

# A Combined Method of Two-model based on Forecasting Meteorological Data for Photovoltaic Power Generation Forecasting

Xiao Zhang, Runjie Shen\* and Yiyang Wang

Faculty of Electronics and Information Engineering, Tongji University, Shanghai, 201804, China

**Abstract.** Under the background of the continuous development of photovoltaic power generation technology, accurate prediction of photovoltaic output power has become an important subject. In this paper, a combined method of two-model based on forecasting meteorological data for photovoltaic power generation forecasting is proposed. To solve the problem of the adaptability of a single model, two different models are used according to the different types of output power characteristics. The K-means clustering algorithm is used to classify different weather types according to the historical meteorological data. After predicting the irradiance and temperature of the period to be predicted and classifying the period into different types, the photovoltaic output power is predicted by a suitable model. The two prediction models are the Wavelet-Decomposition-ARIMA model and EDM-SA-DBN model, which are suitable for periods with larger and smaller fluctuation amplitude of photovoltaic output, respectively. Wavelet decomposition can refine the data with large fluctuations on multiple scales, make the data smooth, and improve the prediction accuracy of the Autoregressive Integrated Moving Average model (ARIMA). The Deep Belief Network (DBN) can effectively process a large number of complex data and deep mining the data features. While the empirical mode decomposition (EMD) can decompose the more stable data and amplify the details in the signal as much as possible. Meanwhile, the simulated annealing algorithm (SA) can avoid the network falling into a local optimal solution and improve the prediction accuracy. This paper uses a large number of photovoltaic power station data for experimental verification. The results show that this combined model has high accuracy and generalization ability.

## 1 Introduction

In recent years, with the increasing awareness of human beings to protect the environment and the continuous development of science and technology, all kinds of clean energy have been gradually paid attention to[1]. Solar energy, as a renewable resource, has the advantages of simpler, more lasting, easier to develop and use, and has more reserves. But the output power of the solar system is easily affected by external environmental conditions and various internal factors. When large high-capacity photovoltaic systems are connected to the grid, the volatility of solar power generation will bring a series of problems and challenges to the distribution network. To integrate solar power plants into the power grid and ensure the safety and reliability of the power grid, it is very important to adjust the balance between the grid-connected power generation of photovoltaic power stations and the generation from other power generation modes in the power grid. Because of the small adjustment range and high cost of thermal power plants, providing accurate short-term prediction for photovoltaic devices

becomes the key factor to integrate photovoltaic devices safely into existing power grids[2].

There are many kinds and methods to predict the output power of photovoltaic power generation at present such as physical methods, statistical methods, neural networks, etc[3]. Physical methods require the use of simulated photovoltaic conversion models, which is difficult to model and can't achieve high precision, without good anti-interference ability and stability[4]. Statistical methods and other types of artificial intelligence algorithms are simpler to model and have better prediction accuracy, although a large amount of historical data is needed for learning and training[5]. But the current data sources are also very easy to obtain. From the available photovoltaic output data, it can be seen that the fluctuation range is very large due to weather factors, so the traditional method using a single prediction model can't predict well in all weather conditions. Usually, a certain method will be more suitable for a specific weather condition, after the classification of weather, for the data type with large fluctuation amplitude, we should find a way to increase its stability so that the accuracy of the prediction model can be improved. At the same time, for

\*Corresponding author's e-mail: shenrunjie@163.com

the data type with small fluctuation amplitude and more stable, we should enlarge its signal details and dig the relationship between the data in depth.

So this paper proposes a combination prediction method that uses the K-means clustering algorithm to classify weather types according to meteorological factors and then uses the Wavelet-Decomposition-ARIMA model and EDM-SA-DBN model to predict different types according to their characteristics.

## 2 Classification using the K-means clustering algorithm

K-means clustering algorithm is a distance-based iterative clustering algorithm[6]. The closer the two objects are, the greater their similarity. The algorithm considers that the cluster is composed of objects close to the distance, so it takes the compact and independent cluster as the final goal.

