Journal of Ambient Intelligence and Humanized Computing (2021) 12:10183–10195
https://doi.org/10.1007/s12652-020-02784-4

ORIGINAL RESEARCH

Reconstructing long-term wind speed data based on measure correlate predict method for micro-grid planning

Masoud Salehi Borujeni[1] · Abbas Dideban[1] · Asghar Akbari Foroud[1]

Received: 8 July 2020 / Accepted: 2 December 2020 / Published online: 3 January 2021

Abstract
One of the most significant uncertain parameters in microgrid planning is wind speed. Wind speed has a complex dynamic behavior and developing models with high accuracy for this parameter can help decrease microgrid costs and can help improve reliability. Previous methods for wind speed modeling have been mostly statistical methods (e.g., the auto-regressive integrated moving average (ARIMA), the multivariate distribution functions, Copula function). Afore-mentioned methods are suitable for modeling stationary data. However, since the wind speed data are not stationary, the measure correlate predict (MCP) method, capable of modeling the non-stationary data, is used in this paper to generate the wind speed scenarios. Based on radial basis function (RBF) artificial neural network, this work has proposed a novel hybrid computational model to improve MCP method's performance. The results indicate that the Hybrid-MCP method is more accurate than the conventional statistical methods, which are used to generate the wind speed scenarios (24 h).

Keywords Scenario generation · Wind speed · Measure-correlate-predict · RBF artificial neural network · Recurrence plot

1 Introduction

The use of wind energy as one of the most feasible renewable energy resources is increasing over time due to its low-cost, practicability, and availability in different regions. The integration of wind turbines in microgrids and the development of wind farms are examples of many wind energy applications (Petersen 2017; Sengar and Liu 2020). To design wind farms and microgrids, wind measurement towers recording wind speed at the height of the wind turbine hub are required to be installed in proper locations. Recording wind speed data for a long period is not possible, and hence the data is typically recorded for 1 year (Kim and Kim 2016). The one-year data record doesn't represent wind speed's long-term behavior, and therefore the use of this data in micro-grid planning leads to an inaccurate estimation of the project profit providing less reliable results (Baringo and Conejo 2013). To decrease the wind speed's uncertainty effect on microgrid planning, different methods have been employed. One of the most important methods is the scenario-based stochastic programming. In this method, the uncertainties are defined as a set of scenarios. Since wind speed data have a daily period, each 24-h data profile of wind speed is considered as a scenario. In microgrid planning, a finite number of scenarios are considered. The scenarios and their accompanied probabilities are required to efficiently model uncertainty variable's behavior as the accuracy of the modeling affects the solutions. The more accurate modeling of the uncertainties results in providing answers to the stochastic programming methods, which are close to the actual optimal values (Niknam et al. 2012; Salehi Borujeni et al. 2017).

To generate the wind speed scenarios, the short-term and long-term behaviors of the wind speed is required to be monitored; the scenarios must be structurally similar to the daily profiles of the wind speed. Furthermore, the mean and variance of the generated scenarios is required to be similar to those of the long-term behavior of the wind speed. In related literature, different methods have been proposed for the scenario generation. The most widely used methods can be divided into two categories. The first category is scenario generation methods based on distance matching.

✉ Asghar Akbari Foroud
aakbari@semnan.ac.ir

Masoud Salehi Borujeni
m.salehi.b@semnan.ac.ir

Abbas Dideban
adideban@semnan.ac.ir

[1] Electrical and Computer Engineering Department, Semnan University, Semnan, Iran

These methods commonly consider the available measured data (training data) as possible scenarios and use distance indexes to eliminate scenarios that fail to meet the distance-based criteria. The most popular methods in this category which are utilized to generate the wind speed scenarios are clustering, backward reduction and forward selection (De Caro et al. 2018; Guan et al. 2018; Li et al. 2020). In these methods, probable scenarios unavailable in the training data are not considered and to overcome this problem, the statistical methods have been developed to generate the wind speed scenario; nevertheless, fundamentals of statistical methods are similar. First, a model is developed based on the available data (one-year gathering of the real data). Next, more scenarios are generated using the model to generate probable scenarios unavailable in the training data, and then their effects are applied to the microgrid optimization (Liu et al. 2018; Salehi Borujeni et al. 2018). One of the simplest methods among these techniques is the use of the probability distribution function (PDF). In this method, first, the probability distribution function of the wind speed data is obtained, and then more scenarios are generated using Monte Carlo method and probability distribution function (Kaplani and Kaplanis 2012; Tabar et al. 2017; Esmaeeli et al. 2020). In this method, the wind speed in different hours is assumed to be independent; however, the underlying assumption is untrue as the wind data show serial linear dependence. To solve this problem, ARIMA method has been proposed for the modeling of wind speed (Conejo et al. 2010; Chen and Rabiti 2017). The use of a multivariate distribution function is another method for wind speed scenario generation. In previous research, the normal multivariate distribution function has been employed. In this method, the dependence among wind speed data is also efficiently modeled (Pinson et al. 2009; Ma et al. 2013). In addition to linear correlation, dependence structure of wind speed data is considered in a few of previous researches. They used different type of copula function to model the dependency structure of wind speed. The result shows that copula model has a good performance to model the wind speed (Salehi Borujeni et al. 2018; Camal et al. 2019; Deng et al. 2020). The restriction of the statistical methods is that they are appropriate for the stationary data, and the mean and variance of the generated data obtained by these methods are completely similar to those of the training data. In other words, the mean and variance of the generated scenarios are not similar to those of the long-term behavior.

Recently, the measure correlate predict (MCP) method has been proposed to predict the long-term wind speed data profile based on the short-term measured data (target data). This method is used in wind farm planning to estimate the project profit and to determine the layout of the wind turbines (Perea et al. 2011; Diaz et al. 2017). In addition to the short-term measured data, reference data are utilized in MCP method. The reference data consist of either re-analysis data or the data reported by meteorology centers near the site, available for a long period. Utilizing the reference data and the measured short-term data (target data), the site's long-term data are simulated using a computational model (Weekws et al. 2015). The computational model determines the relationship between reference data and target data. An appropriate computational model results in accuracy enhancement of the wind speed data modeling. In previous research, different computational models have been presented for MCP, such as linear regression, variance ratio, Weibull scale, least square regression, and artificial neural networks (Weekes and Tomlin 2014; Zhang et al. 2014; Sharma et al. 2019). In wind farm planning, annual power generation of the wind turbines is crucial, and economic analysis is performed based on this parameter. Therefore, in wind farm planning, the focus is on the wind speed's long-term behavior, and the daily power generation profile is typically disregarded.

In microgrid planning, the long-term and short-term behaviors of wind speed are important, and in addition to providing the mean and variance similar to those of the long-term data, scenarios are required to model wind speed's short-term behavior accurately. Our work intends to investigate whether the scenario generation using MCP can provide accurate modeling of the wind's short-term behavior (wind daily profile) and long-term modeling of the behavior. In MCP method, the selection of the reference data is a challenging step. The reference data and the target data are supposed to have similar dynamic and seasonal behaviors as well as linear correlation so that, besides high accuracy in the long-term modeling of wind behavior, higher accuracy in short-term modeling (daily profile) is also achievable. Therefore, the dynamic of the reference wind speed data and the target data is analyzed to highlight the similarities among their behaviors. To do so, the cross-recurrence plot nonlinear data analysis method is used. This analysis is presented as a new criterion for reference data verification. Furthermore, a hybrid computational model is presented for MCP model to increase the model's accuracy. In this model, RBF artificial neural network is employed, and its parameters are optimized using Genetic algorithm resulted in a better performance compared to the following methods: the variance ratio method, the Weibull scale method, and the feed-forward neural network. The results indicate that MCP method with the proposed hybrid model is more accurate than the conventional statistical models utilized for wind speed scenario generation, and MCP method is able to accurately model the short-term and long-term behaviors of wind speed. The main contributions of this paper are follows:

- Accuracy improvement of the short-term and long-term wind speed modeling

- Proposing recurrence plot as a new criterion for re-analysis data verification.
- Presenting a hybrid computational model based on RBF artificial neural network and Genetic Algorithm to improve the MCP method's performance.

