Ultra-short-time prediction technology of wind power station output based on variational mode decomposition and particle swarm optimization least squares vector machine

Runjie Shen[1], Danqiong Hua[1*], Yiying Wang[1], Ruimin Xing[1], Min Ma [1,2]

[1] College of Electronic and Information Engineering, Tongji University, Shanghai, 201800, China
[2]Electric Power Research Institute , State Grid Gansu Electric Power Company，Lanzhou, Gansu, 730000, China

Abstract. Wind power is developing rapidly in the context of sustainable development, and a series of problems such as wind curtailment and power curtailment have gradually emerged. The forecast of power generation output has become one of the hotspots of current research. This paper proposes a wind power plant output ultra-short-time prediction technology based on variational modal decomposition and particle swarm optimization least squares vector machine. Variational Modal Decomposition (VMD) method decomposes the historical output data of wind power plants at multiple levels. At the same time, it explores the impact of various decomposition methods such as EMD decomposition on the prediction accuracy, and uses the least squares support vector machine based on particle swarm optimization algorithm. Predictive summation is performed on each level of data separately to obtain a more accurate prediction effect, which has a certain improvement in prediction accuracy compared with traditional prediction algorithms.

1 Introduction

Wind energy is a kind of renewable energy, which is characterized by a wide range of distribution and low energy density, but it is mostly in a state of free movement in the atmosphere and has low stability. Wind power generation is the most important form of wind energy utilization, and it is also one of the areas with relatively mature technology and the most commercial development prospects in the development and utilization of new energy in the world today.

Due to the inherent random volatility and intermittent nature of wind power, it will bring a series of problems to the system after wind power is connected to the grid. The output of conventional units can be accurately predicted, so the economy of unit combination can be guaranteed [1]. In addition, the anti-peak shaving characteristics of wind power will increase the reserve capacity of the system, and the rapid changes in the output of wind turbines will seriously threaten the reliability of the power system. Therefore, it is of great significance to study wind power output forecasting methods.

At present, many methods have been proposed for wind power output forecasting at home and abroad, such as short-term wind speed and power generation forecasting based on neural network combination prediction [2], short-term wind speed combination prediction based on empirical mode decomposition and

neural network [3], AR_ARIMA Model output prediction. [4]. These single methods are well applied in some specific fields, but the accuracy cannot meet the requirements of the increasingly large wind power grid connection, so a combination method appears to obtain higher prediction accuracy. The combined method makes full use of the advantages of each single method, selects the best model for different situations, and greatly reduces the prediction error. Based on this, this paper proposes a method of applying a hybrid algorithm of variational modal decomposition and particle swarm optimization least squares vector machine to predict the ultra-short-time output of wind farms.

2 Wind power output prediction algorithm flow

2.1 Overall framework of wind power output prediction algorithm

Taking into account the direct correlation between historical output data and wind speed data, historical output data will also have the followability and uncertainty of wind speed data. Directly using a single historical output data to predict will lead to larger forecast errors. In order to improve forecast accuracy, When training the prediction model, to reduce the data complexity, firstly, the original wind speed data of the training set is decomposed by the variational modal

---

[*]Corresponding author's e-mail: 821140020@qq.com

decomposition (VMD) to generate multiple sub-sequences.The forecasting process is shown in Figure 1.
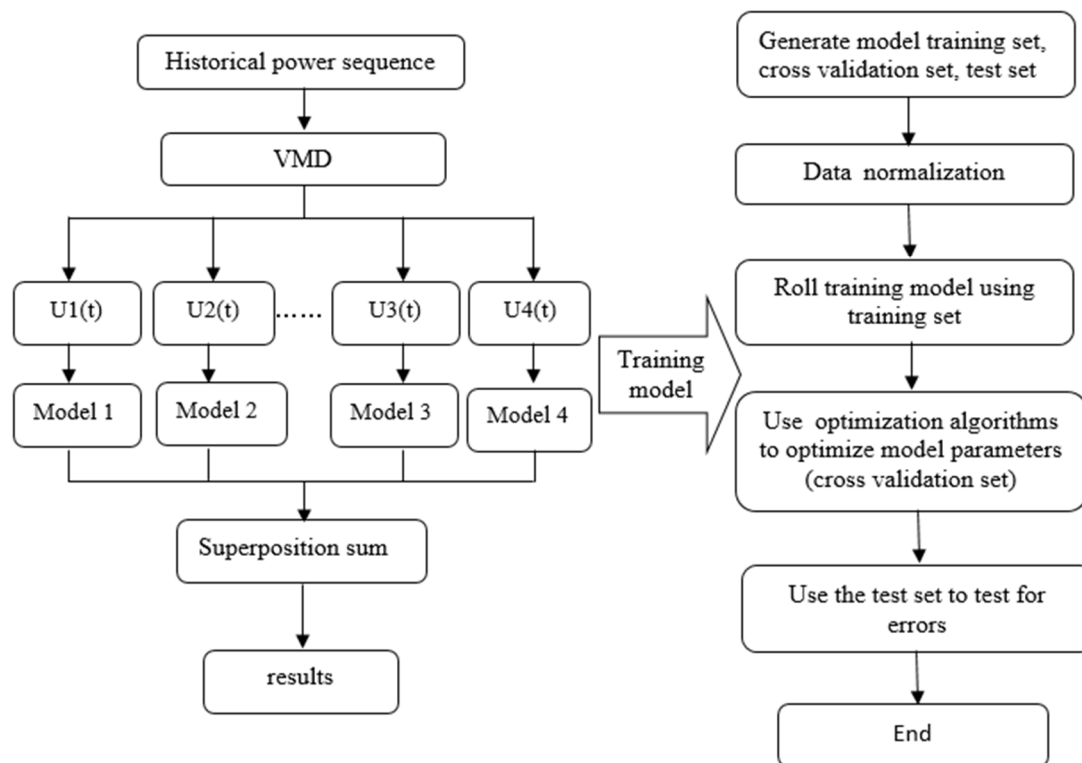

