

## Dynamic characterization and predictability analysis of wind speed and wind power time series in Spain wind farm

N. Bigdeli\* and H. Sadegh Lafmejani

EE Department, Imam Khomeini International University, Qazvin, Iran.

Received 25 April 2014; Accepted 15 September 2015

\*Corresponding author: nooshin\_bigdeli@yahoo.com (N. Bigdeli).

### Abstract

The renewable energy resources such as wind power have recently attracted more researchers' attention. It is mainly due to the aggressive energy consumption, high pollution and cost of fossil fuels. In this era, the future fluctuations of these time series should be predicted to increase the reliability of the power network. In this paper, the dynamic characteristics and short-term predictability of hourly wind speed and power time series are investigated via nonlinear time series analysis methods such as power spectral density analysis, time series histogram, phase space reconstruction, the slope of integral sums, the  $\delta - \varepsilon$  method, the recurrence plot and the recurrence quantification analysis. Moreover, the interactive behavior of the wind speed and wind power time series is studied via the cross correlation, the cross and joint recurrence plots as well as the cross and joint recurrence quantification analyses. The results imply stochastic nature of these time series. Besides, a measure of the short-term mimic predictability of the wind speed and the underlying wind power has been derived for the experimental data of Spain's wind farm.

**Keywords:** *Stochastic Behavior, Recurrence Plot, Recurrence Quantification Analysis, Time Series Analysis, Wind Speed, Wind Power.*

### 1. Introduction

Wind energy is a free, renewable resource of clean energy. Compared with the conventional power plants, wind plants emit no air pollution or greenhouse gases. In fact, wind-based generation is the fastest growing source of renewable energy [1]. However, despite significant environmental benefits, wind power could be highly fluctuating because of the earth's natural atmospheric variability [1]. This variability can put at risk the power system reliability, which in turn requires more backup than the conventional generation in the form of reserve and regulation services. It also poses economical risks for wind farm owners, especially in competitive electricity markets [1]. To fully benefit from a large fraction of wind energy in an electrical grid, it is therefore necessary to predict the electrical energy generated by the wind [2]. The wind production depends on wind speed of the station. Therefore, forecasting the wind speed is an important issue that has increasingly received attentions of many researchers as well [2-7]. In order to forecast the

wind power production, two major questions should be answered. Those are: what is the nature of the wind speed/power time series dynamics and how predictable these time series are? The developed models for wind speed and power dynamics are derived inevitably from these time series. Therefore, one should characterize the nature and predictability of the recorded time series of the wind speed and the related wind power to find the proper model structure as well as the model inputs and to be able to claim the validity of its model.

Based on the above discussions in this paper, we are looking for the nature of the fluctuations of the recorded wind speed and power data to find out whether the fluctuations are from stochastic systems or not. If not, it maybe from a highly nonlinear deterministic or a chaotic systems. Chaotic behavior has been reported in a broad range of scientific disciplines including astronomy, biology, chemistry, ecology, engineering, and physics [8-9] as well as power

market indices [10-12]. On the other hand, stochastic time series contains a collection of random variables and is the probabilistic counterpart to a deterministic system. Some examples of stochastic time series are stock market and exchange rate fluctuations, speech, audio and video signals and medical data such as patient's EEG [12]. Time series analysis is commonly used for responding to the first mentioned question. Different time series approaches have been widely used in the literature for characterizing the properties of natural phenomena [13-15]. In [16], the chaotic behavior of wind time series for averaged weekly wind time series of wind speed has been shown by means of time series analysis tools. Next, in [17], the fractional of this time series has been derived. In [18], a comparative study of different hybrid prediction models has been performed by the authors to forecast the wind power time series from Alberta, Canada. In this study, the time series has been analyzed, briefly based on recurrence plots and correlation analysis to select the proper input sets for the forecasting models. The results of this paper is accordance with the observations in [18] and justifies them, but is more compromising. However, up to authors' knowledge, further study has been done on the characterization and predictability analysis of neither the wind speed nor the wind power. Besides, a question has been remained unanswered for prediction of wind power, that is, how predictable is the wind power in terms of its own time series and in terms of wind speed time series, or both?

In order to answer the above questions, in this paper, at first the experimental wind speed and wind power data from Spain's wind farm has been investigated via different time series analysis methods. The examined methods are power spectral density (PSD) analysis, phase space reconstruction and test of surrogates, test of the fractional dimension and the slope of integral sums, the  $\delta$ - $\epsilon$  method, and the recurrence plots [19]. The results of these analyses not only emphasize the stochastic nature of these time series with a mimic predictability but also show that there exists some type of the seasonality and non-stationarity in the system dynamics. This result implies that a fixed model cannot perform properly, even in case of de-trended input data, and so, multiple or adaptive models should be developed to forecast such time series.

Further, the individual, cross and joint recurrence quantification analyses have been applied to the wind speed and wind power time series' to

investigate the individual and interactive predictability of them. In this context, the degree of synchronization between wind speed and wind power time series dynamics, and the mean and maximum range of validity for the prediction models built based on individual/joint inputs of these time series' will be discussed thoroughly.

The remainder of the paper is organized as follows: In section 2, the experimental data and its time series analysis is provided. In section 3, the interaction analysis of wind speed and wind power time series is performed. The paper is concluded in section 4.

## 2. The experimental data and time series analysis

Seeking for the nature of the wind time series, in this section, the experimental wind data of Spain's wind farm on May 2005 [5] will be closely studied. The available data that are the hourly wind speed and its related wind power time series for 5 weeks (1200 hours) as shown in figure 1; where figure 1(a) shows the wind speed time series and figure 1(b) shows the related wind power of the interested period versus time in hours. According to figures 1 (a and b), no obvious hallmark of periodicity (such as daily or weekly periodicity) can be observed either in the wind speed or in the wind power time series. These time series behavior may come from both chaotic and stochastic dynamics. Besides, the stationarity of it is under question which should be investigated. From one point of view, wind speed and wind power can be considered as system parameter and system operating point, respectively. Based on this interpretation, the wind power values versus wind speed have been plotted in figure 2. From this figure, with increasing the wind speed, some type of bifurcation occurs in the corresponding wind power values. This bifurcation diagram may, however, be due to a chaotic or stochastic nature of a nonlinear process [20-21]. To distinguish between these two types of processes, several methods such as power spectrum analysis, phase space reconstruction, surrogate testing, test of the fractional dimension and the slope of integral sums, the  $\delta$ - $\epsilon$  method, and recurrence plots will be performed using the TISEAN package [22] and CRP toolbox of MATLAB [23] as our tools. Besides, the stationarity of the dynamics will be analyzed via recurrence plots.

### 2.1. The power spectral density

Regarding the behavior of wind speed and wind power in figures 1 and 2, what is important for us,

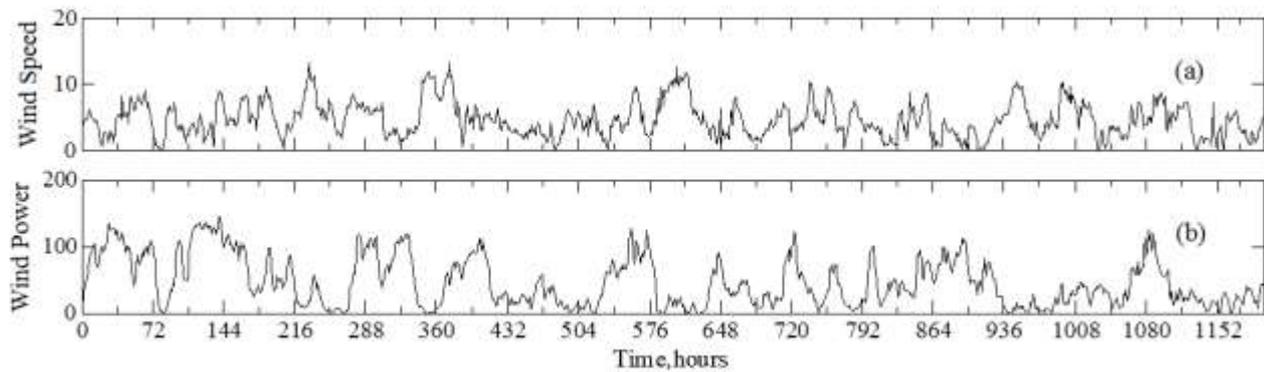


Figure 1. (a) The experimental wind speed time series, and (b) wind power time series.

