IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received July 3, 2018, accepted August 14, 2018, date of publication September 4, 2018, date of current version October 8, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2868478

Ultra-Short-Term Prediction of Photovoltaic Power Based on Periodic Extraction of PV Energy and LSH Algorithm

MAO YANG AND XIN HUANG
School of Electrical Engineering, Northeast Electric Power University, Jilin 132012, China
Corresponding author: Mao Yang (yangmao820@163.com)

This work was supported by the National Key R&D Program of China (Technology and Application of Wind Power/Photovoltaic Power Prediction for Promoting Renewable Energy Consumption, Grant 2018YFB0904200).

ABSTRACT Ultra-short-term photovoltaic power prediction is one of the important measures to reduce the adverse effects of the safe and stable operation of traditional power systems. First, the periodicity of the PV power is taken into account to extract periodic components. For the remaining components, under different weather types, the local sensitive hashing algorithm algorithm is used to achieve rapid classification of photovoltaic power segments, and the European distance is introduced as a measure of the prediction to predict. Using the data of the photovoltaic power station for verification, the results show that the method has a higher prediction accuracy.

INDEX TERMS Photovoltaic power, ultra-short-term prediction, LSH algorithm, European distance.

I. INTRODUCTION

The output power of PV power generation system is affected by many factors, such as climate and components, and has the characteristics of volatility, intermittency and randomness, causing some interference to the power grid in the process of grid connection. In order to ensure the safe and stable operation of the power grid, the PV power generation system must have a certain time width and high-precision power prediction capability. According to the different forecasting duration, the forecast of PV output power can be divided into long-term forecast (duration > 1 month), medium-term forecast (duration 1∼7 days), short-term forecast (4 hours duration) [1]. In the PV system switching mode, load control and energy storage control, the ultra-short-term forecast of PV power plays an important role. Under the background of more and more open electricity trading market, the ultra-short-term forecast of PV power is of great importance to the supervision and regulation of power market.

PV power prediction methods can be divided into two kinds of physical methods and statistical methods [2]. The physical method refers to the calculation of PV output power based on meteorological data such as solar irradiance, temperature, humidity and wind speed, as well as considering the topography and geomorphology, component status, and the output characteristic curve. The method is more comprehensive, and the established model is more complicated. The statistical method seeks the internal law of photovoltaic power source and uses the law to predict the historical power data. This method is easy to implement but has certain limitations on improving the prediction accuracy. At present, there are more studies on the ultra-short-term prediction of PV power both at home and abroad. Commonly used simple prediction methods include time series method, neural network method, support vector machine and Markov chain [3]–[6]. Literature [7] proposed a method based on similar days and wavelet neural network (WNN) for PV power ultra-short-term prediction. However, it is unreasonable to calculate the correlation coefficient between the actual meteorological data and the historical meteorological data of the forecast day. Reference [8] based on the radiation impact factor analysis, the similar day to select predictors and then combined with artificial neural network for short-term prediction of PV power. The selection of similar days has a certain impact on the improvement of prediction accuracy. The literature [9] predicts the PV power by combining the similar day method with the GA-BP neural network. Forecasting the PV power one hour in the future does not have the engineering application value. Literature [10] made use of artificial neural network method and satellite imagery to forecast the solar radiation value, but the estimated time-width was too large to meet the requirements of ultra-short-term forecast of PV power.

TABLE 1. Basic information of PV power plant.

Item	Data
Fixed inclined plane	fixed inclined plane
Component model	JKM255P-60
average altitude	150 m
Installation location	ground
Installed capacity	16000W
Component material	Polysilicon

This paper presents an ultra-short-term prediction method based on photovoltaic power cycle extraction and local sensitive hashing algorithm (LSH). Firstly, the periodic characteristics of photovoltaic power generation are analyzed and extracted, historical data under different weather types are analyzed, and a local ultrasensitive hash algorithm is used to establish the ultra-short-term photovoltaic power forecasting model.

II. PERIODIC ANALYSIS AND EXTRACTION OF PV POWER

The PV power signal can be expanded into Fourier series [11], which can be expressed as the sum of a series of sinusoidal or complex exponential signals of different frequencies.

