Energy Proceedings
Vol 31, 2023
ISSN 2004-2965

Day-ahead Hourly Photovoltaic Power Prediction Based on Multivariate Data Driven Hybrid Physical and Deep Learning Model

Boheng Chen, Zhicong Chen*, Lijun Wu, Peijie Lin, Shuying Cheng

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

*Corresponding Author

ABSTRACT

In order to overcome the negative impact of the discontinuity and fluctuation of photovoltaic (PV) power generation on the power grid, in this study, a multivariate data driven hybrid method for day-ahead hourly PV power curve prediction based on physical model and deep learning model is proposed. The physical model includes Ineichen clear sky model and PV performance model, while the deep learning model is a hybrid model combining two-dimensional grey relational analysis and bi-directional long short-term memory network model (2DGRA-BiLSTM). Firstly, the ideal clear sky global tilted radiation is calculated through the clear sky model, which is used as the input of PV performance model to obtain the ideal PV power under clear sky conditions. Secondly, the improved 2DGRA algorithm is proposed to obtain the best similar day from historical data. Thirdly, under the guidance of ideal clear sky power, the BiLSTM is trained with similarity-physics-informed data to obtain the difference between actual power and ideal clear sky power which is defined as RES-power. Compared with the other methods, results show that the accuracy of the deep learning model combined with physical method is the highest, followed by the deep learning model without physical method, and finally the simple physical model, whether it's in clear sky condition or not.

Keywords: photovoltaic power prediction, clear sky model, PV performance model, grey relational analysis, bi-directional long short-term memory

1. INTRODUCTION

As a renewable energy, solar energy is considered to be one of the most potential new energy in the future because of its cleanability, sustainability, safety and ubiquity. However, PV output power is highly dependent on meteorological conditions, such as solar irradiance, temperature, humidity and wind speed. The uncertainty of weather change has a serious impact on the stability of PV output power. With the increasing penetration of PV power generation in the power grid, the inherent discontinuity and volatility of PV power generation pose a great challenge to the security, stability and economy of the power system. Accurate PV power prediction can eliminate the negative impact in the process of grid connection of PV power generation.

At present, the existing research on the prediction methods of PV power can be mainly divided into physical methods, statistical methods, machine learning methods and deep learning methods. The physical methods is based on the solar irradiance transfer equation, PV module operation equation and other physical equations to realize the modeling between PV power and other physical parameters. Common physical methods include clear sky model [1, 2] and PV performance model [3]. The clear sky model is used to obtain the irradiance under clear sky conditions. PV performance model is used to establish the physical relationship between irradiance and PV power. However, the anti-interference and robustness of physical methods are poor, and the prediction accuracy is limited. The statistical prediction methods establishes the mapping relationship between historical data and target prediction data by fitting the future PV power [4]. Traditional statistical methods include time series method [5, 6], fuzzy logic method [7], regression analysis method [8-10], Markov chain method [11, 12], etc. However, the PV power generation process is still a dynamic and aperiodic complex time series, which will weaken the prerequisite for the application of a large number of correct historical data in statistical methods.

In recent years, a large number of machine learning prediction methods based on artificial intelligence algorithms have been proposed. The development of artificial neural network (ANN) technology has greatly improved the prediction accuracy of PV power [13-15]. Thanks to the powerful nonlinear processing ability of neural network, compared with physical model and traditional statistical model, deep learning model based on artificial intelligence algorithm has made great progress in prediction accuracy, simplicity, stability, universality and automaticity, and has become the mainstream technology of PV power prediction. Many scholars have proposed deep learning methods to improve the prediction accuracy. Lin et al. proposed a hybrid improved Kmeans-GRA-Elman model, which is more accurate than the other eight prediction methods [16]. Wang et al. applied long short-term memory network (LSTM) to an in-dependent day ahead PV power prediction model for the first time [17].

However, most of the existing studies are often limited to using a single method. For example, in order to improve the accuracy, many hybrid deep learning models pursue stacking algorithms too much, but rarely consider that their interpretability and prediction accuracy can be further improved by combining physical formulas. PV output power is a time series that conforms to certain physical laws. If sufficient physical factors are taken into account in its prediction, the prediction results will be more in line with the reality.

