1993. He received the B.S. degree in electronic information engineering from Northeast Electric Power University, China, in 2017, where he is currently pursuing the M.S. degree with the Department of Electrical Engineering.