Figure 1. The overall process of forecasting

The process first uses variational modal decomposition to smooth the historical output data sequence of wind power plants, and decomposes it into a series of modal components with different center frequencies to reduce the complexity of the wind speed sequence; then based on each modal component The sequence established LSSVM sub-models, and applied particle swarm optimization (PSO) to optimize its model parameters to improve prediction accuracy. In the rolling training, the error evaluation index is fed back in real time. If the error evaluation index is not satisfied, the LSSVM sub-model of this level is retrained. Finally, the predicted value of each modal component that meets the error requirement is superimposed to obtain the predicted value of output.

2.2 Research on Wind Farm Output Forecasting Model Based on Variational Mode Decomposition and Particle Swarm Optimization (VMD-PSO-LSSVM)

2.2.1 Variational Mode Decomposition (VMD).
Variational modal decomposition (VMD) was proposed by Dragomiretskiy K. and Zosso D. in 2014 [5]. The algorithm is a complete non-recursive variational modal decomposition model, in which the modes are extracted at the same time. The model looks for a set of modes and their respective center frequencies, so that the modes can reproduce the input signal together, and each All modes are smooth after demodulation into baseband. Since the algorithm uses

non-recursive mode decomposition, compared with the existing mode decomposition model, this model solves the problems of recursive mode decomposition and has good performance. It is aimed at empirical mode decomposition (EMD) for similar frequencies (The shortcoming that the components of <f2<2f1) cannot be separated correctly is solved by the application of this method. Experiments have proved that the variational modal decomposition (VMD) can successfully separate two signals with similar frequencies. More importantly, the model is more robust to sampling and noise.

2.2.1.1 VMD algorithm principle

Before proceeding with the variational modal decomposition (VMD) solution process, three important concepts need to be mastered first: 1) classical Wiener filtering, 2) Hilbert transform, 3) frequency mixing and extrapolation demodulation.
1) Classical Wiener filtering:
   Wienerfiltering is an optimal estimator for stationary processes based on the minimum mean square error criterion. The mean square error between the output of this filter and the expected output is the smallest, so it is an optimal filtering system.
2) Hilbert transform:
   The Hilbert transform of a continuous time signal x(t) is equal to the output response Xh(t) after the signal passes through a linear system

with impulse response. The Hilbert transform is defined as follows:

$$\phi_k(\vec{r}) = (2\pi)^{2/3} \exp\left(i\vec{k}\cdot\vec{r}\right)$$

(1)

3) Frequency mixingand extrapolation demodulation:

Mixing is the process of combining two signals non-linearly to introduce a cross-frequency term in the output. The simplest mixer is multiplication. Multiplying the two real signals with frequencies w1 and w2 will produce mixed frequencies in the output of w1-w2 and w1+w2. This principle can be explained by the following trigonometric identities:

$$2\cos(\omega_1 t)\cos(\omega_2 t) = \cos((\omega_1 + \omega_2)t) + \cos((\omega_1 - \omega_2)t)$$

(2)

The variational modal decomposition (VMD) process is essentially the process of solving the variational problem. Specifically, the historical output data sequence of the wind power station is regarded as a non-stationary signal. The variational problem can be understood as:

1) The structure of the variational problem.

Assuming that each'mode' is a limited bandwidth with a center frequency, look for K modal functions uk(t) (k=1,2,...,K) so that the sum of the estimated bandwidth of each mode is the smallest, and the constraint The condition is that the sum of each mode is equal to the input non-stationary signal. The specific construction steps are as follows:

a) Perform Hilbert transform on the analytical signal of each modal function uk(t) to obtain the unilateral frequency spectrum of each modal function uk(t):

$$\left(\delta(t) + \frac{j}{\pi t}\right) * u_k(t)$$

$$\delta(t) = \begin{cases} 0 & t \neq 0 \\ \infty & t = 0 \end{cases}, \quad \int_{-\infty}^{\infty} \delta(t)dt = 1$$

(3)

b) Mix the analytical signals of each mode-estimate the center frequency $e^{-jw_k t}$ as a reference to modulate the frequency spectrum of each mode to the corresponding base band:

$$\left[\left(\delta(t) + \frac{j}{\pi t}\right) * u_k(t)\right]e^{-jw_k t}$$

(4)

Among them, $e^{-jw_k t}$ is the vector description of the center frequency in the complex plane.

c) Calculate the squared L2 norm of the above demodulated signal gradient and estimate the bandwidth of each modal signal. The constrained variational problem is expressed as follows:

$$\begin{cases} \min\left\{\sum_k \left\|\partial_t\left[\left(\delta(t) + \frac{j}{\pi t}\right) * u_k(t)\right]e^{-jw_k t}\right\|^2\right\} \\ \text{s.t.} \sum_k u_k = f \end{cases}$$

(5)

Among them, {uk}={u1,u2,……,uk}；{wk}={w1,w2,……,wk}.

