Applied Energy 218 (2018) 66–77

Contents lists available at ScienceDirect

Applied Energy

journal homepage: www.elsevier.com/locate/apenergy

Estimate and characterize PV power at demand-side hybrid system

Qian Li[a,b], Zhou Wu[a,b,*], Xiaohua Xia[c]

[a] Key Laboratory of Dependable Service Computing in Cyber Physical Society (Chongqing University) of Ministry of Education, China
[b] College of Automation, Chongqing University, Chongqing 400044, China
[c] Department of Electrical Electronic and Computer Engineering, University of Pretoria, Pretoria 0002, South Africa

HIGHLIGHTS

- A new way is developed to directly perform the forecast of PV power at demand side.
- Effects of temperature, humidity, historical value on PV power forecast are explored.
- Estimation results are qualitatively investigated via data mining approaches.
- Experimental studies show that the new method could achieve more accurate prediction.

ARTICLE INFO

Keywords:
Renewable energy
Distributed generation
Solar irradiation
Echo state network
Data mining

ABSTRACT

Power forecasting, in a hybrid photovoltaic (PV) system, is an important issue regarding to the control and optimization of energy systems. In this work, multi-clustered echo state network (MCESN) models are proposed to directly perform the forecast of PV power generation. Furthermore, data characteristics of measured and estimated PV power are qualitatively investigated via data mining approaches. These characteristics include seasonality, stationarity (or non-stationarity) and complexity analysis. Simulation results indicate that the proposed MCESN model is able to precisely forecast PV power one-hour-ahead. The performance on the 24-h-ahead forecast is competitive with the correlation coefficient 99% for sunny days, and 91–98% for cloudy days. Results of data analysis unveil that critical characteristics between the measured and estimated PV power data are analogous. Comparison studies also show that MCESN could achieve more accurate prediction, compared with auto-regressive moving average (ARMA), back propagation (BP) neural networks.

1. Introduction

In recent years, due to globally increasing energy demand, renewable energy sources(e.g., wind and solar energy) have gained great attention, as they are freely available, omnipresent, and environmental friendly. Thanks to easy accessibility, government's support, and technical development, large-scale photovoltaic (PV) systems have been installed around the world. However, the power generation of PV system is a nonlinear and complex process, depending on time-varying factors, such as, temperature, humidity, wind speed and direction, and historical data of PV system. In order to ensure reliable and efficient operation of PV energy systems, it is essential and urgent to forecast PV power precisely [1,2].

There have been a large number of studies on PV power prediction, in which high accuracy and low computational complexity are two main concerns.

A common approach is to transform PV power prediction into solar irradiance prediction, which consists of two steps. The first step is to forecast solar irradiance, and the second step is to calculate the PV power according to solar irradiation and system parameters. Different models of prediction have been developed by traditional techniques and linear methods, e.g., various clear-day models [3], auto-regressive moving average (ARMA) [4] and other econometric technologies. However, as many statistical assumptions and empirical parameters are involved in these models, it is rather difficult to precisely forecast the dynamic behavior of solar irradiance. Some improved models have been proposed based on advanced technologies in [5–7].

Artificial intelligence (AI) and neural network (NN) provide powerful tools of approximating nonlinear systems. Various AI and NN models have been successfully applied to forecasting solar irradiance in literature. A wavelet-coupled support vector machine (W-SVM) model was adopted to forecast global incident solar radiation [8]. A NN model is proposed to achieve a 24-h-ahead solar irradiance prediction for a PV system [9]. Based on recurrent neural networks (RNNs) and wavelet

* Corresponding author at: Key Laboratory of Dependable Service Computing in Cyber Physical Society (Chongqing University) of Ministry of Education, China.
E-mail address: wuzhsky@gmail.com (Z. Wu).

https://doi.org/10.1016/j.apenergy.2018.02.160
Received 4 December 2017; Received in revised form 23 February 2018; Accepted 24 February 2018

neural networks (WNNs), a new diagonal recurrent wavelet neural network (DRWNN) was established to perform the forecast of hourly and daily global solar irradiance [10]. Advanced approximation techniques based on wavelet analysis [11,12], fuzzy technique [13], and empirical analysis [14] can also be employed to enhance NN models. In addition, some other forecasting approaches have also been proposed, such as, peer-to-peer (P2P) solar forecasting [15], machine learning [16,17], probabilistic approach, and so forth. The predicted values of solar irradiance are used to obtain PV power output. On the one hand, canonical PV formula could be utilized to compute the power output of PV system. On the other hand, some commercial PV simulation softwares, such as HOMER and PVFORM, could be used to forecast PV power based on the forecasted solar irradiance and system parameters.

Echo state networks (ESNs) is an improved and simplified form of RNNS [18]. Unlike classical RNNs, ESNs adopt non-trainable sparse connections in the hidden layer (called dynamic reservoir), and only connections in the output layer need to be trained through linear regression. As a result, the high computational complexity is conquered, and ESNs is much faster than traditional RNNs. ESNs also show obvious advantages in dealing with nonlinear time series and dynamic prediction system due to its high prediction accuracy and efficiency. ESNs have been widely applied to various practical fields, including dynamic pattern classification and recognition [19,20], image processing [21], optimal energy management [22], and especially nonlinear time series prediction [23,24]. To our best knowledge, there exist few results in ESN-based prediction of solar irradiance and PV power.

For a PV hybrid system, one practical issue is the uncertainty of PV power. While considering the external environment and different demand-side features, the PV power cannot be directly calculated from a linear form of solar irradiance. Therefore, recent studies have focused on the direct prediction of PV power [25–28]. In this paper, the uncertain PV power at the demand side will be specifically modeled in a direct approach. In the application of PV hybrid system, few results are reported to evaluate inner rules and hidden patterns of the demand-side PV power. Influenced by many factors, such as seasons, geographic locations, weather and surroundings, the PV power profile presents its own data characteristics, which are closely related to the power generation process [29]. In order to unveil the inner dynamics, data features of measured and estimated PV power are quantitatively analyzed. In this paper, some main data characteristics between measured and forecasted PV power will be studied to check statistical similarity.

The contributions are in three folds. First, the ESN models are established to directly perform the one-hour-ahead and 24-h-ahead forecast in the PV hybrid system. The direct effects of measured temperature, humidity, historical 24-h-lag information are also explored in detail. Comparison between ARMA model, BP neural networks and MCESN have been conducted. Secondly, the estimation performance is evaluated with comprehensive criteria, such as normalized root mean square error (NRMSE), mean absolute error (MAE), root mean square error (RMSE), and correlation coefficient (r). Thirdly, the data characteristics are investigated with respect to descriptive statistics, seasonality, non-stationarity and complexity.

The rest of this paper is organized as follows. In Section 2, background is introduced. Section 3 describes the basic theory of ESN in terms of network structure, mathematical model, and training methods. The experimental design and numerical results are shown in Section 4. The data characteristics of measured and estimated PV power are qualitatively analyzed in Section 5. Finally, the conclusion is presented in Section 6.

2. Uncertainty in the PV hybrid system

The electricity consumption have been increasing in past decades, which could result in over exploration of traditional fossil fuel resources. Therefore, the exploration of renewable energy (RE) resources is necessary to control fossil fuel consumption and pollutant emission.

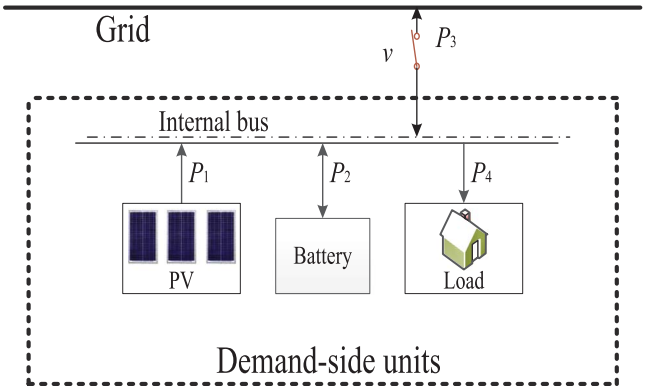


Fig. 1. Schematic of PV hybrid system.

Due to large potential and free availability, wind and solar energy are the popular choices among available RE resources. However, the storage components are required for renewable energy hybrid system due to the intermittent nature. A renewable energy hybrid system is composed of multiple power resources and storage components for stable power supply.

Hybrid renewable energy system (HRES), commonly used for remote power supply, is playing an important role in demand side management with the grid connection, such as, green building and smart community. The PV hybrid system is the most popular application due to easy accessibility, low cost, and high safety. The PV hybrid system consists of PV panel and battery bank that are both connected to the grid, as shown in Fig. 1. As the first priority, the PV power is used to feed the load demand. If the demand is less than the PV power, the surplus PV power will be charged into the battery. If the demand is larger than the PV power, the deficient amount will be then covered by the battery. For saving electricity cost, the battery can be charged by the grid when the electricity has a low price, and be discharged when the electricity has a high price. The grid takes part into the power supply when the load demand cannot be satisfied by the PV and the battery. Note that the PV hybrid system could work in the stand-alone mode and the grid-connected mode, depending on the on/off status of switch v, as shown in Fig. 1.

