A short-term photovoltaic power forecasting model based on a radial basis function neural network and similar days

Zhenlei Xu, Zhichong Chen¹, Haifang Zhou, Lijun Wu, Peijie Lin and Shuying Cheng

College of Physics and Information Engineering, Fuzhou University, Fuzhou, China.

¹Email:zhichong.chen@fzu.edu.cn

Abstract. Intermittence and fluctuation natures of photovoltaic (PV) solar energy pose great challenge on the grid stability and power scheduling. PV power forecasting is an effective measure to alleviate the issue. This study presents an improved model for forecasting one-day-ahead hourly PV power generation using Numerical Weather Prediction (NWP) and historical data, which is based on Radial Basis Function (RBF) neural network and similar day method. Firstly, historical similar days of the same weather type are selected according to the correlation of meteorological data. Secondly, the RBF neural network based forecasting model is trained using the historical data of similar days. Finally, the model is used to forecast the power generation using the NWP data of the forecast day. Experimental results show that the proposed method is accurate and reliable.

1. Introduction
The fast increasing of grid-connected PV power systems poses great challenges on the stability of the power grids, because the solar energy is intermittence and fluctuates frequently. To address this issue, various PV power forecasting approaches have been proposed [1]. Generally, the PV power forecasting models can be classified into two classes: indirect prediction models and direct prediction models [2-4]. The indirect prediction models firstly forecast the solar irradiance and then use PV system model to predict the power output. Therefore, the accuracy of the models strongly depends on the irradiance forecasting. Instead, the direct forecasting models are based on historical data to directly predict PV power generation. Due to the merit that irradiance forecasting is not needed, the direct forecasting models attract a lot of research interests in recent years and many machine learning techniques are applied. Chen et al. combined the artificial neural network (ANN) with the weather classification model to establish a PV power plant output power forecasting model within 24-hour [5]. Similarly, Ref. [6] used the Support Vector Regression (SVR) method and weather classification to predict the power output of photovoltaic power plants. Zhang et al. used the K-Nearest Neighbour (KNN) algorithm and similar day method for PV power prediction [7]. Khan et al. [8] used a back propagation neural network (BPNN) based air quality index model to predict PV power in haze weather, and the results showed that the model, which adds the additional aerosol index and Air quality index variables, has a better performance than KNN model. The ANN and the SVR models have good forecasting effects. The SVR models take a long time to enhance the SVR characteristic parameters and require a large amount of data for fitting and regression, which are appropriate for largescale training samples. Instead, the ANN algorithms have a superior fitting performance and
excellent learning capacity. Therefore, the ANN modelling techniques are commonly used due to their high predictive capacity.

In this study, we propose an improved PV power forecasting method for PV power plants using RBF neural network and similar day method, which achieves a high accuracy. Firstly, we use the coefficient of determination ($R^2$) instead of Euclidean distance (ED) to find several similar days from the historical datasets, which provides the model with more relevant training dataset. Unlike the Historical Similar Mining model that finds similar days in one month before the forecast day [9], we extend the searching range of similar days to one year before, which could provide more training data samples. Then, we classify the weather conditions into four types (sunny, rainy, cloudy, and partially sunny), and the RBF neural network (RBFNN) are used to build the forecasting models for each weather type. Finally, based on the historical data of a real PV power plant, several experiments are carried out to compare different forecasting models and validate the advantage of the proposed $R^2$-RBFNN based forecasting model in terms of training time, accuracy and robustness.

The structure of this paper is organized as follows: section 2 describes the historical and NWP data used for the forecasting, and illustrates the correlation between the meteorological data and the power generation. Section 3 details the proposed forecasting model using similar days and RBFNN. Experiment and result analysis are conducted in Section 4.

2. Historical and NWP data of a real PV plant
In this study, we use electrical and meteorological historical data and NWP data for building the short-term PV power forecasting model [10]. Since the proposed model aims at hourly power forecasting, statistical variables are used instead of the raw data, including the maximum, minimum and average values in each hour. The historical data is selected from a PV power station of the Desert Knowledge Australia Solar Centre (DKASC) for the whole year of 2017, from 0 to 24 o’clock for every five minutes [11]. Since the solar energy is only available at daytime, we selected the data in the time period from 7:00 am to 5:00 pm according to historical data of the selected power station.