2) The solution of variational problems.

a) The second penalty factor C and Lagrangian multiplication operator $\lambda(t)$ are introduced to turn the constrained variational problem into a non-constrained variational problem. The quadratic penalty factor C can ensure the accuracy of signal reconstruction in the presence of Gaussian noise , Lagrangian operator keeps the constraint condition strict, and the extended Lagrangian expression is as follows:

$$L(\{u_k\},\{w_k\},\lambda) := C\sum_{k=1}^{K}\left\|\partial_t\left[\left(\delta(t) + \frac{j}{\pi t}\right)u_k(t)\right]e^{-jw_k t}\right\|_2^2 + \left\|f(t) - \sum_{k=1}^{K}u_k(t)\right\|_2^2 + \left\langle\lambda(t), f(t) - \sum_{k=1}^{K}u_k(t)\right\rangle$$

(6)

b) In VMD, an alternating direction multiplier algorithm is used to solve the listed variational problems. Seek to expand the 'saddle point' of Lagrange's expression by alternately updating u_k^{n+1}、ω_k^{n+1} and λ^{n+1}. Among them, Fourier equidistant transform can be used to transform to frequency domain:

$$\hat{u}_k^{n+1} = \underset{\hat{u}_k \in X}{\arg\min}\left\{C\left\|j\omega\{1 + \text{sgn}(\omega + \omega_k)\hat{u}_k(\omega + \omega_k)\}\right\|_2^2 + \left\|\hat{f}(\omega) - \sum_{k=1}^{K}\hat{u}_k(\omega) + \frac{\hat{\lambda}(\omega)}{2}\right\|_2^2\right\}$$

(7)

Where: ω is the random frequency, and X is the set of u_k all possible values.

ω Will be replaced by $\omega - \omega_k$, and its non-negative frequency interval integral form is:

$$\hat{u}_k^{n+1} = \underset{\hat{u}_k, u_k \in X}{\arg\min}\left\{\int_0^{\infty}\left[4C(\omega - \omega_k)^2|\hat{u}_k(w)|^2 + 2\left|\hat{f}(\omega) - \sum_{k=1}^{K}\hat{u}_k(\omega) + \frac{\hat{\lambda}(\omega)}{2}\right|^2\right]d\omega\right\}$$

(8)

At this point, the solution of the secondary optimization problem is:

$$\hat{u}_k^{n+1}(\omega) = \frac{\hat{f}(w) - \sum_{k=1}^{K}\hat{u}_k(\omega) + \frac{\hat{\lambda}(w)}{2}}{1 + 2C(\omega - \omega_k)^2}$$

(9)

According to the same process, first convert the value of the center frequency to the frequency domain:

$$\omega_k^{n+1} = \underset{\omega_k}{\arg\min}\left\{\int_0^{\infty}(\omega - \omega_k)^2|\hat{u}_k(\omega)|^2 d\omega\right\}$$

(10)

Solution to update method of center frequency:

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega |\hat{u}_k(\omega)|^2 \, d\omega}{\int_0^\infty |\hat{u}(\omega)|^2 \, d\omega} \quad (11)$$

Where: $\hat{u}_k^{n+1}(\omega)$ is the Wiener filter of the current residual $\hat{f}(\omega) - \sum_{k=1}^K \hat{u}_k(\omega)$; ω_k^{n+1} is the center of gravity of the power spectrum of the current modal function; the inverse Fourier transform is performed on $\hat{u}_k(\omega)$, and the actual part is $\{u_k(t)\}$.

2.2.1.2 VMD algorithm flow

1) Initialize $\{\hat{u}_k^1\}$, $\{\omega_k^1\}$, $\{\hat{\lambda}^1\}$, and the number of iterations n is set to 1.

2) Update u_k and ω_k according to equations (9) and (11);

3) Update λ according to equations (12):

$$\hat{\lambda}^{n+1}(\omega) \leftarrow \hat{\lambda}^n(\omega) + \tau[\hat{f}(\omega) - \sum_k \hat{u}_k^{n+1}(\omega)] \quad (12)$$

In the formula, τ is noise tolerance parameter, the default is 0, its meaning is to remove the strong noise contained in the signal. The default value selected in this paper needs to be determined by repeated experiments in consideration of the distortion caused by denoising in actual application.

4) For a given discrimination accuracy e>0, judge whether the convergence condition is satisfied according to formula (13), if it is satisfied, stop the iteration, otherwise n increment to n+1, and return to step 2)

$$\sum_k \left\| \hat{u}_k^{n+1} - \hat{u}_k^n \right\|_2^2 \Big/ \left\| \hat{u}_k^n \right\|_2^2 < e \quad (13)$$

5) According to the given modal number, the corresponding modal sub-sequence is obtained for further construction of the hybrid model.

2.2.2 Particle swarm optimization algorithm (PSO).

The particle swarm optimization algorithm was proposed by Kennedy and Eberhar [6] in 1995. It is an optimization algorithm based on swarm intelligence in the field of intelligent computing. It searches through particles following the optimal particles in the solution space. The particle swarm algorithm imitates the swarming behavior of insects and birds. These groups search for food in a cooperative manner. Each member of the group continuously changes its search mode by learning from its own experience and the experience of other members.

The PSO algorithm first initializes a group of random particles and searches for the optimal value through multiple iterations. In each iteration, the particle updates itself by tracking two "extreme values" [7]: individual extremum p_{best} and global extremum g_{best}. The individual extremum is the optimal position that a particle has passed through, and the global extremum is the best one of the optimal positions reached by all particles in the individual particle swarm. The particle swarm updates its speed and position according to the two extreme values mentioned above. When these two optimal values are found, the example updates its speed and position according to formula (14) and formula (15):

$$v = \omega \times v + C_1 \times Rand() \times (p_{best} - x) + C_2 \times Rand() \times (g_{best} - x) \quad (14)$$

$$x = x + v \quad (15)$$

Among them: v is the velocity of the particle; x is the position of the current particle; $Rand()$ is a random number between (0,1); C_1, C_2 is the learning factor, usually between [0,2], this article chooses C_1=1.5, C_2=1.7.