After classification using the K-means algorithm, several cluster centers, and all data under clustering can be obtained. In multi-dimensional weather data, the type of a certain period can be locked by irradiance and temperature. So by comparing the predicted irradiance value and temperature value with the irradiance value of the cluster center and temperature value of the cluster center, the same is to find the nearest cluster center can be classified. Because the prediction accuracy of irradiance and temperature data using the EMD-SA-DBN model is extremely high, the error is less than 2%, the distance of each cluster center is far away, the cluster is compact, and the clustering is relatively independent, the error of this classification method is almost zero.

## 3 The EMD-SA-DBN forecasting model

Deep neural networks use layer-by-layer training methods, which can avoid the gradient diffusion phenomenon when the residual is transmitted. But because the characteristic of the deep neural network is that it needs to be stratified pre-trained step by step, the learning time is longer, and the parameter selection of neural network is more difficult, and if the selection is improper, it is easy to fall into the local optimal solution, which reduces the accuracy of the prediction model. So the SA algorithm is used to optimize the DBN to avoid falling into the local optimal solution. Meanwhile, to further improve the prediction accuracy of the model, the time series is decomposed by EMD before the SA-DBN model.

### 3.1 The empirical modal decomposition

The advantage of EMD is that it has a strong self-adaptation, which will automatically produce the base function of the appropriate sequence and get the corresponding IMF components according to the characteristics of the original sequence, each component will be predicted separately, and the final prediction effect can be obtained by combining the results. It is based on the time characteristics of the data itself for sequence decomposition, without the need to select the appropriate

basis function and set the feasible number of decomposition layers. This method has great advantages in analyzing natural signals and has a high signal-to-noise ratio. The decomposed components contain local characteristic signals of different time scales of the original signal, which can be restored to the original data after adding these intrinsic modes.

### 3.2 The simulated annealing algorithm

SA algorithm is an optimization algorithm that applies the idea of simulated annealing in physics to the optimization problem, which can effectively solve the local optimal solution problem. At the core of the algorithm is the metropolis algorithm and the annealing algorithm. Metropolis algorithm accepts a solution that is worse than the current solution with a certain probability, it is possible to jump out of this local optimal solution and reach the global optimal solution. But this will greatly increase the operation time, so the annealing algorithm will ensure that the system can converge in a limited time.

The simulated annealing algorithm is independent of the initial value, and the obtained optimal solution is also independent of the initial solution. It is also a global optimization algorithm that can realize parallelism.

### 3.3 The deep belief network

DBN algorithm is flexible, easy to expand, and has the advantage of dealing with a large number of complex data effectively. It can deal with the prediction problem of PV output well, and dig deep into the characteristic relationship between the data. However, DBN has the disadvantage of the local optimum that the neural network may have, it can be optimized by using SA to ensure that the global optimal solution can be obtained and the training efficiency can be improved. And the historical data of the training model is complex and huge, so we can use the method of data decomposition to decompose the original sequence and then model the split-sequence in turn.

## 4 The Wavelet-Decomposition-ARIMA prediction model

For improving the prediction accuracy of the ARIMA model, the wavelet decomposition can be used to process the signal and reduce noise while improving the capture of weak signals. The original time series is first decomposed into two order low frequency, two order high frequency, and one order high-frequency sequences by using the db4 wavelet filter. Each sequence is established and trained. Then the prediction results of each sequence after decomposition are added and compared with the original sequence to obtain the final training effect of the wavelet-decomposition-ARIMA model.

