

A Hybrid Model for Short-term PV Output Forecasting Based on PCA-GWO-GRNN

Leijiao Ge, Yiming Xian, Jun Yan, Bo Wang, and Zhongguan Wang

Abstract—High-precision day-ahead short-term photovoltaic (PV) output forecasting is essential in PV integration to the smart distribution networks and multi-energy system, and provides the foundation for the security, stability, and economic operation of PV systems. This paper proposes a hybrid model based on principal component analysis, grey wolf optimization and generalized regression neural network (PCA-GWO-GRNN) for day-ahead short-term PV output forecasting, considering the features of multiple influencing factors and strong uncertainty. This paper first uses the PCA to reduce the dimension of meteorological features. Then, the high-precision day-ahead short-term PV output forecasting based on GWO-GRNN model is realized. GRNN is used to regressively analyze the input features after dimension reduction, and the parameter of GRNN is optimized by using GWO, which has strong global searching ability and fast convergence. The proposed PCA-GWO-GRNN model effectively achieves a high precision in day-ahead short-term PV output forecasting, which is demonstrated in a case study on a real PV plant in Jiangsu province, China. The results have validated the accuracy and applicability of the proposed model in real scenarios.

Index Terms—Photovoltaic output forecasting, principal component analysis (PCA), grey wolf optimization (GWO), generalized regression neural network (GRNN).

I. INTRODUCTION

PHOTOVOLTAIC (PV) power generation technology is becoming an essential component of the smart grid as PV plants are being connected to the smart distribution network (SDN) on a large scale. The effective way to make full use of PV generation system is to keep the total output power of multi-energy system relatively stable, reduce power fluctuation, improve power quality and reduce the impact on

the power grid. However, the output of PV plant has strong randomness and intermittent because of the significant effect by a number of factors such as solar irradiance, temperature, and humidity. Therefore, large-scale PV integration in SDN has a strong impact on the safety and stability of power system operations. In summary, to reduce the impact of large-scale PV penetration and ensure the secure and economic operation of power systems, it is imperative to achieve an accurately day-ahead short-term PV output forecasting by an effective model [1]–[4].

Existing short-term PV output forecasting approaches can be mostly categorized as physical or statistical methods. The physical method is built on solar irradiance transfer equation, PV module operation equation, and/or other physical equations. This category relies on detailed and precise geographic location information as well as the weather and solar irradiance data of the PV plants to create the model in an often-complicated process [5]. For example, [6] establishes a mathematical model of the PV cell and the PV inverter with high simulation precision to calculate the output of PV system at different irradiances and ambient temperatures. The physical method involves many links and its parameters are difficult to be solved, and the physical formula itself also has some errors, thus the anti-interference ability and robustness of the physical method are poor [7], [8]. The statistical method is based on statistical rules between input and output factors of the PV forecasting model. It does not require complicated geographic information of PV power stations [9], which simplifies the forecasting process by statistical methods such as the Markov chain [10], support vector machine (SVM) [11], and artificial neural networks (ANNs) [12]. Reference [13] establishes five kinds of long short-term memory (LSTM) network models to forecast PV output, which considers the temporal changes in PV plant when constructing the forecasting models. These models have better forecasting accuracy than multiple linear regression (MLR) and bagged regression tree (BRT) models. Reference [14] proposes a hybrid improved multi-verse optimizer algorithm (HIMVO) to optimize the SVM for PV output forecasting. The HIMVO has higher convergence rate, better optimization ability and stability than other optimization algorithms. However, SVM is difficult to be implemented for large-scale training samples. The combined model of PV output forecasting is also one of current research hotspots. In [15], an ensemble technique is proposed, which creates a set of individual recursive arithmetic average models on the forecasting of power out-

Manuscript received: January 3, 2020; accepted: July 3, 2020. Date of Cross-Check: July 2, 2020. Date of online publication: September 9, 2020.

This work was supported by the National Key Research and Development Program of China (No. 2018YFB1500800) and the National Natural Science Foundation of China (No. 51807134).

This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>).

L. Ge and Z. Wang is with the Key Laboratory of Smart Grid of Ministry of Education, Tianjin University, Tianjin 300072, China (e-mail: legendglj99@tju.edu.cn; wang_zg@tju.edu.cn).

Y. Xian (corresponding author) is with the State Key Laboratory of Reliability and Intelligence of Electrical Equipment, Hebei University of Technology, Tianjin 300132, China (e-mail: xianyiming01@sina.com).

J. Yan is with the Concordia Institute for Information Systems Engineering, Concordia University, Montreal, QC H3G 1M8, Canada (e-mail: jun.yan@concordia.ca).

B. Wang is with the School of Electrical and Automation, Wuhan University, Wuhan 430072, China (e-mail: whwdwb@whu.edu.cn).

DOI: 10.35833/MPCE.2020.000004



puts. The data-driven ensemble model can be applied to situations where a large amount of data needs to be processed and performs better than the other forecasting techniques. In [16], a model called short-term PV power dynamic weighted combination forecasting based on the least squares (LS) model is proposed. This model is superior to other single models in forecasting PV power. However, the weighting of the combined model has always been a difficult problem.