This paper is organized as follows: In Sect. 2, the target and the reference data are introduced and their dynamic and seasonal behaviors are studied. Scenario generation based on MCP is explained in Sect. 3. Section 4 provides a comparison between the proposed method and the conventional scenario generation methods. The final section is designated the conclusion.

2 Data

In order to develop a MCP method, two different data sets are needed. Target data which represent short term measured data in the target site and reference data which are available for long-term. In this section, a description is provided for each type of these data sets and also their dynamic and seasonal behaviors are studied.

2.1 Target data

The minimum necessary time duration for recording wind speed data should be to the extent that the seasonal behavior of wind speed is represented. A general recommendation is to consider the data for a 1-year time period (Carta et al., 2013). Wind speed data at 80 m altitude in geographical position of 39° 54′ N–105° 14′ W from 01.01.2011 to 31.12.2016 is considered hourly. These data are related to NWTC M2 tower, extracted from National Renewable Energy Laboratory website (http://www.nrel.gov/midc/nwtc_m2, last visited in January 2017). The data collected in 2011 have been used for training, while the data collected from 2012 to 2016 have been used for test purposes. The mean of wind speed in this region is 4.95 m/s in 2011, with a variance of 14.36 $(m/s)^2$, minimum of 0.26 m/s and a maximum of 27.73 m/s. Figure 1 (a) illustrates wind speed in this region in year 2011.

2.2 Reference data

In addition to the target data, long-term wind speed data is also required to employ MCP model. These data can be obtained either from the meteorological centers near the reference site or from re-analysis data. The re-analysis data have been generated using the observed or measured data from different systems such as meteorological centers or satellites and a numerical weather prediction (NWP) model (Carta et al. 2013). Two important reasons for using re-analysis data in MCP method are as follows: The first reason is its simplicity and availability. The second reason is that the wind speed data are available for a long time period (Brower 2006).

One of the accurate re-analysis data sets is MERRA (Modern Era Retrospective-analysis for Research and Applications) wind speed data. These data have been provided by NASA, available for free (http://www.nasa.gov, last visited in January 2017). In this paper, wind speed data at 50 m altitude for the mentioned site from 01.01.2006 to 31.12.2011 are considered. The mean value of the data in 2011 is 6.52 m/s, its variance is 8.71 $(m/s)^2$, with a minimum of 0.05 m/s and a maximum of 16.94 m/s. Figure 1 (b) illustrates the re-analysis wind speed data in this region in year 2011.

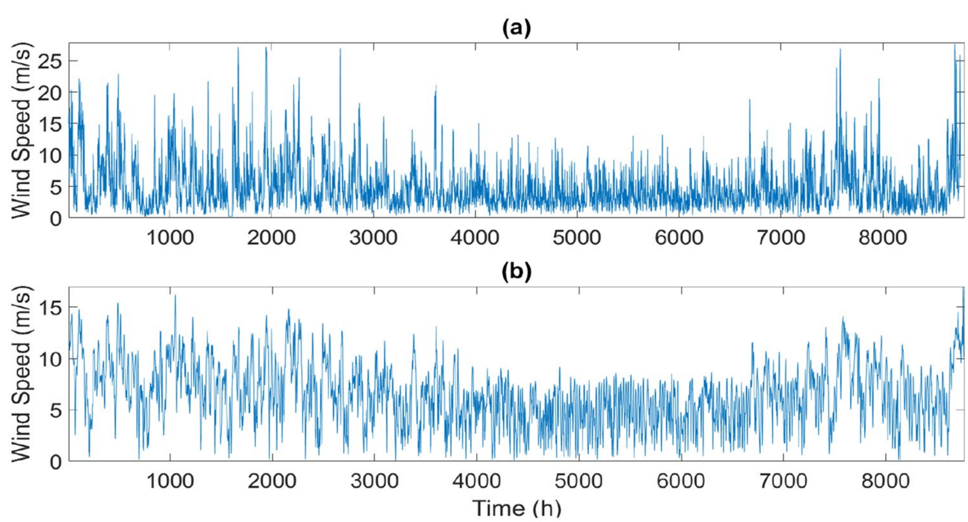

Fig. 1 Wind speed data of the aforementioned site in 2011, **a** target data (at 80 m), **b** re-analysis data (at 50 m)

2.3 Data analysis

As mentioned before, the target data represent wind speed at 80 m altitude, and the re-analysis data represent wind speed at 50 m altitude. Figure 2 shows the histogram and the best-fit probability distribution function of these data.

Shown in Fig. 2, in addition to significant differences between the mean and the variance of these two data sets, the shapes of their histograms are very different, hence making it impossible to use the re-analysis data instead of the target data. It is recommended that the heights above ground level at which the data are collected at the reference and target sites should be similar. It has been shown that the correlation coefficient between the two data series is lower than that obtained when both data series are recorded at the same height (Probst and Cardenas, 2010). Respectively, Fig. 3a illustrates data scattering for re-analysis and target data, and Fig. 3b indicates the linear correlation among the data.

Shown in Fig. 3, most of the linear correlations are associated with the zero time delay, whose value is 0.55.

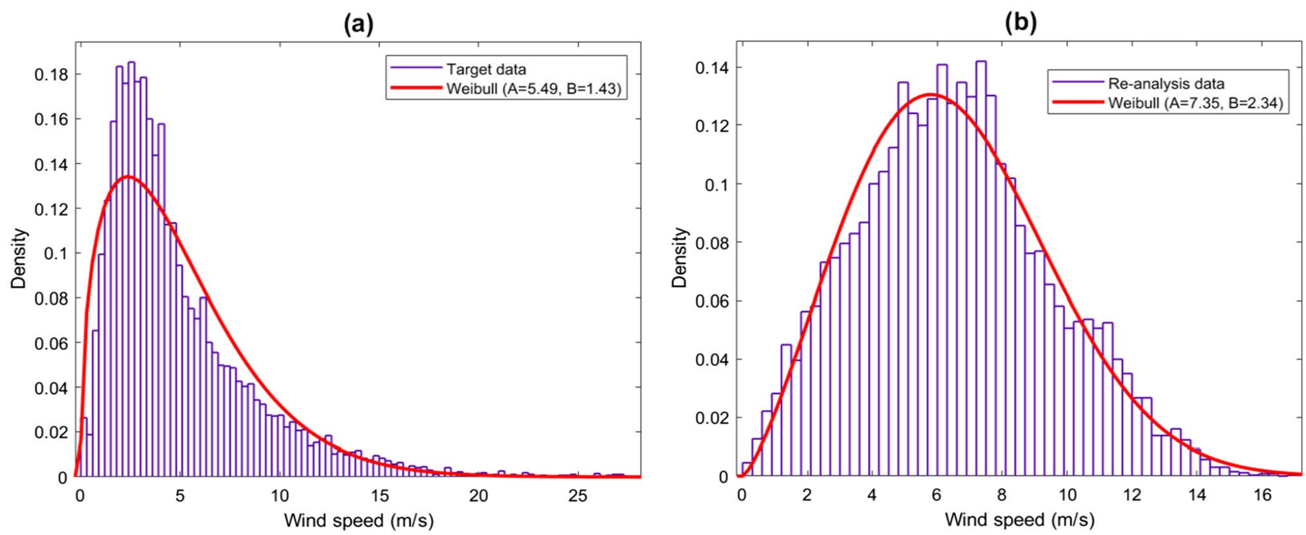

Fig. 2 Histogram and the probability distribution function of **a** target data, **b** re-analysis data