In this study, we use meteorological and historical power data to calculate the clear sky irradiance and ideal clear sky power through the theoretical formulations of physical method, and combines it with the proposed hybrid deep learning model to realize the day ahead hourly multi-output PV power prediction. The main flow chart of this study is shown in Figure 1, which can be divided into two parts.

Firstly, in the part of physical method, the global tilted radiation (RGT) under ideal clear sky condition is obtained by establishing the Ineichen clear sky model, and then the relationship between irradiance and power is established by PV performance model to calculate the ideal clear sky power (P_{AC}), and the cell temperature (T_c) is calculated as well.

Secondly, in the deep learning part we propose the 2DGRA-BiLSTM model. The two-dimensional grey relational analysis (2DGRA) algorithm is used to obtain similar day, and then the similarity-physics-informed data is used to train bi-directional long short-term memory network (BiLSTM) to predict the RES-power between actual power and P_{AC}. Finally, the final predicted power is calculated.

Experiments show that the proposed hybrid physical and deep learning model can effectively predict the hourly power curve next day, and the prediction accuracy is higher than other four models, especially for non clear air conditions.

2. METHODS

2.1 Clear sky model

Clear sky model refers to the irradiance model under the condition of cloudless atmosphere. Solar irradiance is mainly influenced by the presence of clouds, whose presence difficulties irradiance predictions. However, it is possible to approximate the irradiance under clear sky

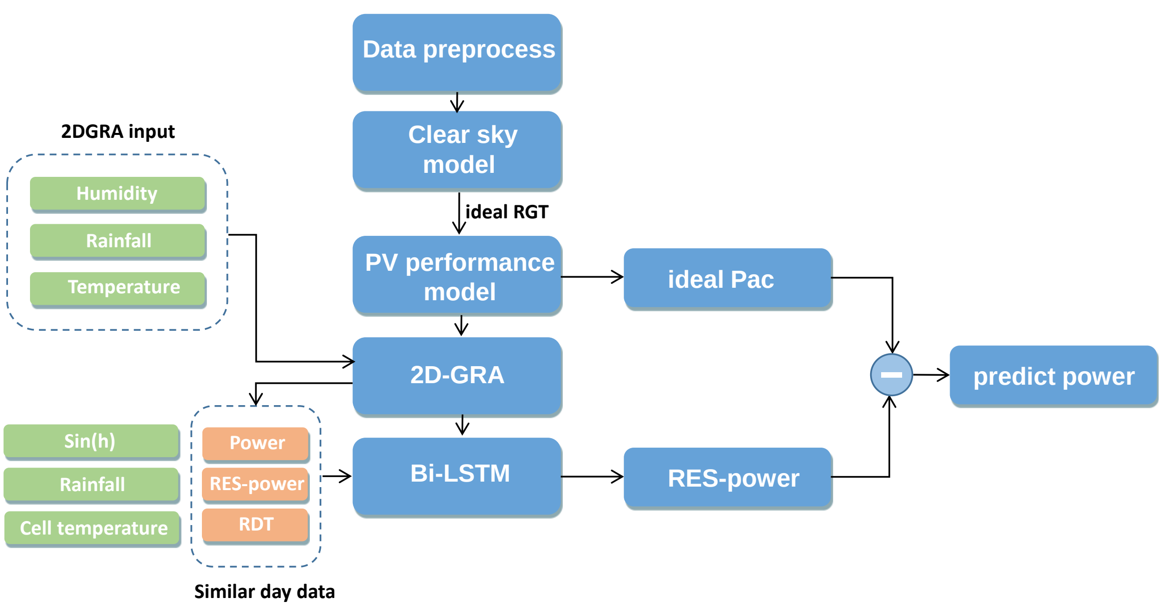

Fig. 1. General flow chart of the proposed hybrid model

conditions, that is, in the absence of clouds [18]. The clear sky model used in this study adopts Ineichen model [2], as given in Eq. (1):

$$\mathrm{GHI} = a_1 \cdot I_0 \cdot \sin(h) \cdot \exp\left(-a_2 \cdot AM \cdot \left(f_{h1} + f_{h2} \cdot (T_L - 1)\right)\right) \tag{1}$$