The finite Fourier decomposition of the PV power time series X_t in the given modeling history time domain D^- can be decomposed as follows:

$$\begin{aligned} X_t &= a_0 + \sum_{i=1}^{N/2-1} [a_i \cos(2\pi it/N) + b_i \sin(2\pi it/N)] \\ &\quad + a_{N/2}\cos(\pi t) \\ &= a_0 + \sum_{i=1}^{N/2} R_i \cos(\omega_i t + \Phi_i) \end{aligned} \tag{1}$$

The cosine terms are orthogonal to each other. This method can be used to decompose the change of PV power $P(t)$ into the angular frequency component of $2\pi/N, 4\pi/N, \ldots, \pi$.

To efficiently represent the energy of the components contained in the signal, a spectrogram can be made. Define energy as formula (2):

$$W(t) = \frac{2\,|FFT(x_t)|}{N} \tag{2}$$

where FFT is a first-order Fast Fourier Transform function, the conversion result is complex. The spectrogram can reflect the periodic characteristics of PV power history data, and high energy corresponds to high periodicity.

Using the historical power data of a certain PV power station in Ashland and the meteorological data related to PV power generation (including the atmospheric irradiance, direct irradiance, diffusion of energy, tilt 15° south of the irradiance, average wind speed, temperature and humidity), the basic information of the power plant is shown in Table 1. The sampling time for the data used is from January 2016 to July 2016 with a sampling interval of 15 min.

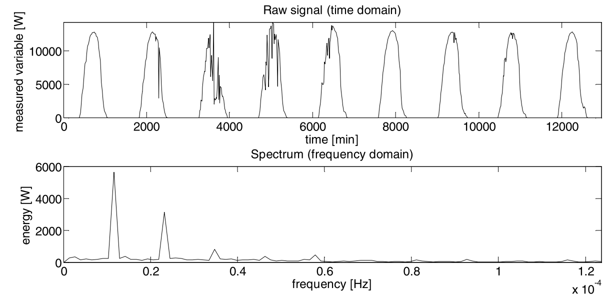

FIGURE 1. PV power curve and spectrum from 2015/01/01 to 14.

TABLE 2. Frequency represented by high energy.

Item	First enegy	Second enegy	Third enegy
Frequency [Hz]	1.587×10^{-5}	2.3175×10^{-5}	3.4762×10^{-5}
Cycle [h]	23.97	11.99	7.99

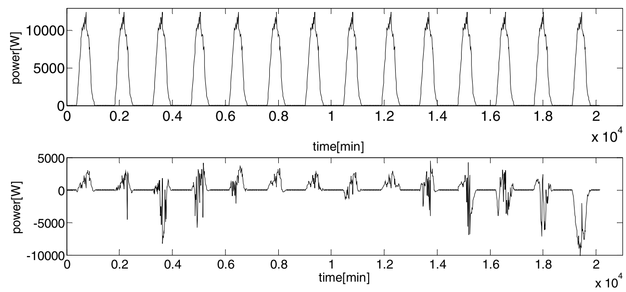

FIGURE 2. PV power periodic components and residual component extractio.

The PV output power spectrum was analyzed from January 1, 2016 to January 14, 2016. The results are as figure 1 and it can be seen that the PV output power has a strong intermittency and volatility with a certain periodicity. It can be seen from Table 2 that the PV output power has a strong daily cycle. The frequency decomposition of PV output power in different regions and different time periods shows that the periods corresponding to the second and third largest energies sometimes differ, but the period corresponding to the maximum energy is always about 24 h, which further proves that PV output power of the periodicity. According to formula (1) on the solar power cycle component extraction, the results shown in Figure 2. The periodic components are regular and need not be predicted anymore. The volatility of the remaining components is relatively strong, which needs to be analyzed and its short-term forecast.

III. PV POWER REMAINING COMPONENT ANALYSIS OF THE FACTORS

The PV power is affected by various meteorological factors and the state of PV modules. In this paper, Pearson correlation coefficient method [12] was used to analyze the residual power of PV power. The correlation between the atmospheric irradiance, direct irradiance, diffusion of energy, tilt 15° south of the irradiance, average wind speed, temperature and

TABLE 3. Pearson correlation coefficient between the various factors.