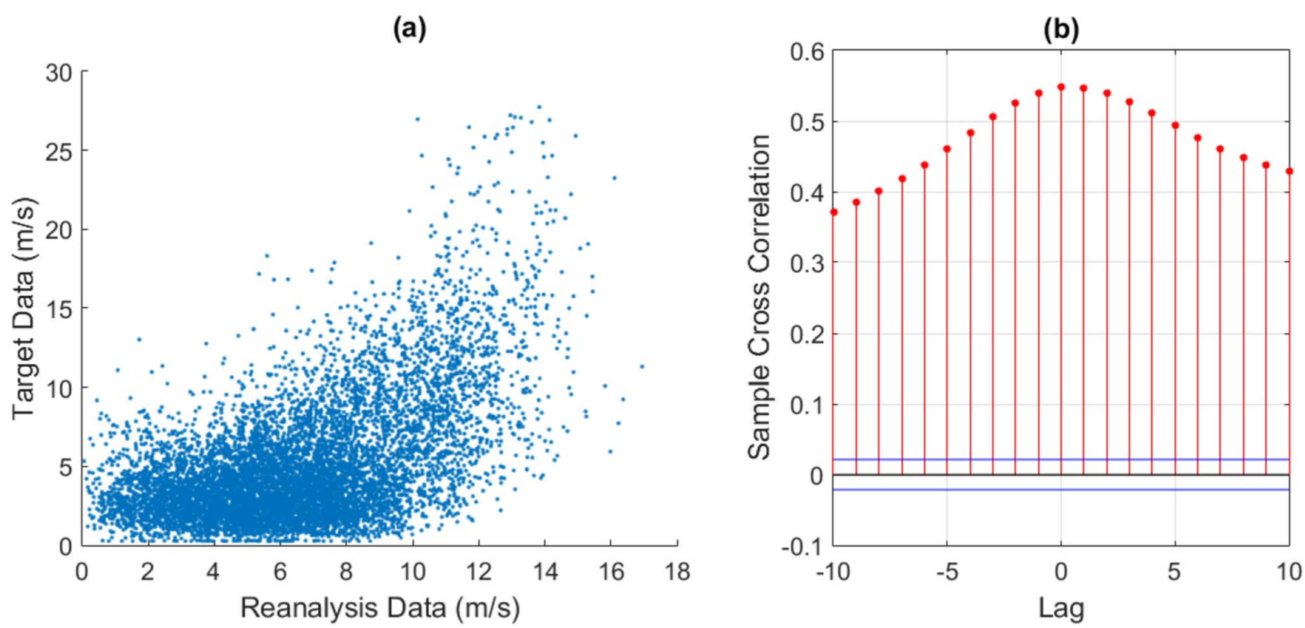

Fig. 3 **a** Scattering of target data and re-analysis data (target data at 80 m and reanalysis data at 50 m), **b** correlations between the target data and re-analysis data

Although no strong linear correlation was observed between the data, nonlinear dynamic analysis will show the possibility of using the reanalysis data model to model the real wind speed data. At first, the power spectral density (PSD) can be utilized to specify the periodic components of each data set. PSD determines how the power distribution of a time series is based on frequency. In general, for a time series with N samples, the set of periods is $\{T_1, T_2, \dots, T_N\}$, where $T_i = \frac{1}{f_i}$ (f_i is the value of the i^{th} frequency). Using the frequency of the PSD peaks, it is possible to obtain the periods of a time series (Ambach 2016). Figure 4 shows the PSD of target data and re-analysis data.

For the both data sets, two PSD peaks were found with frequencies of 0.00011413, and 0.04167 as shown in Fig. 4. The time periods for these frequencies were $\frac{1}{0.00011413} \cong 8760$ h (annually), and $\frac{1}{0.04167} \cong 24$ h (daily), respectively. The other case to investigate is the dynamic behavior of time series of target and re-analysis wind speed data. The more similar their dynamic behaviors, the more effective the use of MCP method in wind speed scenarios' generation. This criterion is more important than the linear correlation criterion as it considers the non-linear relationships among the data. Cross-recurrence plot (CRP) is a tool presented to investigate the dynamic behavior of two time-series. CRP searches for the cases where a state of the first system recurs to one of the other systems. To use CRP analysis, first, the phase space for time series should be reconstructed using delay method. In this method, the vectors are transformed to a new space with respect to the dimension and delay parameters. Given two time series $A = \{\ a_1\ a_2\ \dots\ a_n\ \}$ and $B = \{\ b_1\ b_2\ \dots\ b_n\ \}$, the phase space vectors for the time series of A and B are defined using x and y vectors.