PV output is directly related to solar irradiance, but it is also affected by multiple complex meteorological factors, including temperature, humidity, and precipitation and others [17]. Nevertheless, too many meteorological input features will reduce the sensitivity of the PV output forecasting model to solar irradiance. In the meantime, such variables often exhibit high randomness, and the implementation of data mining is difficult. As a kind of artificial intelligence forecasting algorithm, the ANN algorithm is a computational algorithm based on animals' central nervous systems (especially the brain), which can be used for pattern recognition and machine learning. As a powerful regression tool with a dynamic network structure, the generalized regression neural network (GRNN) belongs to the category of ANNs, which has strong nonlinear mapping capability, high error tolerance and robustness. However, the most critical challenge is how to determine the spread parameter appropriately when the GRNN model is employed to deal with the actual problems. At present, most researchers select the spread parameter value by sufficient experiments or experience, which consumes much time but cannot guarantee the forecasting accuracy. Therefore, in this paper, the spread parameter value of the GRNN model is automatically determined by the grey wolf optimization (GWO) algorithm. Compared with the particle swarm optimization (PSO) algorithm, the GWO algorithm has fewer modulation parameters and is less difficult to be realized. Also, the search ability and search speed of GWO algorithm are stronger and faster than those of the PSO algorithm. Based on the above analysis, this paper proposes a new hybrid model to forecast PV output through a three-step process of principal component analysis (PCA), GWO and GRNN, as shown in Fig. 1. With a focus on day-ahead short-term PV output forecasting, the main contributions of this paper are:

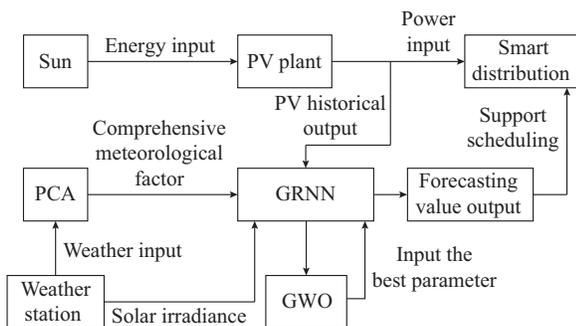


Fig. 1. PV output forecasting model based on PCA-GWO-GRNN.

1) In order to eliminate the collinearity of the input features of meteorological conditions, reduce the dimension of the input features, and avoid the model from overfitting, the

PCA is used to reduce the dimensionality of the input features of meteorological conditions.

2) A day-ahead short-term PV output forecasting model based on GWO-GRNN is proposed. Among them, GRNN is used to fit the complex nonlinear relationship between PV output and input features. And the parameter of GRNN is optimized by GWO with strong global searching capacity. The case study verifies that the model proposed in this paper has strong forecasting accuracy and robustness.

The rest of the paper is organized as follows. The PCA is adopted to reduce the dimension of multi-weather factors and extract the features in Section II. Section III illustrates the basic theories of GRNN and GWO. Section IV presents a case study using a real PV plant in Jiangsu province, China, which verifies the accuracy and robustness of the proposed forecasting model. Section V draws the conclusions.

II. DIMENSION REDUCTION OF METEOROLOGICAL INPUT FEATURES BASED ON PCA

The energy used for PV output is completely derived from the solar irradiance, thus the solar irradiance directly affects the PV output. In addition, the output of PV plants is also affected by many other meteorological factors such as temperature, atmospheric pressure, and humidity. However, too many meteorological input features will cause the forecasting model to have a complicated structure, increase the training burden and affect the learning speed. Moreover, it will reduce the sensitivity of the forecasting model to solar irradiance.

The meteorological input features for PV output forecasting often have strong correlation. In this paper, the PCA is adopted to simplify the meteorological input features into a comprehensive meteorological factor. The PCA mainly finds a small set of linear combination variables to replace the original variables, so as to achieve the purpose of effectively separating the commonality between the data vectors while retaining the original variable information [18]-[21].

Assume that there are n samples and each sample has p variables, we can create an $n \times p$ data matrix. The process of PCA includes six steps as follows.

Step 1: standardize the original data into valid data.

Step 2: calculate the correlation coefficient matrix \mathbf{R} .

Step 3: compute eigenvalues and eigenvectors. Firstly, the characteristic equation $|\lambda \mathbf{I} - \mathbf{R}| = 0$ is solved and the Jacobian method is used to find the eigenvalues $\lambda_j (j = 1, 2, \dots, p)$, where \mathbf{I} is the identity matrix and λ is the eigenvalue. The eigenvalues are arranged in order of size, i.e., $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$. Then, the eigenvectors $\mathbf{a}_j (j = 1, 2, \dots, p)$ corresponding to the eigenvalues λ_j are found.

Step 4: calculate the principal component contribution rate and the cumulative contribution rate. The contribution rate of principal component is formulated as $\eta_i = \lambda_i / \sum_{k=1}^p \lambda_k, i = 1, 2, \dots, m$. The cumulative contribution rate is defined by $\rho_k = \sum_{i=1}^k \lambda_i / \sum_{k=1}^p \lambda_k, i = 1, 2, \dots, m$, where m is the number of principal components. In general, the principal component is tak-

en, corresponding to the eigenvalues with cumulative contribution rate of 85%-95%.

Step 5: calculate the principal component value. The value of each principal component is calculated according to (1).

$$\begin{cases} z_1 = a_{11}x_1 + a_{12}x_2 + \cdots + a_{1j}x_j + \cdots + a_{1p}x_p \\ z_2 = a_{21}x_1 + a_{22}x_2 + \cdots + a_{2j}x_j + \cdots + a_{2p}x_p \\ \vdots \\ z_i = a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{ij}x_j + \cdots + a_{ip}x_p \\ \vdots \\ z_m = a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mj}x_j + \cdots + a_{mp}x_p \end{cases} \quad (1)$$

where a_{ij} ($i=1, 2, \dots, m, j=1, 2, \dots, p$) is the element of the eigenvector matrix; and z_i ($i=1, 2, \dots, m$) is the value of principal component.

Step 6: calculate the comprehensive meteorological factor. The comprehensive meteorological factor F can be obtained from the linear weighted sum of the above m principal components, as shown in (2).

$$F = \eta_1 z_1 + \eta_2 z_2 + \cdots + \eta_i z_i + \cdots + \eta_m z_m \quad (2)$$

where η_i is the contribution rate.