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	Factor 8	Factor 9
Factor 1	1.000	0.871	0.572	0.998	0.967	0.566	0.532	-0.454	0.337
Factor 2	0.871	1.000	0.207	0.876	0.807	0.457	0.543	-0.121	0.471
Factor 3	0.572	0.207	1.000	0.535	0.527	0.400	0.224	-0.135	-0.319
Factor 4	0.998	0.876	0.535	1.000	0.973	0.555	0.531	-0.477	0.358
Factor 5	0.967	0.807	0.527	0.973	1.000	0.493	0.433	-0.121	0.359
Factor 6	0.566	0.457	0.400	0.555	0.493	1.000	0.501	-0.166	0.054
Factor 7	0.532	0.543	0.224	0.531	0.433	0.501	1.000	-0.820	0.213
Factor 8	-0.454	-0.121	-0.135	-0.477	-0.121	-0.166	-0.820	1.000	-0.135
Factor 9	0.337	0.471	-0.319	0.358	0.359	0.054	0.213	-0.135	1.000

TABLE 4. Breakdown of generalized weather type.

Generalized weather type	Weather type
Sunny day	Clear, sunny and cloudy, cloudy and sunny
Cloudy weather	Cloudy
Rain or snow weather	fog, rain, sleet, snow

humidity was calculated. Pearson correlation has also become a product of the cumulative difference is a linear correlation method proposed by British statistician Pearson in the 20th century. Suppose there are two variables X, Y, then the Pearson correlation coefficient between the two variables can be calculated by the formula (3).

$$\rho_{X,Y} = \frac{E(XY) - E(X)E(Y)}{\sqrt{E\left(X^2\right) - E^2(X)}\sqrt{E\left(Y^2\right) - E^2(Y)}} \tag{3}$$

The atmospheric irradiance, direct irradiance, diffusion of energy, tilt 15° south of the irradiance, PV power, average wind speed, ambient temperature, humidity and residual components are defined as factor 1-9. According to the above formula can be obtained between the correlation coefficient of the factors, as shown in the table 3.

It can also be seen from the table that there are also certain coupling relationships between meteorological factors, such as a positive correlation between temperature and irradiance. In order to accelerate the forecasting speed of ultra-short-term power and weaken the coupling relationship between various factors, this paper uses the LSH model to predict the remaining components of optical power.

IV. GENERALIZED WEATHER TYPES CLUSTERING

Climate change is very complicated. Types of weather can be divided into more than 30 types according to the degree of weather, rainfall, wind and temperature. Considering the probability of occurrence of the weather type and considering the feasibility of the study, and in order to predict the PV power more efficiently, this paper classifies the types of weather into three generalized weather types: clear, cloudy and rain and snow. Table 4 shows the specific common types of weather included in each generalized type of weather.

Forecast the weather for the next 4 hours every 15 minutes based on weather forecast information.There are many prediction methods for the residual variables. For example, the Group-LASSO model proposed in [13] can be obtained by using this model plus the ultra-short-term photovoltaic power prediction mode [14]. In order to solve this kind of problem more effectively and reasonably, this paper adopts the LSH method.

V. LSH ALGORITHM

In applications such as information retrieval, data mining, and recommendation systems, one of the problems that we often encounter is facing massive high-dimensional data and finding nearest neighbors. If a linear search is used, then the efficiency for low-dimensional data is acceptable, and for high-dimensional data, it is very time consuming. In order to solve such problems, people designed a special hash function, making two data with high similarity map to the same hash value with higher probability, and making two data with very low similarity be extremely low. The probability is mapped to the same hash value. We call this function LSH (local sensitive hash). The most fundamental role of LSH is to be able to efficiently handle the nearest neighbor problem of massive high-dimensional data.

We will be a family of hash functions like this

$$H = \{h : S \to U\} \tag{4}$$

If for any function H in h, the following two conditions are satisfied:

$$\begin{cases} \text{If } d(O_1, O_2) < r_1, & \textit{Then } \Pr[h(O_1) = h(O_2)] \geq p_1 \\ \text{If } d(O_1, O_2) > r_2, & \textit{Then } \Pr[h(O_1) = h(O_2)] \geq p_2 \end{cases} \tag{5}$$

it will be Called sensitive with (r_1, r_2, p_1, p_2). Where $O_1, O_2 \in S$, which represents two data objects with multidimensional attributes; $d(O_1, O_2)$ is the degree of dissimilarity for 2 objects, which is 1-similarity. In fact, the above two conditions put it plainly that the probability of mapping to the same hash value is sufficiently large when they are similar enough, while the probability of mapping the same hash value is small enough when they are not similar.