$$x_k = \{a_k, a_{k+\tau}, a_{k+2\tau}, \dots, a_{k+(d-1)\tau}\}, \tag{1}$$

$$y_k = \{b_k, b_{k+\tau}, b_{k+2\tau}, \dots, b_{k+(d-1)\tau}\}, \tag{2}$$

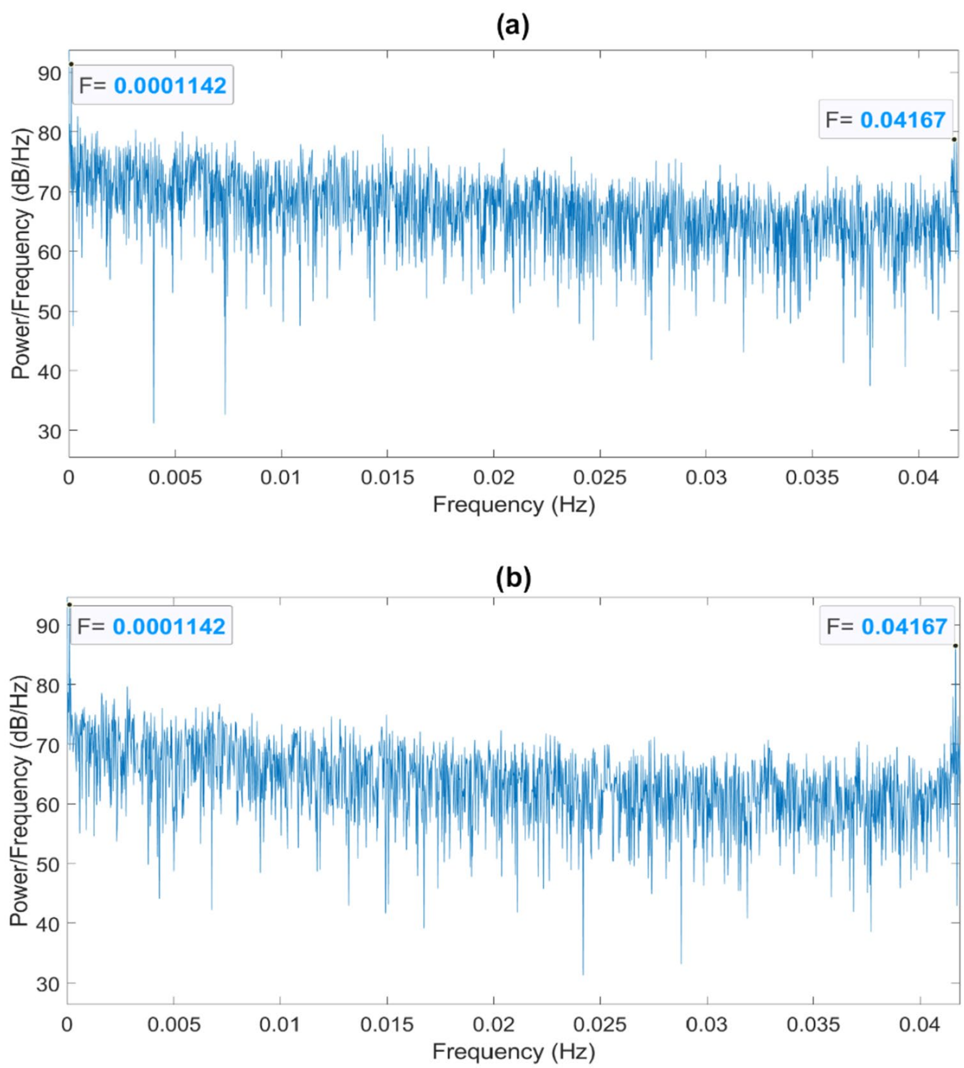

Fig. 4 PSD of wind speed for periods between 24 h up to 1-year **a** target data, **b** re-analysis data

where d and τ represent dimension and delay, respectively. It is required to indicate both time series in a phase space. If the embedding parameters of the two systems are not the same, greater parameters are required to be considered. After phase space reconstruction, CRP matrix is calculated using Eq. (3) (Brower 2006; Marwan et al. 2007).

$$R_{ij} = \|x_i - y_j\|. \tag{3}$$

In this paper, the false nearest neighbor method is used for dimension calculation, extracted from Kennel et al. (1992), and mutual information method is used for delay calculation (Marwan and Kurths 2002). For this purpose, CRP toolbox MATLAB is used (http://tocsy.agnld.uni-potsdam.de, last visited in September 2016). In various papers, several characteristics of CRP matrix have been investigated. However, in this paper, the values of the main diagonal of CRP matrix, representing the dynamic behavior of the two systems without time delay, are considered. Figure 5 illustrates the diagonal elements of CRP matrix for the target and re-analysis data. The data has a dimension and delay value of 9 and 20, respectively. The normalized entries of the main diagonal are between zero and one. If the values were closer to zero, the dynamic behavior of the two series would be more similar.

As illustrated in Fig. 5, over several time periods during a year, the dynamic behavior of the two systems is extremely close to each other. Generally, the mean difference between the dynamic behaviors of these two systems is 0.25, a value close to zero implying the similarity of dynamic relationship for these two systems. This indicates that although the linear correlation between these two sets of data is low, they have similar dynamic behaviors; therefore, due to the equal periodic components as well as similar dynamic behaviors the re-analysis data can be used to model real wind speed data.

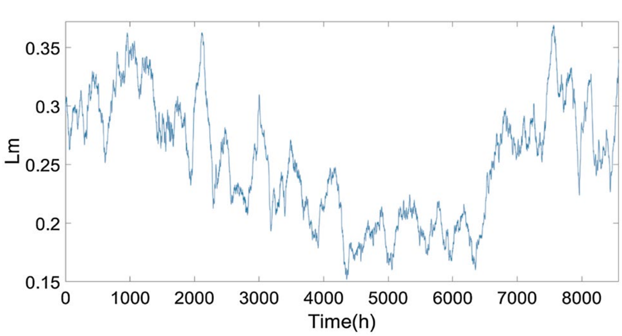

Fig. 5 Diagonal elements of CRP matrix for target and re-analysis wind speed data

3 Scenario generation

One of the effective methods for microgrid optimization is scenario-based stochastic programming. In this method, the uncertainties are modeled as a set of scenarios. Higher accuracy of the scenario generation results in the stochastic programming method's answers that are close to the actual optimal values. In this paper, a measure-correlate-predict (MCP) method is developed to generate 24-h scenarios of wind speed and this method is presented in this section.

3.1 Introduction of MCP method

Measure-correlate-predict (MCP) is proposed to determine the long-term behavior of a site using the re-analysis data. At a target site, short-term wind speed data, usually during a one-year period, is recorded by a wind measurement tower $\left(W_{Site}^{short}\right)$. These data are not enough for determining long-term behavior of the site; therefore, the re-analysis wind speed data at this site are used. Re-analysis data consists of two groups including short-term data whose time period is proportional to W_{Site}^{short} data $\left(W_{MERRA}^{short}\right)$, and long-term data indicating wind speed in a 5-year period $\left(W_{MERRA}^{long}\right)$. Firstly, a computational model is trained using the short-term data of W_{Site}^{short} and W_{Site}^{short} to determine the relationship between the short-term data obtained by the target and the re-analysis data. The training method of the computational model is explained in the following section. Secondly, the long-term re-analysis data $\left(W_{MERRA}^{long}\right)$ are imported to the computational model, and the long-term data of the site $\left(W_{Site}^{long}\right)$ are simulated using the computational model. Then, the simulated data can be used to determine the long-term behavior of wind speed at the target site (Carta and Velázquez, 2011). Figure 6 illustrates the fundamentals of this method.

After the simulation of the long-term wind speed at the site from 2006 to 2010 using MCP model, these data are considered as wind speed scenarios (24-h scenarios). Sequentially, the output power of a wind turbine is calculated using the modeled wind turbine as indicated by Eq. (4). The mathematical model of a wind turbine is as follows:

$$p_W = \begin{cases} 0 v \le v_c \\ av^3 - bP_r v_c < v < v_r \\ P_r v_r \le v \le v_o \\ 0 v > v_o \end{cases}, \tag{4}$$

where v_c is the cut-in speed that the turbine starts to work. From v_c to v_o, the turbine power remains constant. For the speeds greater than v_o, the turbine is turned off to avoid damages. The coefficients a and b are calculated as follows (Mohammadi et al. 2012):

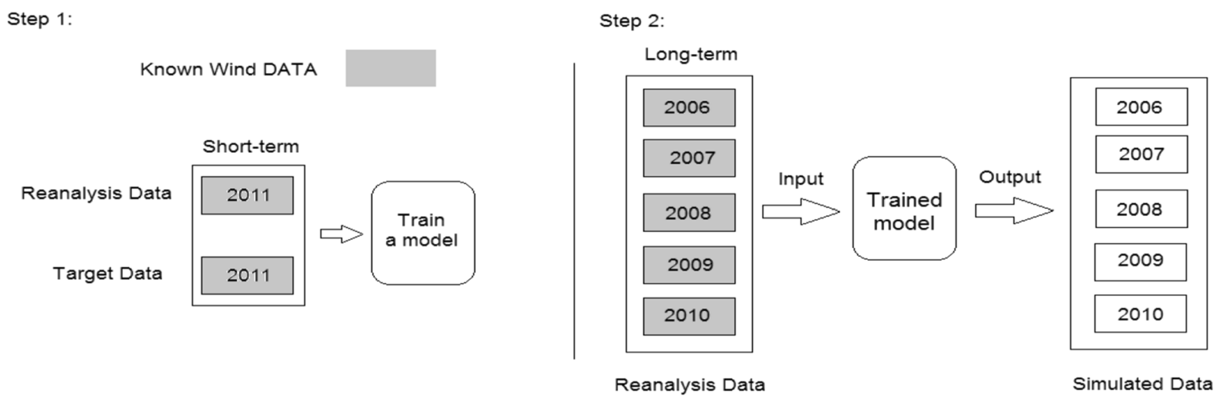

Fig. 6 Schematic of MCP method

$$a = \frac{P_r}{v_r^3 - v_c^3}. \tag{5}$$

$$b = \frac{v_c^3}{v_r^3 - v_c^3}. \tag{6}$$

Finally, the number of scenarios is required to be reduced to decrease the calculation burden in microgrid planning. In this paper, the number of scenarios is reduced to 10 using forward scenario selection method (Marwan and Kurths, 2002).

