Photovoltaic Power Forecasting Based on EEMD and a Variable-Weight Combination Forecasting Model

Hui Wang 1,*, Jianbo Sun 1 and Weijun Wang 2

1 School of Economics and Management, North China Electric Power University, Beijing 102206, China; 1162206088@ncepu.edu.cn
2 Department of Economics and Management, North China Electric Power University, Baoding 071003, Hebei, China; wwjhd@ncepu.edu.cn
* Correspondence: 50601860@ncepu.edu.cn; Tel.: +86-130-516-623-00

Received: 28 June 2018; Accepted: 22 July 2018; Published: 26 July 2018

Abstract: It is widely considered that solar energy will be one of the most competitive energy sources in the future, and solar energy currently accounts for high percentages of power generation in developed countries. However, its power generation capacity is significantly affected by several factors; therefore, accurate prediction of solar power generation is necessary. This paper proposes a photovoltaic (PV) power generation forecasting method based on ensemble empirical mode decomposition (EEMD) and variable-weight combination forecasting. First, EEMD is applied to decompose PV power data into components that are then combined into three groups: low-frequency, intermediate-frequency, and high-frequency. These three groups of sequences are individually predicted by the variable-weight combination forecasting model and added to obtain the final forecasting result. In addition, the design of the weights for combination forecasting was studied during the forecasting process. The comparison in the case study indicates that in PV power generation forecasting, the prediction results obtained by the individual forecasting and summing of the sequences after the EEMD are better than those from direct prediction. In addition, when the single prediction model is converted to a variable-weight combination forecasting model, the prediction accuracy is further improved by using the optimal weights.

Keywords: EEMD; variable-weight combination forecasting; harmonic mean; photovoltaic (PV) power generation forecasting

1. Introduction

Compared with other energy sources, solar energy has advantages of universality, cleanliness, extendibility, and sustainability, and it is one of the most ideal renewable energy sources. In solar energy applications, PV power generation is one of the most important forms [1]. Because PV power generation does not need to consume any resources and will not cause damage to the environment, it will be the best way for human beings to develop electric power business in the future. Whether it is from the perspective of protecting the Earth’s environment, or the sustainable development of the Earth’s resources, or in order to solve the problem of human beings’ growing demand for electricity, PV power generation is of great significance and has a profound impact on social development. With the advancement of science and technology in recent years, PV power generation has developed rapidly, and it accounts for an increasing proportion of power generation. The PV system can be applied to rural household power systems, large-scale PV power plants in remote areas, micro grids, and so on [2–4], and it plays an important role.

Although PV power generation has many advantages, there are also some problems in PV power generation systems, such as how to maximize power production [5], how to achieve relay coordination,
and how to coordinate fault current limiters \cite{6–9}. However, the most serious problem is that the power output of PV power generation is not stable, and it is greatly affected by the intensity of sunlight. PV panels do not generate electricity when there is no light at night. However, once the weather changes during the day, the impact on the power output is also great. When there is a cloud on a sunny day, the output power of the PV power generation will fluctuate greatly. Therefore, in order to utilize solar energy better and vigorously develop PV power generation, it is necessary to master the law of PV power generation and accurately forecast the output power of PV power generation. Accurate PV power forecasting \cite{10–13} can not only provide a reliable basis for system planning and dispatching, but also plays a vital role in system optimization, effective use of energy, and safe and stable operation of the power grid.

The time series prediction method \cite{14,15} was widely applied in the early stages of PV power generation forecasting, but its prediction accuracy is poor. The grey prediction method \cite{16,17} has also been used for forecasting, but the results are not stable. New emerging machine learning algorithms, such as support vector machines (SVM) \cite{18,19} and random forest \cite{20–22}, have better prediction accuracies. The neural network algorithm \cite{23,24} has better prediction efficacy and can reflect changes in PV power generation under complex weather conditions due to its strong learning ability.