In order to select effective meteorological factors for building the model, we analyze the correlation between all the available meteorological factors and the power generation using the historical data by the $R^2$ method. The result indicated that the air temperature, weather humidity, global horizontal radiation, and diffuse horizontal radiation are most relevant factors to power generation. In addition, the forecasting values of these factors are easy to retrieve from the NWP model. The correlation between the selected four factors and power generation is illustrated in Figure 1.

![Figure 1. Meteorological factors and power relationship.](image-url)
3. Improved forecasting model

The proposed PV power forecasting model is based on RBFNN and similar days, which is used to forecast one day ahead power generation. Firstly, the NWP meteorological data and weather type for the subsequent forecasting day are extracted from the NWP data source. Secondly, based on hourly statistic of four meteorological factor, the coefficient of determination $R^2$ is used as the similar function find some similar days for the same type of weather in historical data of the PV plant. Then, we train the RBFNN based forecasting model based the meteorological data and the output power data of the similar days. Lastly, the forecasting power is obtained using the RBFNN model and NWP meteorological data, and the performance is evaluated. The brief flowchart of the forecasting model is shown in Figure 2.

![Figure 2. Forecasting model flowchart.](image1)

![Figure 3. Similar day model flow.](image2)

3.1. Improved similar day model

A similar day method is to find some historical days similar to the forecast day in few previous days or a month by comparing multiple meteorological factors [12]. However, the searching range is not enough for a training dataset of neural networks. In this paper, the search range is the days of the same weather type as the forecasting day in the previous year. When forecasting the output power of PV
power plants, the model factors are mainly extracted from meteorological factors. Air temperature, weather humidity, global horizontal radiation and diffuse horizontal radiation are selected as the main influencing factors of PV power plant output power as described in Eq. (1), where $T_{max}, T_{min}, T_{ave}$ is the maximum, minimum, and average of the air temperature in the $k$-th day; $H_{max}, H_{min}, H_{ave}$ is the maximum, minimum, and average of the weather humidity in the $k$-th day; $G_{max}, G_{min}, G_{ave}$ is the maximum, minimum, and average values of the global horizontal radiation $k$-th day; $D_{max}, D_{min}, D_{ave}$ is the maximum, minimum, and average values of the diffuse horizontal radiation on the $k$-th day.

$$S_k = [T_{max}, T_{min}, T_{ave}, H_{max}, H_{min}, H_{ave}, G_{max}, G_{min}, G_{ave}, D_{max}, D_{min}, D_{ave}]$$ (1)

The Euclidean distance (ED) is the most commonly used criterion for determining similar days. However, in this study, we choose the coefficient of determination ($R^2$) to describe the overall similarity of various meteorological factors between the forecast day and a historical day, as given in Eq. (2), because the $R^2$ is commonly used to evaluate the similarity between the two continuous curves, and the meteorological factors are kind of variable varying continuously. Figure 3 shows the flow of similar day model.

$$\bar{y} = \frac{1}{m} \sum_{i=1}^{m} y_i$$
$$R^2 = 1 - \frac{\sum_{i=1}^{m} (y_i - f(x))^2}{\sum_{i=1}^{m} (y_i - \bar{y})^2}$$ (2)

3.2. RBFNN based forecasting model

The RBFNN is composed of three layers [13]. The first layer consists of input nodes, and the number of input nodes is equal to the dimension $m$ of the input vector $x$. The second layer is the hidden layer which consists of the nodes directly connected to the input nodes. The output of the $i$-th hidden layer nodes is $\phi_i = \Phi_i(x)$, where $X_i=[x_1, x_2, \ldots, x_m]$ is the center of the basis function. The final output of the RBFNN is the linearly weighted sum of the outputs of the hidden layer nodes. The structure of the RBFNN for the proposed forecasting method is shown in figure 4, in which the selected four meteorological factors serves as the input variables and the output is the forecast power.