III. GRNN BASED ON GWO

Recently, ANNs have entered the PV output forecasting realm, and a number of approaches are developed based on Elman neural network [22], back propagation (BP) neural network [23], and LSTM neural network [24], etc. However, these networks are all global approximation networks, and one or more weights of these networks have effect on each output, resulting in a slower learning speed. In addition, these networks have randomness when determining the weights, resulting in an uncertain relationship between the input and output after each training, which makes the forecasting results different. Moreover, the determination of the number of layers of these networks and the number of hidden layer nodes lack theoretical guidance. As an improved algorithm of radial basis function (RBF) neural network, GRNN is a local approximation neural network [25]. The establishment of GRNN has a clear theoretical basis, there is no need to perform a cyclic training process, and there is no need to adjust the connection weights between neurons during the training process. Therefore, GRNN has the advantages of satisfactory robustness and fast calculation rate. And GRNN also has a stronger nonlinear mapping ability and a flexible network structure with high fault tolerance [26]-[28]. As a result, GRNN is a suitable tool for forecasting the PV output with many factors and complex randomness [29].

The GWO algorithm is a swarm intelligence algorithm proposed by [30] with good self-organizing learning ability, simple parameters, easy implementation, and good global searchability [31], [32]. And by comparing with other four famous meta-heuristic algorithms (PSO algorithm, gravity search algorithm, differential evolution algorithm and fast evolutionary programming algorithm) on 29 test functions, the simulation test proves its superiority. In this paper, the GWO algorithm is adopted to the selection of the spread parameter of GRNN to enhance the performance of GRNN, which can

also be found in our previous study [33].

A. GRNN

GRNN is composed of the input, the pattern, the summation, and the output layers. The corresponding input and output vectors can be denoted by $\mathbf{X} = [X_1, X_2, \dots, X_n]$ and $\mathbf{h} = [h_1, h_2, \dots, h_m]$, respectively. The GRNN structure is shown in Fig. 2.

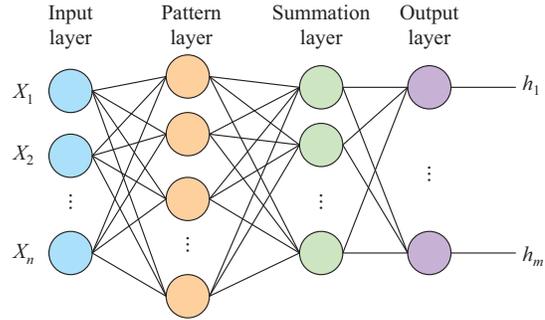


Fig. 2. GRNN structure.

The number of input layer neurons is the same as the input dimension of the training samples, and each neuron transmits the input data directly to the pattern layer.

The number of pattern layer neurons is consistent with the number of training samples, and the transfer function is the RBF:

$$K_i = \exp\left(-\frac{(\mathbf{X} - X_i)^T (\mathbf{X} - X_i)}{2\sigma^2}\right) \quad (3)$$

where K_i is the transfer function; and σ is the spread parameter.

The summation layer utilizes two summation ways: one is to calculate the weighted sum of the output of each neuron in the pattern layer; the other is to calculate the arithmetic sum of the outputs of the neurons in the pattern layer. The two types of formulas are shown in (4) and (5), respectively.

$$S_{Nj} = \sum_{i=1}^n h_{ij} \exp\left(-\frac{(\mathbf{X} - X_i)^T (\mathbf{X} - X_i)}{2\sigma^2}\right) \quad j=1, 2, \dots, m \quad (4)$$

$$S_D = \sum_{i=1}^n \exp\left(-\frac{(\mathbf{X} - X_i)^T (\mathbf{X} - X_i)}{2\sigma^2}\right) \quad (5)$$

where S_{Nj} is the weighted sum; S_D is the arithmetic sum; $j=1, 2, \dots, m$; and h_{ij} is the j^{th} element in the i^{th} training sample, and the value of j is 1 during PV output forecasting.

The output layer adopts a linear function to output the result, and the estimation of the corresponding neuron j is:

$$h_j = \frac{S_{Nj}}{S_D} \quad (6)$$

GRNN has only one parameter that needs to be determined, i.e., the spread parameter σ . If σ is too large, the forecasted value will approximate the mean of the target value in all training samples. If σ is too small, the generalization ability

of the forecasting model will be limited. Therefore, in order to determine the best value of σ , the GWO is applied to find the optimal value and improve the forecasting accuracy of GRNN.

B. GWO Algorithm

The GWO algorithm imitates the leadership hierarchy and hunting mechanism of grey wolves in nature. Compared with the PSO algorithm, the GWO algorithm does not depend on the setting of the parameters, and has both stronger search ability and faster search speed. α , β , δ , and ω are employed to simulate the leadership hierarchy as shown in Fig. 3 [33]. The primary stages of grey wolf hunting are as follows: ① searching and tracking the prey; ② chasing and encircling the prey until it stops moving; ③ attacking the prey.

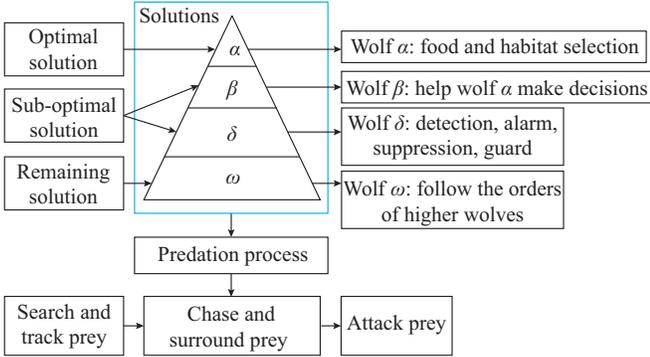


Fig. 3. Leadership hierarchy of wolf group.

To simulate the social hierarchy of wolves in GWO algorithm, we define three wolves α , β , and δ , and denote the remaining wolves as ω , according to the hierarchy. Among them, wolf α is the optimal solution, wolves β and δ are the sub-optimal solutions, and the remaining wolves ω are candidate solutions. The wolf pack approaches the optimal solution in the search space through the initial solution of three individual wolves α , β , and δ . The locations of wolves are then updated and evolved, while the distance from the prey is updated until the optimal solution is obtained. Figure 4 shows the positioning of the grey wolf and its prey, and the parameters involved in the equation to update the position of the grey wolf in the search space [30]. The position of the prey would be in a random place within a circle that is defined by the positions of α , β , δ in the search space. R is the radius of that circle.