### 4.1 The wavelet decomposition

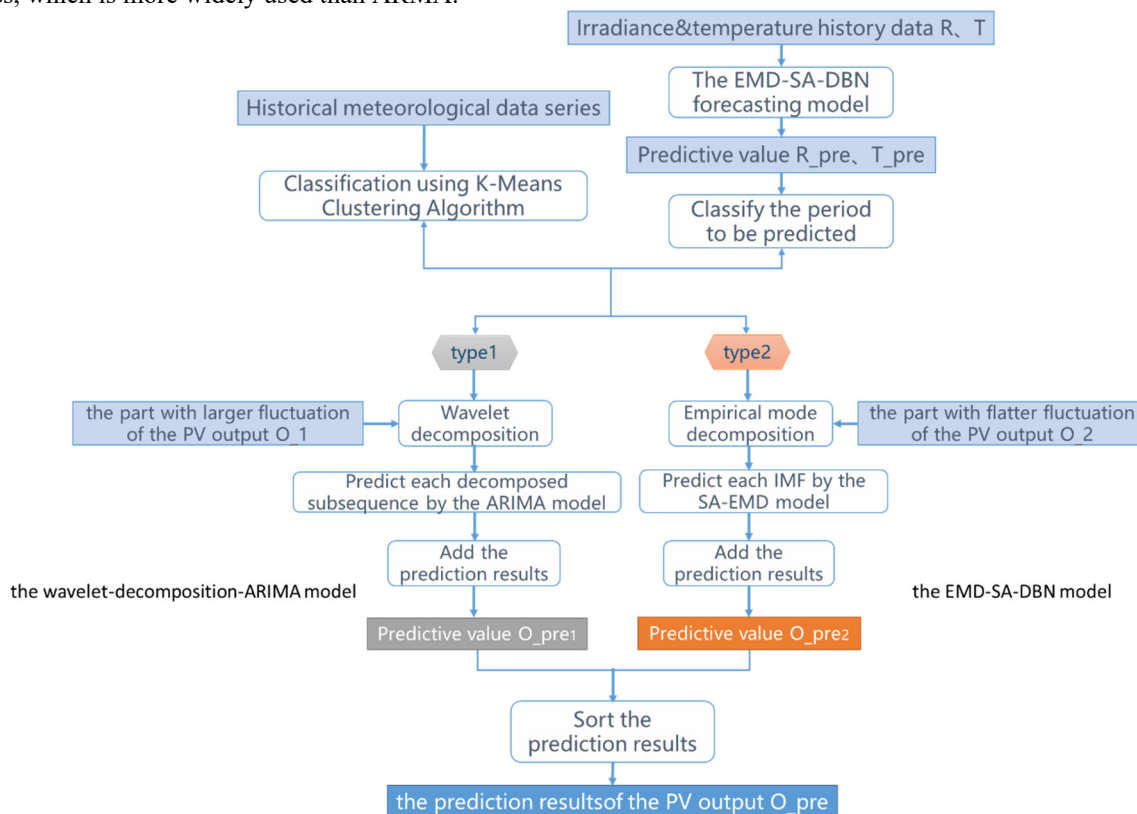
Wavelet decomposition is to realize multi-scale subdivision of different frequency bands by stretching and translation operation, which has strong adaptability and can amplify any detail of the signal. After the wavelet function is selected, the coefficients of the corresponding wavelet filter are also obtained. Low-pass filter and high-pass filter are constructed with wavelet filter coefficients. The low-pass filter can be regarded as a smoothing filter. After the two filters are obtained, the horizontal multi-resolution signals can be obtained by the recursive decomposition algorithm.

### 4.2 The Autoregressive Integrated Moving Average model

ARIMA is a traditional method of time series forecasting based on statistical regression analysis. It contains three parameters (p,d,q), which are the number of autoregressive terms (p) in the AR, the number of moving average terms (q) in the MA, and the difference order (d) that makes the time series a stationary sequence. So to some extent, ARIMA is the combined evolutionary form of three models. ARIMA model can increase the stationarity of the time series because of the difference process, which is more widely used than ARMA.

## 5 The Combined method of two-model for photovoltaic power generation forecasting

The prediction effect of the data classified by the K-means clustering method will be better than that before classification. So the irradiance and the temperature are predicted by using the EDM-SA-DBN model, and the two predicted weather factors are used to treat the prediction period. The specific classification method is to compare the predicted irradiance and temperature values with the irradiance and temperature values of the cluster centers formed by previous data sets and to classify into the categories represented by the nearest cluster centers. Then the Wavelet-Decomposition-ARIMA model is used to predict the types with large fluctuation amplitude of photovoltaic output and the EDM-SA-DBN model is used to predict the types with small fluctuation amplitude of photovoltaic output power. Finally, the final prediction value can be obtained by arranging the predicted values according to the original time order, and the prediction effect of the model can be evaluated by comparing it with the real value. The program block diagram of the overall model is shown in figure 1.