In recent years, numerous experts and scholars have applied decomposition-reconstruction prediction models to predict targets that are obvious and have related fluctuations, such as wind speed and PV power generation. Two commonly used methods are wavelet analysis \cite{25–28} and empirical mode decomposition (EMD) \cite{29–31}. Both methods can decompose the original waveform and are able to improve the prediction accuracy; however, both have drawbacks. Wavelet analysis has poor adaptability, and EMD has problems such as end-point and over-envelope effects. Ensemble empirical mode decomposition (EEMD) \cite{32–37} was developed to improve EMD by weakening the impact of modal aliasing and has been applied in several fields of prediction research.

Since the last century, experts and scholars have conducted research into combination forecasting \cite{38–40}. These studies have determined that the combination forecasting method has better prediction accuracy than a single method and can relieve cases in which a single method has large prediction errors at individual points. Combination models were usually used to predict variables with smaller data sizes at that time, and most of the methods that had been used were simple time series methods. Researchers then realized that the accuracy of a prediction result can be improved by giving different weights to different methods. Recently, several studies have assigned different weights to each time point of different methods (i.e., variable-weight combination forecasting) \cite{41–44}; however, the determination of the weights is a major problem in these studies.

Based on previous studies, this paper presents a PV power generation forecasting method based on EEMD and a variable-weight combination forecasting model to improve PV prediction accuracy. First, the EEMD method is applied to decompose PV power generation data, and then multiple components are converted into three sequences: low-frequency, intermediate-frequency, and high-frequency. The variable-weight combination forecasting model is then applied to these three groups of sequences used to perform the prediction to obtain the final prediction results. A comparison between the case study indicates that the EEMD method can decompose the entire waveform into multiple small components, which is more conducive to forecasting and improves the prediction accuracy, and the variable-weight combination forecasting model provides a significant improvement in the prediction accuracy in comparison with the single method. The solution of the weights in predictions using the harmonic mean (HM) method gives the best results.

2. PV Power Prediction Modelling Theory

2.1. EEMD

EEMD is an improvement to EMD, which is a new noise-assisted data analysis method. Its core function is to incorporate Gaussian white noise into the signal for multiple EMD decompositions,
and the outcomes of multiple decompositions are averaged to obtain the final decomposition result. Due to the influence of noise, the model can prevent scale mixing in the decomposition. Because of the zero homogeneity of Gaussian white noise, the noise cancels out after the averaging algorithm, which makes the resulting decomposition sequence more valid. Ultimately, the EEMD method provides a significant improvement in the decomposition effect of the EMD method by notably reducing modal aliasing.

Two important parameters in the EEMD algorithm are the amplitude of the white noise (i.e., the ratio of the added white noise to the standard deviation of the original signal amplitude) and the total number of repeated EMD decompositions. Currently, there is no formula for the selection of these two parameters. Based on various literature and the experimental data from this paper, the amplitude is set from 0.1 to 0.4, and the number of decompositions is set at 150.

### 2.2. Method of the Optimal Weight Determination

This paper uses two methods to determine the optimal weights and discusses which is more suitable for application to variable-weight combination forecasting.

The first method is the HM method. The optimal weight of each method at each time point is obtained by the calculation of each error with the HM formula:

$$ k_i = \frac{1}{\frac{1}{e_1} + \frac{1}{e_2} + \cdots + \frac{1}{e_n}}, (i = 1, 2, \ldots, n) $$  \hspace{1cm} (1)

where $e_i$ is the error of model $i$ at each time point.

The second method is the quadratic programming (QP) method. We set $E$ to be the sum of the squared errors of the variable-weight combination forecasting model:

$$ E = \sum_{t=1}^{N} e_t^2 = \sum_{t=1}^{N} \sum_{i=1}^{m} \sum_{j=1}^{m} k_{it} k_{jt} e_{it} e_{jt} $$  \hspace{1cm} (2)

The optimal weights of the variable-weight combination forecasting model can be expressed by the following optimization problems:

$$ \begin{cases} 
\min E = \sum_{t=1}^{N} \sum_{i=1}^{m} \sum_{j=1}^{m} k_{it} k_{jt} e_{it} e_{jt} \\
\sum_{i=1}^{m} k_{it} = 1 \\
\sum_{j=1}^{m} k_{jt} = 1 \\
k \geq 0 \\
i, j = 1, 2, \ldots, m, t = 1, 2, \ldots, N 
\end{cases} $$  \hspace{1cm} (3)

where $e_{it}$ is the error of model $i$ at time point $t$, $k_{it}$ is the weights of model $i$ at time point $t$.