Where:

$$a_1 = 5.09 \cdot 10^{-5} \cdot altitude + 0.868$$

$$a_2 = 3.92 \cdot 10^{-5} \cdot altitude + 0.0387$$

$$f_{h1} = \exp(-altitude/8000)$$

$$f_{h2} = \exp(-altitude/1250)$$

GHI represents the global clear sky radiation reaching the ground on a horizontal surface. h is the solar altitude angle. T_L is the Linke turbidity coefficient, this experiment assumes T_L = 2 in clear sky condition. I_0 is the extra-terrestrial radiation calculated with a yearly varying term in Eq. (2) [19]. AM is air mass determined by the following Eq. (3).

$$I_0 = 1367.7\left(1 + 0.033 \cdot cos\left(\frac{2\pi}{365} \cdot N_{date}\right)\right) \tag{2}$$

$$AM = \frac{\exp(-z/z_h)}{sin(h) + 0.50572 \cdot (h + 6.07995)^{-1.6364}} \tag{3}$$

Through the Ineichen model, we can calculate the GHI under clear sky conditions. In order to further establish the relationship between irradiance and PV power, we also need to convert it into RGT. Therefore, we use geometric method to deduce the following conversion Eq. (4) to approximate RGT.

$$\mathrm{RGT} = GHI \cdot \frac{sin(h_{max} + \beta)}{sin(h_{max})} \tag{4}$$

Where h_{max} is the noon solar altitude angle of the day, β is the angle between the PV array and the ground.

2.2 PV performance model

PV performance models convert RGT or GHI into PV power [20]. In order to obtain the ideal PV power by clear sky model, this study adopts the formulas in [3]. The model is mainly divided into two parts. One part establishes the relationship between RGT and T_c (Eq. (5)).

$$\mathrm{T_c} = \mathrm{T_a} + 0.018 \cdot \mathrm{RGT} \tag{5}$$

The other part establishes the relationship between RGT and PV AC power P_{AC} (Eqs. (6)-(7)).

$$\mathrm{P_{AC}} = \eta_{\mathrm{eff}} \cdot \mathrm{P_{DC}} \tag{6}$$

$$\mathrm{P_{DC}} = \mathrm{n_m} \cdot \eta_{\mathrm{c}} \cdot \mathrm{PF} \cdot \mathrm{RGT} \cdot \mathrm{A}_m \tag{7}$$

Where η_{eff} is the Inverter efficiency, T_a stands for ambient temperature, T_c is the cell temperature, n_m is the number of modules in the PV array. The pack factor (PF) is the ratio of the total area of PV-cells (A_c) over the area of the PV-module (A_m) and is given as the following Eq. (8).

$$\mathrm{PF} = \frac{\mathrm{A_c}}{\mathrm{A_m}} \tag{8}$$

The η_c is the cell efficiency and satisfies the following relationship in Eq. (9):

$$\eta_{\mathrm{c}} = \eta_{\mathrm{n},c} \cdot \mathrm{a} \cdot \left[\mathrm{b} \cdot \frac{\mathrm{RGT}}{\mathrm{G_0}} + \left(\frac{\mathrm{RGT}}{\mathrm{G_0}}\right)^{\mathrm{c}}\right] \cdot \left[\mathrm{d} + \mathrm{e} \cdot \frac{\mathrm{T_c}}{\mathrm{T_r}} + \mathrm{f} \cdot \frac{\mathrm{AM}}{\mathrm{AM_0}} + \left(\frac{\mathrm{AM}}{\mathrm{AM_0}}\right)^{\mathrm{g}}\right] \tag{9}$$

where, G_0 = 1000 Wm^2, T_r = 25 ^{0}C, AM_0= 1.5, The parameters a, b, c, d, e, f and g are regression coefficients with values of 1.249, 0.241, 0.193, 0.244, 0.179, 0.037 and 0.073, respectively [3].The nominal efficiency ($\eta_{n,c}$) of PV-cells is given as Eq. (10).