3.2 Computational models used in MCP method

In this section, the well-known computational models including Weibull scale method and variance ratio method, provided to be used for comparison purposes, are introduced. Then, our new computational model based on RBF neural network is explained.

3.2.1 The Weibull scale method (WSM)

WSM is developed based on the two-parameter Weibull distribution function. This method is used when the distribution function of wind speed is Weibull; Weibull probability distribution function is shown by Eq. (7):

$$f(x;c,k) = \frac{k}{c}\left(\frac{x}{c}\right)^{k-1} \exp\left[-\left(\frac{x}{c}\right)^k\right], \tag{7}$$

where $x \geq 0$, k and c are the shape and scale parameters approximated using the maximum likelihood estimation method. WSM method obtains Weibull parameters of the long-term data of the site as follows:

$$k_{site}^{long} = \frac{k_{site}^{short}}{k_{MERRA}^{short}} \times k_{MERRA}^{long}, \tag{8}$$

and

$$c_{site}^{long} = \frac{c_{site}^{short}}{c_{MERRA}^{short}} \times c_{MERRA}^{long}, \tag{9}$$

where the superscript long/short refers to the long/short term wind speed data (Zhang et al. 2014).

3.2.2 The variance ratio method (VRM)

Another computational model is the variance ratio method that models the long-term wind speed of the target site using the linear relationship in Eq. (10).

$$W_{Site}^{long} = \mu_{W_{Site}^{short}} - \frac{\sigma_{W_{Site}^{short}}}{\sigma_{W_{MERRA}^{short}}}\mu_{W_{MERRA}^{short}} + \frac{\sigma_{W_{Site}^{short}}}{\sigma_{W_{MERRA}^{short}}} W_{MERRA}^{long}, \tag{10}$$

where $\mu_{W_{Site}^{short}}$, $\mu_{W_{MERRA}^{short}}$, $\sigma_{W_{Site}^{short}}$, and $\sigma_{W_{MERRA}^{short}}$ are the mean and variance of the target and re-analysis short-term data, respectively. When Eq. (10) generates negative wind speed, the null value is considered instead (Dinler 2013).

3.2.3 The hybrid model based on RBF neural network

Intelligent algorithms such as artificial neural network and support vector regression are implemented in several studies as MCP computational model. Generally, Intelligent algorithms are more accurate than WSM and VRM methods (Weekes and Tomlin 2014; Zhang et al. 2014). One of the artificial neural network methods is RBF neural network. RBF neural network is widely used in classification, signal processing, and modeling problems because of its simple topology, rapid convergence and high accuracy. One of the positive characteristics of RBF neural network is the possibility of using faster linear training algorithms for network training purposes. The higher speed of training enables the use of optimization algorithms to calculate neural network parameters to achieve higher accuracy (Jia et al. 2016). In this paper, a hybrid method based on RBF neural network is utilized, and Genetic Algorithm optimization is used to improve the RBF's performance.

RBF network is assumed to have a single output, the wind speed of a tower at time t ($W_R(t)$). In Sect. 3.2, it is discussed that the phase space of re-analysis wind speed data has the dimension and delay values of 9 and 20, respectively. The neural network's inputs are considered based on the phase space of re-analysis wind speed data, and hence we have considered a neural network with 9 inputs and one output. The input vector is represented as follows:

$$I(t) = \left[\ W_A(t)\ \ W_A(t-20)\ \ W_A(t-40)\ \ \ldots\ \ W_A(t-140)\ \ W_A(t-160)\ \right]. \tag{13}$$

RBF neural network's structure consists of three layers. The first layer includes input neurons. The second layer is a hidden layer transferring data from the input to the hidden space using radial basis function. The final layer is the output layer executing a linear transform. Figure 7 illustrates the structure of RBF neural network.

For the hidden neurons, Gaussian radial basis function is usually selected, which can be introduced as follows:

$$\varnothing_i(x) = \varnothing\left(\|x - c_i\|, \sigma_i\right) = \exp\left(\frac{\|x - c_i\|^2}{\sigma_i{}^2}\right), \tag{11}$$

where $\varnothing_i$ is the output of the *ith* neuron, x is the input vector, c is the center, and $\sigma$is spread of Gaussian radial basis function. In this paper, a single output network is considered, and the output of the neural network is calculated using Eq. (12).

$$y = \sum_{i=1}^{m} w_i \varnothing_i(x), \tag{12}$$

where m represents the number of neurons and w_i is the weight between the i$^{\text{th}}$ hidden neuron and the output neuron (Liu et al. 2016).

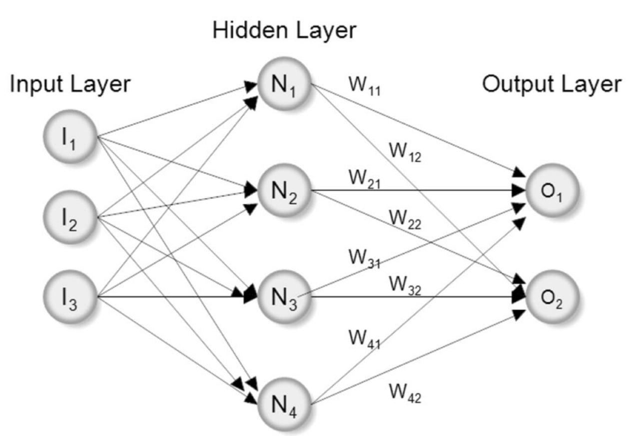

Fig. 7 Structure of RBF neural network

RBF network has several adjustable parameters required to be adjusted at training step. These parameters include c as the center, σspread as Gaussian radial basis function, the number of hidden neurons and the weight vector (w). In the literature, fuzzy and non-fuzzy clustering methods are frequently considered for obtaining centers of Gaussian radial basis function. In this approach, the input data are clustered into a specific number of clusters, and the center of each cluster is considered as the center of each radial basis function. In this paper, the K-means clustering method is used to determine the centers of Gaussian radial basis function. After the determination of the centers (c), the parameter of σ for the i^{th} neuron is calculated using the following equation:

$$\sigma_i = \varepsilon \times d_i, \tag{14}$$

where d_i is the minimum distance between the *ith* cluster and other clusters, and ε is the overlap coefficient (Jia et al. 2016; Yuqing et al. 2016). Another parameter required to be adjusted at the training step is the weight vector (w). Equation (15) is used to calculate the weight vector as follows:

$$w = O_T \times (O_H)^{-1}, \tag{15}$$

where O_T is the desired output, and O_H is the hidden layer output (Liu et al. 2016).

To improve the accuracy of RBF neural network, two parameters including the number of neurons in hidden layers and overlap coefficient (ε) are required to be determined. To achieve this purpose, the cross-validation method is employed. In this method, the data are divided into n sections. Sequentially, in each step, one of the sections is considered as the validation data, and the remaining n-1 sections are considered as the training data. Then, using the training data, the model is trained. The trained model is tested using the validation data, and the validation error is also calculated. The sum of validation errors is considered as objective function (Zhang et al. 2009). The optimum number of neurons in hidden layers and overlap coefficient ε is obtained by minimizing the objective function.