In the PV hybrid system, a critical problem arisen is the power flow control, which refers to scheduling the power flow between each component for satisfying requirements of cost saving and safety. Let P_1 denote the PV power generation, and P_2 denote the charging/discharging power of battery. Let P_3 denote the grid power flow, and P_4 denote the load demand. With respect to cost, the electricity cost can be expressed as

$$J = \int_{t=0}^{T} \rho(t) P_3(t), \tag{1}$$

where $\rho(t)$ is the real time price of electricity, and J is the electricity cost. With respect to safety, the power balance should be first satisfied as

$$P_1 + P_2 + P_3 = P_4, \tag{2}$$

Power flow control methods have significant effects on electricity cost and operational safety at demand side. In literature, rule-based and optimization-based methods are proposed to reduce the cost and enhance the safety. However, the uncertainty of PV power has presented several challenges on the power flow control. First, the uncertainty could violate the condition of power balance and risk the security of grid and demand-side units. Secondly, the uncertainty could influence actual energy consumption, so that the electricity cost might deviate from the reference one.

In this paper, the prediction of uncertain PV power is specifically studied at the demand side, as solar irradiation at a certain location (a weather station or solar farm) cannot be directly used in the PV power

for other distant customers. A PV panel usually consists of several PV cells to convert solar irradiation into direct current power. With a number of PV panels, the hourly PV power output can be simply formulated as:

$$P_{pv}(t) = \eta_{pv}(t) I_{pv}(t) A_c, \tag{3}$$

where $P_{pv}(t)$ is the hourly power output from the PV panels; $\eta_{pv}(t)$ is the efficiency of solar generation; $I_{pv}(t)$ is the hourly solar irradiation incident on the PV panels (kW h/m^2); A_c is the total size of PV panels.

Many researchers have studied the prediction of solar irradiation, and several kinds of methods have been proposed. The PV power can be linearly derived from the solar irradiation, if customers have the same characteristics. Considering different demand-side characteristics, such as, location, weather, external environment, the efficiency $\eta_{pv}(t)$ is time-varying. For example, when partial shading occurs due to cloud and other objects, the efficiency will decrease. Therefore, the uncertain PV has to be modeled specifically at the demand side, while the solar irradiation can only be regarded as a reference. In this study, distributed generation at a university of South Africa is investigated, and the uncertain PV power is directly modeled with an approach of echo state networks. Note that the proposed approach can also be extended to the prediction of solar irradiation.

3. Echo state neural network

As a kind of neural networks, the ESN has a typical architecture that is composed of an input layer, a hidden layer (referred to as a dynamical reservoir), and an output layer, as shown in Fig. 2(a). In the ESN, the input signal, the output signal, and reservoir states are denoted as $\mathbf{u}(t)$, $\mathbf{y}(t)$, $\mathbf{x}(t)$, respectively. For the task of PV power prediction, the input signal could be current and historical values of PV power, temperature, humidity, and other meteorological indicators. The output signal is the future PV power that needs to be predicted, and the reservoir states are states of neurons in the dynamic reservoir, i.e., the hidden layer. First, the ESN has adopted a dynamic reservoir to transfer the input signal into a high-dimensional state vector, which is expected to include all characteristics. Then, an optimal combination of states is chosen for representing output dynamics that is task-related. In other words, the output signal, extracted from the reservoir, is expected to match the desired target signal. In the rest of the paper, vectors are denoted by boldface lowercase letters, e.g., $\mathbf{x}$, while matrices are denoted by boldface uppercase letters, e.g., $\mathbf{X}$.

In the reservoir, there are a large number of neurons with sparse connections. The weight of each connection is randomly initialized, and remains unchanged in the process of training and testing. Inspired by the nature of biological neural system, such as small-world and modular characteristics, a multi-clustered structure of reservoir was designed in the authors' recent study [30]. Compared with the traditional ESN with a random structure, the multi-clustered ESN (MCESN) achieved more accurate prediction. As illustrated in Fig. 2(b), the MCESN has a similar architecture with the traditional ESN, and their difference is the structure of reservoir. In this paper, the MCESN is adopted for the prediction of PV power. The multi-clustered structure is generated according to Kaisers clustering algorithm [31]. All neurons in the reservoir are divided into two different kinds of neurons, i.e., pioneer neurons and normal neurons. The pioneer neurons, with mutual connections, are the critical neurons that determine the number of clusters. The normal neurons have connections within a cluster according to spatial distance between neurons and associated time windows probability model. Note that the spatial distance is defined as the Euclidean distance in the graphic space, and the time window size determines the value of the probability function and affects the connection probability between neurons. The procedure for reservoir generation is given as the following steps [30]:

Step 1: The reservoir is initialized by a small number (denoted as n) of pioneer neurons, which are bi-directionally connected to each other.

Step 2: A random neuron is added and categorized into the nearest cluster, which is determined by the evaluation of the nearest pioneer neuron. This neuron has a probability to connect each node belonging to the same cluster. The probability is calculated based on the spatial distance and the time window size. Any new neuron that fails to establish a connection will be given up. Step (2) is repeated until the number of existing nodes reaches the defined reservoir size (denoted as N).

Step 3: Each node is connected with itself with a self-connecting probability.

Step 4: The reservoir connection matrix $\mathbf{W}^{res}$ is calculated as follows:

$$\mathbf{W}^{res} = \begin{pmatrix} \mathbf{W}_{1,1} & \cdots & \mathbf{W}_{1,i} \\ \vdots & \ddots & \vdots \\ \mathbf{W}_{i,1} & \cdots & \mathbf{W}_{i,i} \end{pmatrix} \tag{4}$$

where $\mathbf{W}_{i,i}$ is the weight matrix of the ith cluster ($i = 1,\ldots,n$); $\mathbf{W}_{i,j}$ are the weight matrix between the ith and jth cluster. Fig. 3 shows the topology of 200 nodes in the two-dimensional graphic plane [0,1]. The clustered phenomenon is obvious, and it is also clear that the intra-cluster connections are more intensive than the inter-cluster connections.

Assume that the MCESN has K, N, and L neurons in the input, hidden, and output layer, respectively. There exist connection weights from the input units to reservoir (denoted as $\mathbf{W}^{in}$,$\mathbf{W}^{in} \in \mathbb{R}_{N\times K}$), reservoir connection weights collected in an $N \times N$ weight matrix $\mathbf{W}^{res} \in \mathbb{R}_{N\times N}$, and connection weights from the reservoir to the readout neurons given in a $L \times N$ output weight matrix $\mathbf{W}^{out} \in \mathbb{R}_{L\times N}$. For $\mathbf{W}^{in}$ and $\mathbf{W}^{res}$, each component is a random number in the MCESN. Furthermore, the connection weights projected back from the readout neurons to the reservoir units are given in an $N \times L$ feedback weight matrix $\mathbf{W}^{back} \in \mathbb{R}_{N\times L}$. The update of the reservoir states is expressed as follows:

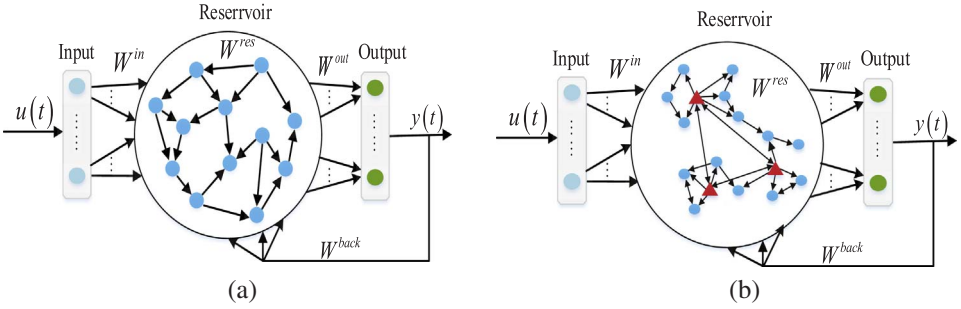


Fig. 2. Network architecture: (a) regular echo state network model with random reservoir structure; (b) multi-clustered echo state network, where the triangles denote the pioneer neurons.