The PSO algorithm is a random and parallel optimization algorithm, which does not require the optimized function to be differentiable, differentiable, continuous, etc., and has a fast convergence speed, a simple algorithm, and easy programming. The flowchart is shown in Figure 2, and the algorithm flow is as follows:

1) Initialize the particle swarm, including the size of the swarm N, the position of each particle x_i and v_i;

2) Calculate the fitness value of each particle $F_n[i]$;

3) For each particle, compare its fitness value $F_n[i]$ with the individual extreme value $p_{best}(i)$, if $F_n[i] > p_{best}(i)$, replace $p_{best}(i)$ with $F_n[i]$;

4) For each particle, compare its fitness value $F_n[i]$ with the global extremum g_{best}, if $F_n[i] > g_{best}$, replace g_{best} with $F_n[i]$;

5) According to the formula(14) and (15),update the particle speed v_i and position x_i;

6) If the end condition is met (the error is good enough or the maximum number of cycles is reached) then exit, otherwise return 2).
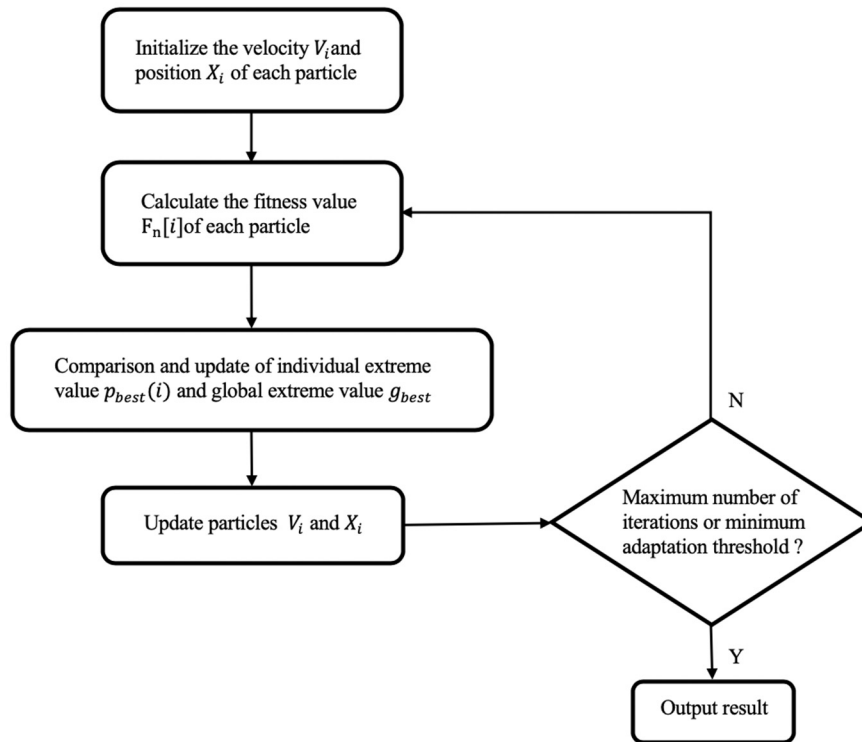
Figure 2. Flow chart of particle swarm optimization algorithm.

2.2.3 Least Squares Support Vector Machine (LSSVM). Support Vector Machine (SVM) is a data-based machine learning method originally established by Vapnik [8] on the basis of statistical learning theory. Least squares vector machine (LSSVM) is a support vector machine (SVM) Expansion and improvement. The least squares vector machine (LSSVM) replaces the inequality constraints of the support vector machine (SVM) optimization problem with the equality constraints, which improves the convergence speed of the algorithm, so the solution process of the least squares vector machine (LSSVM) is linear The calculation efficiency of equation solving is higher than that of support vector machine (SVM) [9].

The principle of the regression model of the least squares support vector machine (LSSVM) is as follows. The sample set can be expressed as $\{x_i, y_i\}_{i=1}^{l}$, $x_i \in R^n$ is the input vector of the i-th sample, $y^i \in R$ is the target value of the i-th sample, and l is the number of training samples. Non-linear mapping $\phi(\cdot)$ maps the samples to the high-dimensional feature space. The LSSVM model in the feature space can be expressed as:

$$f(x) = w^T \phi(x) + b \quad (16)$$

Where: $\phi(\cdot)$ is the non-linear transformation mapping function; w is the vector of feature space weight coefficients; b is the offset.

The objective function of Least Squares Support Vector Machine (LSSVM) can be described as:

$$\min J \left(\omega^T \xi\right) = \frac{1}{2}\omega^T\omega + \frac{\gamma}{2}\sum_{i=1}^{l}\xi^2 \quad i = 1, 2, \ldots, l \quad (17)$$

Where: ξ is the error variable; $\gamma > 0$ is the penalty coefficient.

Introduce the Lagrangian function to solve it and get:

$$L(\omega, B, \xi, \alpha) = J(\omega, \xi) - \sum_{i=1}^{l}\alpha_i[\omega^T\varphi(x_i) + b + \xi - y_i] \quad (18)$$

The commonly used kernel functions of Least Squares Support Vector Machine (LSSVM) mainly consist of polynomial kernel functions and linear kernel functions and radial basis kernel functions (RBF). Due to its strong generalization ability, the radial basis kernel function has relatively few parameters and fewer numerical constraints, it can reduce model complexity and improve model training speed. Therefore, this paper chooses radial basis function as the kernel function of least squares support vector machine (LSSVM), and its expression is:

$$R(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|}{2\sigma^2}\right) \quad (19)$$

Therefore, in the LSSVM model based on the radial basis function (RBF), the kernel width σ of the radial basis function (RBF) and the penalty factor C are the main parameters that affect its performance. In order to

improve the prediction accuracy, the above-mentioned particle swarm optimization is used The algorithm optimizes these two parameters [10].