The four meteorological factors are air temperature, weather humidity, global horizontal radiation and diffuse horizontal radiation. The forecast power is denoted by the function in Eq. (3) where the $x_i : x_4$ represents the above four meteorological factors, and $y$ is the forecasting power of the PV.

$$y = f(x_1, x_2, x_3, x_4)$$ (3)

![Figure 4. Proposed RBFNN structure.](image-url)
The Gaussian activation function of the RBFNN is given in Eq. (4). Thus, the output of the RBFNN can be obtained by the Eq. (5), where $x_p$ is the $p$-th input sample and $h$ is the number of nodes in the hidden layer.

$$R(x_p - c_i) = \exp\left(-\frac{1}{2\sigma^2} \| x_p - c_i \|^2 \right)$$

$$y_j = \sum_{i=1}^{h} w_i \exp\left(-\frac{1}{2\sigma^2} \| x_p - c_i \|^2 \right) j = 1, 2, \cdots, n$$

The procedure of the learning algorithm for the RBFNN is detailed in Figure 5, in which there are three types of coefficients to be solved, including the centre of the basis function, the variance, and the weight of the hidden layer to the output layer. In addition, there are two parameters for the learning algorithm of RBFNN, including the spread and number of neurons. In order to optimize the forecasting model, different sets of values are used for different weather type, as given in Table 1.

**Step 1:** Calculate the center of the basis function based on K-means clustering method:
1) Randomly select $h$ training samples as cluster center $c_i$.
2) Group the input training sample set by nearest neighbor rule.
3) Re-adjust the cluster center. If the new cluster centers no longer changes, the resulting $c_i$ is the final base function center of the RBF neural network, otherwise return 2 for the next round of solution.

**Step 2:** Solve variance $\sigma_i$:

$$\sigma_i = \frac{c_{\max}}{\sqrt{2h}} \quad i = 1, 2, \cdots, h$$

where $c_{\max}$ is the maximum distance between the selected centers.

**Step 3:** Calculate the weight between the hidden layer and the output layer by least squares method:

$$w = \exp\left(\frac{h}{c_{\max}} \| x_p - c_i \|^2 \right) \quad p = 1, 2, \cdots, P; i = 1, 2, \cdots, h$$

**Figure 5.** RBFNN learning algorithm.

| Parameter       | Cloudy | Partially Sunny | Rainy | Sunny |
|-----------------|--------|-----------------|-------|-------|
| Spread          | 78.6   | 98              | 90    | 79    |
| Number of neurons | 80     | 100             | 98    | 68    |

4. Experiments and results

In this section, based on the DKASC datasets and the software of MATLAB R2017a, we select a typical day from for each type of weather as the day to be forecast for evaluating the proposed power forecasting method, and 14 similar days are selected according to the ranking of the $R^2$ similarity. And, three performance metrics are used for the evaluation by comparing the forecast power and actually measured power, including the MSE, the $R^2$ and residual. The MSE is calculated by Eq. (6) [14]:

$$MSE = \frac{1}{m} \sum_{i=1}^{m} (y_i - \hat{y}_i)^2$$

In order to verify the advantage of the proposed method, we comp are the ED based similar day method ED-RBF and the $R^2$ based similar day method ($R^2$-RBF, $R^2$-Elman, $R^2$-BP). The forecast and actual power values for the four weather types are plotted together in Figure 6. It is obviously demonstrated that all of the methods are quite accurate, since the two power curves are very close to
each other. But, the power forecasting results of R²-RBF method are slightly better than the ED-RBF method, R²-Elman and R²-BP results for all weather types.

**Figure 6.** Comparison of the hourly forecasting and measured power.

**Table 2.** Predictive model parameters.

| Method  | Index | Cloudy     | Partially sunny | Rainy     | Sunny    |
|---------|-------|------------|-----------------|-----------|----------|
| ED-RBF  | MSE   | 0.007481   | 0.021422        | 0.010900  | 0.013000 |
| R²-RBF  | MSE   | 0.006796   | 0.007542        | 0.006900  | 0.009698 |
| R²-Elman| MSE   | 0.011285   | 0.021134        | 0.003773  | 0.155990 |
| R²-BP   | MSE   | 0.012749   | 0.027721        | 0.001891  | 0.011540 |
| ED-RBF  | R²    | 0.99198    | 0.99366         | 0.99131   | 0.99725  |
| R²-RBF  | R²    | 0.99310    | 0.99816         | 0.99655   | 0.99885  |
| R²-Elman| R²    | 0.99554    | 0.99690         | 0.99835   | 0.99850  |
| R²-BP   | R²    | 0.99027    | 0.99373         | 0.99875   | 0.99933  |

Furthermore, the MSE and R² values are given in Table 2. As shown in Table 2, the MSE values of the proposed R²-RBF forecasting model are less than 1% for all the four weather types, which are all less than that of ED-RBF model. And, the R² values of the proposed R²-RBF are all higher than 0.993, which are higher than that of ED-RBF as well. In comparison with the R²-Elman and R²-BP models, the proposed R²-RBF forecasting model achieves very competitive R² value. But, the MSE reflects the accuracy of the prediction. Obviously, the proposed method ensures global fitness while providing
higher prediction accuracy. In addition, the Figure 7 shows the hourly residuals of the predicted power for the four weather types, which validate the superiority of the proposed method as well.