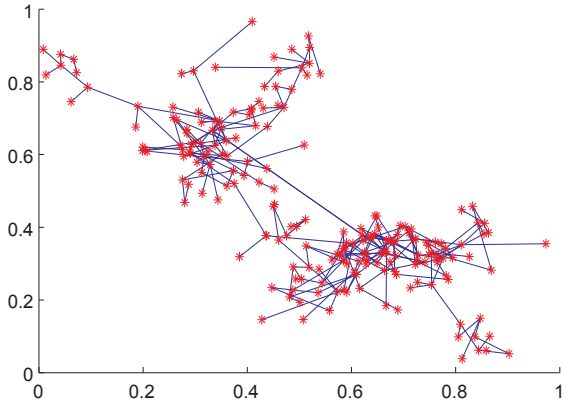


Fig. 3. Two-dimensional projection of multi-clustered network with cluster size $n = 2$.

$$\mathbf{x}(t+1) = f(\mathbf{W}^{in}\mathbf{u}(t+1) + \mathbf{W}^{res}\mathbf{x}(t) + \mathbf{W}^{back}\mathbf{y}(t) + \mathbf{v}(t)), \tag{5}$$

where f is the activation function of each reservoir neuron (usually defined as a sigmoid or Fermi function), and $\mathbf{v}(t)$ is noise signals. The Fermi function is adopted as the hidden neurons function in the paper. The network output is calculated as

$$\mathbf{y}(t+1) = f^{out}(\mathbf{W}^{out}\mathbf{x}(t+1)), \tag{6}$$

where f^{out} is the activation function of the output units. Note that the identity function is adopted in this paper. In the MCESN, the main task is to determine the output weight matrix $\mathbf{W}^{out}$ by training the networks.

At the training stage, the teaching signal, i.e., the future PV power, is given in prior, and the reservoir states can be updated according to Eq. (5). Regression methods could be employed to calculate the output weight matrix. Let l_{tr} represent the length of training datasets, and $\mathbf{X}$ represent the internal state matrix. The corresponding teacher signal vector matrix Λ is denoted as

$$\Lambda = \begin{bmatrix} d_1(1) & d_2(1) & \dots & d_L(1) \\ d_1(2) & d_2(2) & \dots & d_L(2) \\ \vdots & \vdots & \vdots & \vdots \\ d_1(l_{tr}) & d_2(l_{tr}) & \dots & d_L(l_{tr}) \end{bmatrix}_{l_{tr}\times L} \tag{7}$$

and the internal states matrix $\mathbf{X}$ is collected as

$$\mathbf{X} = \begin{bmatrix} x_1(1) & x_2(1) & \cdots & x_N(1) \\ x_1(2) & x_2(2) & \cdots & x_N(2) \\ \vdots & \vdots & \vdots & \\ x_1(l_{tr}) & x_2(l_{tr}) & \cdots & x_N(l_{tr}) \end{bmatrix}_{l_{tr}\times N} \tag{8}$$

where $d(t)$ is the teacher signal, i.e., the future PV power at the training stage.

According to the classical pseudo-inverse method, the output weight matrix $\mathbf{W}^{out}$ is computed as

$$(\mathbf{W}^{out})^T = \mathbf{X}^{+}\Lambda, \tag{9}$$

where $\mathbf{X}^{+}$ is defined as generalized inverse matrix of $\mathbf{X}$.

To overcome the over-fitting phenomenon, a ridge regression training method [32] is applied as

$$(\mathbf{W}^{out})^T = (\mathbf{X}^T\mathbf{X} + \rho\mathbf{I})^{-1}\mathbf{X}^T\Lambda, \tag{10}$$

where $\mathbf{I}$ denotes the identity matrix, ρ is the regularization parameter which should be determined through a large number of experiments for the specific learning tasks.

4. Experimental design and estimation results

In this study, a MCESN model is established to forecast the hourly PV power at the PV hybrid system, installed in University of Pretoria at South Africa. The PV system comprises a large number of equal PV modules with rated power 250 W, providing the cooling, heating, and electrical needs for the campus. The historical data, mainly including temperature, humidity, and PV power, is collected for the year 2014. The meteorological sensors are installed for measuring temperature and humidity, the Danfoss Comlynx Monitor logger [33] is used for recording these PV power, temperature, and humidity. As an example of recorded data, Fig. 4 shows the profiles of hourly PV power $P_{pv}(t)$, temperature (T), and humidity (H) from January 1st 2014 to December 31st 2014.

4.1. ESN setup

To generate the multi-clustered reservoir, the parameter settings are given in Table 1 based on [30]. Weight matrices $\mathbf{W}^{in}$ and $\mathbf{W}^{back}$ are sampled from a uniform distribution over $[-1,1]$, and the spectral radius of $\mathbf{W}^{res}$ is set as 0.8 [34]. The ridge regression training method is adopted to obtain the output weights in the current study. The prediction accuracy is indicated by the normalized root mean square error (NRMSE) [18], which can be expressed as

$$NRMSE = \sqrt{\sum_{t=1}^{l_t}(y(t)-d(t))^2/l_t\sigma^2}, \tag{11}$$

where $y(t)$ is the forecast of PV power; $d(t)$ is the actual PV power; l_t is the number of samples; and σ^2 is the variance of the actual PV power. In this application, 60% of data is used for training, and the remaining data is used for testing.

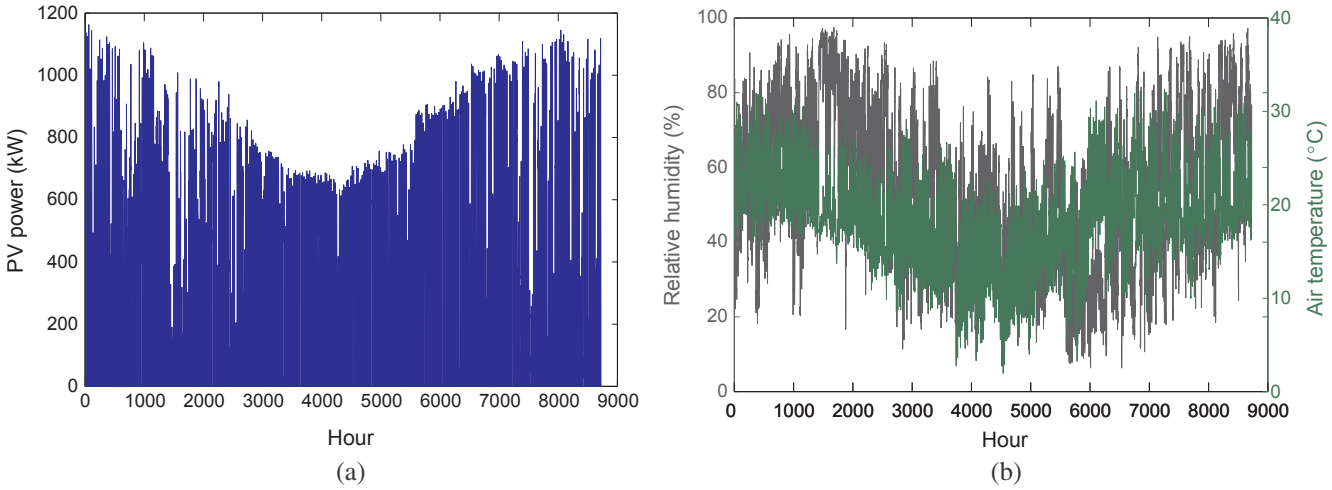


Fig. 4. Recorded data set used in this study: (a) hourly PV power data; (b) corresponding air temperature, humidity.

Table 1
Model parameters for multi-clustered network.

Parameter meaning	Values
Reservoir size	200
Cluster number	2
Time window size	0.3
Self-connecting probability	0.8
Connection probability coefficient1	6
Connection probability coefficient2	10

4.2. One-hour-ahead prediction

In this section, the feasibility and prediction performance of MCESN is evaluated in the one-hour-ahead prediction. As the PV power of each month shows different characteristics, the hourly PV power is modeled for each month in this paper. The PV power at a certain time, denoted as $P_{pv}(t-1)$, is regarded as the input signal, and the PV power at the subsequent hour is regarded as the teacher signal. Take sub-data in summer (January) and winter (July) as two examples, respectively. Results of MCESN are presented in terms of the actual and predicted values at the training and testing stages, as shown in Fig. 5. It can be observed that the prediction output could well match the actual output, and that large fluctuations could be feasibly discovered.

For each month, the prediction accuracy is evaluated in the terms of training and testing NRMSE, respectively. The average NRMSE over 20 independent runs is calculated and shown in Fig. 6. As a result, it can be seen that the prediction accuracy is the lowest in summer.