$$\eta_{\mathrm{n},c} = \frac{\mathrm{P_{MPP(STC)}}}{\mathrm{A_c} \cdot G_{STC}} \tag{10}$$

2.3 Two-dimensional grey relational analysis algorithm

Grey relational analysis (GRA) algorithm obtains the sequence with the highest correlation degree to the target sequence in the comparison sequence family by calculating the geometric similarity between the comparison sequence and the target sequence. In PV power prediction, GRA algorithm is often used to search the historical similar days to improve the prediction accuracy [16]. However, the traditional GRA algorithm can only be used to process one-dimensional sequences. In contrast, the two-dimensional matrix using multiple meteorological features can more accurately obtain the historical similar days of the days to be predicted. Therefore, this study expands the calculation of geometric similarity from one-dimensional to two-dimensional, and introduces 2DGRA algorithm to calculate the geometric similarity between comparison matrix and target matrix (one matrix contains multiple meteorological feature sequences), then obtaining the historical best similar day. The calculation process of 2DGRA correlation coefficient is as the following Eq. (11).

$$\xi_i(j,k) = \frac{\min_i \min_j \min_k |y_0(j,k) - y_i(j,k)| + \rho \cdot \max_i \max_j \max_k |y_0(j,k) - y_i(j,k)|}{|y_0(j,k) - y_i(j,k)| + \max_i \max_j \max_k |y_0(j,k) - y_i(j,k)|} \tag{11}$$

Where ξ_i(j, k) is the correlation coefficient between target matrix y_0(j, k) and comparison matrix y_i(j, k). ρ represents the resolution coefficient, here ρ is 0.5. After calculating the average value of correlation coefficient (i.e. the average value of points in the matrix) as the quantitative representation of comparison matrix and target matrix, the definition of correlation degree is given as the following Eq. (12).

$$r_i = \frac{1}{jk}\sum_j^m \sum_k^n \xi_i(j,k) \tag{12}$$

Where, i is the number of matrix clusters, that is, the number of days of data. j is the number of matrix rows, that is, the resolution of meteorological features in a day. k is the matrix columns, that is, the number of meteorological features.

2.4 Bi-directional long short-term memory network

In the deep learning model, LSTM performs well in time series prediction because of their better dealing with the correlation of time series data, big data processing ability and no gradient disappearance. While BiLSTM with bi-directional learning characteristics makes it have higher accuracy and faster learning speed in sequence prediction. However, PV power prediction belongs to the category of one-way time series. We cannot know the backward information with one-way prediction, but when multi-point prediction is carried out, that is the day ahead hourly power curve prediction pursued in this study, we can realize the front and rear connection between multiple prediction points by using the two-way transmission of BiLSTM, making the prediction sequence more continuous and higher hourly prediction accuracy.

Therefore, the P_{AC} obtained in chapter 2.2 is subtracted from the actual power to obtain RES-power as training label, while the daily rainfall, T_c, sin(h), similar day power, similar day RES-power and similar day radiation diffuse tilted (RDT) were used as training inputs, to train the BiLSTM network.

3. RESULTS

3.1 Experimental data

The training set of this experiment adopts the historical power and meteorological data from Desert Knowledge Australia Solar Center (DKASC) website, while the test set adopts the historical power data from DKASC website and the numerical weather prediction (NWP) data from Wunderground website. After data integration and preprocessing, the historical data are divided into clear sky and non clear sky according to the weather, and 80% of them are set as the training set and 20% as the test set. The processed data provide various input parameters for clear sky model, PV performance model, 2DGRA and BILSTM in turn. The experimental results of each part will be described below.

3.2 PV performance model output

As shown in the Figure 2. The P_{AC} calculated from the PV performance model is basically consistent with the actual power under clear sky conditions. However, there is an obvious gap under the non clear sky conditions, which is caused by the RGT gap. The power gap which is called as the RES-power will be the goal of our prediction in the deep learning section.