The definition of similarity is different according to the actual situation. For different similarity measurement methods, the algorithm design of local sensitive hash is different. We mainly look at two different LSHs under the two most common similarities:

1. Min-hash when measuring data similarity using Jaccard coefficient

2. P-stable hash when using Euclidean distance to measure data similarity

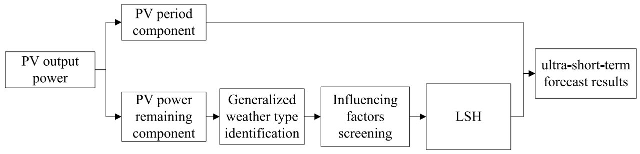

FIGURE 3. Flow chart of ultra-short-term prediction method for PV power based on feature extraction of similar segments.

This paper adopts the second method. For the specific model construction, please refer to the literature [15].

After completing the LSH function family construction for p-stable distribution, the logic of AND then OR is used to implement the construction of the hash table and query the nearest neighbor. Assuming $P = \Pr[h_{a,b}(v_1) = h_{a,b}(v_2)]$ (a and b are function parameters), the probability that two pieces of data are considered neighbors is:

$$1-\left(1-P^k\right)^L \tag{6}$$

where k is the serial number of the randomly selected function; L is the sequence number between the function groups

A similar segment of photovoltaic power can be found more efficiently by using a local sensitive hashing algorithm.

In summary, the method used in this paper can be represented by Figure 3.

VI. CASE ANALYSIS

A. PV POWER ULTRA-SHORT-TERM FORECASTING EVALUATION INDEX

The error assessment indicators used were root mean square error (RMSE), mean relative error (MRE) and pass rate (QR). The RMSE reflects the degree of dispersion between the predicted value and the actual value. The average relative error reflects the degree of difference between the predicted value and the actual value. The pass rate can reflect the proportion of the predicted number of predicted points that reach the national target.

$$E_{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{P_{Mi}-P_{Pi}}{Cap_i}\right)^2} \tag{7}$$

$$MRE = \frac{\sum_{i=1}^{n}|P_{Mi}-P_{Pi}|}{Cap_i \times n} \tag{8}$$

$$B_i = \begin{cases} 1 & \left(1-\dfrac{|P_{Mi}-P_{Pi}|}{Cap_i}\right) \geq 0.75 \\ 0 & \left(1-\dfrac{|P_{Mi}-P_{Pi}|}{Cap_i}\right) < 0.75 \end{cases}$$

$$QR = \frac{1}{n}\sum_{i=1}^{n} B_i \times 100\% \tag{9}$$

where P_{Mi} is the actual power at i; P_{Pi} is the predicted power at i; Cap_i is the average daily power-on capacity; and n is the number of daily samples.

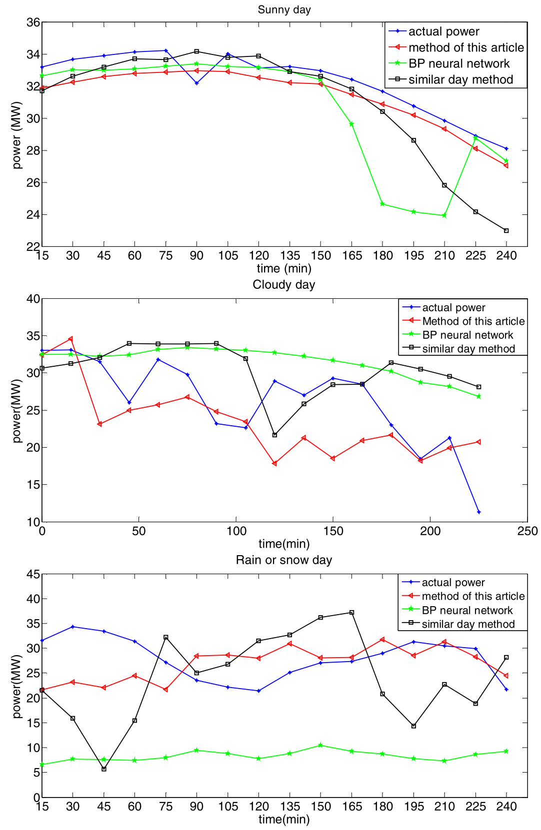

FIGURE 4. Comparison of prediction results at a determine time under different weather types.

TABLE 5. Evaluation results of each predict method under different weather types.