**Figure 1.** Flow chart of The Combined method of two-model based on forecast meteorological data for photovoltaic power generation forecasting.

## 6 Experimental validation

K-means algorithm was used to divide the original data set of meteorological data and photovoltaic power generation

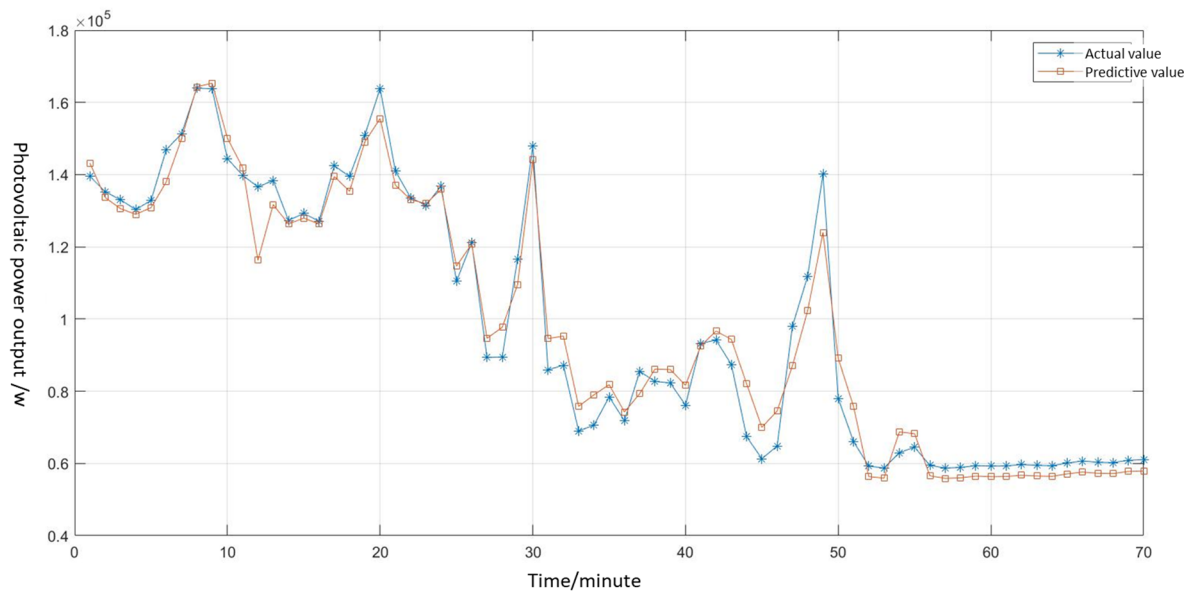
data in January 2015 into three types according to the distribution of meteorological data. Three cluster-centers' data were obtained in table 1. According to the characteristics of each type of meteorological data, they were roughly named cloudy, partly cloudy, and sunny.

**Table 1.** Distribution of cluster centers

	Cloudy type	Partly cloudy type	Sunny type
Radiation	97.530	268.512	505.216
Temperature	-0.087	1.555	2.936
Humidity	48.09	39.55	31.65
Cloud	65	46	14
Wind speed	2.80	2.84	2.80
Pressure	818.93	818.74	819.71

After the experiment, we know that the best PV time series prediction model for cloudy days is the wavelet-decomposition-ARIMA prediction model, the prediction

error is only 5.44%. Partly-cloudy weather type and sunny weather type can be respectively predicted by using the EMD-SA-DBN model. The prediction error is 4.85% and 4.29%. Comparing the distribution law of experimental data sets of various day types, it is speculated that this may be because the output of photovoltaic power generation of cloudy type is generally at a lower level, and there was no excessive floating. So it is more suitable for the wavelet-decomposition-ARIMA model with lower decomposition layers and decomposition using independent wavelet functions. And the output of photovoltaic power generation data of partly-cloudy and sunny types fluctuates greatly and is not consistent. This Verifies the model design ideas mentioned above.