In addition, in order to directly analyze the factors that may affect PV power, a reasonable input layer of MCESN should be designed. In this study, measured temperature (T) and humidity (H) are used as examples to analyze the direct effect on PV power. The historical values of temperature and humidity are also regarded as the input signals. Besides the model previously derived, 3 other models are evaluated, as illustrated in Table 2. In the first model, the input signal includes the PV power. In the second model, the input signal includes the PV power and temperature. In the third model, the input signal includes the PV power and humidity. In the fourth model, the input signal includes the PV

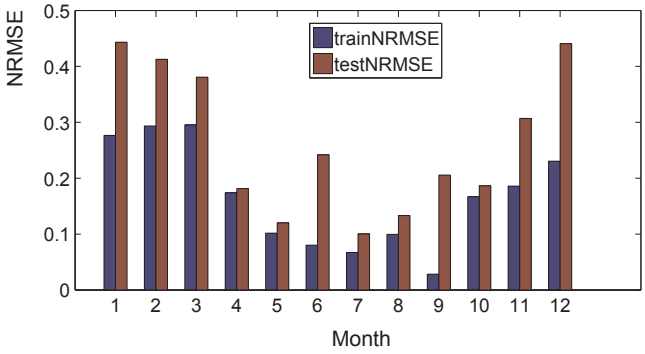
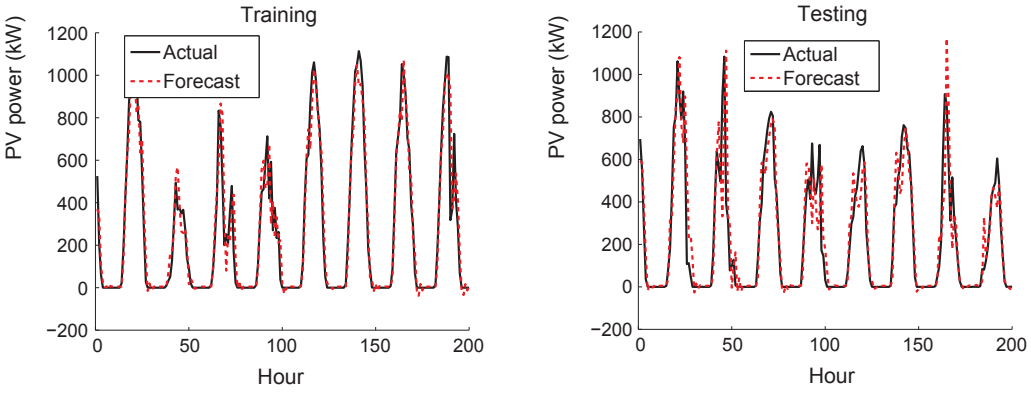


Fig. 6. NRMSEs comparison of training and testing set for each month.

Table 2
Different input and output for MCESN model considering temperature (T) and humidity (H).

Model	MCESN	MCESN + T	MCESN + H	MCESN + T + H
Input	$P_{pv}(t-1)$	$P_{pv}(t-1),T(t-1)$	$P_{pv}(t-1),H(t-1)$	$P_{pv}(t-1),T(t-1),H(t-1)$
Output	$P_{pv}(t)$	$P_{pv}(t)$	$P_{pv}(t)$	$P_{pv}(t)$



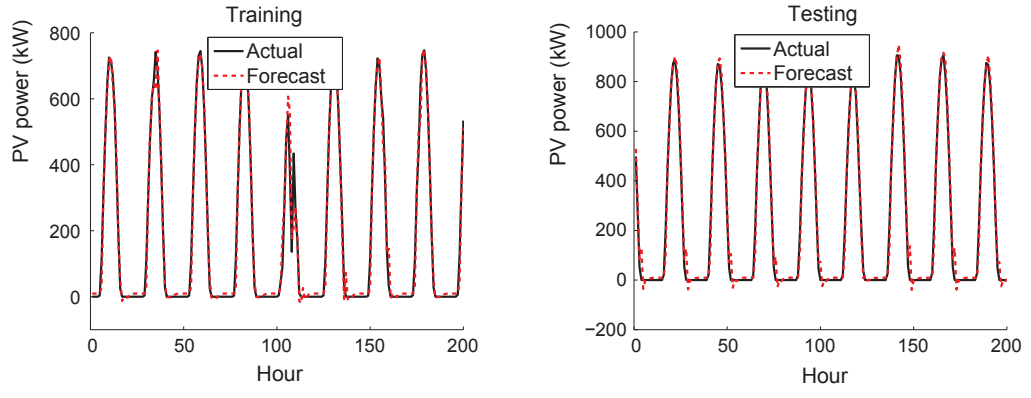


Fig. 5. One-hour-ahead prediction results by MCESN versus actual values for 200 training and testing points: (a) January; (b) July.

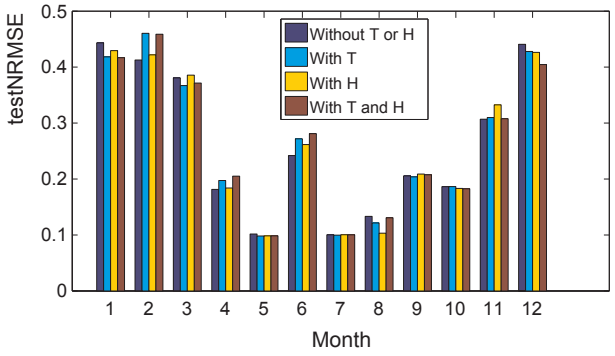


Fig. 7. Testing NRMSEs comparison under four diverse input-output models.

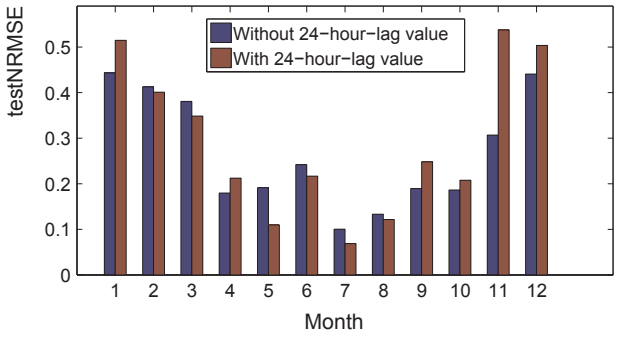


Fig. 9. Testing NRMSEs comparison without/with considering historical 24-h-lag information.

power, temperature, and humidity. In these different situations, the input and output signals are listed in Table 2. For each model, the prediction accuracy is reported with respect to the NRMSE, as shown in Fig. 7. It can be observed that the input effects of setting measured temperature and humidity are minor, as each model has similar accuracy. In the same way, the direct effect of other measured factors (cloud cover, geographic location) could also be analyzed.

Furthermore, in order to evaluate the periodic phenomenon, the hourly data of PV power is represented as a 24∗365 matrix, in which the component at the mth column and the sth row represents the PV power at the mth hour of the sth day (m = 1,...,24, and s = 1,...,365). The 2-D matrix is plotted as a surface mesh shown in Fig. 8. The daily profile of PV power has a periodic pattern, so the effects of 24-h-lag information on the prediction accuracy are further explored. The multiple inputs are selected as $P_{pv}(t-1)$ and $P_{pv}(t-24)$, and the single output is selected as $P_{pv}(t)$. The testing NRMSE without/with the lag information is presented in Fig. 9, where the blue bar represents the results without the lag information and the red bar represents the results with the lag information. It can be obtained that the 24-h-lag information has positive effects on the prediction accuracy in winter and negative effects in summer. The reason behind this phenomenon is that there exist intensive fluctuations that cause greater prediction error, as shown in Fig. 10.

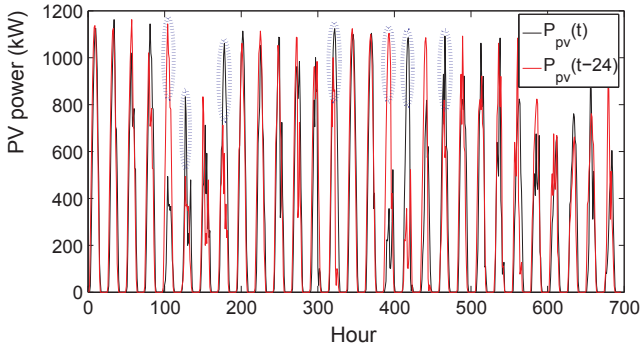


Fig. 10. Comparison of current signal and 24-h-lag information in January.

4.3. 24-h-ahead prediction

The MCESN approaches are utilized to predict the hourly PV power with a good accuracy. However, the hourly PV power is insufficient for certain cases of daily schedule and optimization. Therefore, 24-h-ahead forecast of PV power is further investigated. According to [9], MCESN permits to estimate 24-h-ahead of PV power based on the actual mean value of daily current PV power, daily air temperature, and the day of each month.

Experimental results are presented in Fig. 11 to compare the forecasted profiles and the measured profiles for 4 sunny days (July 19th–22nd). As can be seen, the forecast profiles of PV power can approximate the measured profiles with well accuracy. The scatter plots of prediction results are given in Fig. 12. Most points are close to the diagonal line with the coefficient of determination R^2 = 0.99.

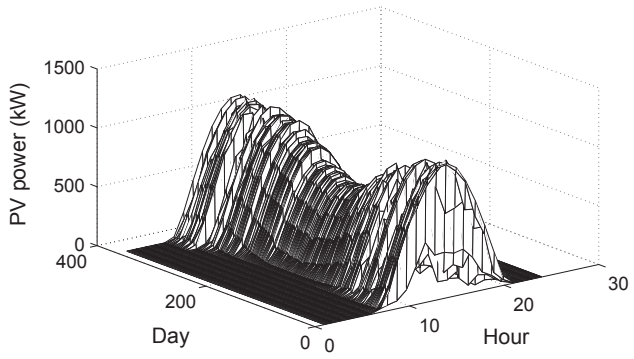


Fig. 8. 2-D surface plot of hourly measured PV power data.