3 Analysis of Wind Farm Output Forecast Results Based on Variational Mode Decomposition and Particle Swarm Optimization (VMD-PSO-LSSVM)

3.1 data collection

The research data in this article comes from Fujian Huadian Kemen Wind Power Station. Kemen Wind Power Station is located in the eastern coastal area of Fujian. The wind speed, output and other data change obviously with the seasons and weather. There is a certain seasonal regularity while maintaining the followability of the wind speed series. , Non-linear and other characteristics.

3.2 Data normalization processing

Data normalization is a method of processing data at the input of a neural network. It uses a specific normalization function to transform data sets that are originally in different value ranges or a larger range of values. The purpose is to cancel each dimension. The data sees the magnitude difference, so as to avoid the large difference in the magnitude of the input and output data sets, which may cause large network prediction errors. This paper uses the maximum-minimum method, the Mapmaxmin

function, as the data normalization method. The function form is as follows:

$$x_k = (x_k - x_{min})/(x_{max} - x_{min}) \qquad (20)$$

Where x_{min} is the minimum value in the data sequence, X_{max} is the maximum value in the data sequence.

3.3 Rolling forecast strategy

In order to improve the prediction accuracy, it is necessary to maintain the time-series correlation of the data at the prediction time. Therefore, the historical data used in the training model must be close to the prediction time in real time. Therefore, the VMD-PSO-LSSVM model is trained in real time by rolling training, and each prediction takes 4 hours as a step. Each step includes 16 time points (one prediction point every 15 minutes, 16 in 4 hours in total). When the next next step is predicted, the training set is updated, and the 16 time points predicted by the previous step are added to the end of the training set. And remove the true value of the 16 time points at the front end of the training set, and use the updated data set to retrain the model, which not only ensures the stability of the number of training sets and saves model training time, but also continues the time series correlation and data The authenticity guarantees the prediction accuracy.

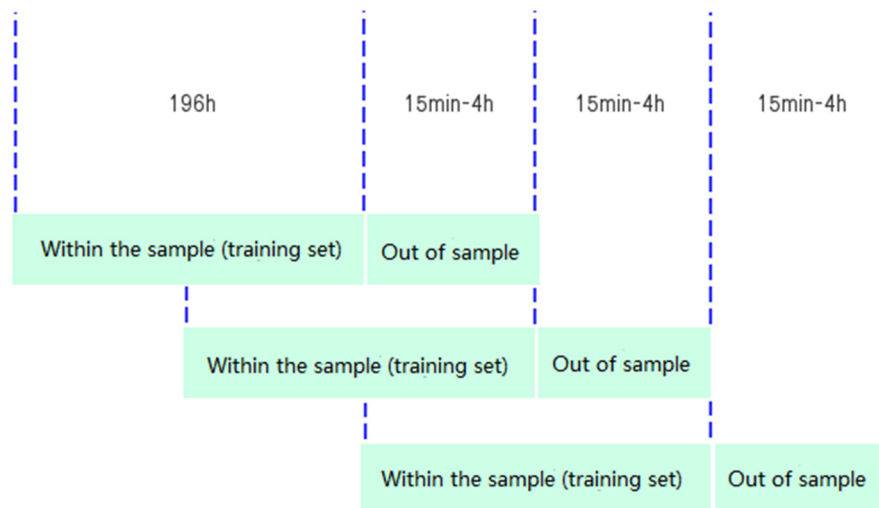The data rolling training update method is shown in Figure 3:

Figure 3. Data rolling training update method diagram

3.4 Result analysis

3.4.1 Comparison of data by EMD and VMD

The empirical mode decomposition method EMD is used to decompose historical output data of wind farms. The

decomposition level of EMD cannot be manually set in advance [11]. The number of decompositions is automatically obtained in the process of recursive algorithm. The experimental data sequence is decomposed into 7 The intrinsic mode function (IMF, intrinsic mode function) and a residual difference component r are shown in Figure 4.

In order to facilitate analysis and comparison, the K value of VMD decomposition is selected as 7. This highlights an advantage of VMD decomposition. Since the VMD algorithm is solved by a non-recursive process, the number of decomposition levels can be set artificially in advance to simplify model design. The remaining set values of VMD algorithm are the same as in section 2.2.1, and the decomposition result is shown in Figure 5.