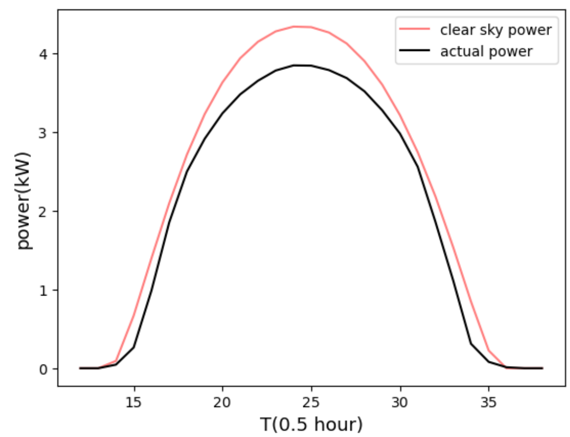

Fig. 2. (a) the clear sky power P_{AC} curve vs. the actual power curve under clear sky conditions

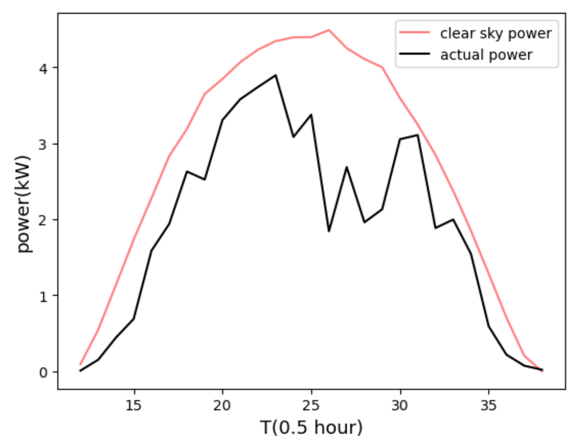

Fig. 2. (b) the clear sky power P_{AC} curve vs. the actual power curve under non clear sky conditions

3.3 BiLSTM output

The trained BiLSTM network can better fit the curve of power gap, which is compared with the actual RES-power in Figure 3. Finally, the predicted power is obtained by subtracting RES-power from P_{AC}. The predicted power comparison with other methods in clear and non-clear skies is shown in Figure 4. It is obvious that deep learning model with physics is more accurate than that without physics, while the BiLSTM is more accurate

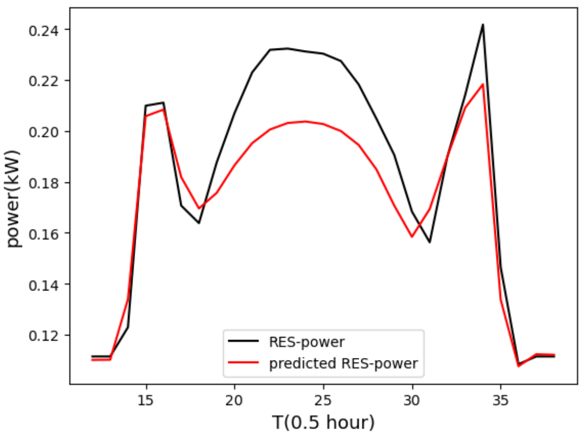

Fig. 3. (a) the predicted RES-power curve vs. the actual RES-power curve under clear sky conditions

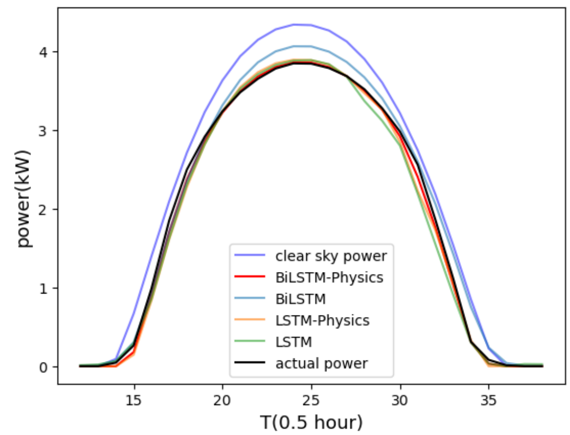

Fig. 4. (a) the predicted power curve vs. the actual power curve under clear sky conditions