To quantify the prediction performance, several different statistical criteria, i.e., root mean square error (RMSE), correlation coefficient r, and the mean absolute error (MAE), are calculated for different case studies. These statistical results are listed in Table 3. It can be observed that the RMSE for the sunny day is smaller than the cloudy day, and that the correlation coefficient for the sunny day is larger than the cloudy day. The results indicate that the MCESN model delivers less accuracy on the cloudy days. One possible reason is that the weather information, such as, rain and cloud, which is missing in this study, is required for the prediction task of cloudy days.

4.4. Comparisons of MCESN and other typical models

In order to validate the effectiveness of proposed method, two popular models, i.e., auto-regressive moving average (ARMA) and BP neural networks, are selected in the comparison study.

In Fig. 13 and Table 4, the MAE, RMSE, r values between measured and forecasted profiles are compared with respects of ARMA, BP and MCESN. Obviously, MCESN has the highest precision, while ARMA performs the worst. This demonstrates that MCESN has obviously better performance to deal with nonlinear PV power prediction task.

For the PV hybrid system, future PV power is essential information for most problems of design and operation, e.g., sizing and power flow dispatching. For power flow dispatching, day-ahead optimal control is usually applicable to minimize the electricity cost of customers, who already install the hybrid PV system at demand side. Under a certain pricing policy, the PV power prediction affects the optimal dispatching strategy and its associative cost. For example, the PV hybrid system with the proposed MCESN could be used in a time-of-use (TOU) program, which is a typical demand response program to alleviate peak burden. In TOU, the electricity prices are fixed in advance for the

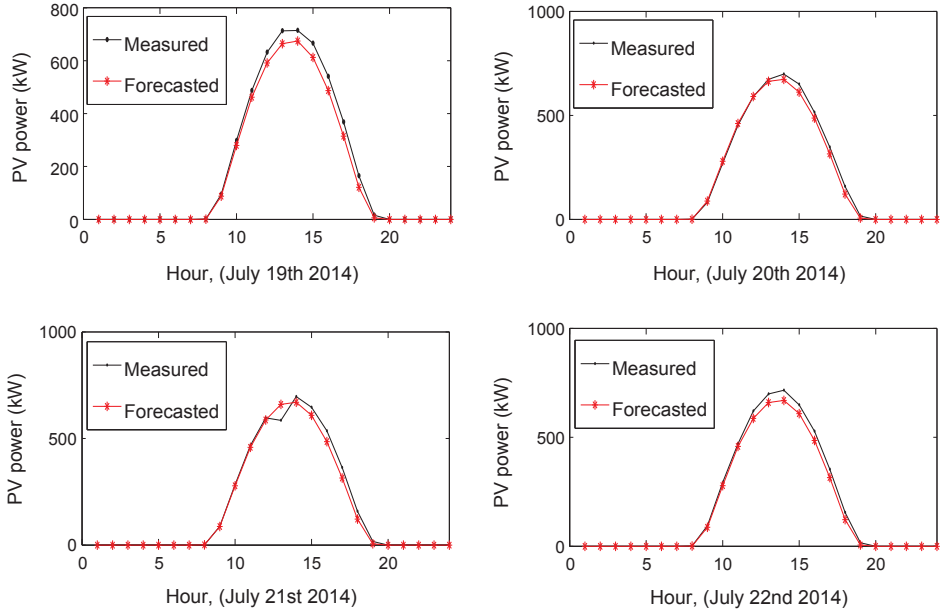


Fig. 11. Comparison between measured and forecasted PV power values 24-h-ahead during period July 19th–22nd (sunny days).

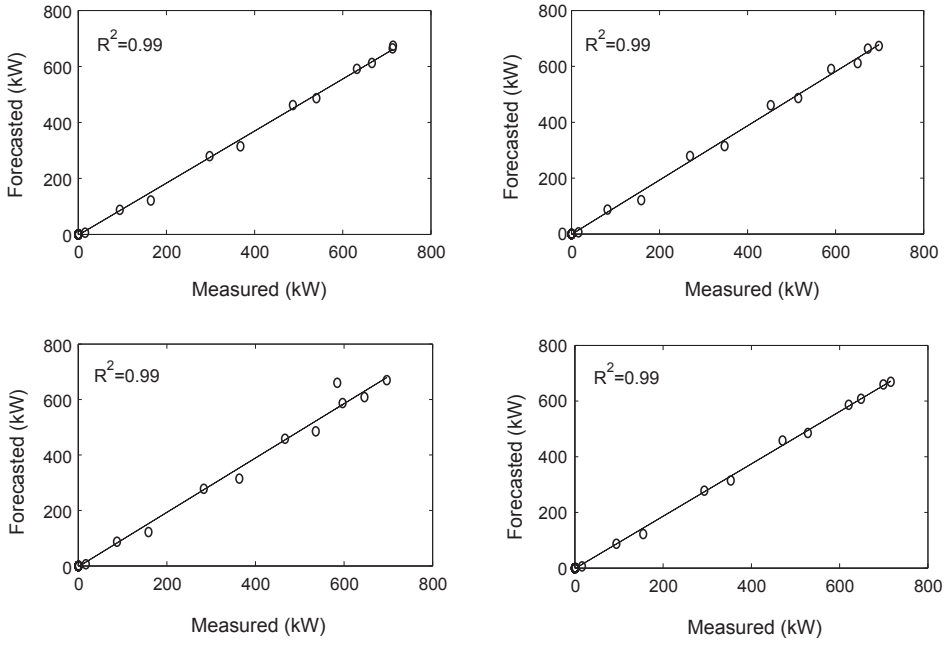


Fig. 12. Scatter plots comparison between measured and forecasted PV power values 24-h-ahead for 4 sunny days (July 19th–22nd).

Table 3
Statistical test between measured and forecasted PV power values 24-h-ahead for 4 sunny days: July 19th–22nd 2014 and 4 cloudy days: November 19th, December 23rd, January 20th, February 20th.

Seasons	Days	MAE (kW)	r	RMSE (kW)
Winter	July 19th	16.33	0.9993	26.72
	July 20th	8.61	0.9988	15.49
	July 21st	12.97	0.9957	24.54
	July 22nd	13.32	0.9996	21.83
Summer	November 19th	65.16	0.9567	113.61
	December 23rd	97.58	0.9225	158.19
	January 20th	74.04	0.9153	120.34
	February 20th	37.14	0.9864	60.12

customer reference. Note that future PV power and load demand could be forecasted using MCESN.

5. Analysis of data characteristics

The prediction performance has been evaluated by quantifying the difference between the predicted results and the measured results. However, internal dynamic of the measured results are not essentially the same with the predicted results. Therefore, some data characteristics, including descriptive statistics, seasonality, stationarity (or non-stationarity), and complexity, are qualitatively investigated in this section. The one-hour-ahead forecast is taken as an example to analyze these characteristics.

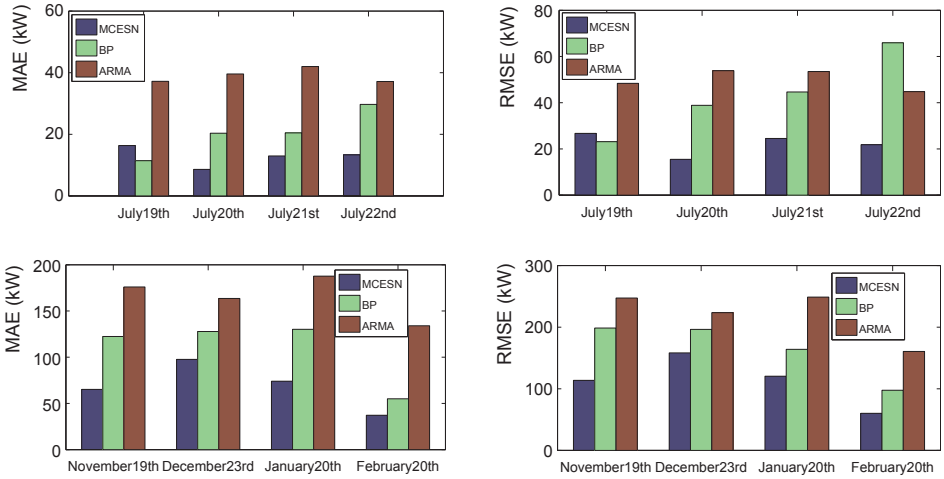


Fig. 13. MAE, RMSE comparison between measured and forecasted PV power values for 4 sunny days: July 19th–22nd 2014 and 4 cloudy days: November 19th, December 23rd, January 20th, February 20th.

Table 4
Correlation coefficient (r) comparison between measured and forecasted PV power values for 4 sunny days: July 19th–22nd 2014 and 4 cloudy days: November 19th, December 23rd, January 20th, February 20th.