From the comparison of the decomposition results, it is generally believed that the changes in high-frequency components reflect the random influence of historical output data. In the decomposition performance of the high-frequency parts, the range of EMD algorithm amplitude changes [-10,10] is greater than the VMD algorithm amplitude The change range [-3,3] should be large, indicating that the VMD algorithm should be more thorough in the decomposition of the high-frequency part of the data. Low-frequency components mostly indicate the trend of data. In terms of low-frequency performance, the two decomposition methods have obvious indications of the trend of output data. However, EMD decomposition has more low-frequency items, and VMD decomposition has fewer low-frequency items. If you directly train each sub Modal data will greatly increase the amount of training, increase time cost, and affect prediction performance. In the literature [12], the approximate entropy method is used to reconstruct the EMD decomposition sequence, which to a certain extent solves the problem of large training volume when there are many sub-modalities decomposed by the EMD algorithm. Therefore, for a sequence with strong correlation, if the time requirement is met, the decomposition sequence can no longer be reconstructed, so that the VMD algorithm can also simplify the algorithm to a certain extent.
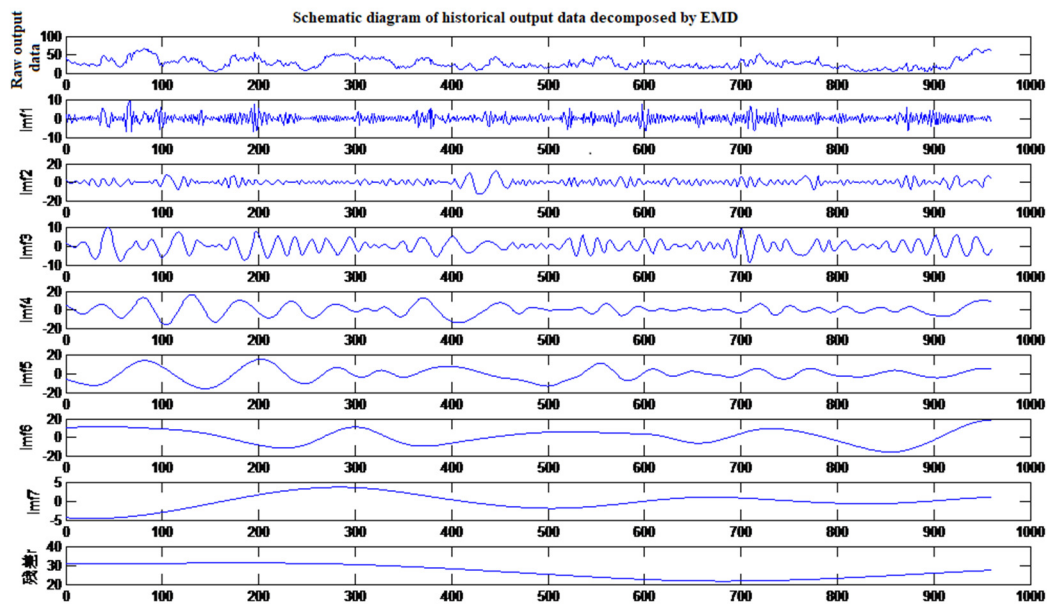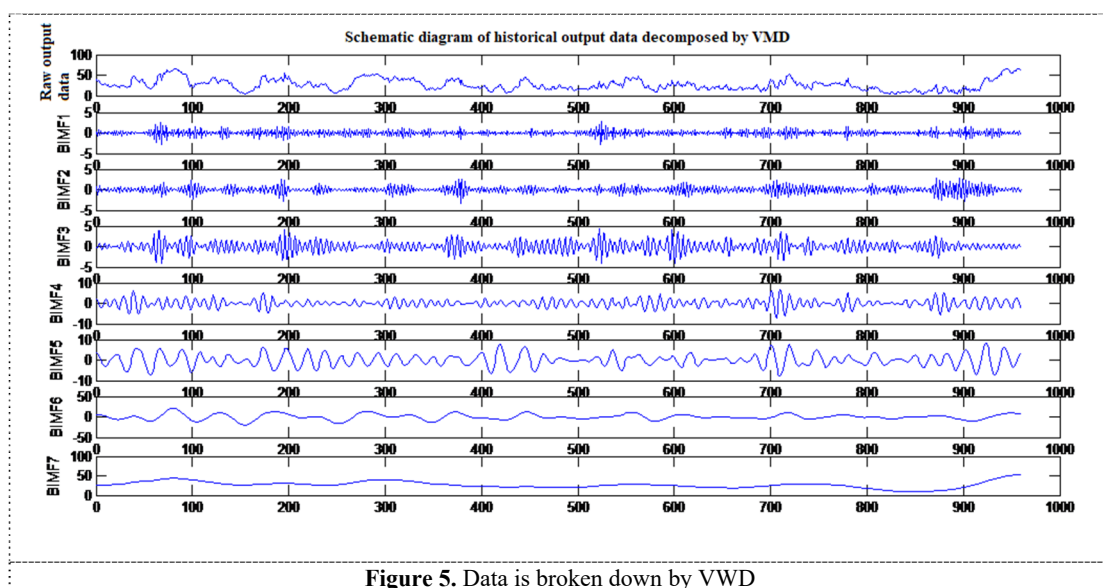

Figure 4. Data is broken down by EMD

Figure 5. Data is broken down by VWD

3.4.2 Comparison of prediction effects of various models

In order to verify the prediction performance of the model proposed in this paper, five models are established on the training data set, and the rolling training method is used to predict the output data of wind power stations in the future 15min to 4h, and then the prediction results of each model are compared and analyzed. The overall data set comes from the 15-minute interval records of Fujian Huadian Kemen Wind Power Station in 2018. The prediction and evaluation indicators used root mean square error (RMSE), mean absolute percentage error (MAPE) and mean absolute error (MAE). The five models are:

1) The particle swarm algorithm optimizes a single model of the least squares support vector machine (PSO-LSSVM); the experimental results are shown in Figure 6, and the evaluation indicators are RMSE = 3.6906, MAE = 2.5453, and MAPE = 9.38%. Without decomposing the original output data, directly using the least squares support vector machine of the particle swarm optimization algorithm to predict the performance of the model has a serious lag, but compared to the original spatial cross search method, the particle swarm optimization algorithm improves the algorithm calculation Speed and improve prediction accuracy.
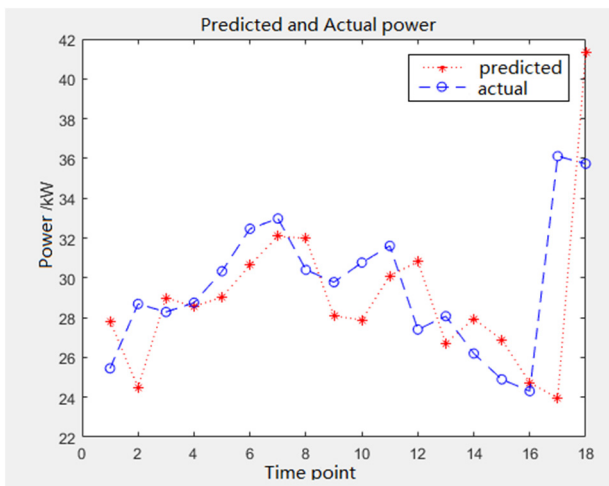