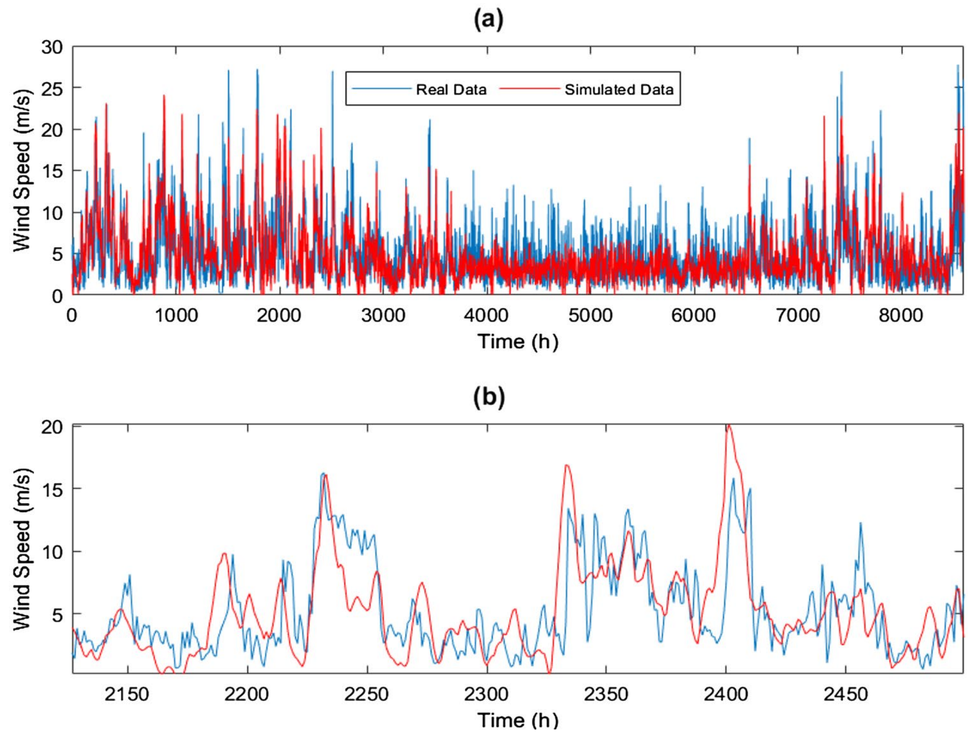

Fig. 8 **a** Results of wind speed data modeling using the proposed MCP method, **b** zoomed out of **a**

$$F = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m}\frac{|O_{validation}(i,j) - O_{real}(i,j)|}{O_{real}(i,j)}, \tag{16}$$

where $O_{validation}$ is the network output for validation data, O_{real} is the desired output, n is the number of sections ($n=5$) and m represents the number of data in each section. To optimize the objective function specified by Eq. (16), Genetic Algorithm with an initial population of 40, single point cross over function and 0.2 mutation rate is used (In 50 repetitions with validation data, the best accuracy of the proposed method was obtained with these values). The results achieved by Genetic Algorithm optimization indicate 53 as the number of optimum hidden neuron and the overlap coefficient of 4.4.

4 Results

In this paper, MCP method is used to generate a set of 24-h wind speed scenarios. For this purpose, a hybrid model based on RBF neural network is developed (detailed in the Sect. 3.2.3). In order to investigate the accuracy of the proposed method, at first, a comparison is made among MCP methods. After that, the results of scenario generation of wind speed by the proposed method and conventional statistical methods of scenario generation are investigated. The results are presented in the following two sub-sections.

Table 1 Results of wind speed data modeling using different computational models

	VRM	WSM	Neural Network (feed-forward)	Hybrid model
MSE	13.21	11.13	10.98	9.60
L_m	0.25	0.15	0.13	0.11

4.1 Comparison of the various MCP methods

In order to make a comparison among MCP methods to verify the accuracy of the proposed method, WSM and VRM algorithms along with the feed-forward neural network has been used to model the wind speed, and the obtained results are demonstrated in Table 1. To compare the results, two criteria including MSE (Mean Square Error) and L_m are used. The lower the value of the MSE, the lower the difference between the simulated and real data and the lower the value of L_m, the more similarity between dynamic behavior of the simulated and real data.

As shown in Table 1, MSE and L_m of the mentioned method have lower values compared to the conventional MCP methods, indicating the higher accuracy of this method. MSE and L_m in this method are 12.6% and 15.4% lower than those of the artificial feed-forward neural network with three hidden layers having 6, 8 and 6 neurons in each

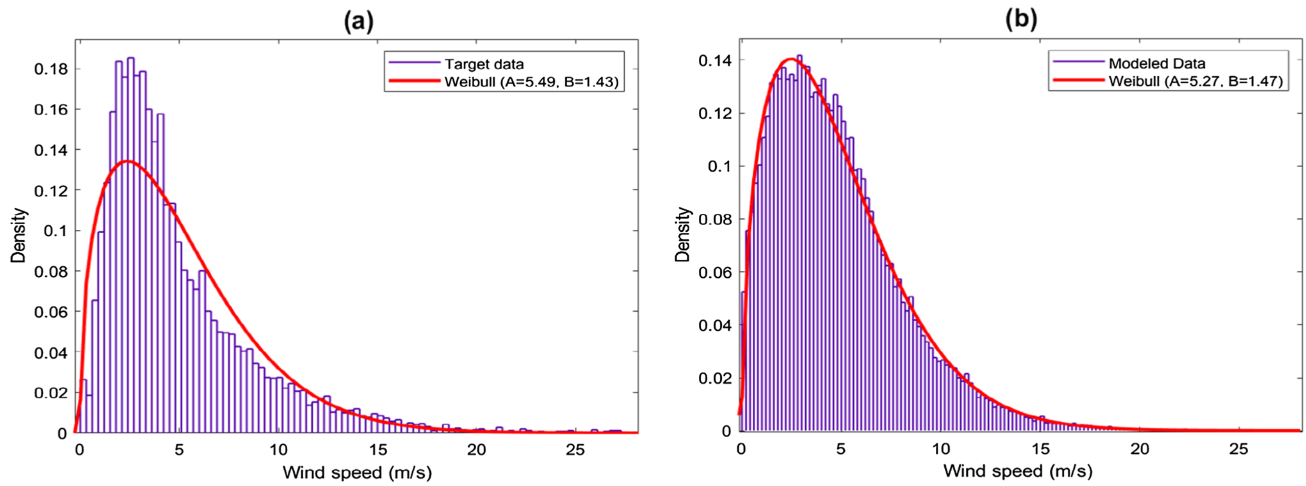

Fig. 9 Histogram and the probability distribution function of **a** target data, **b** modeled data by the proposed MCP model

layer, respectively (this arrangement has the best accuracy between 20 tested different arrangement). Figure 8 illustrates the results of wind speed data modeling of the hybrid model for one year.

On the other hand, in order to investigate the performance of the proposed MCP method to model the statistical elements of wind speed, Fig. 9 shows the histogram and the best-fit probability distribution function of the modeled data and target data.

A comparison of Figs. 2 and 9 shows that although there is a significant difference between the histogram of the reanalysis data and the target data, the proposed MCP is able to simulate the wind speed data with a histogram similar to the actual data.