The optimal weight of each time point for each method can be obtained by solving the problem.

### 2.3. Variable Weight Prediction Modelling

The original data used for the prediction are divided into three parts: a training set $(x_1, y_1)$, an auxiliary set $(x_2, y_2)$, and a test set $(x_3, y_3)$. Models $a$, $b$, and $c$ are obtained using the decision tree, SVM, and ensemble methods to conduct fitting of the training set. By using these models for the prediction of the auxiliary set $x_2$, the prediction results, $y_{2a}$, $y_{2b}$, and $y_{2c}$, of the auxiliary set are obtained. These are then subtracted from $y_2$ to give the errors $e_a$, $e_b$, and $e_c$, respectively. By using the HM or QP methods the optimal weights for the three prediction methods, $w_{2a}$, $w_{2b}$, and $w_{2c}$, can be obtained. The ensemble method is then used to fit $x_2$ with $w_{2a}$, $w_{2b}$, and $w_{2c}$ and predict the test set $x_3$ to obtain the weights $w_{3a}$, $w_{3b}$, and $w_{3c}$ of each time point for each method. In addition, models $a$, $b$, and $c$ are used to predict the test set $x_3$ and obtain the prediction results $y_{3a}$, $y_{3b}$, and $y_{3c}$. Finally,
the variable weights of the three models are combined using the formula \( y = y_{3a} \times w_{3a} + y_{3b} \times w_{3b} + y_{3c} \times w_{3c} \), and the final prediction result is obtained. The design logic is shown in Figure 1.

Figure 1. Flowchart of the variable-weight combination forecasting model.

3. PV Power Prediction Modelling Based on EEMD and the Variable-Weight Combination Forecasting Model

The prediction modelling procedure in this paper begins with the selection of the input variables. The lead time for the input variables is adjusted to one hour to make the prediction results practical. Modelling is then conducted using the EEMD and the variable-weight combination forecasting model, and the prediction results are output and evaluated.

3.1. Input Variable Selection

This paper presents the forecasting of hourly PV power generation. The lead time is 1 h, and there are 13 input variables: time, total column liquid water, total column ice water, surface pressure, relative humidity at 1000 mbar, total cloud cover, 10 m U wind component, 10 m V wind component, 2 m temperature [45], surface solar rad down, surface thermal rad down, top net solar rad, and total precipitation. Because time is a state variable, it is excluded from the SVM modelling.

3.2. Modelling Procedure

EEMD was used first to decompose the original PV power generation data to obtain multiple intrinsic mode function (IMF) components (IMF_1, IMF_2) and the residual component (RC). The IMF components and residual component were divided into high-frequency (HF), intermediate-frequency
The variable-weight combination forecasting (VWCF) method was then applied to predict these three sequences. The sum of the three forecasting results gives the final prediction result. The flowchart of this process is shown in Figure 2.

3.3. Forecasting Results Evaluation

This paper uses the mean absolute error (MAE) and the mean square error (MSE) to evaluate the forecasting results obtained from the models. MAE and MSE are calculated using the following equations:

\[
MAE = \frac{\sum_{i=1}^{N} |\hat{y}_i - y_i|}{N}
\]

\[
MSE = \frac{\sum_{i=1}^{N} (\hat{y}_i - y_i)^2}{N}
\]

where \(\hat{y}_i\) is the predicted value, \(y_i\) is the actual value, and \(N\) is the sample size.