The distance D between a wolf and the prey should be determined in advance before hunting:

$$D = |C\Phi_p(i) - \Phi(i)| \quad (7)$$

where $\Phi_p(i)$ and $\Phi(i)$ are the locations of the prey and the wolf at the iteration i , respectively; and $C = (c_i)$ is the coefficient vector.

$$c_i = 2r_1 \quad (8)$$

where r_1 is the spatial distance coefficient in $[0, 1]$.

As the distance between the individual wolf and the prey decreases, the position of the individual wolf is constantly updated by:

$$\Phi(i+1) = \Phi_p(i) - \zeta D \quad (9)$$

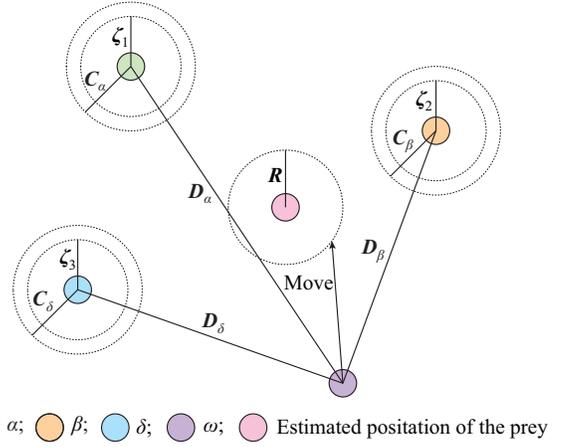


Fig. 4. Location of grey wolf and prey in search space.

where ζ is the coefficient vector.

$$\zeta = 2ar_2 - a \quad (10)$$

where a decreases from 2 to 0 as the number of iterations increases; and r_2 is the same random coefficient as r_1 in $[0, 1]$.

Wolves α , β , and δ are assumed to be the first wolves closest to the prey in the wolf pack. The positions of the remaining wolves ω are updated by:

$$D_\alpha = |C_\alpha \Phi_\alpha(i) - \Phi(i)| \quad (11)$$

$$D_\beta = |C_\beta \Phi_\beta(i) - \Phi(i)| \quad (12)$$

$$D_\delta = |C_\delta \Phi_\delta(i) - \Phi(i)| \quad (13)$$

$$\Phi_1 = \Phi_\alpha - \zeta_1 D_\alpha \quad (14)$$

$$\Phi_2 = \Phi_\beta - \zeta_2 D_\beta \quad (15)$$

$$\Phi_3 = \Phi_\delta - \zeta_3 D_\delta \quad (16)$$

$$\Phi(i+1) = \frac{\Phi_1 + \Phi_2 + \Phi_3}{3} \quad (17)$$

where Φ_α , Φ_β , and Φ_δ are the current positions of wolves α , β , and δ , respectively; C_α , C_β , and C_δ are the coefficient vectors of wolves α , β , and δ , respectively; ζ_1 , ζ_2 , and ζ_3 are the coefficient vectors; and D_α , D_β , and D_δ are the distances between the individual wolves α , β , δ and the head wolf in the remaining wolves ω , respectively.

The GWO algorithm locates the range of prey (optimal solution) through the positions of wolves α , β , and δ , as it gradually reduces the distance from the prey before finally catching it. Compared with other intelligent algorithms that search for the optimal solution, the GWO algorithm is capable of a multi-position search, which significantly improves the global search capacity.

C. Process of Proposed Model

This paper proposes a hybrid model based on PCA-GWO-GRNN for the PV output forecasting, which can be divided into the following five steps.

Step 1: data preprocessing.

Step 2: dimension reduction. The PCA is adopted to simplify the meteorological input features into a comprehensive

meteorological factor.

Step 3: sample selection. The historic weather type, temperature, and day of year (DOY) are used as indicators to identify similar days for GRNN training. According to the meteorological conditions on the day of forecasting with the weather forecast, the samples with the same weather type as the day of forecasting are selected from the historical day to constitute the set A . Samples in set A whose daily maximum temperature is within ± 3 °C from the day of forecasting are selected to form set B . Similarly, the samples in set B , whose DOY is within 30 days from the day of forecasting, are selected to form set C , which is called the set of similar days and used for parameter optimization and model training.

Step 4: parameter optimization. The samples in set C are divided into 10 folds for cross-validation, among which one fold is selected as the validation set, and the other 9 folds are combined as the training set. Then, we can use the forecasting error of GRNN as the fitness function of GWO algorithm to optimize the parameter σ of GRNN. The relevant initial parameters of the GWO algorithm are set as follows: the number of wolves is 20, the number of iterations is 50, and the variable dimension is 1. Finally, the optimal value among the 10 validations will be chosen.

Step 5: offline training. After determining the optimal parameter of the GRNN model, the training sample set is used for offline training of the GRNN model. The PV output forecasting model is then obtained after importing the input feature data on the day of forecasting to the trained GRNN model. The flowchart of short-term PV output forecasting based on PCA-GWO-GRNN is shown in Fig. 5.

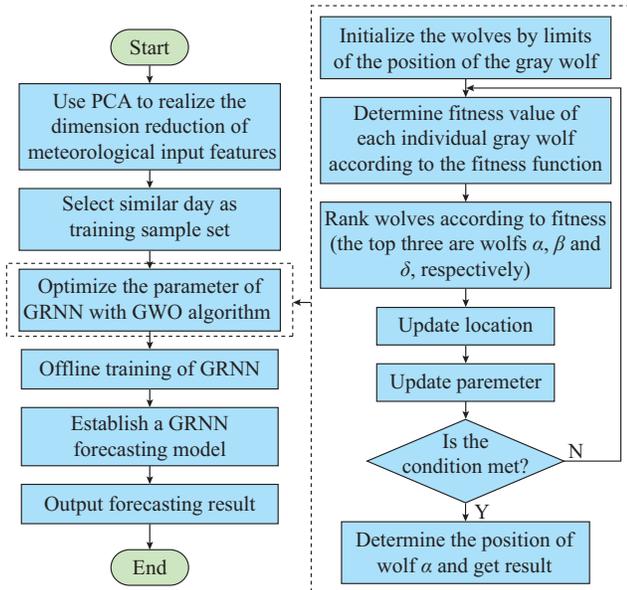


Fig. 5. Flowchart of short-term PV output forecasting based on PCA-GWO-GRNN.