5 Results of scenario generation of wind speed

In Sect. 4.1, it has been shown that the MCP method with the hybrid computational model has a higher accuracy than other MCP methods in wind speed modeling. In this section, the results of scenario generation of wind speed by MCP methods and conventional statistical methods are investigated. The conventional methods, suggested in several papers to generate wind speed scenarios (output power of a wind turbine), are as follows: Fuzzy-Cmeans clustering (De Caro et al. 2018), ARIMA (Conejo et al. 2010), Monte Carlo (Conejo et al. 2010), multivariate distribution function (Salehi Borujeni et al. 2017) and Copula Salehi Borujeni et al. 2018). In the methods, a statistical model is obtained using the available data (365 data for 24 h), and by the use of this model, more scenarios are generated. The selected methods are well-known and accurate methods in wind speed modeling which are able to model the dependence among wind speed data efficiently. In this paper, the conventional statistical methods are used to compare and contrast in terms of the accuracy of MCP method in generating wind speed scenarios. For each of these models, the tower wind speed data (target data) collected in 2011 is used and 5-year wind speed data, equivalent to the number of MCP scenarios, is generated based on each method. In this paper, after the generation of scenarios, the output power of the wind turbine is calculated, and then the number of scenarios is reduced to 10 by using of scenario forward selection. These scenarios and their probabilities indicate annual wind speed behavior. To compare the proposed scenario generation methods, Kantorovich distances factor is used. This index represents the distances among the generated scenarios and the original data. The lower its value, the greater the accuracy of the generated scenarios (Salehi Borujeni et al. 2017). Kantorovich distance is defined by Eq. (17):

$$KD\left(S, S'\right) = \frac{1}{T} \sum_{n=1}^{N} \sum_{t=1}^{T} p(n).\|S(n) - S'(t)\|, \tag{17}$$

where S is the generated scenarios, S' is the real data (in this paper, the output power of the wind turbine is calculated using the tower wind speed data from 2012 to 2016, and it is considered as test data). p is the probability of each scenario, N is the number of scenarios, T represents real data, and $\|.\|$ is an Euclidean norm (Ma et al. 2013). Table 2 represents the results from the scenario generation of the output power of the wind turbine.

As shown in Table 2, the mean and variance of the scenarios generated using the statistical methods are similar to each other, while they are different from the mean and variance of the output power of the wind turbine relating to the test data (data from 2012 to 2016). This cause of this difference is due to the fact that the generated scenarios are

Table 2 Results from scenario generation of wind turbine's output power

Methods	Average of Scenarios (kWh)	Average of Data (kWh)	Variance of Scenarios ($(kWh)^2$)	Variance of Data ($(kWh)^2$)	KD
PDF	0.217	0.177	0.105	0.091	1.91
ARIMA	0.217	0.177	0.106	0.091	1.74
Multivariate Gaussian	0.218	0.177	0.106	0.091	1.72
Copula	0.196	0.177	0.104	0.091	1.70
Fuzzy-Cmeans	0.198	0.177	0.096	0.091	1.65
VRM	0.212	0.177	0.094	0.091	1.69
WSM	0.160	0.177	0.081	0.091	1.66
Neural Network	0.162	0.177	0.084	0.091	1.62
Hybrid model	0.164	0.177	0.086	0.091	1.59

only based on the data of year 2011, and their mean and variance are similar to the corresponding values of the data in year 2011, which are 0.199 and 0.104, respectively. On the other hand, the mean and variance of the scenarios generated using MCP methods are so close to the real values. This state is achieved due to the consideration of the long-term wind speed data (from 2006 to 2010) in generating the output power of wind turbine scenarios. This is one of the advantages of MCP method over statistical methods. The next important issue required to be considered is KD factor. This factor represents the difference among the generated scenarios and the real data, and it is suitable for comparing the short-term behavior of wind speed (daily profiles). The lower the factor, the higher the accuracy achieved by the generated scenarios. The KD factor of the scenarios generated using PDF is 1.91, which is higher than those of all other methods. The low accuracy of scenario generation in this method is due to the fact that the dependence between the wind speed data is not modeled by the PDF. ARIMA method is another method that is frequently used to generate wind speed scenarios. In this method, the dependence among wind speed data are modeled efficiently. The results indicate that the accuracy of this method is higher than that of single-variate distribution function, and its KD factor is 9% lower. Another method considered for comparison purposes is the multivariate probability distribution function. This method is also able to model the correlated data, and it is more accurate than the two afore-mentioned methods. In statistical methods, the Copula function has the best accuracy to generate the wind speed scenarios. In addition to modeling the dependence between data, this method also has the ability to model tail dependence. Another method that has been used to generate the scenarios is Fuzzy-Cmeans clustering method. This method which belongs to the distance matching category, has a better performance than the statistical methods with KD = 1.65. But this method has not been very accurate in modeling long-term behavior. As shown in Table 2, MCP methods have a better accuracy than the conventional statistical methods. KDs of VRM and WSM are 1.69 and 1.66 respectively. The results show that the NN-MCP method with a KD of 1.62 has a better accuracy than last two methods. Nevertheless, as the results of Table 1 show the Hybrid-MCP has the highest accuracy among the MCP methods to model the wind speed, the results of Table 2 indicate it is also more accurate to generate wind speed scenarios. In addition, KD factor of the Hybrid-MCP is 7%, 4% lower than that of the conventional statistical methods and Distance matching method, proving its higher accuracy. One of the most important reasons representing the advantage of this method over statistical methods is the ability to model the seasonal behavior of wind speed. The statistical method is appropriate for stationary data and since the variations of the mean and variance of the wind speed is not constant over time, these methods are less accurate.

Now we can answer the following questions: whether the accuracy of MCP method's scenario generation is more than those of the conventional statistical methods for the output power of the wind turbine scenario generation?" By comparing the accuracy of the long-term behavior modeling employing the mean and variance of the short-term behavior using KD factor, it is indicated that the MCP method is able to provide an appropriate model for the short-term and long-term wind speed (the output power of the wind turbine). Advantages of MCP method for the wind speed scenario generation are as follows:

MCP method can be used for non-stationary data as well as stationary data.

The re-analysis data are available for free.

In models developed based on MCP method, the long-term behavior of wind speed is considered; therefore, the mean and variance obtained from this method are closer to the real values.

The short-term behavior is modeled by this method more efficiently lower KD factor), hence the accuracy of generated scenarios by this method is higher than those of the conventional statistical methods.

Thus, this work suggests that microgrid planners use the proposed method to model the wind speed uncertainty to improve and enhance the reliability of microgrid.

6 Conclusions

One of the most important uncertain parameters in microgrid planning is wind speed. The improved accuracy of the wind speed's scenario generation leads to better planning and greater reliability of the microgrid. In this paper, MCP method is used to generate wind speed scenarios. For this purpose, a hybrid method is presented based on RBF artificial neural network and Genetic Algorithm optimization. The results indicate that the generation of wind speed scenarios by MCP method is at least 7% more accurate than those of the conventional statistical methods including single and multivariate probability distribution function, ARIMA and Copula model. Moreover, the mean and variance of the scenarios generated by MCP method are closer to the real data. Higher accuracy of modeling and the absence of restrictions for MCP method indicate that microgrid planners are able to use this work's proposed method for microgrid planning.

Funding Not applicable.

Code availability The codes have been developed in MATLAB software and can be submitted to reviewers (if needed). CRP Toolbox has also been used to analyze wind data which is referenced in the text.

Compliance with ethical standards

Conflict of interest Not applicable.

Availability of data and material All data are available for free in the mentioned references.