Weather Type	Model	Error index RMSE	MRE	QR
Sunny day	Method of this paper	5.17%	0.0215	96.42%
	BP neural network	8.15%	0.0879	84.38%
	Similar day method	5.81%	0.0235	95.92%
Cloudy day	Method of this paper	9.09%	0.0977	95.70%
	BP neural network	15.34%	0.1242	82.56%
	Similar day method	22.17%	0.1350	73.96%
Rain or snow weather	Method of this paper	12.61%	0.1023	92.75%
	BP neural network	15.78%	0.1279	86.54%
	Similar day method	14.50%	0.1151	90.63%

B. ULTRA-SHORT-TERM FORECAST RESULTS OF PV POWER UNDER DIFFERENT WEATHER TYPES

To verify the effectiveness of the proposed method, BP neural network method [16] and similar day method [17] are used as a comparison method. The graph below shows the comparison of power curves of different methods for predicting the next four hours when the time is 12:00 at different weather types. It can be seen from the figure that the prediction effect has a great relationship with the type of weather. In the case of larger climatic fluctuations, the prediction effect is worse. The similarity segment method based on periodic extraction adopted in this paper is better than other methods in forecasting different weather types.

Table 5 shows the statistical indicators of the one-day forecast result assessment indicators using different PV power

TABLE 6. Evaluation results of each predict method under different weather types.

Weather Type	Model	Error index RMSE	MRE	QR
Sunny day	Method of this paper	4.23%	0.0202	97.42%
	BP neural network	6.44%	0.0616	88.24%
	Similar day method	4.86%	0.0196	98.35%
Cloudy day	Method of this paper	9.92%	0.1158	92.52%
	BP neural network	13.02%	0.1533	80.35%
	Similar day method	17.43%	0.1741	75.23%
Rain or snow weather	Method of this paper	11.55%	0.0993	90.24%
	BP neural network	16.54%	0.1603	87.85%
	Similar day method	15.89%	0.1159	86.52%

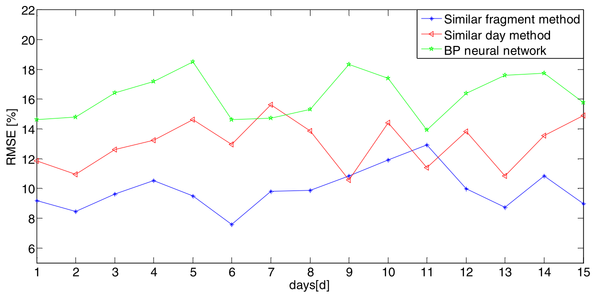

FIGURE 5. Distribution of RMSE under cloudy weather.

prediction methods for each typical weather type. It can be seen from the table that, similar to the weather method in Japan under large-scale climate change, the prediction effect is not good; the BP neural network method can predict the output power fluctuation caused by climate change to a certain extent in the forecast results, but the overall prediction accuracy is not high. In this paper, the prediction method is used to segment the residual components of the output power, and the meteorological information is used to find the segment similar to the residual component of the power to be predicted, so that the high-precision prediction result can be obtained more effectively. Under the sunny conditions, the root mean square error decreased by at least 0.64% compared with the traditional prediction model, the average relative error decreased by at least 0.2%, and the decline of other weather indicators was more obvious.

In order to further illustrate the effectiveness of the proposed method, the data from a small distributed PV power station in a university in Jilin Province are used to simulate the calculation. The installed capacity of the PV power plant is 10 kW, sampling time is from June 2014 to December 2014, sampling interval is 15 min. Table 6 is also a typical weather type, the use of different methods of PV power forecasting method for 15 days when the forecasting index tables. In Figure 5, the daily distribution of root mean square error for each method under cloudy weather conditions is listed. It can be seen from Table 6 and Figure 5 that the method used in this paper has a certain improvement on the prediction accuracy compared with the comparative method, which further proves the effectiveness of the method used in this paper in the prediction of PV power.

VII. CONCLUSION

Considering the characteristics of PV output power, this paper periodically analyzes the PV power time series and extracts date components to lay a foundation for further prediction.

PV output power is influenced by the state of each meteorological factor and composition, and there is a certain coupling relationship between meteorological factors. The LSH algorithm can quickly and efficiently classify different power segments and predict the predicted segments.

It is predicted that the difficulty of photovoltaic power generation under different weather types is different. Under different types of weather conditions, the PV power ultra-short-term prediction model based on the periodic extraction of PV energy and the LSH algorithm has higher prediction accuracy than other methods.