		Correlation coefficient r		
Seasons	Days	ARMA	BP	MCESN
Winter	July 19th	0.9866	0.9957	*0.9993*
	July 20th	0.9836	0.9928	*0.9988*
	July 21st	0.9824	0.9835	*0.9957*
	July 22nd	0.9845	0.9690	0.9996
Summer	November 19th	0.8686	0.9227	*0.9567*
	December 23rd	0.8273	0.9143	*0.9225*
	January 20th	0.8213	0.9048	*0.9153*
	February 20th	0.9020	0.9440	*0.9864*

5.1. Descriptive statistics

The histogram between the measured and predictive PV power values are firstly studied. The histogram for January (in the summer) and July (in the winter) is given in Fig. 14(a) and (b), respectively. From Fig. 14, it can be seen that the PV power distribution of the forecasted results is similar with that of the measured results. The distribution of January is also different with that of July, which indicates there exist varying dynamics between seasons. The mean and standard deviation of monthly PV power are computed in Table 5. It can be concluded that the mean, standard deviation of the forecasted results are close to those metrics of the measured results. In addition, statistical test between measured and forecasted values is conducted, e.g., F-test and T-test, and the results are reported in Table 5. Note that 0 means two data sets are statistically similar, and 1 means they are significantly different. The F-test results indicate that there is no significant difference between measured values and forecasted values for most months except March. The T-test results also show that there is no significant difference between measured and forecasted PV power values for all months. Therefore, it can be concluded that the forecasted values is similar with the measured valued.

5.2. Periodicity and stationarity

In order to explore the periodic or seasonal characteristics, a surface mesh and a gray image are plotted in Fig. 15. When the region is brighter, the PV power is more intensive, and vice versa. The profiles of PV power show seasonally periodic, although some fluctuations occur in summer (January, November, and December). There is a wider white blob during the summer compared with the winter, as the period from dawn to dusk is longer. Meanwhile, autocorrelation coefficients of measured and forecasted data are plotted in Fig. 16, which can indicate the cyclical pattern has a period of 24 h and non-stationarity. Both the measured and forecasted autocorrelation coefficients values with lag of 24 h are far higher than those with other lags, further demonstrating the 24-h-lag information has strongly positive correlation. Note that non-stationarity means that the statistical properties of PV power dynamics remain diverse during the data generation process. It can be concluded that internal dynamic characteristics, with respect to periodicity and stationarity, keep similar between the measured and forecasted results.

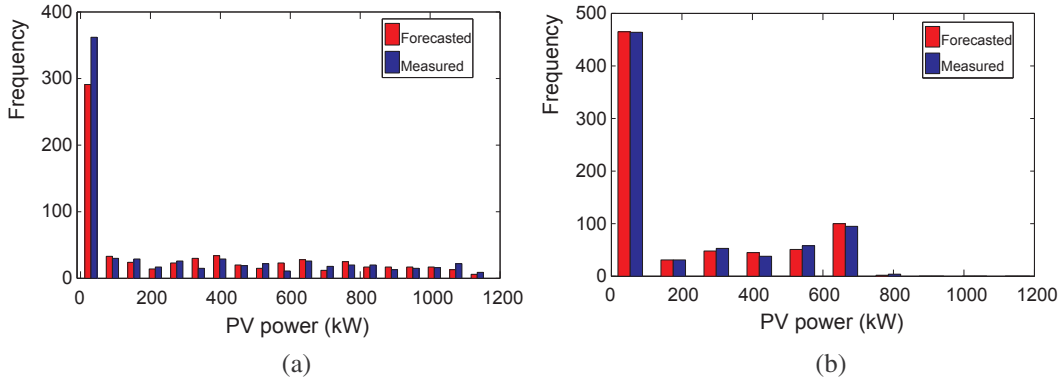


Fig. 14. Histogram comparison between forecasted and measured hourly PV power: (a) January; (b) July.

Table 5
Monthly mean and standard deviation comparison between the forecasted and actual PV power.

The measured values			The forecasted values			Statistical test		
Month	Mean (kW)	Std (kW)	Month	Mean (kW)	Std (kW)	Month	*F*-test	*T*-test
1	269.92	346.92	1	280.39	343.94	1	0	0
2	253.65	328.49	2	253.18	319.09	2	0	0
3	173.96	259.98	3	163.76	231.44	3	1	0
4	201.59	282.26	4	202.75	277.53	4	0	0
5	186.78	260.31	5	188.83	260.06	5	0	0
6	174.72	244.89	6	177.59	245.64	6	0	0
7	183.38	256.32	7	185.28	256.15	7	0	0
8	209.44	287.96	8	211.61	296.99	8	0	0
9	261.14	340.75	9	262.92	343.72	9	0	0
10	287.49	370.75	10	285.03	363.40	10	0	0
11	236.96	332.01	11	235.03	320.09	11	0	0
12	265.97	355.24	12	273.59	348.61	12	0	0

5.3. Complexity

The complexity characteristic could reflect the complex state between regular and irregular relationships. Different techniques have been applied to measure the data complexity, including the phase-space reconstruction method [35], the G-P algorithm [36], and so on. A simple and fast method, i.e., visibility graph method [37], is used to analyze the complexity of forecasted and measured results in this study. The basic idea of the algorithm is to map a time series signal into an associated graph, and graph theory can be employed to characterize the associated graph. The visibility graph method can reflect the structure of the mapped time series according to [37].

For the visible graph method, scatter diagrams and corresponding degree distributions are shown in Figs. 17 and 18. Besides forecasted

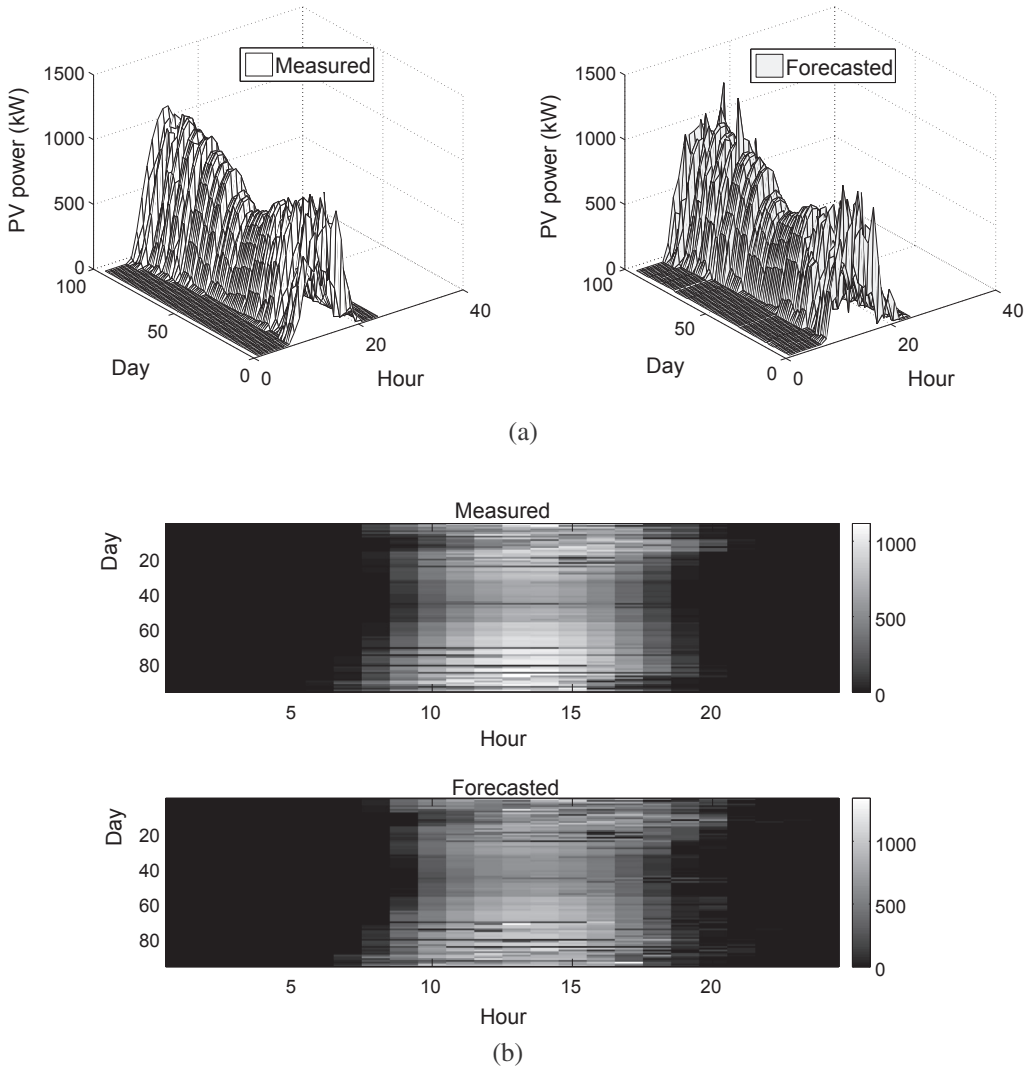


Fig. 15. (a) 2-D surface plot comparison between the estimated and measured PV power. (b) Image visualization comparison between the estimated and measured PV power.