References

Ambach D (2016) Short-term wind speed forecasting in Germany. J Appl Stat 43:351–369

Baringo L, Conejo AJ (2013) Correlated wind-power production and electric load scenarios for investment decisions. Appl Energy 101:475–482

Brower MC (2006) The use of NCEP/NCAR reanalysis data in MC. Proceedings of the European wind energy conference & exhibition, Athens, Greece, 27 February–2 March

Camala S, Tengb F, Michiorria A, Kariniotakisa G, Badesab L (2019) Scenario generation of aggregated wind, photovoltaics and small hydro production for power systems applications. Appl Energy 242:1396–1406

Carta JA, Velázquez S (2011) A new probabilistic method to estimate the long-term wind speed characteristics at a potential wind energy conversion site. Energy 36:2671–2685

Carta JA, Velázquez S, Cabrera P (2013) A review of measure-correlate-predict (MCP) methods used to estimate long-term wind characteristics at a target site. Renew Sustain Energy Rev 27:362–400

Chen J, Rabiti C (2017) Synthetic wind speed scenarios generation for probabilistic analysis of hybrid energy systems. Energy 120:507–517

Conejo AJ, Carrión M, Morales JM (2010) Decision making under uncertainty in electricity markets. Springer, New York

CRP toolbox for Matlab provided by TOCSY, last visited in September 2016. http://tocsy.agnld.uni-potsdam.de

De Caro F, Vaccaro A, Villacci D (2018) Adaptive wind generation modeling by fuzzy clustering of experimental data. Electronics 47:1–16

Deng J, Li H, Hu J, Liu Z (2020) A New Wind Speed Scenario Generation Method Based on Spatiotemporal Dependency Structure. Renewable Energy

Díaz S, Carta JA, Matias JM (2017) Comparison of several measure-correlate-predict models using support vector regression techniques to estimate wind power densities. A case study. Energy Convers Manag 140:334–354

Dinler A (2013) A new low-correlation MCP (measure-correlate-predict) method for wind energy forecasting. Energy 63:152–160

Esmaeeli M, Golshannavaz S, Siano P (2020) Determination of optimal reserve contribution of thermal units to afford the wind power uncertainty. J Ambient Intell Hum Comput 11:1565–1576

Global Modeling and Assimilation Office (GMAO) (2008) tavg1_2d_slv_Nx: MERRA 2D IAU Diagnostic, Single Level Meteorology, Time Average 1-hourly V5.2.0, Greenbelt, MD, USA, Goddard Earth Sciences Data and Information Services Center (GES DISC), last seen visited: January 2017). http://www.nasa.gov

Guan L, Wen B, Zhan X, Zhou B, Zhao W (2018) Scenario generation of wind power based on longitudinal-horizontal clustering strategy. IEEE Innovative Smart Grid Technologies - Asia (ISGT Asia) pp 934–939

Heitsch H (2003) Omisch WR (2003) Scenario reduction algorithms in stochastic programming. Comput Optim Appl 24:187–206

Jager D, Andreas A (1996) NREL National Wind Technology Center (NWTC): M2 Tower; Boulder, Colorado (Data); NREL Report No. DA-5500–56489. http://www.nrel.gov/midc/nwtc_m2. Last visited in January 2017

Jia W, Zhao D, Ding L (2016) An optimized RBF neural network algorithm based on partial least squares and genetic algorithm for classification of small sample. Appl Soft Comput 48:373–384

Kanase-Patil AB, Saini RP, Sharma MP (2011) Sizing of integrated renewable energy system based on load profiles and reliability index for the state of Uttarakhand in India. Renewable Energy 36:209–221

Kaplani E, Kaplanis S (2012) A stochastic simulation model for reliable PV system sizing providing for solar radiation fluctuations. Appl Energy 97:970–981

Kennel MB, Brown R, Abarbanel HDI (1992) Determining embedding dimension for phase-space reconstruction using a geometrical construction. Phys Rev A 45:3403–3411

Kim JY, Kim KY (2016) Reconstructing long-term wind data at an offshore met-mast location using cyclostationary empirical orthogonal functions. J Wind Eng Ind Aerodyn 156:146–158

Li J, Zhou J, Chen B (2020) Review of wind power scenario generation methods for optimal operation of renewable energy systems. Appl Energy 280:115992

Liu Y, Huang H, Huang T, Qian X (2016) An improved maximum spread algorithm with application to complex-valued RBF neural networks. Neurocomputing 216:261–267

Liu Z, Hajiali M, Torabi A, Ahmadi B, Simoes R (2018) Novel forecasting model based on improved wavelet transform, informative feature selection, and hybrid support vector machine on wind power forecasting. J Ambient Intell Hum Comput 9:1919–1931

Ma XY, Sun YZ, Fang HL (2013) Scenario generation of wind power based on statistical uncertainty and variability. IEEE Trans Sustain Energy 4:894–904

Marwan N, Kurths J (2002) Nonlinear analysis of bivariate data with cross recurrence plots. Phys Lett A 302:299–307

Marwan N, Romano MC, Thiel M, Kurths J (2007) Recurrence plots for the analysis of complex systems. Phys Rep 438:237–329

Niknam T, Massrur HR, Bahmani Firouzi B (2012) Stochastic generation scheduling considering wind power generators. J Renew Sustain Energy 4:063119

Perea AR, Amezcua J, Probs O (2011) Validation of three new measure-correlate-predict models for the long-term prospection of the wind resource. J Renew Sustain Energy 3:023105

Petersen EL (2017) In search of the wind energy potential. J Renew Sustain Energy 9:052301

Pinson P, Madsen H, Nielsen HA, Papaefthymiou G, Klockl B (2009) From probabilistic forecasts to statistical scenarios of short-term wind power production. Wind Energy 12:51–62

Probst O, Cardenas D (2010) State of the art and trends in wind resource assessment. Energies 3:1087–1141

Salehi Borujeni M, Akbari Foroud A, Dideban A (2017) Accurate modeling of uncertainties based on their dynamics analysis in microgrid planning. Sol Energy 155:419–433

Salehi Borujeni M, Akbari Foroud A, Dideban A (2018) Wind speed scenario generation based on dependency structure analysis. J Wind Eng Ind Aerodyn 172:453–465

Sengar S, Liu X (2020) Ensemble approach for short term load forecasting in wind energy system using hybrid algorithm. IOP Publishing journal of ambient intelligence and humanized computing

Sharma PK, Warudkar V, Ahmed S (2019) Application of lidar and measure correlate predict method in offshore wind resource assessments. J Clean Product 215:534–543

Tabar VS, Ahmadi Jirdehi M, Hemmati R (2017) Energy management in microgrid based on the multi objective stochastic programming incorporating portable renewable energy resource as demand response option. Energy 118:827–839

Weekes SM, Tomlin AS (2014) Data efficient measure-correlate-predict approaches to wind resource assessment for small-scale wind energy. Renew Energy 63:162–171

Weekes SM, Tomlin AS, Vosper SB, Skea AK, Gallani ML, Standen JJ (2015) Long-term wind resource assessment for small and medium-scale turbines using operational forecast data and measure-correlate-predict. Renew Energy 81:760–769

Yuqing S, Junfei Q, Honggui H (2016) Structure Design for RBF neural network based on improved K-means algorithm. In: Control and decision conference (CCDC)

Zhang W, Li C, Zhong B (2009) LSSVM parameters optimizing and non-linear system prediction based on cross validation, Fifth International Conference on Natural Computation

Zhang J, Chowdhury S, Messac A, Hodge BM (2014) A hybrid measure-correlate-predict method for long-term wind condition assessment. Energy Convers Manag 87:697–710

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.