4. Empirical Analysis

4.1. Data Source and Parameter Initialization

This paper uses the PV power data from the CEFCom2014 competition, which were downloaded from the internet site http://www.gefcom.org. The time period was from 1 April 2012 to 29 June 2012. Because PV power is not generated at night, only the periods between 4 a.m. and 7 p.m. on these days are used as input data, and the periods between 5 a.m. to 8 p.m. are used as output data. A total of 1440 data points are used as the overall data set. The original data are divided into three groups of sequences: 960 for the training set, 320 for the auxiliary set, and 160 for the test set. Because the SVM method is used here, the original data must be normalized. This article uses MATLAB as a simulation tool.

4.2. EEMD and Variable Weight Combination Model

Figure 3 shows the decomposition results of the PV series using EEMD.
Ten sequence components are formed after the EEMD. The first three components are added to form the high-frequency sequence, components 4–6 are added to form the intermediate-frequency sequence, and the last four sequences are added to form the low-frequency sequence.

The “TREE + SVM + ENSEMBLE” combination prediction model is used to perform forecasting of the high-frequency, intermediate-frequency, and low-frequency sequences (the QP and HM methods are used for the weight selection), and the sum of the three forecasting results gives the final prediction result.

4.3. Comparison of the Models and Forecast Results

Based on the modelling, Figure 4 shows a comparison between the prediction results, and Table 1 shows a comparison of the forecasting evaluation indicators of each model.

Figure 3. Decomposition results of the PV series using EEMD.

Figure 4. Comparison of forecasting results of the models. QP: quadratic programming method; HM: harmonic mean method.
Table 1. Comparison of the evaluation indicators of the models. MAE: mean absolute error; MSE: mean square error.

<table>
<thead>
<tr>
<th>Model</th>
<th>MAE</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>TREE</td>
<td>0.0761</td>
<td>0.0199</td>
</tr>
<tr>
<td>EEMD + TREE</td>
<td>0.0682</td>
<td>0.0114</td>
</tr>
<tr>
<td>EEMD + VWCF (QP)</td>
<td>0.0697</td>
<td>0.0123</td>
</tr>
<tr>
<td>EEMD + VWCF (HM)</td>
<td>0.0622</td>
<td>0.0107</td>
</tr>
</tbody>
</table>

Table 1 reveals that the use of the HM method for the variable-weight combination forecasting provides better results for the weight selection; the reason will be discussed in the next section. Compared to the model, the prediction results were worst when the decision tree method was used alone. When EEMD was used first and followed by the decision tree method, MAE decreased by 10.38% and MSE decreased by 42.7%. However, when using the proposed method (i.e., EEMD to decompose the data followed by a variable-weight combination forecasting model (HM)) compared with the EEMD + TREE, the MAE decreased by 8.79% and the MSE decreased by 6.14%, which further improves the PV power generation prediction accuracy.

The predictions using recent existing approaches and EEMD + VWCF (HM) are compared in Table 2.

Table 2. Comparison of recent existing approaches and EEMD + VWCF (HM).

<table>
<thead>
<tr>
<th>Model</th>
<th>MAE</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random Forest</td>
<td>0.0833</td>
<td>0.0140</td>
</tr>
<tr>
<td>Back Propagation (BP) Neural Network</td>
<td>0.0816</td>
<td>0.0157</td>
</tr>
<tr>
<td>EMD + TREE</td>
<td>0.0713</td>
<td>0.0130</td>
</tr>
<tr>
<td>EMD + VWCF (HM)</td>
<td>0.0684</td>
<td>0.0127</td>
</tr>
<tr>
<td>EEMD + VWCF (HM)</td>
<td>0.0622</td>
<td>0.0107</td>
</tr>
</tbody>
</table>

In Table 2, we can see that EEMD + VWCF (HM) is better than recent existing single approaches like Random Forest and Back Propagation (BP) Neural Network both in MAE and MSE. When using EMD + VWCF (HM) compared with EMD + TREE, the MAE decreased by 4.06%, which shows that VWCF (HM) is a better method. Additionally, the comparison of EEMD + VWCF (HM) and EMD + VWCF (HM) shows that EEMD is an improvement to EMD.