Figure 6. Single PSO-LSSVM model prediction effect diagram

2) The EMD-PSO-LSSVM prediction model is reconstructed after the sub-modes are decomposed by EMD, and the PSO-LSSVM model is reconstructed; the experimental results are shown in Figure 7. The evaluation indicators are RMSE = 2.6301, MAE = 2.1619, MAPE = 7.77%. The experimental results prove that Empirical Mode Decomposition (EMD), as a widely used method of complex data processing, has a powerful role in the field of data prediction. Compared with the wind speed-power curve

optimization fitting algorithm model and the single PSO-LSSVM algorithm model, Model 2) improves the prediction accuracy and better predicts the overall trend change of the output sequence, but the model is not satisfactory in detail.
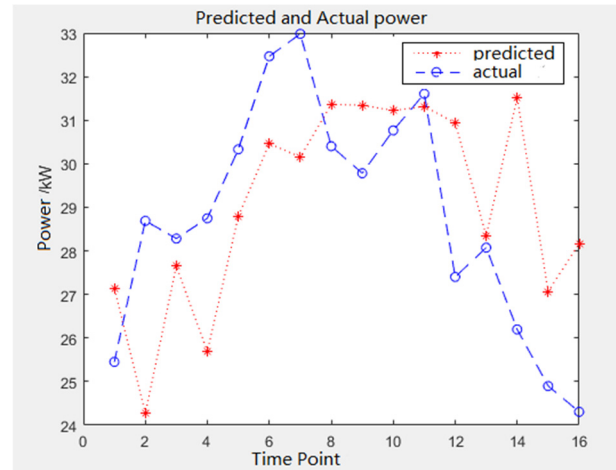

Figure 7. Based on EMD-PSO-LSSVM prediction model renderings

3) After using VMD decomposition, the sub-modes are respectively established PSO-LSSVM and then reconstructed VMD-PSO-LSSVM prediction model; the experimental results are shown in Figure 8, each evaluation index is RMSE = 1.9661, MAE = 1.5906, MAPE = 5.63%. Compared with the optimization of the first three prediction models, this model has a certain improvement. It improves the hysteresis of the single PSO-LSVVM model formally, and has a significant improvement in accuracy compared to the wind speed-power fitting optimization model.
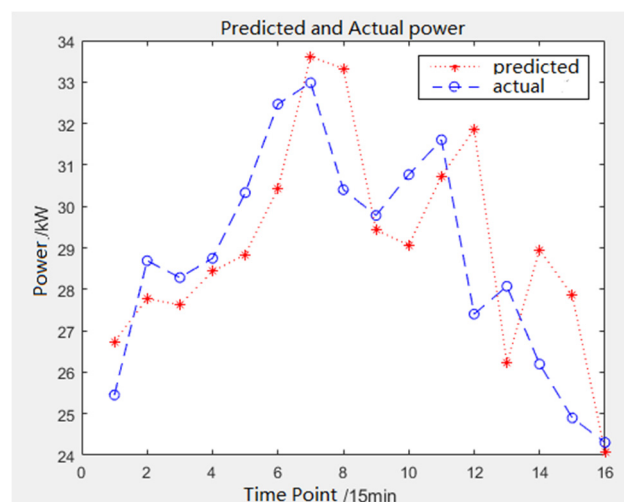

Figure 8. Based on VMD-PSO-LSSVM prediction model renderings

4) After using the variational modal decomposition (VMD), the GA-BP neural network re-summing prediction model is established for each sub-mode. This model is similar to the VMD-PSO-LSSVM model, except that the deep learning network chooses the wavelet network. Networks such as BP are also widely used in the direction of short-term data prediction. As far as this experiment is concerned, the results are shown in Figure 9(b), and the evaluation indicators are as follows: MAPE=6.87%, MAE=2.1012, RMSE=3.3848. It can be seen from the results