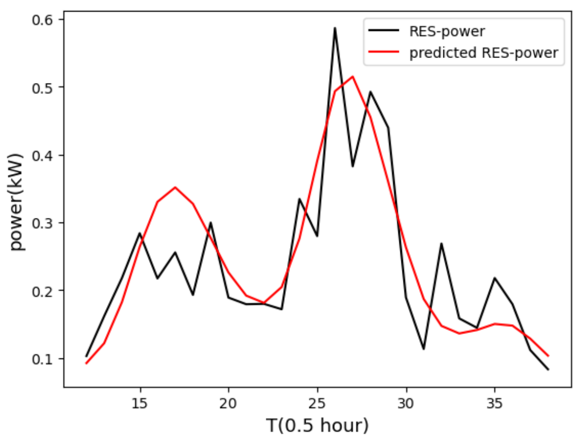

Fig. 3. (b) the predicted RES-power curve vs. the actual RES-power curve under non clear sky conditions

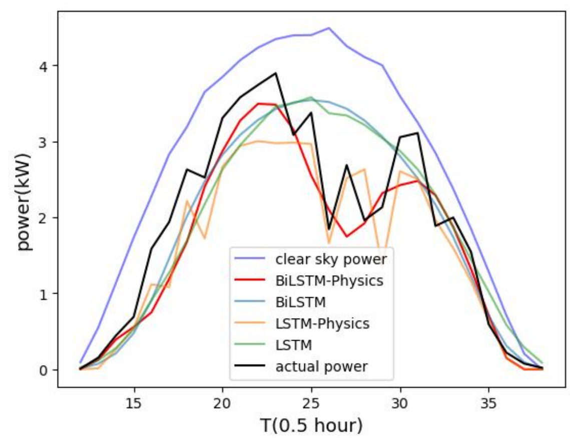

Fig. 4. (b) the predicted power curve vs. the actual power curve under non clear sky conditions

than the LSTM. The prediction accuracy of different forecasting stages are show in Table 1. The proposed method has the highest accuracy with 99.84% R^2 and 0.06 RMSE in clear sky and 84.24% R^2 and 0.45 RMSE in non clear sky. This shows that BiLSTM combined with physical method can have higher performance in hourly power curve prediction, and the performance improvement is more obvious under non-clear sky conditions.

4. CONCLUSION

This study combines physical method and deep learning method to predict the hourly PV power curve of next day. The physical method includes clear sky model and PV performance model, while the deep learning method is composed of 2DGRA and BiLSTM algorithm. The proposed hybrid model is trained and tested with historical power and meteorological data. Comparing with the four other approaches, experiments show that the proposed hybrid method combining deep learning and physical method can effectively predict the power curve of the next day with 99.84% R^2 in clear sky and 84.24% R^2 in non clear sky. The prediction accuracy under the guidance of physical method is better than that using pure deep learning algorithm, especially in non clear weather. For the prediction of power curve, the fitting

Table 1 The prediction accuracy of different forecasting stages.

Conditions	Metrics	Actual power	P_{AC}	LSTM	LSTM-Physics	BiLSTM	BiLSTM-Physics
non clear sky	R^2(%)	100	53.84	77.04	80.25	80.52	84.24
	RMSE	0	1.02	0.59	0.48	0.57	0.45
clear sky	R^2(%)	100	95.72	99.23	99.49	98.90	99.84
	RMSE	0	0.35	0.13	0.11	0.17	0.06

ability of BiLSTM with bidirectional transmission characteristics is higher than that of one-directional LSTM. To sum up, in terms of PV power prediction, the combination of physical method can improve the accuracy of deep learning model and has certain interpretability. The follow-up research can continue to study how to better grasp the balance between physical theorems and deep learning algorithms.

ACKNOWLEDGEMENT

The authors would like to acknowledge the supports by the Fujian Provincial Department of Science and Technology of China (Grant Nos. 2021J01580), the Fuzhou Science and Technology Planning Project (Grant Nos. 2021-P-030 and 2021-P-059), and the Fujian Provincial Economic and Information Technology Commission of China (Grant No. 82318075).