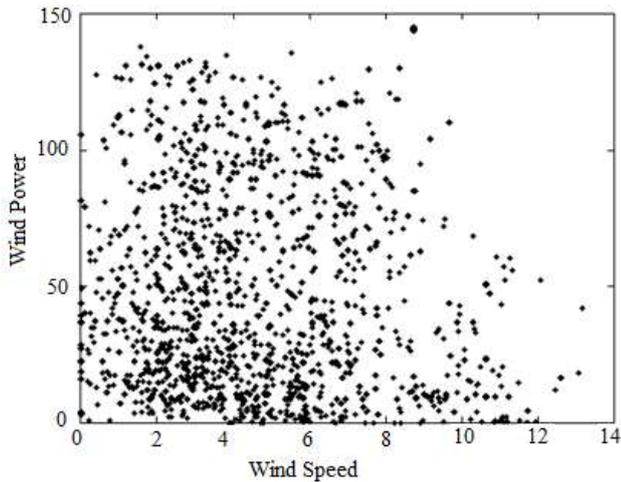


Figure 2. Bifurcation diagram of wind speed versus windpower.

is the answer to this question: “what is the nature of such complex behavior of wind, deterministic chaos or stochastic nature?” To answer this question, first of all, we analyze the power spectral density (PSD) of the data as shown in figure 3. The power spectral density (PSD) of wind speed and power provides information on the character of fluctuations in the time series data. The (PSD) describes how the power of a time series is distributed with frequency. The graph has been derived based on periodogram PSD estimation method. As expected, in figures 3 (a) and (b), there are no regular sharp peaks which is the representative of aperiodic nature of wind speed and wind power time series. Lack of periodic components in these time series implies low predictability of these time series as it would be examined in the next sections. To analyze the results of figure 3 more closely, one should note that existence of the higher harmonics in the spectra indicates that the processes underlying the time series are not linear processes, but there is some kind of nonlinearity [21, 24]. As another point, consider the frequency content of the plots. A broadband dense spectrum which also preserves these properties in small frequency ranges is often

considered as hallmark of chaos. Spectrum of a chaotic system is not solely comprised of discrete frequencies, but has a continuous broadband nature [25]. In case of our data, such a broadband spectrum preserving its properties in small frequency ranges is not observed (See Figures 3 (c) and (d)). These observations imply the stochastic nature of wind speed and wind power. However, since such a type of broadband spectrum may be due to either chaotic or stochastic nature of a time series, some stronger tests should be carried out to distinguish strictly between these two types of dynamics.

## 2.2. The histogram analysis

Histogram is a graphical representation that shows the distribution of data. It consists of tabular frequencies shown with discrete intervals by adjacent rectangles. The total area of the histogram is equal to the number of data. Next, the experimental probability density function (PDF) of a data can be derived when the y-axis values of the histogram (or number of data in each histogram bin) is normalized by the total number of data. The experimental PDF graphs of the wind speed and wind power data have been illustrated in figure 4. In order to investigate the nature of the wind speed and wind power dynamics, it has been tried to fit a proper random distribution to the experimental PDF via the statistics toolbox of MATLAB [23]. Also, it is observed that the Weibull distribution described by (1) is well fitted to the histogram of the data. That is:

$$f(x|a,b) = \frac{b}{a} \left(\frac{x}{a}\right)^{b-1} e^{-\left(\frac{x}{a}\right)^b} \quad (1)$$

where,  $f(\cdot)$  is the Weibull PDF and  $a, b$  are its corresponding parameters. Trying to fit the Weibull PDF to the wind speed and wind power time series, the following parameter are found: For wind speed:  $a_1=5.1040, b_1=1.7057$ ; for wind power  $a_2=48.9156, b_2=1.1341$ . The well matching of the found PDF's with the experimental PDF of the wind speed and wind power time series may be seen in figures 4(a) and (b), respectively.

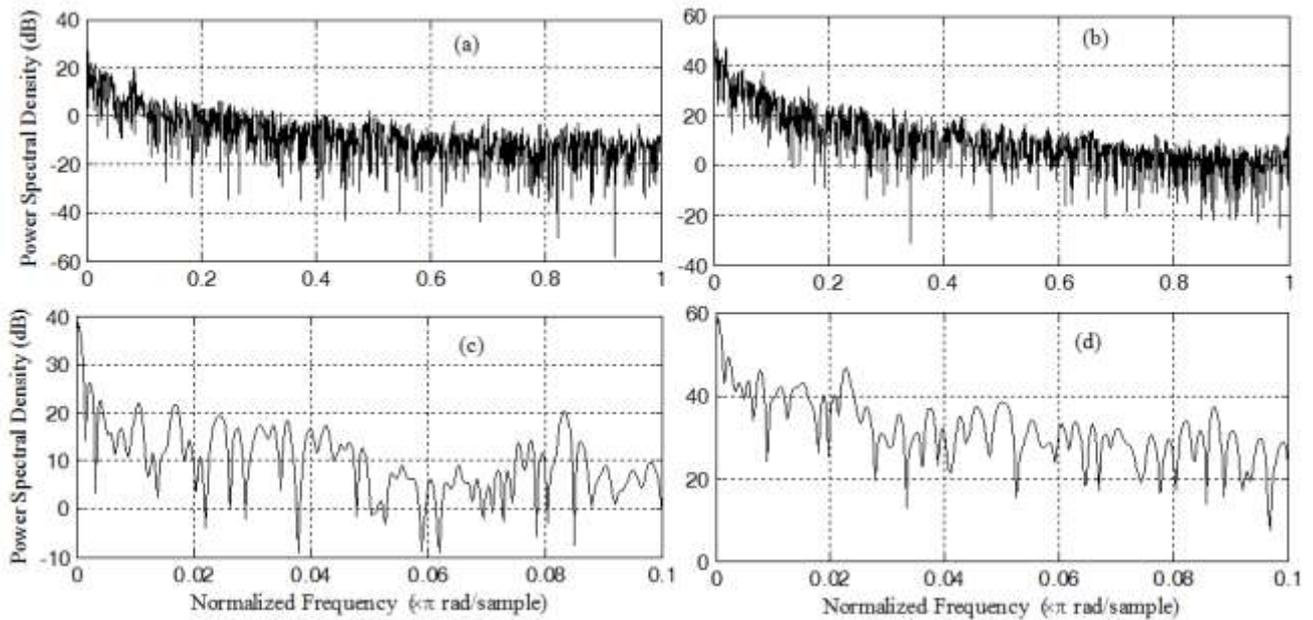


Figure 3. Power spectral density for (a) The wind speed, (b) The wind power, (c) and (d) Figs. 3(a) and 3(b) zoomed out.

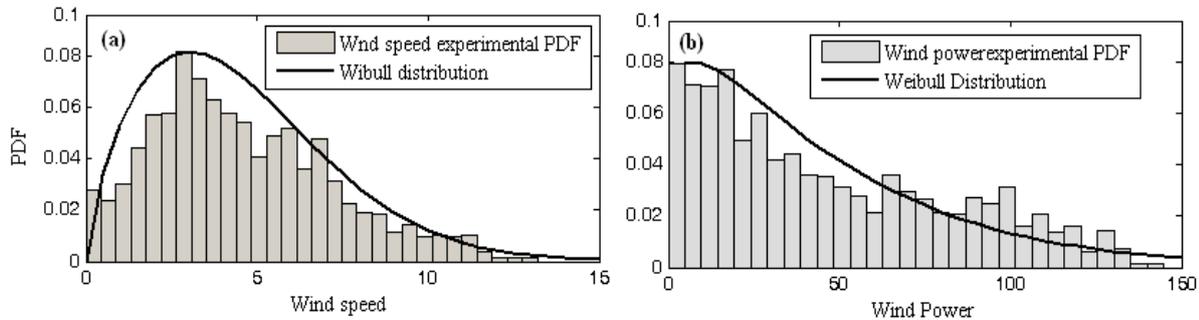


Figure 4. Experimental PDF graphs of (a) wind speed, (b) wind power in comparison with the fitted Weibull distributions.

This observation is another hallmark of stochastic behavior of the wind speed and wind power time series.

### 2.3. Embedding delay determination

The first step in analyzing a time series is to reconstruct its embedded phase space via methods of delays [22]. In this method, vectors in a new space, *the embedding space*, are formed from time delayed values of the scalar measurements [22]. For implementing this method, two important parameters should be determined firstly: the embedding delay and embedding dimension. In order to find out the embedding delay, two methods may be employed. In the first approach, the first zero cross or first cutoff (corresponding to 95% confidence level) of the autocorrelation function (ACF) is the embedding delay [26]. In the second approach, the first minimum of the average mutual information (MI) is the embedding delay [26]. In this paper, both approaches have been employed to the data and the results are illustrated in figure 5. For these time series, the first approach is inconclusive

because the first zero cross of ACF plot doesn't reveal an integer embedding delay as shown in figures 5 (a) and (c). According to figure 5(b), the first minimum is occurred at 6, so the embedding delay for the wind speed is  $\tau = 8$  hours and from figure 5(d), the first minimum is occurred at 8, so for the wind power, the embedding delay is  $\tau = 12$  hours. These embedded delays will be used in recurrence plots and recurrence quantification analysis.