Figure 7. Residual curve.

5. Conclusions
In this study, based on similar day method and RBFNN, an improved PV power generation forecasting model is proposed for one-day-ahead hourly power prediction, which uses the historical power and meteorological data, as well as the NWP weather type and meteorological forecasting data. To improve the reliability of the forecasting, the hourly statistic data are used instead of the raw data. Firstly, 14 top historical similar days of the same weather type are selected according to the $R^2$ correlation of four meteorological factors including air temperature, weather humidity, global horizontal radiation, and diffuse horizontal radiation. Secondly, the RBFNN regression based forecasting model is trained using the historical power and meteorological data of the selected similar days. Finally, the model is used to forecast power generation with NWP data. Experimental results show that the proposed $R^2$-RBF method is accurate and reliable. In comparison to the ED-RBF, $R^2$-Elman and $R^2$-BP forecasting models, the proposed $R^2$-RBF forecasting method has better robustness and higher prediction.

Acknowledgement
The authors would like to acknowledge the financial supports by the National Natural Science Foundation of China (Grant Nos. 61601127, 51508105 and 61574038), the Fujian Provincial Department of Science and Technology of China (Grant Nos. 2016H6012 and 2018J01774), and the Fujian Provincial Economic and Information Technology Commission of China (Grant Nos. 830020 and 83016006).

References
[1] Antonanzas J, Osorio N, Escobar R, Urraca R, Martinez-de-Pison F J & Antonanzas-Torres F 2016 Review of photovoltaic power forecasting Solar Energy 136 78-111
[2] Ramsami P & Oree V 2015 A hybrid method for forecasting the energy output of photovoltaic systems Energy Conversion and Management 95 406-413
[3] Lin P, Peng Z, Lai Y, Cheng S, Chen Z & Wu L 2018 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 177 704-717

[4] Raza M Q, Nadarajah M & Ekanayake C 2016 On recent advances in PV output power forecast Solar Energy 136 125-144

[5] Chen C, Duan S, Cai T & Liu B 2011 Online 24-h solar power forecasting based on weather type classification using artificial neural network Solar Energy 85(11) 2856-2870

[6] Shi J, Lee W J, Liu Y, Yang Y & Wang P 2012 Forecasting power output of photovoltaic systems based on weather classification and support vector machines IEEE Transactions on Industry Applications 48(3) 1064-1069

[7] Zhang Y, Beaudin M, Taheri R, Zareipour H & Wood D 2015 Day-ahead power output forecasting for small-scale solar photovoltaic electricity generators IEEE Transactions on Smart Grid 6(5) 2253-2262

[8] Khan I, Zhu H, Yao J, Khan D & Iqbal T 2017 Hybrid Power Forecasting Model for Photovoltaic Plants Based on Neural Network with Air Quality Index International Journal of Photoenergy 2017

[9] Monteiro C, Santos T, Fernandez-Jimenez L A, Ramirez-Rosado I J & Terreros-Olarte M S 2013 Short-term power forecasting model for photovoltaic plants based on historical similarity Energies 6(5) 2624-2643

[10] Chen J L, Li G S & Wu S J 2013 Assessing the potential of support vector machine for estimating daily solar radiation using sunshine duration Energy conversion and management 75 311-318

[11] Lima F J, Martins F R, Pereira E B, Lorenz E & Heinemann D 2016 Forecast for surface solar irradiance at the Brazilian Northeastern region using NWP model and artificial neural networks Renewable Energy 87 807-818

[12] Yang H T, Huang C M, Huang Y C & Pai Y S 2014 A weather-based hybrid method for 1-day ahead hourly forecasting of PV power output IEEE Trans. Sustain. Energy 5(3) 917-926

[13] Kisi O, Tombul M & Kermani M Z 2015 Modeling soil temperatures at different depths by using three different neural computing techniques Theoretical and applied climatology 121(1-2) 377-387

[14] Oneto L, Laureri F, Robba M, Delfino F & Anguita D 2018 Data-driven photovoltaic power production nowcasting and forecasting for polygeneration microgrids IEEE Systems Journal 12(3) 2842-2853