**Figure 2.** Comparison of the predictive value of PV output from the combined method and the actual value.

Therefore, a data set of 8638 minutes in January 2015 was clustered by the K-means algorithm according to meteorological data. After classification, we got 3674 groups data of cloudy type, 2253 groups data of partly-cloudy type, and 2710 groups data sunny day type. From 07:43 to 14:32 on January 6, 2015, 410 sets of data were selected for the experiment. The final three types of predictive data in the original order of the experimental results are shown in figure 2, the prediction accuracy is MAPE=4.13%. Among the 70 min data predicted by the experiment, 17 minutes were cloudy, 40 minutes were partly-cloudy and 13 minutes were sunny. It can be seen from table 2 that the accuracy is improved either directly compared to the single timing model or compared to the multi-input model.

**Table 2.** Comparison of the effect of various predictive models with the combined method

Predictive model	MAPE
PCA-GA-BP model	23.59%
PCA-WNN model	18.97%
PSO-LSTM model	18.90%
Wavelet-decomposition-ARIMA model	12.16%
EMD-SA-DBN model	5.59%
The combined method of two-model based on forecasting meteorological data	4.13%

## 7 Conclusions

The instability of the photovoltaic power generation system will bring great security hidden trouble to the power grid when connected to the grid, so this paper combines meteorological data to classify the daily types by clustering method and selects the best prediction models of each day type. Finally, a combined method of two-model for photovoltaic power generation forecasting based on forecasting meteorological data is proposed. This combination prediction method uses different prediction models for different characteristics of input data: the EMD-SA-DBN model is used to predict the part with flatter fluctuation, and the wavelet-decomposition-ARIMA model is used to predict the part with larger fluctuation. This method takes into account the timing of the output power data of photovoltaic power generation and the influence of meteorological data on the output of photovoltaic power generation and obtains better results in the short-time prediction experiment of photovoltaic power generation. The prediction error (evaluation index is MAPE) of the model is 4.13%, compared with the single timing prediction model and multidimensional meteorological input prediction model, the prediction accuracy is improved. The innovation of this model is:

- Use the simulated annealing algorithm to optimize the parameters of the DBN to avoid falling into a local optimal solution and improve the model accuracy.
- Classify the data with different characteristics, and adopt a suitable data decomposition method and prediction model to smooth the data with large fluctuation and detail the data with small fluctuation, to be able to give the right remedy to the case, better analyze and predict the data and improve the prediction accuracy.
- Use meteorological data to classify historical data, and combined with the predicted value of meteorological data to classify the prediction period, making full use of meteorological data. With high accuracy of the natural timing data prediction, the classification has almost no error, and the problem of classification difficulties on no meteorological data for the prediction period can be easily solved.

## References

1. Jibrán, K., Arsalan, M.H. (2016) Solar power technologies for sustainable electricity generation a review. *J. Renewable and sustainable energy reviews.*, 55: 414-425.
2. Hong-hua, X. (2007) The Study on Development of PV Technology in China. *J. Power System Technology.*, 31(20): 77-81.
3. Changwei, L., Jinghua, L., Bo, C., Yujin, H., Shanyang, W. (2019) Summary of Research on Photovoltaic Power Output Forecast Technology. *J. Journal of Electrical Technology.*, 34(06): 1201-1217.
4. Antonanzas, J., Osorio, N., Escobar, R. (2016) Review of photovoltaic power forecasting. *Solar Energy.*, 136: 78-111.
5. Wenxiu, Z., Siyu, C., Zhongchen, H., Hengqing, Z., Lanhong, Z. (2016) Summary of Power Prediction Technology for Photovoltaic Power Generation. *J. Power World.*, 01: 45-48.
6. Lu, S.X., Lin, G.Y., Liu, H.L. (2019) A weekly load data mining approach based on hidden markov model. *J. IEEE Access.*, 7: 34609-34619.