### 2.4. Embedding dimension determination

As mentioned before, embedding dimension determination is the other parameter to be determined for phase space reconstruction. The dimension, where a time delay reconstruction of the system phase space provides a necessary number of coordinates to unfold the dynamics from overlaps on itself caused by projection, is called the embedding dimension [27]. A common method to determine the minimal sufficient embedding dimension  $m$  is the false nearest neighbor (FNN) method proposed by Kennel et al. [27].

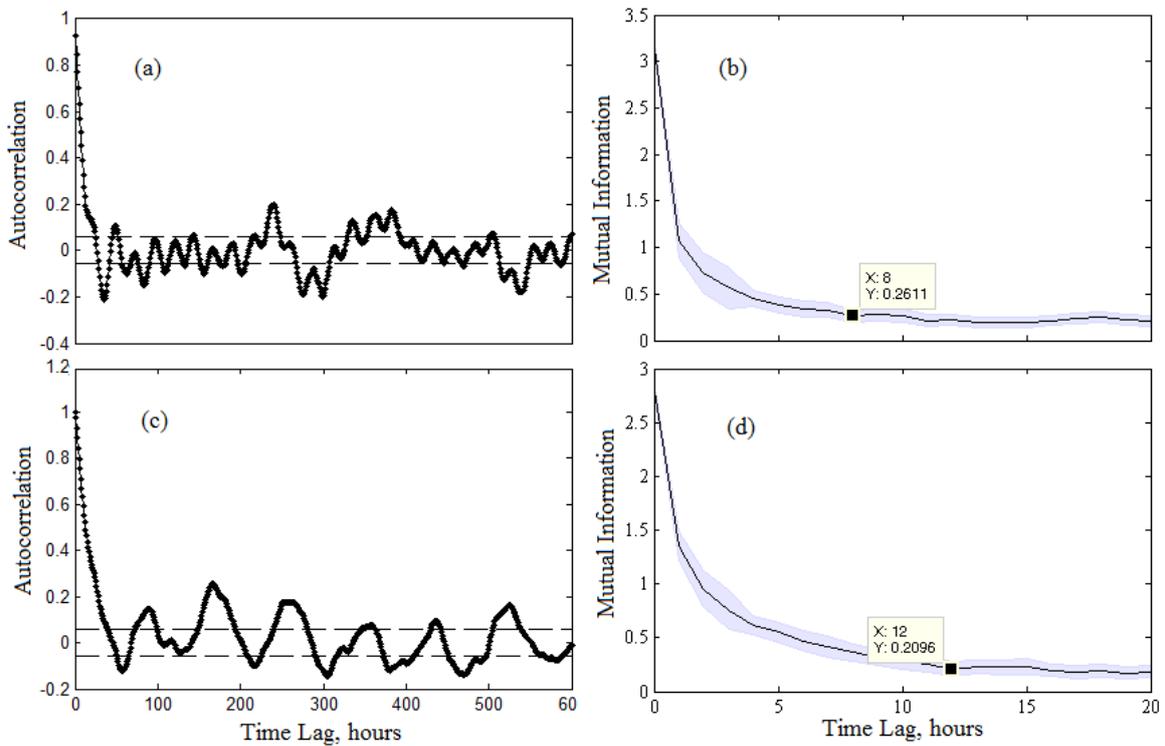


Figure 5. (a) and (b) Wind speed autocorrelation and average mutual information, (c) and (d) Wind power autocorrelation and average mutual information.

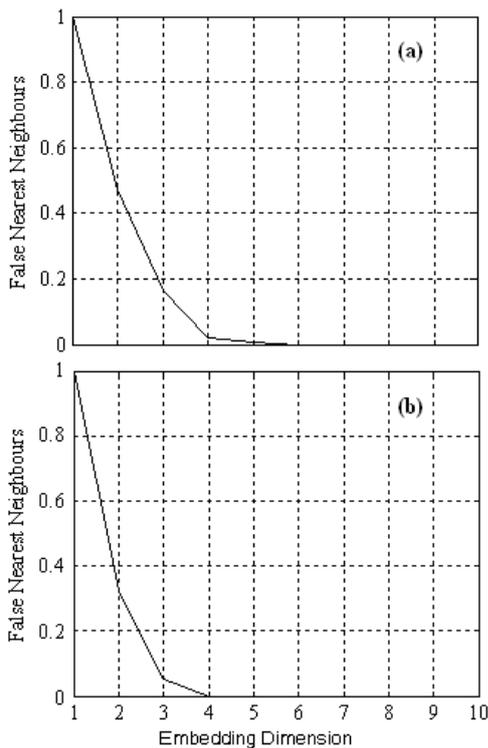


Figure 6. The fraction of false nearest neighbors versus the embedding dimension for (a) Wind speed, (b) Wind power time series.

The detailed description of this algorithm is at [17]. Figure 6 shows the fraction of false nearest neighbors for the considered wind speed and wind power time series. As we see in figure 6, the fraction falls down to zero at  $m = 6$  for wind speed

and  $m = 4$  for wind power. Therefore, the embedding dimension is 6 and 4 for the wind speed and wind power time series, respectively.

### 2.5. The phase space reconstruction and test of Surrogates

Mapping time series data into a phase space allows one to view the temporal series in a spatial manner. The distinguishing feature of chaotic processes is their sensitive dependence on initial conditions and highly irregular behavior that makes prediction difficult except in short term. This feature is the so-called “strange attractors” associated with chaotic processes which often have a complex, fractal structure [27]. Based on these properties, one of the most interesting procedures for checking the presence of chaos is based on the ability of recovering the strange attractor of a system in the phase space and especially observation of the so-called butterfly effect [27]. The three-dimensional phase spaces of the wind speed data (*WS*) with a delay time of  $\tau = 8$  hours, and that of wind power (*WP*) with a delay time of  $\tau = 12$  have been shown in figures 7(a) and (b), respectively. As seen, the strange attractor and the butterfly effect are not observed in the reconstructed phase space. Instead, a random like trajectory is demonstrated in the graphs. To develop a better appreciation whether the data set is chaotic or stochastic, one can comparatively assess the phase space maps of the

original data with their corresponding surrogated data sets. Surrogate data sets have Fourier decompositions with the same amplitude of the original data set but with random phase components. The method of surrogate data serves as a null hypothesis whose objective is to examine the hypothesis that the original data have come from a random process [8, 28-30]. The method may be used as a reference for visual comparison between the original and random data sets' phase

space. These two phase spaces would not look like if the data is a randomly spreading out cloud [8]. The phase space map of the surrogated wind speed and wind power data are shown in figure 8. As it is observed, the surrogated phase space maps do not look like the original ones. Besides, not strange attractor or butterfly effect is observed in figures 8 (a) and (b). This observation emphasizes the stochastic nature of our data once again.

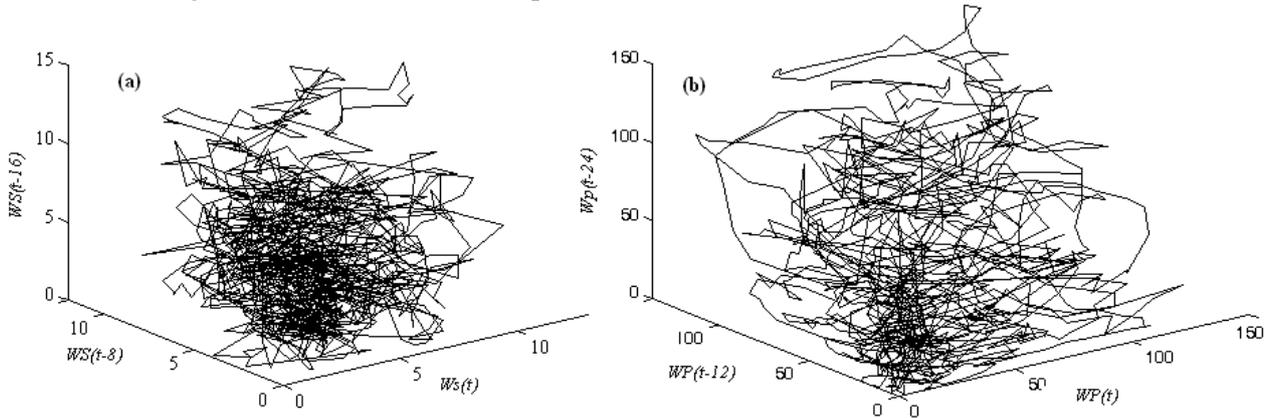


Figure 7. The phase space plot of (a) the wind speed time series ( $\tau = 8$ ), (b) the wind power time series ( $\tau = 12$ ).