IV. CASE STUDY

In this case study, the focus is on verifying the accuracy and robustness of the proposed model for day-ahead short-term PV output forecasting [34], [35]. To this end, the actual PV output data from a PV plant in Jiangsu province, China

from January 1, 2018 to December 31, 2018 are collected, which have a 15-minute interval from 08:00 to 17:00. In order to verify the accuracy and superiority of the proposed model based on the PCA-GWO-GRNN, three forecasting models, including GWO-GRNN, PCA-LSTM, and PCA-PSO-BP models, are used for PV output forecasting. The forecasting results of the four models are compared and verified.

A. Forecasting Accuracy Evaluation

In this paper, after obtaining the final PV output forecasting value, the nominal mean absolute error (nMAE) and root mean square error (RMSE) are used to evaluate the forecasting accuracy [36], [37], as shown in (18) and (19), respectively.

$$E_M = \frac{1}{n} \sum_{t=1}^n \frac{|\hat{y}_t - \hat{y}_{st}|}{y_z} \times 100\% \quad (18)$$

$$E_R = \sqrt{\frac{1}{n} \sum_{t=1}^n (\hat{y}_t - \hat{y}_{st})^2} \quad (19)$$

where n is the number of forecasting points; \hat{y}_t is the PV output forecasting value at time t ; \hat{y}_{st} is the actual PV output at time t ; and y_z is the installed PV capacity.

B. Data Preprocessing

Taking the features of PV into account, the irradiance, temperature, atmospheric pressure, wind speed, relative humidity, and precipitation are used as input features. The input variables $x_1 - x_8$ and output variable y of the forecasting model are shown in Table I. And all input features are normalized to $[-1, 1]$.

TABLE I
INPUT AND OUTPUT VARIABLES OF FORECASTING MODEL

Variable	Definition
x_1	Temperature at forecasting time
x_2	Atmospheric pressure at forecasting time
x_3	Relative humidity at forecasting time
x_4	Wind speed at forecasting time
x_5	Precipitation at forecasting time
x_6	Solar irradiance at forecasting time
x_7	Solar irradiance at 15 min before forecasting time
x_8	Solar irradiance at 15 min after forecasting time
y	PV output at forecasting time

C. Principal Component Analysis of Meteorological Factors

According to Table I, five meteorological variables x_1 , x_2 , x_3 , x_4 , and x_5 are analyzed by principal component analysis, and five index variables are obtained. The eigenvalues of the covariance matrix of the five variables are shown in Table II.

As shown in Table II, the cumulative contribution rate of the first three principal components reaches 84.20%, thus the first three principal components are taken as the input features of the forecasting model in this paper. According to the PCA model in Section II, the proportion coefficient of principal components is shown in Table III.

TABLE II
EIGENVALUE OF COVARIANCE MATRIX OF FIVE VARIABLES

Component	Eigenvalue	Contribution rate (%)	Cumulative contribution rate (%)
1	1.88	37.67	37.67
2	1.25	25.05	62.72
3	1.07	21.48	84.20
4	0.71	14.19	98.39
5	0.08	1.61	100.00

TABLE III
PROPORTION COEFFICIENT OF PRINCIPAL COMPONENT

Principal component	x_1	x_2	x_3	x_4	x_5
1	0.69	-0.70	-0.02	0.13	0.11
2	-0.12	-0.13	0.78	-0.35	0.49
3	-0.16	0.08	-0.07	0.76	0.63

The sample values of the first three principal components can be obtained as follows:

$$\begin{cases} z_1 = 0.69x_1 - 0.7x_2 - 0.02x_3 + 0.13x_4 + 0.11x_5 \\ z_2 = -0.12x_1 - 0.13x_2 + 0.78x_3 - 0.35x_4 + 0.49x_5 \\ z_3 = -0.16x_1 + 0.08x_2 - 0.07x_3 + 0.76x_4 + 0.63x_5 \end{cases} \quad (20)$$

The weighted summation is based on the weighted contribution rate of each principal component to obtain a comprehensive meteorological factor:

$$F = 0.3767z_1 + 0.2505z_2 + 0.2148z_3 \quad (21)$$

The input features for the proposed model are the comprehensive meteorological factor and the solar irradiance at forecasting time and 15 min before and after the forecasting time.

D. Forecasting Result and Discussion

In order to verify the accuracy and superiority of the PCA-GWO-GRNN model, GWO-GRNN, PCA-LSTM, PCA-PSO-BP and the proposed model are used to forecast the output of the PV plant from July 4 to 31, 2018 (28 days in four weeks). Among them, the similar day selection of the four models adopts the same way as proposed in this paper, and the input and output variables are also the same. The neural network and its optimization algorithm are implemented using MATLAB 2018(b) and PCA is implemented using SPSS 19.

The 28 consecutive forecasting days include 6 sunny days, 12 cloudy days, 1 day of overcast day, and 9 rainy days. For each of the four weather types, 1 day is selected for qualitative analysis. The actual output and forecasting results of the four models are shown in Figs. 6-9.

Figure 6 shows the forecasting and actual curves of PV output on a sunny day. The variation of actual PV output curve is well-regulated. The forecasting curves of the four models are all close to the actual curve and the result of the proposed model is closest to the actual PV output.

Figure 7 shows the forecasting and actual curves of PV output on a cloudy day. Compared with the sunny day, the

thickness and movement of clouds in cloudy weather are difficult to forecast. Between 14:00 and 17:00, the PV output changes very drastically, and the four forecasting curves and the actual curve have a large deviation. For the period with large forecasting error, the forecasting curve of PCA-GWO-GRNN model is closer to the actual curve, which can significantly reduce the forecasting error.