The performance of the proposed forecasting approach was validated by using a large database from the Desert Knowledge Australia Solar Center (DKASC) website and Wunderground website.

REFERENCE

[1] Rigollier C, Bauer O, Wald L. On the clear sky model of the ESRA — European Solar Radiation Atlas — with respect to the heliosat method. Solar Energy 2000;68:33-48.

[2] Ineichen P, Perez R. A new airmass independent formulation for the Linke turbidity coefficient. Solar Energy 2002;73:151-7.

[3] Ayompe LM, Duffy A, McCormack SJ, Conlon M. Validated real-time energy models for small-scale grid-connected PV-systems. Energy 2010;35:4086-91.

[4] Sharadga H, Hajimirza S, Balog RS. Time series forecasting of solar power generation for large-scale photovoltaic plants. Renewable Energy 2020;150:797-807.

[5] Fliess M, Join C, Voyant C. Prediction bands for solar energy: New short-term time series forecasting techniques. Solar Energy 2018;166:519-28.

[6] Shireen T, Shao C, Wang H, Li J, Zhang X, Li M. Iterative multi-task learning for time-series modeling of solar panel PV outputs. Applied Energy 2018;212:654-62.

[7] Halabi LM, Mekhilef S, Hossain M. Performance evaluation of hybrid adaptive neuro-fuzzy inference system models for predicting monthly global solar radiation. Applied Energy 2018;213:247-61.

[8] Massidda L, Marrocu M. Use of Multilinear Adaptive Regression Splines and numerical weather prediction to forecast the power output of a PV plant in Borkum, Germany. Solar Energy 2017;146:141-9.

[9] Sheng H, Xiao J, Cheng Y, Ni Q, Wang S. Short-Term Solar Power Forecasting Based on Weighted Gaussian Process Regression. IEEE Transactions on Industrial Electronics 2018;65:300-8.

[10] Bessa RJ, Trindade A, Silva CSP, Miranda V. Probabilistic solar power forecasting in smart grids using distributed information. International Journal of Electrical Power & Energy Systems 2015;72:16-23.

[11] Sanjari MJ, Gooi HB. Probabilistic Forecast of PV Power Generation Based on Higher Order Markov Chain. IEEE Transactions on Power Systems 2017;32:2942-52.

[12] Ghasvarian Jahromi K, Gharavian D, Mahdiani H. A novel method for day-ahead solar power prediction based on hidden Markov model and cosine similarity. Soft Computing 2020;24:4991-5004.

[13] Ding M, Wang L, Bi R. An ANN-based Approach for Forecasting the Power Output of Photovoltaic System. Procedia Environmental Sciences 2011;11:1308-15.

[14] Cervone G, Clemente-Harding L, Alessandrini S, Delle Monache L. Short-term photovoltaic power forecasting using Artificial Neural Networks and an Analog Ensemble. Renewable Energy 2017;108:274-86.

[15] Mellit A, Sağlam S, Kalogirou SA. Artificial neural network-based model for estimating the produced power of a photovoltaic module. Renewable Energy 2013;60:71-8.

[16] Lin P, Peng Z, Lai Y, Cheng S, Chen Z, Wu L. 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 2018;177:704-17.

[17] Wang F, Xuan Z, Zhen Z, Li K, Wang T, Shi M. A day-ahead PV power forecasting method based on LSTM-RNN model and time correlation modification under partial daily pattern prediction framework. Energy Conversion and Management 2020;212:112766.

[18] Antonanzas J, Osorio N, Escobar R, Urraca R, Martinez-de-Pison FJ, Antonanzas-Torres F. Review of photovoltaic power forecasting. Solar Energy 2016;136:78-111.

[19] Reno MJ, Hansen CW, Stein JS. Global horizontal irradiance clear sky models : implementation and analysis. 2014.

[20] Durisch W, Bitnar B, Mayor J-C, Kiess H, Lam K-h, Close J. Efficiency model for photovoltaic modules and demonstration of its application to energy yield estimation. Solar Energy Materials and Solar Cells 2007;91:79-84.