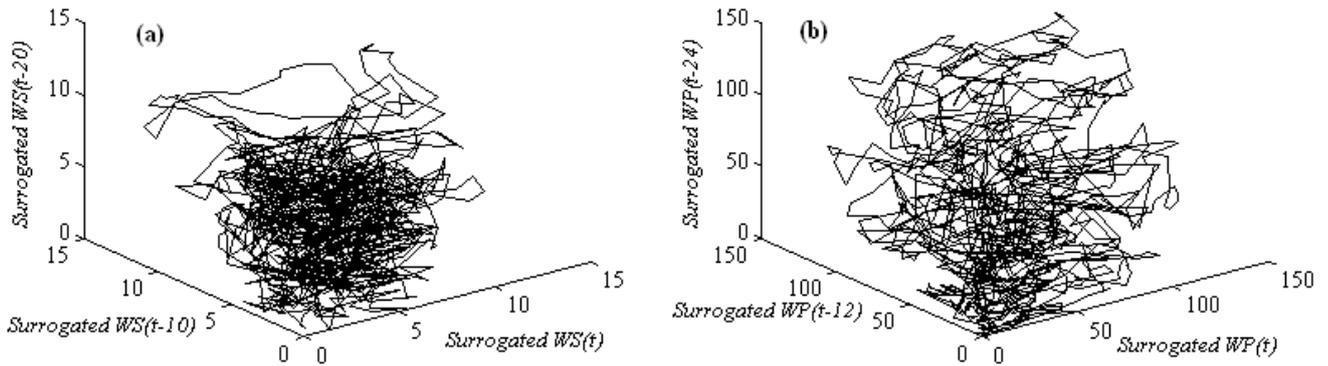


Figure 8. (a) The phase space plot of the surrogated the wind speed ( $\tau = 10$ ); and (b) that of the wind power ( $\tau = 12$ ).

### 2.6. The slopes of local integral sums

One of the fascinating features of chaotic systems is the fractal dimension of the attractor. Correlation dimension  $D_2$  [8] is a measure for fractal dimension, which is formulated below:

$$D_2 = \lim_{r \rightarrow 0} \frac{d \log C(m, r)}{d \log(r)} \quad (2)$$

where,  $C(m, r)$  is defined by:

$$C(m, r) = \text{constant} \times \sum_{i=1}^N \sum_{j=i+\mu}^N \Theta(r - |\vec{x}(i) - \vec{x}(j)|) \quad (3)$$

where,  $\vec{x}(i), \vec{x}(j)$  denote states embedded in reconstructed phase space with embedding dimension  $m$  [31].

$\Theta(\cdot)$  is the Heaviside step function (which is 1 for a positive argument and 0 elsewhere) applied to count the number of pair of points within radius

$\mu$  is the Theiler correction employed to exclude temporally correlated points [28].

Seeking for fractal attractors, the plots of local slopes of the logarithm of the correlation integrals with respect to the logarithm of  $r$  can be investigated. Evidence for a fractal attractor is given if the local slopes are constant for a large enough range of small radii but does not change for embedding dimension higher than a minimal value [20]. In the local slopes plot, if the system has the chaotic behavior, by increasing the embedding dimension, the curve will be saturated and the level at which most of the curves settle down defines the fractal dimension [22]. Figures 9(a) and (b) show the local slopes of the correlation integral versus  $r$  for wind speed and wind power time series data, respectively. The embedding delay used for these plots is  $\tau = 8$  and

$\tau = 12$  for wind speed and wind power data, respectively. Embedding dimension varies from 1 to 10 from bottom to top of the graphs. As seen in figure 9, there is no indication of a low-dimensional fractal attractor in the graphs. This is another hallmark of stochastic behavior of the wind speed and wind power dynamics.

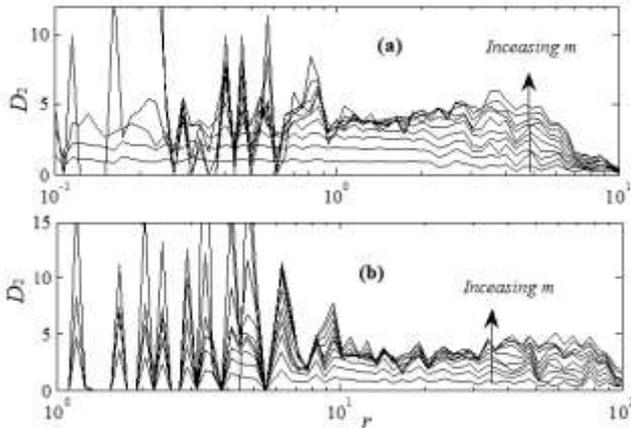


Figure 9. Local slopes of the correlation integrals, for  $m = 1, \dots, 10$ , for (a) Wind speed and (b) Wind power versus  $r$ .

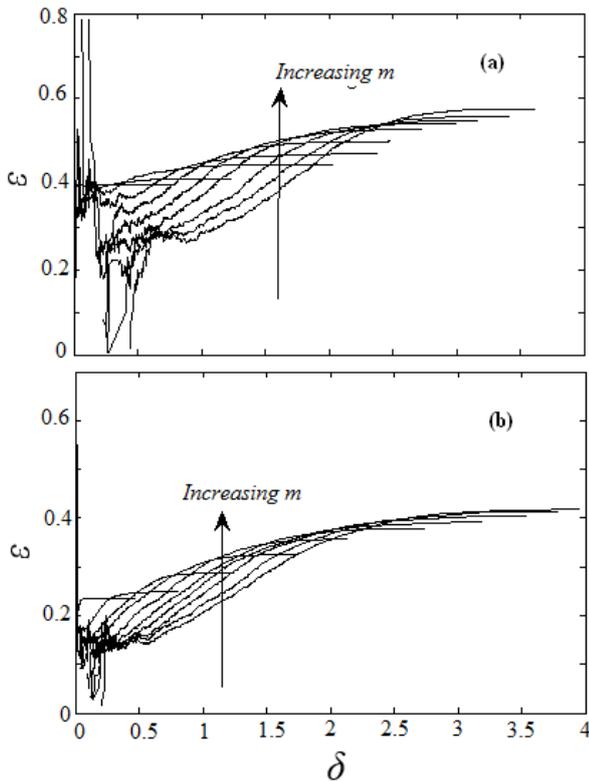


Figure 10. Results for the  $\delta$ - $\epsilon$  method for (a) wind speed and (b) wind power.

### 2.7. The $\delta - \epsilon$ method

The  $\delta - \epsilon$  method was first introduced in [32], where a detailed discussion of the method is given. According to this method, a deterministic dynamics embedded in a sufficiently high-dimensional state space should induce a continuous mapping from past to present states

and the size of the neighborhoods is increased to investigate the continuity. For deterministic processes,  $\epsilon$  is expected to decrease to zero for decreasing  $\delta$  for sufficiently high embedding dimensions [26], but for stochastic processes and processes which are covered by a significant amount of additive observational noise, a non-zero intercept for  $\epsilon$  is expected. Figures 10 (a) and (b) show the results of this method for the wind speed and the wind power time series. This method has been employed in the wind speed and power time series by  $\tau = 8$  and  $\tau = 12$  respectively. The embedding dimensions vary from 1 to 10 from top to bottom of the graphs. As seen in figures 10 (a) and (b), for small  $\delta$ ,  $\epsilon$  does not tend to zero. These results emphasize the stochastic nature of the both wind speed and wind power dynamics.