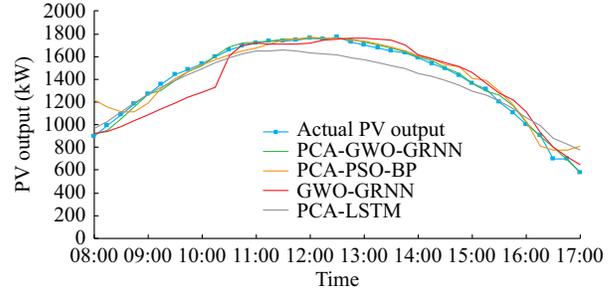


Fig. 6. Forecasting results on a sunny day.

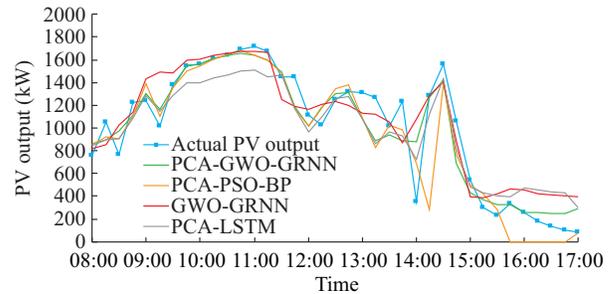


Fig. 7. Forecasting results on a cloudy day.

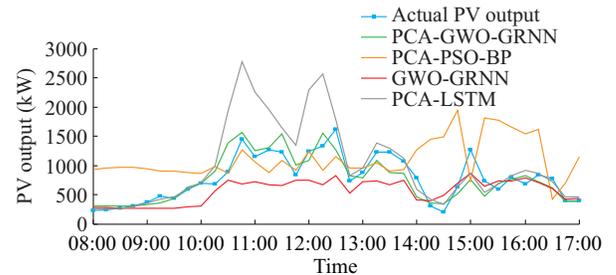


Fig. 8. Forecasting results on an overcast day.

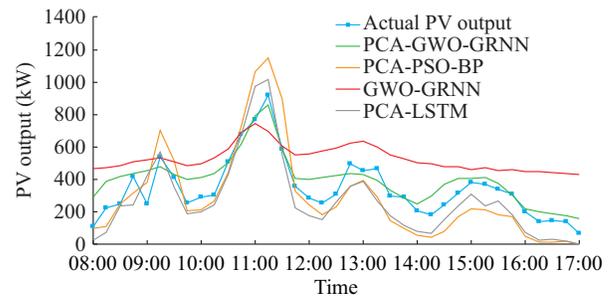


Fig. 9. Forecasting results on a rainy day.

Figure 8 shows the forecasting and actual curves of PV output on an overcast day. Since the number of training samples on overcast days is small, and the forecasting accuracy of BP neural network and LSTM neural network for small

sample data is insufficient, the forecasting results of the PCA-LSTM model and PCA-PSO-BP model differ greatly from the actual curve. The forecasting curve of the proposed model can better reflect the overall change trend of the actual power.

Figure 9 shows the forecasting and actual curves of PV output on a rainy day. The PV output on rainy days has more uncertainty and randomness. The forecasting results of all four models are significantly different from the actual curve. However, the curve of proposed model is closer to the actual curve.

The comparison of accuracy about Figs. 6-9 is shown in Table IV. As shown in Table IV, the average value of nMAE of the proposed model is 2.55%, which is reduced by 2.41%, 1.86% and 3.30%, compared with the GWO-GRNN, PCA-LSTM, and PCA-PSO-BP models, respectively. The average RMSE is 122.32 kW, which is also significantly lower than the other three models. The results indicate that the proposed model has a superior performance in day-ahead short-term PV output forecasting.

TABLE IV
COMPARISON OF ACCURACY

Weather	GWO-GRNN		PCA-LSTM		PCA-PSO-BP		Proposed model	
	E_M (%)	E_R (kW)	E_M (%)	E_R (kW)	E_M (%)	E_R (kW)	E_M (%)	E_R (kW)
Sunny	2.37	112.97	2.38	98.12	1.57	84.31	0.61	29.58
Cloudy	4.61	218.78	4.77	195.87	4.28	234.98	3.44	172.79
Overcast	7.31	341.57	7.62	468.07	14.12	621.46	4.09	194.39
Rainy	5.52	220.90	2.88	113.38	3.42	139.78	2.07	92.50
Average	4.96	223.55	4.41	218.86	5.85	270.13	2.55	122.32

To further analyze the model performance, due to the large dispersion of the daily error of the PV output forecasting, the forecasting error of different weather types for 28 days is shown in Table V.

TABLE V
FORECASTING ERROR FOR DIFFERENT WEATHER TYPES IN 28 DAYS

Weather	Time (day)	GWO-GRNN		PCA-LSTM		PCA-PSO-BP		Proposed model	
		E_M (%)	E_R (kW)	E_M (%)	E_R (kW)	E_M (%)	E_R (kW)	E_M (%)	E_R (kW)
Sunny	6	3.32	161.95	3.66	168.87	3.23	172.61	2.66	143.99
Cloudy	12	5.51	318.04	3.97	173.75	2.72	148.43	2.64	140.77
Overcast	1	7.31	341.57	7.62	468.07	14.12	621.46	4.09	194.39
Rainy	9	5.11	232.20	3.90	214.24	8.82	448.78	2.89	148.31
All	28	4.98	265.46	4.01	204.03	5.20	307.17	2.78	146.13

According to the error statistics in Table V, for different weather types, the forecasting accuracy of the PCA-GWO-GRNN model is better than that of the other three models. Although the forecasting error of PV output on an overcast day is 4.09%, which is much worse than the forecasting accuracy of other weather types, its nMAE and RMSE are still

higher than those of the other three models. In the whole case, compared with the other three models, the PCA-GWO-GRNN model has lower nMAE and RMSE. Therefore, the proposed model has higher accuracy.