Table 3 shows that the use of the QP method in the forecasting is significantly better than the use of the HM method when the weights of the three methods at each time point can be accurately predicted. However, because the weights obtained by the QP method are too extreme, the prediction model is not general. As a result, the final prediction accuracy of each sequence is essentially the same as that of the HM method. The final prediction results obtained after summing the three sequences show that the MAE and MSE of the QP method are worse than those of the HM method because its model is less general due to the extreme prediction, which leads to the low accuracy of the final prediction result.

Table 3. Comparison of forecasting results of QP and HM methods.

<table>
<thead>
<tr>
<th>MAE of Different States</th>
<th>QP</th>
<th>HM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Optimal forecasting of LF</td>
<td>0.0192</td>
<td>0.0267</td>
</tr>
<tr>
<td>Last forecasting of LF</td>
<td>0.0308</td>
<td>0.0313</td>
</tr>
<tr>
<td>Optimal forecasting of MF</td>
<td>0.0278</td>
<td>0.0381</td>
</tr>
<tr>
<td>Last forecasting of MF</td>
<td>0.0544</td>
<td>0.0556</td>
</tr>
<tr>
<td>Optimal forecasting of HF</td>
<td>0.0287</td>
<td>0.0460</td>
</tr>
<tr>
<td>Last forecasting of HF</td>
<td>0.0723</td>
<td>0.0717</td>
</tr>
<tr>
<td>Last forecasting</td>
<td>0.0697</td>
<td>0.0622</td>
</tr>
</tbody>
</table>
4.4. Discussion on the Effects of Input Variables

Since there are 13 input variables, and the influence of each variable on the prediction results is different, the degree of influence of each variable is discussed here. We can use the random forest algorithm to determine the impact of each variable as follows.

In Figure 5 we can see that the first six influential variables can be found and arranged in order: surface thermal rad down, time, top net solar rad, surface pressure, relative humidity at 1000 mbar, and 2 m temperature.

![Figure 5. Analysis of the effects of each input variable.](image)

Table 4 shows that when only the six aforementioned variables are selected, in comparison to the results obtained from using all 13 variables, the prediction accuracy (in terms of MAE) of each model is reduced. The reason may be that although some variables have a small degree of influence, they can still affect the accuracy of the prediction in some cases. Therefore, each variable should be retained.

<table>
<thead>
<tr>
<th></th>
<th>MAE 6 Variables</th>
<th>MAE 13 Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>TREE</td>
<td>0.0761</td>
<td>0.0900</td>
</tr>
<tr>
<td>EEMD + TREE</td>
<td>0.0682</td>
<td>0.0826</td>
</tr>
<tr>
<td>EEMD + VWCF (HM)</td>
<td>0.0622</td>
<td>0.0773</td>
</tr>
</tbody>
</table>

5. Conclusions

This paper proposed a PV power generation forecasting method based on EEMD and a variable-weight combination forecasting model. The sequences decomposed by EEMD are predicted using a variable-weight combination forecasting model individually and integrated to obtain the final forecasting result. The following conclusions are drawn from simulation experiments. After the waveform is decomposed by EEDM and reintegrated, the prediction accuracy of the PV power generation forecast model is significantly higher than that of the direct prediction; this reduces the impact of the non-stationary characteristics of the sequence on the prediction results. Variable-weight combination forecasting solves the problems that are present in single prediction methods, such as inaccurate predictions and large prediction errors at individual points. The HM method is most suitable for obtaining the weights of variable-weight combination forecasting models. The final prediction results show that the integration of EEMD with a variable-weight combination forecasting model can effectively improve the accuracy of PV power generation forecasting.
Author Contributions: In this research activity, all the authors were involved in the data collection and preprocessing phase, model construction, empirical research, results analysis and discussion, and manuscript preparation. All authors have approved the submitted manuscript.

Funding: The paper is supported by “the 111 Project (B18021)”.

Acknowledgments: We would like to thank the editor and reviewers for their constructive comments and helpful suggestions, which helped us to improve the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References


© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).