### 2.8. Recurrence Plots (RP)

Since its introduction by Eckman and Ruelle the recurrence plot has emerged as a useful tool in the analysis of nonlinear, non-stationary time series and useful for finding hidden correlations in highly complicated data [19]. With RP, one can graphically detect hidden patterns and structural changes in data or see similarities in patterns across the time series under study [12]. The RPs exhibits characteristic large scale and small scale patterns. Large scale patterns can be characterized as homogeneous, periodic, drift and disrupted, that obtain the global behavior of the system (noisy, periodic, auto-correlated, etc.) [19]. The RP is derived directly from the distance matrix  $D = D_{i,j}, i, j = 1, \dots, N$  ( $N$  is the length of the data series or trajectory):

$$D_{i,j} = \|\vec{x}_i - \vec{x}_j\| \quad (4)$$

By applying a threshold  $\epsilon$ :

$$R_{i,j} = \Theta(\epsilon - D_{i,j}) \quad (5)$$

where,  $\Theta$  is the Heaviside function. And,

$$\Theta(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } x \geq 0 \end{cases} \quad (6)$$

And if  $\vec{x}_i \approx \vec{x}_j$  then  $R_{i,j} = 1$  if not, then  $R_{i,j} = 0$ . One assigns a “black” dot to the value one and a “white” dot to the value zero. The two-dimensional graphical representation of  $R_{i,j}$  and then it is called RP [19]. The visual inspection of RPs reveals (among other things) the following typical small scale structures: single dots, diagonal lines as well as vertical and horizontal lines [19, 33]; in addition, even bowed lines may occur [19]. Single, isolated recurrence points can

occur if states are rare, if they do not persist for any time, or if they fluctuate heavily [33]. A diagonal line occurs when the trajectory visits the same region of the phase space at different times [33]. RP gives the reader a first impression of the patterns of recurrences which will allow studying dynamical systems and their trajectories like periodic systems, stochastic random systems, and chaotic ones [19]. Long and non-interrupted diagonals are related to periodic motion and the period of oscillation is equal to the vertical distance between these lines. If the diagonals are shorter, it seems that the RP is related to the chaotic systems. The RP with so many single black points with erratic distribution is related to the uncorrelated stochastic signal. All together, we find that the shorter the diagonals are in the RP, the less predictability the system has [12].

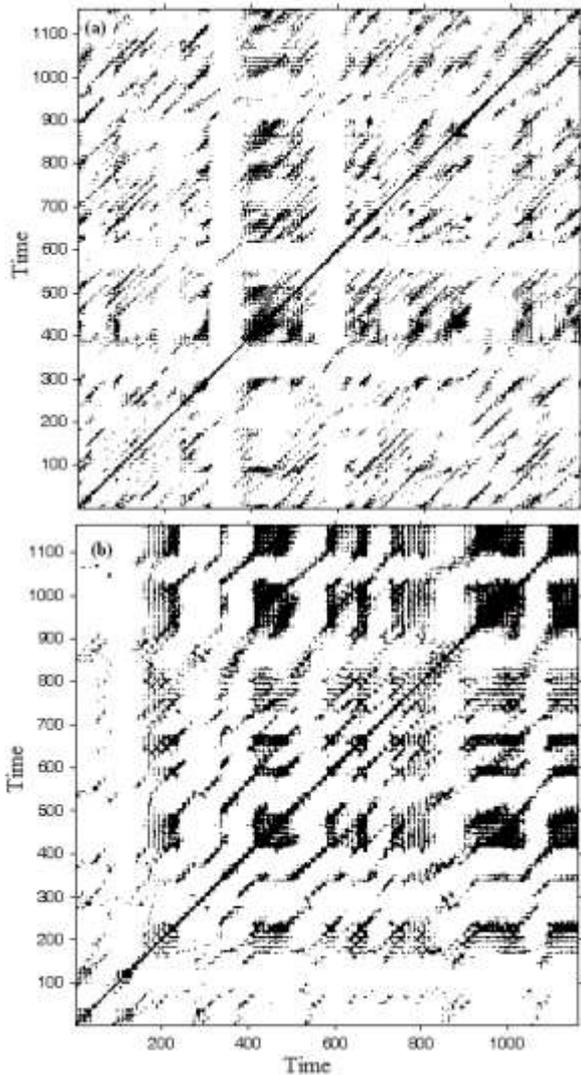


Figure 11. (a) RP of wind speed time series, (b) RP of wind power time series.

The RP of the wind speed and wind power time series has been shown in figures 11 (a) and (b), respectively. The figure has been generated via

the CRP toolbox of MATLAB [34]. In order to plot the RP, the parameter  $m$  has been considered 6, 4 for wind speed and wind power, respectively, and the parameter  $\tau$  has been assumed 8 for wind speed and 12 for wind power. Besides, the Euclidean norm and the threshold of 1.5 have been assumed for plotting the recurrence plots. According to figure 11(a), very short diagonals in wind speed RP imply stochastic nature of wind speed. Moreover, the presence of distributed dark regions as well as white ribbons is indicative of seasonality and unstationarity in the hourly wind speed time series. Altogether, very short-term predictability is concluded from the RP of wind speed. According to figure 11(b), the same properties is observed in the RP of wind power unless the density of dark points is more in some regions with respect to that of the wind speed. Via the RP, the stochastic nature, seasonality and unstationarity is emphasized for wind power as well, but it seems due to its correlated dynamics, the predictability is enhanced here.

### 2.9. Recurrence quantification analysis (RQA)

In order to go beyond the visual impression yielded by RPs, several measures of complexity which quantify the small scale structures in RPs have been proposed known as recurrence quantification analysis (RQA) [19]. These measures are based on the recurrence point density and the diagonal and vertical line structures of the RP [12]. The simplest RP measure is the recurrence rate ( $RR$ ) and it is defined as the percentage of dark pixels in recurrence plots by:

$$RR(\varepsilon) = \frac{1}{N^2} \sum_{i,j=1}^N R_{i,j}(\varepsilon) \times 100 \quad (7)$$

where,  $N$  is the length of the time series. The more periodic the dynamic is, the larger the recurrence rate will be [19]. The next measure is  $DET$  (% determinism) which is defined as the percentage of recurrence points that form diagonal structures (of at least length  $l_{min}$ ) as:

$$DET = \frac{\sum_{l=l_{min}}^N lp(l)}{\sum_{l=1}^N lp(l)} \quad (8)$$

where,  $p(l) = p(l, \varepsilon)$  is the histogram of diagonal lines of length  $l$  with the recurrence threshold of  $l_{min}$ . The threshold  $l_{min}$  excludes the diagonal lines which are formed by the tangential motion of the phase space trajectory. A diagonal line of length  $l$  means that a segment of the trajectory is rather close during  $l$  time segment of the trajectory at a different time.  $DET$  plots may be interpreted as a signature of determinism and so predictability in

the time series data. The average diagonal line length:

$$L = \frac{\sum_{l=l_{min}}^N lp(l)}{\sum_{l=l_{min}}^N p(l)} \quad (9)$$

is the average time that two segments of the trajectory are close to each other, and can be interpreted as the mean prediction time [19]. Another RQA measure considers the length  $L_{max}$  of the longest diagonal line found in the RP [19]:

$$L_{max} = \max(\{l_i\}_{i=1}^{N_l}) \quad (10)$$

where,  $N_l = \sum_{l \geq l_{min}} p(l)$  is the total number of diagonal lines. This measure is related to the exponential divergence of the phase space trajectory. The faster the trajectory segments diverge, the shorter are the diagonal lines and the lower is the measure  $L_{max}$ .  $L_{max}$  is also an estimator of lower limit of the sum of the positive Lyapunov exponents [24]. Positive Lyapunov exponents gauge the rate at which trajectories diverge, and are the hallmark for dynamic chaos. Altogether, the shorter is the  $L_{max}$ , the less predictable is the signal. The Shannon entropy of the probability distribution of the diagonal line lengths  $p(l)$  is defined as:

$$ENTR = - \sum_{l=l_{min}}^N p(l) \ln p(l) \quad (11)$$

$ENTR$  reflects the complexity of the RP in respect of the diagonal lines, e.g. for uncorrelated noise the value of  $ENTR$  is rather small, indicating its low complexity [19].

The percentage of the recurrence points forming vertical structures is another RQA measure and is known as  $LAM$  (% Laminarity) and is defined as [19]:

$$LAM = \frac{\sum_{v=v_{min}}^N vp(v)}{\sum_{v=1}^N vp(v)} \quad (12)$$

where,  $p(v) = p(v, \epsilon)$  is the histogram of vertical lines of length  $v$  that exceeds a minimal length  $v_{min}$  and with recurrence threshold of  $\epsilon$ . Laminarity ( $LAM$ ) represents the occurrence of laminar states in the system without describing the length of these laminar phases.  $LAM$  will decrease if the RP consists of more single recurrence points than vertical structures. The average length of vertical structures is called trapping time ( $TT$ ), which estimates the mean time that the system will stay at a specific state or how long the state will be trapped [19]. It is given by:

$$TT = \frac{\sum_{v=v_{min}}^N vp(v)}{\sum_{v=v_{min}}^N p(v)} \quad (13)$$

Finally, the maximal length of the vertical lines in the RP:

$$v_{max} = \max(\{v_l\}_{l=1}^{N_v}) \quad (14)$$

can be regarded, analogously to the standard measure  $V_{max}$  ( $N_v$  is the absolute number of vertical lines). In contrast to the RQA measures based on diagonal lines, these measures are able to find chaos-chaos transitions [19]. Hence, they allow for the investigation of intermittency, even for rather short and non-stationary data series.