that, overall, the model performs well in the trend of output prediction, and the effect is not as good as the model (3). At the same time, the prediction results of the simple GA-BP model are compared, as shown in Figure 9(a): MAPE=7.05%, MAE=2.1769, RMSE=3.3951. The VMD-GA-BP model is more effective, illustrating the superiority of the basic forecasting strategy based on historical output data and the VMD data decomposition method used to process the original historical output data.
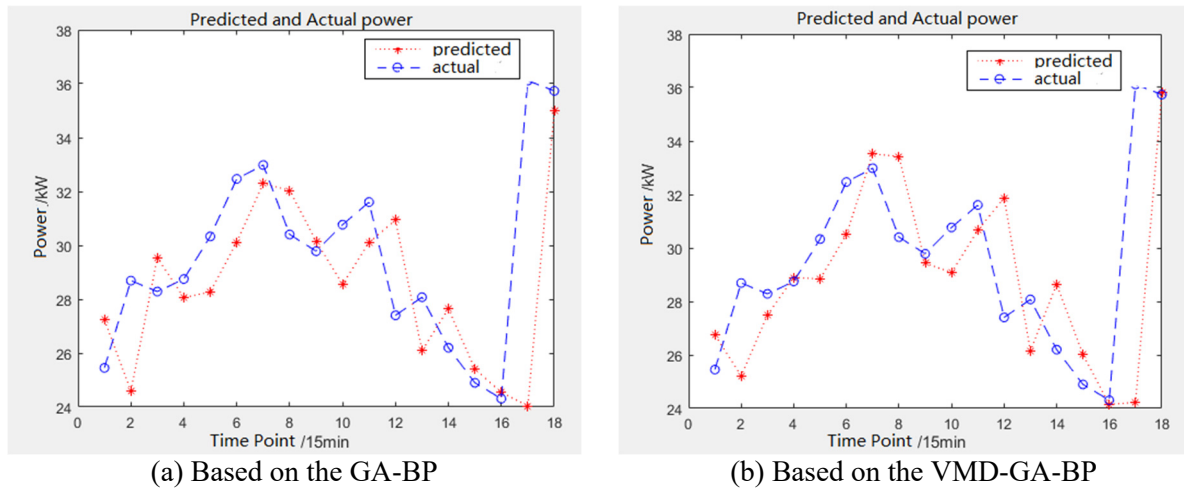

(a) Based on the GA-BP   (b) Based on the VMD-GA-BP

Figure 9. Based on GA-BP and VMD-GA-BG prediction model renderings

Comparison of evaluation criteria for multiple prediction models is shown in Table 1

Table 1. Comparison of effects of multiple forecasting models

	PSO-LSSVM	EMD-PSO-LSSVM	VMD-PSO-LSSVM	GA-BP	VMD-GA-BP
MAPE	9.38%	7.77%	5.63%	7.05%	6.87%
MAE	2.5453	2.1619	1.5906	2.1769	2.101
RMSE	3.6906	2.6301	1.9661	3.3951	3.3848

The comparison experiment proved the excellent performance of the least square support vector machine optimized by the particle swarm optimization algorithm based on the variational modal decomposition in the prediction field. Among the comparisons of the five models, the proposed VMD+PSO+LSSVM algorithm performed the best The prediction performance of the ultra-short time point shows the best prediction ability in the evaluation standard. The data simulation test proves that the prediction accuracy of the model (using MAPE as the evaluation index) can be effectively controlled at about 5%. Satisfactory results have been achieved in the prediction field.

variational modal decomposition method on the data sequence with strong follow-up and uncertainty is explored. The particle swarm algorithm is used to optimize the parameters of the vector machine model. The comparison test proves that it is based on the variational modal decomposition. The least squares support vector machine optimized with particle swarm has excellent performance in the field of prediction. At the same time, this paper confirms to a certain extent the excellent performance of the variational modal decomposition method for random sequence decomposition [13], which is of great significance for the development of short-term or ultra-short-term output prediction systems for wind power plants.

4 Summary

This paper proposes an ultra-short-time prediction model of wind power station output based on variational modal decomposition and particle swarm optimization of least squares support vector machine. The effect of the

References

1. Shan Gao，Ningyu Zhang．A review of different methodologies for solving the problem of wind

power's fluctuation[J] . Sustainable Power Generation and Supply, SUPERGEN '09. International Conference，2009，1(5)：1－7.

2. Yang Jiangping. Short-term wind speed and power forecasting in wind farm based on ANN combination forecasting[D]. Chongqing ： Chongqing University，2012.

3. Hu J , Wang J , Zeng G . A hybrid forecasting approach applied to wind speed time series[J]. Renewable Energy, 2013, 60:185-194..

4. Tai-Hua C , Lu W , Wei M A . Wind Speed Prediction Based on AR,ARIMA Model[J]. East China Electric Power, 2010.

5. K. Dragomiretskiy, D. Zosso, Variational mode decomposition, IEEE Trans. Signal Process. 62 (3) (2014) 531-544.

6. Qian Tao,Hui-you Chang,Yang Yi,Chun-qin Gu,Wen-jie Li. A rotary chaotic PSO algorithm for trustworthy scheduling of a grid workflow[J]. Computers and Operations Research,2010,38(5).

7. Guo-Feng Fan,Li-Ling Peng,Xiangjun Zhao,Wei-Chiang Hong. Applications of Hybrid EMD with PSO and GA for an SVR-Based Load Forecasting Model[J]. Energies,2017,10(11).

8. Vapnik V.The nature of statistical learning theory[M].NewYork:Spring-Verlag, 1995.

9. Yang Deyou，Cai Guowei. The short- term wind speed forecasting for wind farm based on EEMD and LS_SVM［J］. Journal of Northeast Dianli University，2015，35（3）：44—49

10. Ji X , Yang B , Tang Q . Acoustic Seabed Classification Based on Multibeam Echosounder Backscatter Data Using the PSO-BP-AdaBoost Algorithm: A Case Study From Jiaozhou Bay, China[J]. IEEE Journal of Oceanic Engineering, 2020, PP(99):1-11.

11. Guo Z , Zhao W , Lu H , et al. Multi-step forecasting for wind speed using a modified EMD-based artificial neural network model[J]. RENEWABLE ENERGY, 2012, 37(1):241-249.

12. Huang L Z , Guo X M , Ding X R . Heart sound recognition based on EMD approximate entropy and SVM[J]. Journal of Vibration and Shock, 2012, 31(19):21-25.

13. Zan T , Pang Z , Wang M , et al. Research on Early Fault Diagnosis of Rolling Bearing Based on VMD[C]// 2018 6th International Conference on Mechanical, Automotive and Materials Engineering (CMAME). IEEE, 2018.