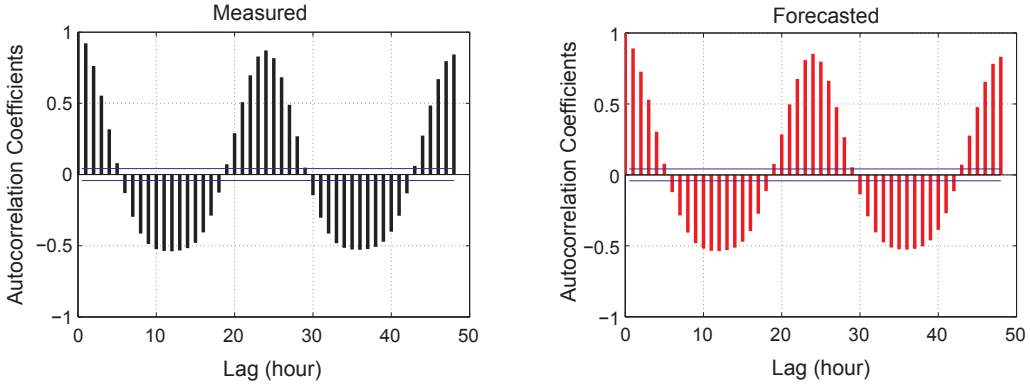


Fig. 16. Autocorrelation coefficients comparison between the measured and forecasted PV power.

and measured singles, several time series are modeled for comparison, such as, a random sequence uniformly distributed in [0,1], and a chaotic sequence generated from the Mackey-Glass system (MGS). From Fig. 17, we can conclude that both forecasted and measured signals present more intensive intra-cluster connections compared with random series. In Fig. 18, the degree distribution of random sequence fits an exponential distribution, while the degree distribution of forecasted and measured signals fits the Gauss-like distribution, which is similar to the distribution of MGS chaotic series. The Gauss-like distribution shows certain chaotic property of PV power.

Note that the scatter diagrams and degree distributions differ from each month. To further evaluate the seasonal complexity, the basic graph metrics, including average path length (AP), clustering coefficient (CC), and average degree (AD), are calculated for each month. The formulas of AP and CC are given in the following equations:

$$AP = \frac{1}{\mathrm{O}(\mathrm{O}-1)} \sum_{\varsigma \neq \zeta} \vartheta_{\varsigma,\zeta}, \tag{12}$$

$$CC = \frac{1}{\mathrm{O}} \sum_{\varepsilon=1}^{\mathrm{O}} \frac{2e_{\varepsilon}}{\xi_{\varepsilon}(\xi_{\varepsilon}-1)}, \tag{13}$$

where $\vartheta_{\varsigma,\zeta}$ denotes the shortest length between point ς and ζ of time series, O is the length of sequence; ξ_{ε}, e_{ε} represent the degree of point ε and the actual number of edges among the points connected to point ε.

In Table 6, AP, CC and AD are calculated for different time series. It can also be seen that the AP and AD of measured and forecasted signals are between those of MGS and random signals, which could indicate certain small-world properties between random and chaos. In Table 7, AP, CC and AD are calculated for each month. The measured results are comparable with respect to these three metrics. The small values of AP, CC, and AD in summer mean the high randomness, which can explain the poor performance in summer.

In this section, several main data features, including descriptive statistics, seasonality, non-stationarity and complexity of measured and forecasted results are qualitatively analyzed. Experimental results show that the measured and forecasted signals have similar dynamics and complexity. Some linear models, such as, moving average (MA), auto-regressive (AR), and auto-regressive moving average (ARMA) may not be suitable for modeling the demand-side PV power precisely. Therefore, the MCESN model is proposed to predict the demand-side PV power due to its nonlinear mapping capacity.

6. Conclusions

For the PV power forecast in the demand-side hybrid system, this paper presents a direct approach for one-hour-ahead prediction and 24-h-ahead prediction based on multi-clustered echo state network

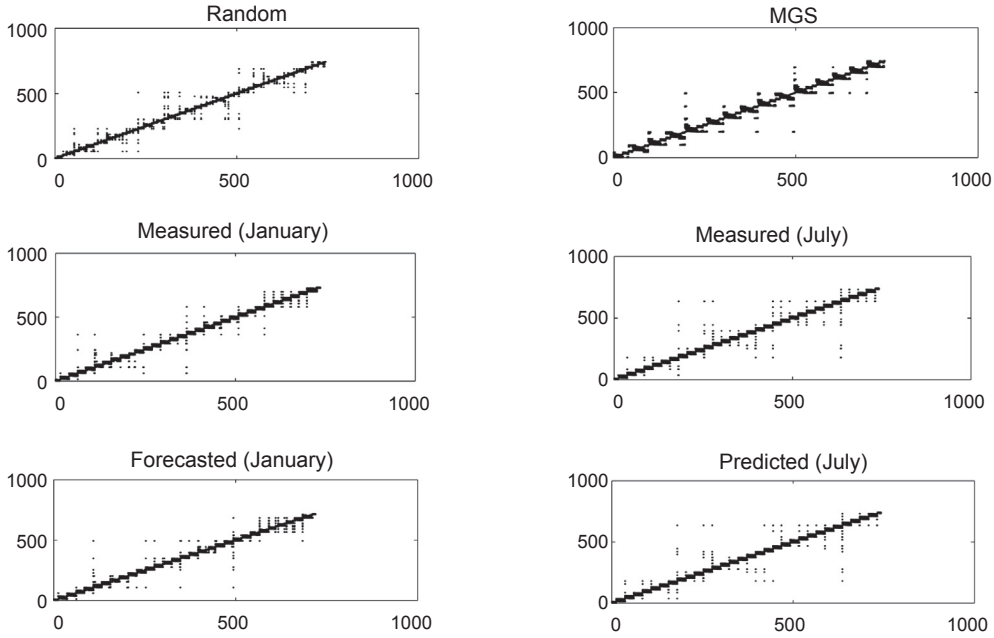


Fig. 17. Network topologies comparison from six time series according to visible graph method.

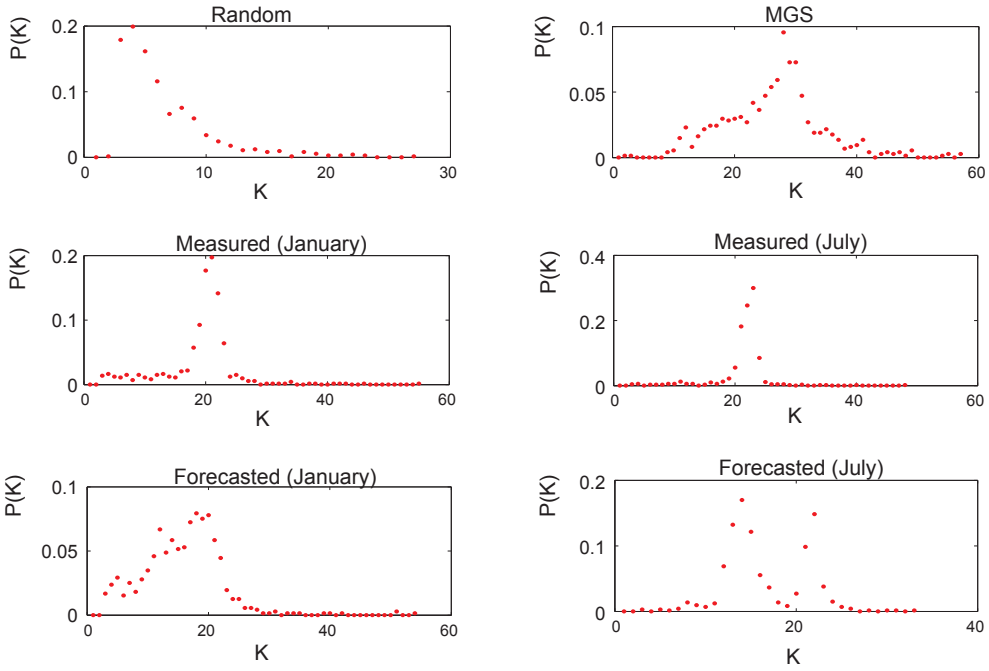


Fig. 18. Network degree distributions comparison from six time series according to visible graph method.

Table 6
Complexity characteristics comparison from six time series according to visible graph method.

Data sequences	AP	CC	AD
Random	5.4455	0.7443	5.4993
MGS	4.3041	0.6862	25.2128
Measured (January)	4.6102	0.8382	18.1641
Forecasted (January)	4.5138	0.7692	14.5063
Measured (July)	4.6141	0.8858	20.8652
Forecasted (July)	4.8181	0.7427	15.6137

Table 7
Complexity characteristics comparison from each month of PV power according to visible graph method.