Furthermore, since for periodic dynamics the measures quantifying vertical structures are zero, chaos-order transitions can also be identified [19]. Furthermore, since for periodic dynamics the measures quantifying vertical structures are zero, chaos-order transitions can also be identified [19]. The RQA measures of the wind speed and wind power time series with  $v_{min} = l_{min} = 6$  and  $\epsilon = 1.5$  have been brought in figures 12 and 13, respectively. According to figure 12(a), the  $RR$  plot shows that the wind speed data has weak periodical dynamic as we find out from RP. Predictability for wind speed time series is noticeably small regarding  $DET$  measure in figure 12 (b), but figure 12(c) shows that the mean value of the  $L$  measure of this data is about 8 which is rather low. Thus, the mimic predictability of the wind speed time series can be concluded. As seen in figure 12(d),  $L_{max}$  falls to the low value of 3 in some regular periods. This behavior in conjunction with those observed in figures 12(f) and (h) emphasizes the seasonality and unstationarity of the wind speed. Figure 12 (e) shows the  $ENTR$  of wind speed. From this figure, noticeable  $ENTR$  is representative of highly complex dynamics of wind speed. Referring to figure 12(g), an almost flat  $TT$  plot locating at about 6, is representative of the predictability about  $8 \times 6 = 48$  hours ( $\tau = 8$ ) for wind speed. Investigation of figure 11 shows similar results for the RQA measures of the wind power time series. According to this figure, as  $RR$ ,  $DET$ ,  $L$  and  $L_{max}$  have been increased, it is concluded that the wind power is more predictable than wind speed. However, a flat plot of  $L_{max}$  as well as lower variations in  $LAM$  and  $v_{max}$  are representative of lower degrees of dynamic transitions or equivalently more stationarity in the wind speed dynamics. The plot of  $TT$  is less flat but its mean value is larger than that of wind speed (it is about 12). In this case, the mean prediction time is about  $12 \times 12 = 148$  hours ( $\tau = 12$ ).

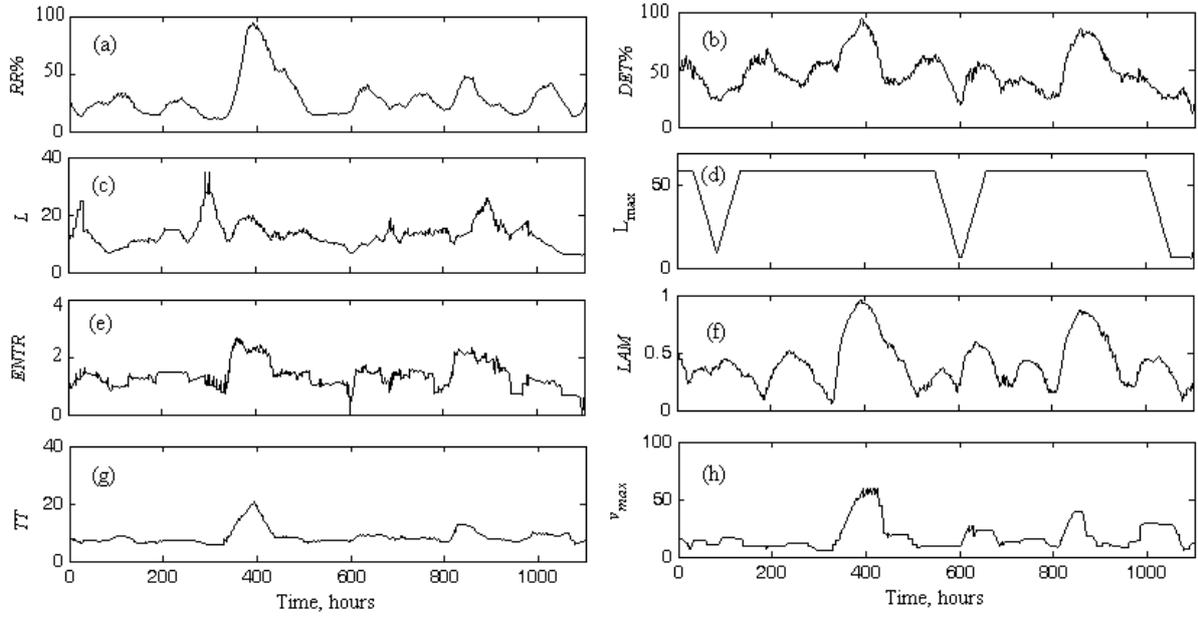


Figure 12. RQA measures for wind speed (a)  $RR\%$ , (b)  $DET\%$ , (c)  $L$ , (d)  $L_{max}$ , (e)  $ENTR$ , (f)  $LAM\%$ , (g)  $TT$ , and (h)  $V_{max}$ .

### 3. Interaction analysis of wind speed and wind power

#### 3.1. The correlation analysis

One of the most common tools for determination of effective inputs for a forecasting model is the correlation analysis between the underlying time series. The cross correlation between two time series  $X = \{x_i, i = 1, \dots, N\}$  and  $Y = \{y_i, i = 1, \dots, N\}$  can be defined as:

$$C_{X,Y}(m) = \frac{E\{x_n y_{n+m}^*\} - E(X)E(Y^*)}{\sigma_X \sigma_Y}, m = 1, \dots, N \quad (15)$$

where, in (15),  $E(\cdot)$  and  $*$  stand for mathematical expectation and complex conjugate, respectively. Indeed, the cross correlation is employed to examine the linear correlation of two time series. The more slower decays the correlation plot, the more linearly predictable is the  $Y$  in terms of  $X$ . Figure 13 shows the cross correlation of wind power versus the lagged wind speed time series. From this figure, one may conclude there exists an almost noticeable but decreasing correlation among the wind power time series and almost up to one weak lagged wind speed time series, which may be employed for forecasting this time series.

#### 3.2. The cross recurrence plots (CRP)

As stated earlier, the correlation analysis is a linear tool which reflects the linear dependency of two time series. In the literature, there are some nonlinear tools which investigate the bi-variate dependencies of the time series as well. The CRP is a bi-variate extension of the RP and was introduced to analyze the dependencies between two different systems by comparing their states [19] which can be considered as a generalization

of the linear cross-correlation function. Suppose we have two dynamical systems, each one is represented by the embedded state trajectories  $\vec{x}_i$  and  $\vec{y}_i$  in  $m$ -dimensional phase space. Analogous to the RP (Equations (4) and (5)), the corresponding cross recurrence matrix is defined by [19]:

$$CR_{i,j}^{x,y} = \Theta(\epsilon - \|\vec{x}_i - \vec{y}_j\|), \quad (16)$$

$$i = 1, \dots, N, j = 1, \dots, M$$

where,  $N$  and  $M$  are the lengths of the trajectories of  $\vec{x}_i$  and  $\vec{y}_i$  which are not required to be identical. However, the both systems are required to be represented in the same  $m$ -dimensional phase space, because a CRP looks for those times when a state of the first system recurs to one of the other system [12]. The graphical representation of the matrix CR is called CRP. In this graph, long diagonals are representative of similarity or correlation of the two dynamics. A measure based on the lengths of such lines can be used to find nonlinear interrelations between two systems, which cannot be detected by the common cross-correlation function. The more similar/correlated the two time series are the longer the diagonals and the higher density of dark dots around the main diagonal of the graph [12]. Figure 15(a) shows the CRP of the wind speed and wind power. According to that figure, the diagonals are short and the density of the dark dots is low. Therefore, the correlation between two time series is weak as implied before. The transitions as well as white ribbons in this graph are representative of existence a non-stationary nonlinear correlation between these two dynamics.