REFERENCES

[1] Y. Gong, L. Zonxiang, and Q. Ying, "PV power prediction technology," *Automat. Electr. Power Syst.*, vol. 40, pp. 140–151, 2016.
[2] F. Xu, J. Tong, and S. Cai, "Model cloud modeling for distributed ultra-short-term PV power forecasting," *J. Sol. Energy.*, vol. 37, pp. 1748–1755, 2016.
[3] Y. Zhu and J. Tian, "Application of least square support vector machine in photovoltaic power forecasting," *Power Syst. Technol.*, vol. 35, pp. 54–59, Jul. 2011.
[4] M. Ding, L. Wang, and R. Bi, "A short-term prediction model to forecast output power of photovoltaic system based on improved BP neural network," *Power Syst. Protection Control.*, vol. 40, pp. 93–99, Nov. 2012.
[5] D.-Q. Yang, Y. Wang, and Y.-J. Jiao, "Generation forecasting for photovoltaic system based on wavelet neural networks," *Renew. Energy Resour.*, vol. 31, pp. 1–5, Jul. 2013.
[6] M. Ding and N. Xu, "Short-term output power prediction method of PV power generation system based on Markov chain," *Power Syst. Technol.*, vol. 35, pp. 152–157, 2011.
[7] R. Song, F. Liu, and M. Dongmei, "Ultra-short-term prediction model of PV power generation based on similar days and WNN," *J. Electr. Instrum. Meas.*, vol. 54, pp. 75–80, 2017.
[8] A. Mellit and A. M. Pavan, "A 24-h forecast of solar irradiance using artificial neural network: Application for performance prediction of a grid-connected PV plant at Trieste, Italy," *Sol. Energy*, vol. 84, no. 5, pp. 807–821, 2010.
[9] B. Geng, G. Zhenyan, and B. Hengyuan, "Prediction of PV power generation based on GA-BP neural network with similar date," *Electr. Power Syst. Automat.*, vol. 29, pp. 118–123, 2017.
[10] O. Şenkal and T. Kuleli, "Estimation of solar radiation over Turkey using artificial neural network and satellite data," *Appl. Energy*, vol. 86, nos. 7–8, pp. 1222–1228, 2009.
[11] J. Wang, Z. Song, and R. Ran, "Short-term PV power generation rolling rolling based on SVM," *J. Power Syst. Automat.*, vol. 28, pp. 9–13, 2016.
[12] C. Li, P. Cao, and J. Li, "A review of large-scale distributed PV grid-connected reactive voltage control method," *J. Northeastern Electr. Power Univ.*, vol. 37, pp. 82–88, 2017.
[13] J. Zhang and B. Zhang, "Application of scoring model based on group-LASSO method for generalized semi-parametric additively," *J. Math. Statist. Manage.*, vol. 35, pp. 517–524, 2016.
[14] B. Liu, Z. He, and J. Hao, "Current status and development trend of wind power generation," *J. Northeast Electric Power Univ.*, vol. 36, pp. 7–13, 2016.
[15] M. Datar, N. Immorlica, P. Indyk, and V. S. Mirrokni, "Locality-sensitive hashing scheme based on p-stable distributions," in *Proc. 29th Annu. Symp. Comput. Geometry*, 2004, pp. 253–262.
[16] C. Zhang, Y. Tang, and Y. Li, "Application of neural network to small sample solar forecasting method," *Electr. Power Automat. Equip.*, vol. 37, pp. 101–106, 2017.
[17] H. Liu and M. Yang, "Identification and prediction of wind power ramp based on wavelet transform and support vector machine," *J. Northeast Electr. Power Univ.*, vol. 36, pp. 30–35, 2016.

MAO YANG was born in Jilin, China, in 1982. He received the Ph.D. degree in control theory and control engineering from Jilin University, Changchun, China, in 2010.

He is currently an Associate Professor with the Department of Electrical Engineering, Northeast Electric Power University. His research interests include power system load forecasting and new energy generation.

XIN HUANG was born in Henan, China, in 1993. He received the B.Sc. degree in mathematics and applied mathematics from Northeast Electric Power University, Jilin, China, in 2016.

He is currently pursuing the M.Sc. degree in electrical engineering, Northeast Electric Power University. His research interests include PV power ultra-short-term forecast.