V. CONCLUSION

The model proposed in this paper based on PCA-GWO-GRNN solves the problem of large number of input features and strong randomness in the day-ahead short-term PV output forecasting. This paper makes the following contributions:

1) PCA is adopted to reduce the dimension of meteorological input features and extract variables containing more than 85% of the original information. It can simplify the dimensionality of the input features of the model while ensuring accuracy.

2) GRNN can well fit the complex nonlinear relationship between PV output and input features, and further improve the ability to fit regression by introducing GWO algorithm to optimize its parameter. Thus, the proposed model is an appropriate mathematical tool to achieve high-precision PV output forecasting. Furthermore, because of the good forecasting performance, the proposed model can also provide reference for wind power output, power load and heat load forecasting in the future.

3) The results show that the forecasting model proposed in this paper fully excavates the effective information in the input features with high robustness and forecasting accuracy, which can offer effective solutions to the day-ahead short-term PV output forecasting and provide a basis for the optimal operation of multi-energy systems.

REFERENCES

- [1] B. Elsinga and W. van Sark, "Short-term peer-to-peer solar forecasting in a network of photovoltaic systems," *Applied Energy*, vol. 206, pp. 1464-1483, Nov. 2017.
- [2] M. N. Akhter, S. Mekhilef, H. Mokhlis *et al.*, "Review on forecasting of photovoltaic power generation based on machine learning and metaheuristic techniques," *IET Renewable Power Generation*, vol. 13, no. 7, pp. 1009-1023, Nov. 2019.
- [3] X. Zhang, Y. Li, S. Lu *et al.*, "A solar time based analog ensemble method for regional solar power forecasting," *IEEE Transactions on Sustainable Energy*, vol. 10, no. 1, pp. 268-279, Jan. 2019.
- [4] C. Feng, M. Cui, B. Hodge *et al.*, "Unsupervised clustering-based short-term solar forecasting," *IEEE Transactions on Sustainable Energy*, vol. 10, no. 4, pp. 2174-2185, Oct. 2019.
- [5] C. Wan, J. Zhao, Y. Song *et al.*, "Photovoltaic and solar power forecasting for smart grid energy management," *CSEE Journal of Power and Energy Systems*, vol. 1, no. 4, pp. 38-46, Dec. 2015.
- [6] C. Cui, Y. Zou, L. Wei *et al.*, "Evaluating combination models of solar irradiance on inclined surfaces and forecasting photovoltaic power generation," *IET Smart Grid*, vol. 2, no. 1, pp.123-130, Mar. 2019.
- [7] C. Lai, J. Li, B. Chen *et al.*, "Review of photovoltaic power output prediction technology," *Transactions of China Electrotechnical Society*, vol. 34, no. 6, pp.1201-1217, Mar. 2019.
- [8] E. Oglari, A. Dolara, G. Manzolini *et al.*, "Physical and hybrid methods comparison for the day ahead PV output power forecast," *Renewable Energy*, vol. 113, pp. 11-21, Dec. 2017.
- [9] M. N. Akhter, S. Mekhilef, H. Mokhlis *et al.*, "A review on forecasting of photovoltaic power generation based on machine learning and metaheuristic techniques," *IET Renewable Power Generation*, vol. 13, no. 7, pp. 1009-1023, Feb. 2019.
- [10] B. Chen and J. Li, "Combined probabilistic forecasting method for photovoltaic power using an improved Markov chain," *IET Generation, Transmission & Distribution*, vol. 13, no. 19, pp. 4364-4373,