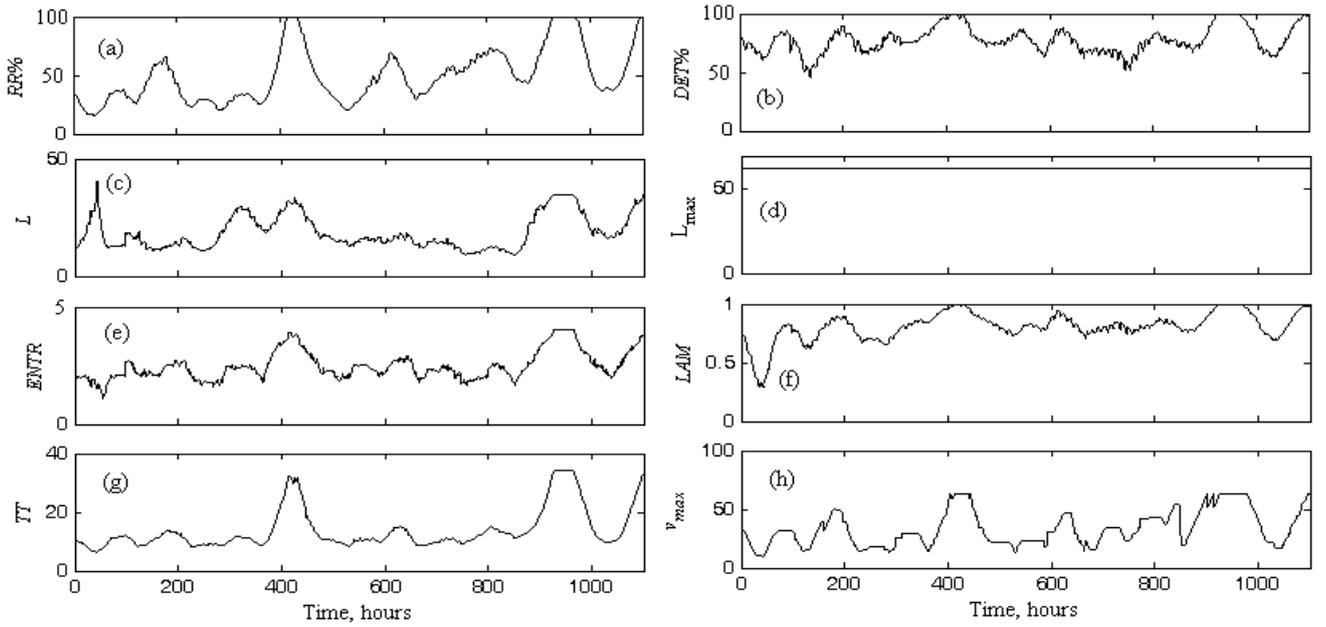


Figure 13. RQA measures for wind power (a)  $RR\%$ , (b)  $DET\%$ , (c)  $L$ , (d)  $L_{max}$ , (e)  $ENTR$ , (f)  $LAM\%$ , (g)  $TT$ , and (h)  $V_{max}$ .

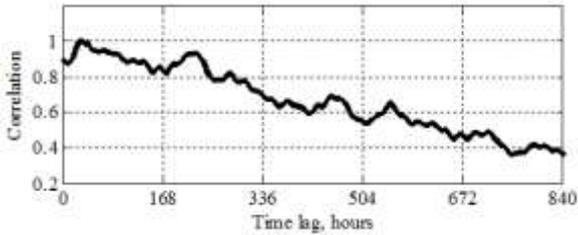


Figure 14. The cross correlation of wind speed and wind power time series.

Therefore, it is concluded that in constructing a forecasting model for wind power prediction, the seasonality and unstationarity should be accounted for by adaptive tuning of the model or constructing seasonal models.

### 3.3. Cross recurrence quantification analysis (CRQA)

Similar to RQA measures, cross RQA measures (CRQA) are defined. For example, the  $RR$  of CRPs is known as  $CC_2(\varepsilon)$  and is defined as [19]:

$$CC_2(\varepsilon) = \frac{1}{N^2} \sum_{i,j=1}^N CR_{i,j}(\varepsilon) \times 100 \quad (17)$$

Other CRQA measures are defined similar to RQA measures as in Section 2.9 (Equations (8) to (13)). Figure 16 shows the CRQA measures of wind speed and wind power. The results admit the last conclusions and yield very short term cross predictability of the wind power in terms of wind speed time series.

### 3.4. The joint recurrence plot (JRP)

In order to compare different systems' dynamics, another extension of RP has been developed named joint recurrence plot (JRP). In this method,

the recurrences of each time series to its trajectory in its respective phase spaces is considered separately to find out times when both of them recur simultaneously, i.e. when a joint recurrence occurs. In this way, the individual phase spaces of both systems are preserved. This type of comparison, especially when we have two physically different systems, makes more sense. JRP of two time series  $\vec{x}_i$  and  $\vec{y}_i$  embedded in  $m_x$  and  $m_y$  dimensional phase spaces is defined as [19]:

$$JR_{i,j}^{x,y} = \Theta(\varepsilon_x - \|\vec{x}_i - \vec{x}_j\|) \Theta(\varepsilon_y - \|\vec{y}_i - \vec{y}_j\|), \quad (18)$$

$i, j = 1, \dots, N$

where,  $\varepsilon_x$  and  $\varepsilon_y$  are the corresponding thresholds for time series  $\vec{x}_i$  and  $\vec{y}_i$ . Figure 15(b) shows the JRP of the wind speed and wind power time series. As seen in this figure, the same pattern as Figure 15(a) is observed, but the longer diagonals are formed in the JRP plot.

### 3.5. Joint recurrence quantification analysis (JRQA)

Similar to RQA measures, joint RQA measures (JRQA) are defined and the  $RR$  of JRPs of  $n$  systems with thresholds  $(\varepsilon^1, \dots, \varepsilon^n)$  known as  $JC_2(\varepsilon)$  is defined as [25]:

$$JC_2(\varepsilon^1, \dots, \varepsilon^n) = \frac{1}{N^2} \sum_{i,j=1}^N \prod_{k=1}^n R_{i,j}^{\varepsilon^k}(\varepsilon^k) \quad (19)$$

Other JRQA measures are defined similar to RQA measures as in Section 2.9 (Equations (8) to (13)). Figure 17 shows the JRQA measures of wind speed and wind power time series. From this figure, it is observed that joint recurrence of the

two time series performs more degree of predictability with respect to that of CRP. As seen in this figure, the JRQA measures perform almost flat which is representative of stationarity in the recurrence of the two time series, which corresponds to similarity in recurrence dynamics of the two time series. That is, it is concluded that in constructing the prediction model for wind power, finding the recurrence dynamic model of the wind speed may be so effective.

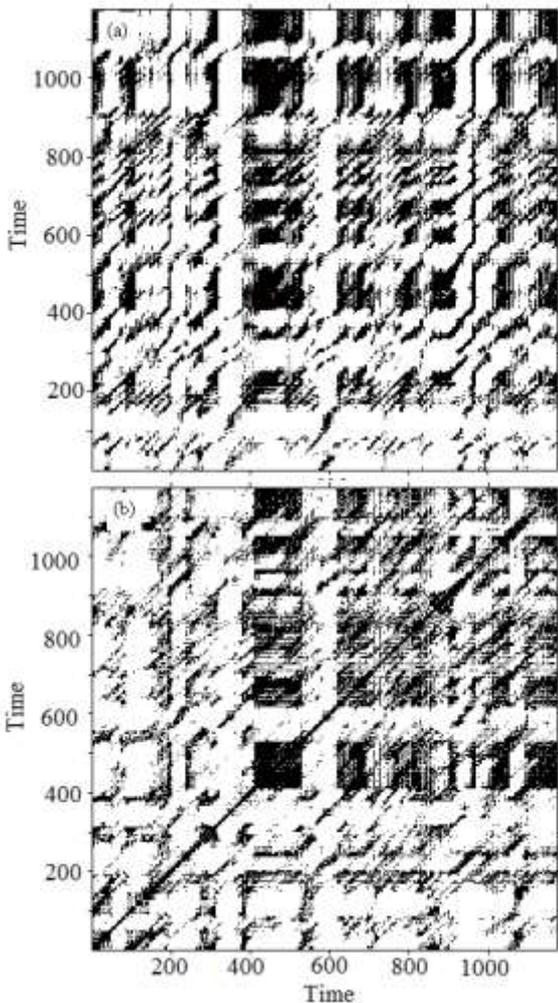


Figure 15. (a) CRP and (b) JRP of wind speed and wind power time series'.

#### 4. Conclusion

The characterization and predictability analysis of wind speed and wind power time series has been considered in this paper. The employed data was the experimental data from Spain's wind farm on May 2005 [5]. The data analysis procedure includes histogram plots, power spectral density (PSD) analysis, the phase space reconstruction and test of surrogates, the slope of integral sum, the  $\delta - \varepsilon$  method, recurrence plots (RPs) and recurrence quantification analysis. The analyses are representative of seasonal unstationary stochastic behavior and short term predictability

of wind speed and power time series. In order to investigate the interactive behavior, the mentioned wind speed and wind power, the bi-variate linear and nonlinear analysis methods such as cross correlation analysis, cross recurrence plots, joint recurrence plots, CRQA as well as JRQA were performed as well. The analysis results show that a noticeable similarity exists in the recurrence dynamics of these two time series, which is almost stationary. Nevertheless, the correlation of these two time series is mimic, seasonal and unstationary.