The measured values				The forecasted values			
Month	AP	CC	AD	Month	AP	CC	AD
1	4.6102	0.8382	18.1641	1	4.5138	0.7692	14.5063
2	4.8815	0.8467	18.3612	2	4.3329	0.7696	14.3134
3	4.5481	0.8599	19.3315	3	4.2975	0.7865	14.6334
4	4.7880	0.8639	20.0139	4	4.4859	0.7748	15.6384
5	5.1523	0.8853	20.9757	5	4.1334	0.7722	16.5714
6	5.5073	0.8843	20.7549	6	4.6996	0.7676	15.8217
7	4.6141	0.8858	20.8652	7	4.8181	0.7427	15.6137
8	5.1591	0.8761	20.5175	8	4.3979	0.7534	15.5283
9	4.6141	0.8858	20.8652	9	4.0719	0.7386	16.0056
10	5.0368	0.8483	19.4536	10	4.5815	0.7421	15.3244
11	4.2022	0.8326	18.3477	11	4.5901	0.7680	13.9026
12	4.2323	0.8449	18.4339	12	4.2502	0.7797	14.2587

(MCESN). The proposed approach can achieve competitive performance of prediction. The effects of measured temperature, humidity, and 24-h-lag information are also studied in the MCESN model. The results show that consideration of temperature and humidity information has negligible effects on the prediction accuracy, and that the historical 24-h-lag information has positive effects on the prediction accuracy in winter and negative effects in summer. The simulation results also indicate that the proposed model could perform accurate 24-h-ahead prediction for sunny days with the correlation coefficient being 99%, and acceptable precision for cloudy days with the correlation coefficient being in the range 91–98%. MCESN could achieve more accurate prediction, compared with ARMA, BP neural networks. Finally, several data characteristics of measured and estimated PV power are qualitatively analyzed. Experimental results show that the seasonality, non-stationarity, complexity, and descriptive statistics characteristics are analogous between measured and estimated values.

There are some open issues for the PV hybrid system. One issue is big data analysis in the PV power forecast. Additional factors, such as cloud cover, sunshine duration, should be considered in the ESN model. Some advanced neural networks, such as convolutional neural network (CNN) and long short-term memory (LSTM) can also be studied for large and complicated applications. Another issue is the load modeling, which is closely related to customer behavior and demand response. The ESN model will be investigated for the load forecast. Furthermore, after day-ahead PV power output and load demand are forecasted, power flow dispatching in the PV hybrid system will be studied under different demand-side programs, e.g., the time-of-use program. Energy efficiency and economic performance must be considered in some rule-based or optimization-based strategies.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.apenergy.2018.02.160.

References

[1] Roldn MI, Monterreal R. Heat flux and temperature prediction on a volumetric receiver installed in a solar furnace. Appl Energy 2014;120:65–74.
[2] Gulin M, Vaak M, Peric N. Dynamical optimal positioning of a photovoltaic panel in all weather conditions. Appl Energy 2013;108:429–38.
[3] Younes S, Muneer T. Clear-sky classification procedures and models using a world-wide data-base. Appl Energy 2007;84(6):623–45.
[4] Mellit A, Benghanem M, Arab AH, Guessoum A. A simplified model for generating sequences of global solar radiation data for isolated sites: using artificial neural network and a library of markov transition matrices approach. Solar Energy 2005;79(5):469–82.
[5] Hassan MA, Khalil A, Kaseb S, Kassem MA. Exploring the potential of tree-based ensemble methods in solar radiation modeling. Appl Energy 2017;203:897–916.
[6] Hassan GE, Youssef ME, Mohamed ZE, Ali MA, Hanafy AA. New temperature-based models for predicting global solar radiation. Appl Energy 2016;179:437–50.

[7] Wong L, Chow W. Solar radiation model. Appl Energy 2001;69(3):191–224.
[8] Deo RC, Wen X, Qi F. A wavelet-coupled support vector machine model for forecasting global incident solar radiation using limited meteorological dataset. Appl Energy 2016;168:568–93.
[9] Mellit A, Pavan AM. A 24-h forecast of solar irradiance using artificial neural network: application for performance prediction of a grid-connected pv plant at trieste, Italy. Solar Energy 2010;84(5):807–21.
[10] Cao J, Lin X. Study of hourly and daily solar irradiation forecast using diagonal recurrent wavelet neural networks. Energy Convers Manage 2008;49(6):1396–406.
[11] Capizzi G, Napoli C, Bonanno F. Innovative second-generation wavelets construction with recurrent neural networks for solar radiation forecasting. IEEE Trans Neural Netw Learn Syst 2012;23(11):1805–15.
[12] Cao S, Weng W, Chen J, Liu W, Yu G, Cao J. Forecast of solar irradiance using chaos optimization neural networks. Power and energy engineering conference, Asia-Pacific. 2009. p. 1–4.
[13] Cao J, Lin X. Application of the diagonal recurrent wavelet neural network to solar irradiation forecast assisted with fuzzy technique. Eng Appl Artif Intell 2008;21(8):1255–63.
[14] Shah ASBM, Yokoyama H, Kakimoto N. High-precision forecasting model of solar irradiance based on grid point value data analysis for an efficient photovoltaic system. IEEE Trans Sust Energy 2015;6(2):474–81.
[15] Elsinga B, van Sark WG. Short-term peer-to-peer solar forecasting in a network of photovoltaic systems. Appl Energy 2017;206:1464–83.
[16] Lou S, Li DHW, Lam JC, Chan WWH. Prediction of diffuse solar irradiance using machine learning and multivariable regression. Appl Energy 2016;181:367–74.
[17] Salcedo-Sanz S, Deo RC, Cornejo-Bueno L, Camacho-Gmez C, Ghimire S. An efficient neuro-evolutionary hybrid modelling mechanism for the estimation of daily global solar radiation in the sunshine state of australia. Appl Energy 2018;209:79–94.
[18] Jaeger H, Haas H. Harnessing nonlinearity: predicting chaotic systems and saving energy in wireless communication. Science 2004;304(5667):78–80.
[19] Ozturk MC, Principe JC. An associative memory readout for ESNS with applications to dynamical pattern recognition. Neural Netw 2007;20(3):377–90.
[20] Gallicchio C, Micheli A. Tree echo state networks. Neurocomputing 2013;101:319–37.
[21] Meftah B, Lzoray O, Benyettou A. Novel approach using echo state networks for microscopic cellular image segmentation. Cognit Comput 2016;8(2):237–45.
[22] Shi G, Liu D, Wei Q. Echo state network-based q-learning method for optimal battery control of offices combined with renewable energy. IET Control Theory Appl 2017;11(7):915–22.
[23] Zhong S, Xie X, Lin L, Wang F. Genetic algorithm optimized double-reservoir echo state network for multi-regime time series prediction. Neurocomputing 2017;238:191–204.
[24] Li D, Min H, Wang J. Chaotic time series prediction based on a novel robust echo state network. IEEE Trans Neural Netw Learn Syst 2012;23(5):787–99.
[25] Utpal KumarDas Mea, SoonTey Kok. Forecasting of photovoltaic power generation and model optimization: a review. Renew Sust Energy Rev 2018;81(1):912–28.
[26] Long H, Zhang Z, Su Y. Analysis of daily solar power prediction with data-driven approaches. Appl Energy 2014;126:29–37.
[27] Alessandrini S, Monache LD, Sperati S, Cervone G. An analog ensemble for short-term probabilistic solar power forecast. Appl Energy 2015;157:95–110.
[28] Ferlito S, Adinolfi G, Graditi G. Comparative analysis of data-driven methods online and offline trained to the forecasting of grid-connected photovoltaic plant production. Appl Energy 2017;205:116–29.
[29] Tang L, Yu L, He K. A novel data-characteristic-driven modeling methodology for nuclear energy consumption forecasting. Appl Energy 2014;128:1–14.
[30] Xue F, Li Q, Zhou H, Li X. Reservoir computing with both neuronal intrinsic plasticity and multi-clustered structure. Cogn Comput 2017;9(3):400–10.
[31] Nisbach F, Kaiser M. Developmental time windows for spatial growth generate multiple-cluster small-world networks. Euro Phys J B 2007;58(2):185–91.
[32] Wyffels F, Schrauwen B, Stroobandt D. Stable output feedback in reservoir computing using ridge regression. In: International conference on artificial neural networks; 2008. p. 808–17.
[33] Smith MR. Photovoltaic system performance monitoring guidelines for measurement, data exchange and analysis. Int Organ 1998;29(3):211–8.
[34] Venayagamoorthy GK, Shishir B. Effects of spectral radius and settling time in the performance of echo state networks. Neural Netw 2009;22(7):861–3.
[35] Packard NH, Crutchfield JP, Farmer JD, Shaw RS. Geometry from a time series. Phys Rev Lett 1980;45(9):712–6.
[36] Grassberger P, Procaccia I. Measuring the strangeness of strange attractors. Phys D Nonlinear Phenom 1983;9(1):189–208.
[37] Lacasa L, Luque B, Ballesteros F, Luque J, Nuo JC. From time series to complex networks: the visibility graph. Proc Natl Acad Sci USA 2008;105(13):4972–5.