- Oct. 2019.
- [11] J. Shi, W. Lee, Y. Yang *et al.*, "Forecasting power output of photovoltaic system based on weather classification and support vector machine," *IEEE Transactions on Industry Applications*, vol. 48, no. 3, pp. 1064-1069, May 2012.
- [12] C. Huang and P. Kuo, "Multiple-input deep convolutional neural network model for short-term photovoltaic power forecasting," *IEEE Access*, vol. 7, pp. 74822-74834, Jun. 2019.
- [13] M. Abdel-Nasser and K. Mahmoud, "Accurate photovoltaic power forecasting models using deep LSTM-RNN," *Neural Computing & Applications*, vol. 31, no. 7, pp. 2727-2740, Jul. 2019.
- [14] L. Li, S. Wen, and M. Tseng, "Renewable energy prediction: a novel short-term prediction model of photovoltaic output power," *Journal of Cleaner Production*, vol. 228, pp. 359-375, Aug. 2019.
- [15] L. Liu, M. Zhan, and Y. Bai, "A recursive ensemble model for forecasting the power output of photovoltaic systems," *Solar Energy*, vol. 189, pp. 291-298, Sept. 2019.
- [16] M. Yang and L. Meng, "Short-term photovoltaic power dynamic weighted combination forecasting based on least squares method," *IEEE Transactions on Electrical and Electronic Engineering*, vol. 14, no. 12, pp. 1739-1746, Dec. 2019.
- [17] H. Zhou, Y. Zhang, L. Yang *et al.*, "Short-term photovoltaic power forecasting based on long short term memory neural network and attention mechanism," *IEEE Access*, vol. 7, pp. 78063-78074, Jun. 2019.
- [18] A. Basilevsky, *Statistical Factor Analysis and Related Methods: Theory and Applications*. New York: Wiley, 1994, pp. 351-352.
- [19] G. A. Licciardi, R. Dambreville, J. Chanussot *et al.*, "Spatiotemporal pattern recognition and nonlinear PCA for global horizontal irradiance forecasting," *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 2, pp. 284-288, Jul. 2014.
- [20] F. M. Bianchi, E. de Santis, A. Rizzi *et al.*, "Short-term electric load forecasting using echo state networks and PCA decomposition," *IEEE Access*, vol. 3, pp. 1931-1943, Oct. 2015.
- [21] X. Yao, Z. Wang, and H. Zhang, "A novel photovoltaic power forecasting model based on echo state network," *Neurocomputing*, vol. 325, pp. 182-189, Jan. 2019.
- [22] P. Lin, Z. Peng, Y. Lai *et al.*, "Short-term power prediction for photovoltaic power plants using a hybrid improved Kmeans-GRA-Elman model based on multivariate meteorological factors and historical power datasets," *Energy Conversion and Management*, vol. 177, pp. 704-717, Dec. 2018.
- [23] H. Zhu, W. Lian, L. Lu *et al.*, "An improved forecasting method for photovoltaic power based on adaptive BP neural network with a scrolling time window," *Energies*, vol. 10, no. 10, 1542-1547, Oct. 2017.
- [24] J. Ospina, A. Newaz, and M. O. Faruque, "Forecasting of PV plant output using hybrid wavelet-based LSTM-DNN structure model," *IET Renewable Power Generation*, vol. 13, no. 7, pp. 1087-1095, May 2019.
- [25] K. Nose-Filho, A. D. P. Lotufo, and C. R. Minussi, "Short-term multi-nodal load forecasting using a modified general regression neural network," *IEEE Transactions on Power Delivery*, vol. 26, no. 4, pp. 2862-2869, Oct. 2011.
- [26] L. Yi, N. Dongxiao, and H. Wei-Chiang, "Short term load forecasting based on feature extraction and improved general regression neural network model," *Energy*, vol. 166, pp. 653-663, Jan. 2019.
- [27] R. Hu, S. Wen, Z. Zeng *et al.*, "A short-term power load forecasting model based on the generalized regression neural network with decreasing step fruit fly optimization algorithm," *Neurocomputing*, vol. 221, pp. 24-31, Jan. 2017.
- [28] Z. Ming, X. Song, W. Zhijie *et al.*, "Short-term load forecasting of smart grid systems by combination of general regression neural network and least squares-support vector machine algorithm optimized by harmony search algorithm method," *Applied Mathematics & Information Sciences*, vol. 7, pp. 291-298, Feb. 2013.
- [29] L. Liu, Y. Zhao, D. Chang *et al.*, "Prediction of short-term PV power output and uncertainty analysis," *Applied Energy*, vol. 28, pp. 700-711, Jul. 2018.
- [30] S. Mirjalili, S. M. Mirjalili, and A. Lewis, "Grey wolf optimizer," *Advances in Engineering Software*, vol. 69, no. 3, pp. 46-61, Mar. 2014.
- [31] K. Li, G. Cheng, X. Sun *et al.*, "A nonlinear flux linkage model for bearingless induction motor based on GWO-LSSVM," *IEEE Access*, vol. 7, pp. 36558-36567, May 2019.
- [32] S. Dai, D. Niu, and Y. Li, "Daily peak load forecasting based on complete ensemble empirical mode decomposition with adaptive noise and support vector machine optimized by modified grey wolf optimization algorithm," *Energies*, vol. 11, pp. 1-25, Jan. 2018.
- [33] L. Ge, Y. Xian, J. Yan *et al.*, "A FA-GWO-GRNN method for short-term photovoltaic output prediction," in *Proceedings of 2020 IEEE PES General Meeting*, Montreal, Canada, Aug. 2020, pp. 1-9.
- [34] Y. Li, H. Zhang, X. Liang *et al.*, "Event-triggered-based distributed cooperative energy management for multienergy systems," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 4, pp. 2008-2022, Apr. 2019.
- [35] H. Zhang, Y. Li, D. Gao *et al.*, "Distributed optimal energy management for energy internet," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 6, pp. 3081-3097, Dec. 2017.
- [36] L. Fu, Y. Yang, X. Yao *et al.*, "A regional photovoltaic output prediction method based on hierarchical clustering and the mRMR criterion," *Energies*, vol. 12, no. 20, pp. 3817-3826, Oct. 2019.
- [37] P. Du, G. Zhang, P. Li *et al.*, "The photovoltaic output prediction based on variational mode decomposition and maximum relevance minimum redundancy," *Applied Sciences*, vol. 9, no. 17, pp. 3593-3599, Sept. 2019.

Leijiao Ge received the Ph.D. degree in electrical engineering from Tianjin University, Tianjin, China, in 2016. He is currently a lecturer in the School of Electrical and Information Engineering, Tianjin University. His main research interests include situational awareness of smart distribution network, cloud computing and big data.

Yiming Xian received the B.E. degree in electrical engineering and automation from Lanzhou Jiaotong University, Lanzhou, China, in 2016. Currently, he is a postgraduate student of the State Key Laboratory of Reliability and Intelligence of Electrical Equipment (School of Electrical Engineering, Hebei University of Technology), Tianjin, China. His current research interest is power load and new energy resources output forecasting.

Jun Yan received the B.Eng. degree in information and communication engineering from Zhejiang University, Hangzhou, China, in 2011, and the M.S. and Ph.D. (with Excellence in Doctoral Research) degrees in electrical engineering from The University of Rhode Island, Kingston, USA, in 2013 and 2017, respectively. He is currently an assistant professor in the Concordia Institute for Information Systems Engineering, Concordia University, Montreal, Canada. He was a recipient of the IEEE International Conference on Communications (ICC) Best Paper Award in 2014 and the IEEE International Joint Conference on Neural Networks (IJCNN) Best Student Paper Award in 2016, among others. His research interests include computational intelligence and cyber-physical security, with applications in smart grids, smart cities, and other smart critical infrastructures.

Bo Wang received the Ph.D. degree in computer science from Wuhan University, Wuhan, China, in 2006. He did a postdoctoral research with the School of Electrical and Automation, Wuhan University, from 2007 to 2009, where he is currently a professor with the School of Electrical Engineering. His research interests include power system online assessment, big data, and integrated energy systems.

Zhongguan Wang received the B.S. and the Ph.D. degree from the Department of Electrical Engineering, Tsinghua University, Beijing, China, in 2014 and 2019, respectively. He is currently an associate professor with the School of Electrical Engineering and Automation, Tianjin University, Tianjin, China. His current research interests include distributed control and optimization, microgrid analysis and operation, and active distribution system hierarchical control and energy management.