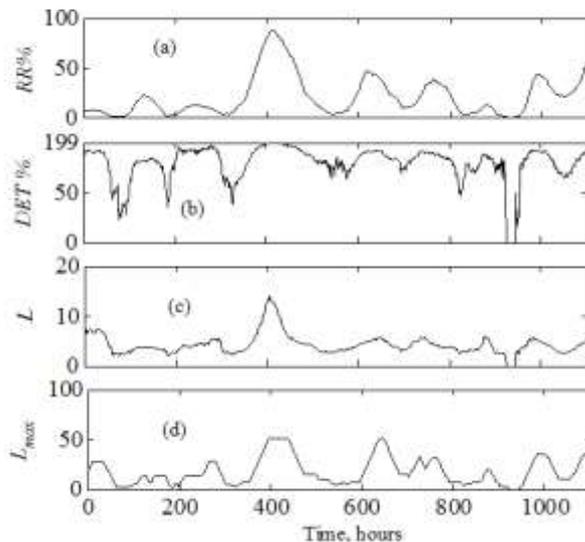


Figure 16. CRQA measures for wind speed and wind power time series (a) RR%, (b) DET%, (c) L, (d) Lmax.

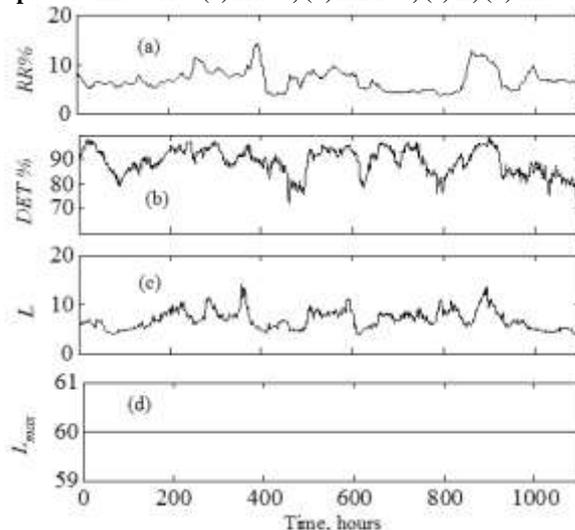


Figure 17. JRQA measures for wind speed and wind power time series (a) RR%, (b) DET%, (c) L, (d) Lmax.

#### References

[1] Amjady, N., Keynia, F. & Zareipour, H. (2011). Short-term wind power forecasting using ridgelet neural network. *Elect PowSystRes*, vol. 81, no. 12, pp. 2099-2107.

- [2] Foley, AM., et al. (2012). Current methods and advances in forecasting of wind power generation. *Renewable Energy*, vol. 37, no. 1, pp. 1-8.
- [3] Fan, S., et al. (2009). Forecasting the wind generation using a two-stage network based on meteorological information. *IEEE Trans Energy Conv*, vol. 24, no. 2, pp. 474-482.
- [4] Jiang, Y., Song, Z. & Kusiak, A. (2013). Very short-term wind speed forecasting with Bayesian structural break model. *Renewable Energy*, vol. 50, pp. 637-647.
- [5] Amjady, N., Keynia, F. & Zareipour, HR. (2011). A new hybrid iterative method for short-term wind speed forecasting. *Europ Trans Elect Pow*, vol. 21, no. 1, pp. 581-595.
- [6] Hong, YY., Chang, HL. & Chiu, CS. (2010). Hour-ahead wind power and speed forecasting using simultaneous perturbation stochastic approximation (SPSA) algorithm and neural network with fuzzy inputs. *Original. Energy*; vol. 35, no. 9, pp. 3870-3876.
- [7] Amjady, N., Keynia, F. & Zareipour, HR. (2011). Wind Power Prediction by a New Forecast Engine Composed of Modified Hybrid Neural Network and Enhanced Particle Swarm Optimization, *IEEE Trans Sustain Energy*, vol.2, no.3, pp. 265-276.
- [8] Ng, WW., Panu, US. & Lennox, WC. (2007). Chaos based analytical techniques for daily extreme hydrological observations. *J Hydro*, vol. 342, no. 1-2, pp. 17-41.
- [9] Bigdeli, N. & Haeri, M. (2009). Time-series analysis of TCP/RED computer networks, an empirical study. *Chaos, SolitFract*, vol. 39, no. 2, pp. 784-800.
- [10] Bigdeli, N., Afshar, K. & Amjady, N. (2009). Market data analysis and short-term price forecasting in the Iran electricity market with pay-as-bid payment mechanism. *Elect PowSystRes*, vol. 79, no. 6, pp. 888-898.
- [11] Bigdeli, N. & Afshar, K. (2009). Chaotic behavior of price in the power markets with pay-as-bid payment mechanism. *Chaos SolitFract*, vol. 42, no. 4, pp. 2560-2569.
- [12] Bigdeli, N. & Afshar K. (2009). Characterization of Iran electricity market indices with pay-as-bid payment mechanism. *Physica A*, vol. 388, no. 8, pp. 1577-1592.
- [13] Khoa, TQD., Yuichi, N. & Masahiro, N. (2009). Recognizing brain motor imagery activities by identifying chaos properties of oxy-hemoglobin dynamics time series. *Chaos SolitFract*, vol. 42, no. 1, pp. 422-429.
- [14] Sakai, K., Noguchi, Y. & Asada, S. (2008). Detecting chaos in a citrus orchard; reconstruction of nonlinear dynamics from very short ecological time series. *Chaos SolitFract*, vol. 38, no. 5, pp. 1274-1282.
- [15] Andrade, KC., et al. (2012). Statistical evaluation of recurrence quantification analysis applied on single trial evoked potential studies. *Clinical Neurophysiology*, vol. 123, no. 8, pp. 1523-1535.
- [16] Karakasidis, TE. & Charakopoulos, A. (2009). Detection of low-dimensional chaos in wind time series. *Chaos SolitFract*, vol. 41, no. 4, pp. 1723-1732.
- [17] Chang, TP., et al. (2012). Fractal dimension of wind speed time series. *Applied Energy*, vol. 93, pp. 742-749.
- [18] Bigdeli, N., et al. (2013). A comparative study of optimal hybrid methods for wind power prediction in wind farm of Alberta, Canada. *Journal of Renewable and Sustainable Energy Reviews*, vol. 27, pp. 20-29.
- [19] Marwan, MC., et al. (2007). Recurrence plots for the analysis of complex systems. *Physics Reports*, vol. 438, no. 5-6, pp. 237-329.
- [20] Bigdeli, N., et al. (2007). Characterization of complex behaviors of TCP/RED computer networks based on nonlinear time series analysis methods. *Physica D: Nonlin Phenomena*, vol. 233, no. 2, pp. 138-150.
- [21] Timmer, J., et al. (2000). Pathological tremors: deterministic chaos or nonlinear stochastic oscillators?. *Chaos*, vol. 10, no. 1, pp. 278-288.
- [22] Hegger, R., Kantz, H. & Schreiber, T. (1999). Practical implementation of nonlinear time series methods: the TISEAN package. *Chaos*, vol. 9, no. 2, pp. 413-38.
- [23] The MathWorks, MATLAB [Online]. Available: <http://www.mathworks.com>.
- [24] Grassberger, P. & Procaccia, I. (1983). Measuring the strangeness of strange attractors. *Physica D: nonlinear phenomena*, vol. 9, pp. 189-208.
- [25] Tung, W-w., et al. (2005). Direct characterization of chaotic and stochastic dynamics in a population model with strong periodicity. *Chaos SolitFract*, vol. 24, no. 2, pp. 645-652.
- [26] Strozzi, F., Zaldivar, JM. & Zabilut, JP.(2002). Application of non-linear time series analysis techniques to high frequency currency exchange data. *Physica A*, vol.312, no. 3-4, pp. 520-538.
- [27] Kennel, MB., Brown, R. & Abarbanel, HDI. (1992). Determining embedding dimension for phase-space reconstruction using a geometrical reconstruction. *Phys Rev A*, vol. 45, no. 2, pp. 3403-3411.
- [28] Theiler, J., et al. (1992). Testing for nonlinearity in time series: the method of surrogate data. *Physica D*, vol. 58, no. 1-4, pp. 77-94.
- [29] Schreiber, T. (1998). Constrained randomization of time series data. *Phys Rev Lett*, vol. 80, no. 10, pp. 2105-2108.

[30] Engbert, R. (2002). Testing for nonlinearity: the role of surrogate data. *Chaos SolitFract*, vol. 13, no. 1, pp. 79-84.

[31] Takens, F. (1981). Detecting strange attractors in turbulence. *Lect Note Math*, Springer, New York, vol. 898, pp. 366-381.

[32] Kaplan, D. & Glass, L. (1995). *Understanding nonlinear dynamics*. Springer.

[33] Marwan, N. & Kurths, J. (2005). Line structures in recurrence plots. *PhysLett A*, vol. 336, no. 4-5. pp. 349-357.

[34] CRP toolbox for Matlab provided by TOCSY:<http://tocsy.agnld.uni-potsdam.de>.