ON THE EXISTENCE AND CHARACTERIZATION OF EXTREME EVENTS IN WIND DATA

Constantino A. García
Department of Information and Communications Systems Engineering, University San Pablo CEU, Madrid, ES-28668, Spain, constantino.garciama@ceu.es , ORCID: 0000-0002-9739-2252

Pedro G. Lind
Department of Computer Science, Oslo Metropolitan University, P.O. Box 4 St. Olavs plass, N-0130 Oslo, Norway, pedro.lind@oslomet.no, ORCID: 0000-0002-8176-666X

Matthias Wächter
Institute für Physik and ForWind, Carl-von-Ossietzky, Universität Oldenburg, DE-26111 Oldenburg, Germany, matthias.waechter@uni-oldenburg.de, ORCID: 0000-0001-9212-1122

Abraham Otero
Department of Information and Communications Systems Engineering, University San Pablo CEU, Madrid, ES-28668, Spain, aotero@ceu.es , ORCID: 0000-0003-4568-2933

Joachim Peinke
Institute für Physik and ForWind, Carl-von-Ossietzky, Universität Oldenburg, DE-26111 Oldenburg, Germany, joachim.peinke@uni-oldenburg.de, ORCID: 0000-0002-0775-7423

Cite this paper as: *García, C.A., Lind, P., Wächter, M., Otero, A., Peinke, J. On the existence and characterization of extreme events in wind data. 7th Eur. Conf. Ren. Energy Sys. 10-12 June 2019, Madrid, Spain.*

Abstract: We present evidence that quantitative definitions of extreme events in wind, so called gusts, may be flawed and unable to grasp true extreme situations of wind velocity in short time intervals. In particular, we show that the same statistics of particular pattern shapes in wind data are found in surrogate data generated from the original series of measurements. We apply a pattern recognition algorithm to a two-parameter "Mexican-hat" (duration and amplitude) and find the same probability of occurrence for both original measurements and surrogate data. The distribution of the corresponding values of the duration and amplitude present deviations that can be explained by the filtering of n-point correlations in the original series.

***Keywords**:* *Wind data, Gusts, Extreme Events*

Nomenclature	
EOG	Extreme Operating Gust
EDC	Extreme Direction Change
ECG	Extreme Coherent Gust
EWS	Extreme Wind Shear

1. INTRODUCTION

Though being one of the most promising sources of renewable energy [1], wind has a challenging feature: it is highly intermittent. Due to its highly fluctuating velocity it is difficult to account for an estimate of power output within a given time period [2]. Indeed, in just a few seconds one can observe an increase of the wind velocity from a few meters per second up to almost 20 meters per second. Such sudden occurrence of strong wind, so-called gusts, drives a corresponding high fluctuation of the power output of one wind turbine [3], which therefore appeals for a better understanding of such strong fluctuations in wind data. For wind power production, approaches to evaluate wind speed persistence [4] as well as wind speed synthetic generation of wind fields [5], were proposed for better reproduce power production in wind turbines and wind farms. In particular, Markov models have been proposed for reproducing the statistics of wind velocities together with power output in Portuguese wind farms [6, 7]. Methods for modeling wind gusts [8] as well as some numerical procedures for detecting them [5] have also been proposed. However, from a physical perspective, a gust is still loosely defined: it results from a local, short-term and sudden wind speed variation in the turbulent atmosphere, but, up to our knowledge, a closed quantitative definition is not established.

In this paper we address the question of the statistical and functional features of gusts. Assuming that a gust results from the joint happening of a sudden increase in the wind velocity followed by a similarly fast decrease, i.e. it yields a functional shape for the gust given by the so-called "Mexican-hat". Other shapes are equally possible, e.g. a high wind speed that occurs in a short time (high acceleration) leads to a gust shape given by a so-called "Front". According to the 2005 International Electrotechnical Comission standards [9] there are four different types of extreme wind events: (i) the extreme operating gust (EOG) typically modeled with a Mexican-hat function of varying amplitude (maximum velocity) and width (duration); (ii) the extreme direction change (EDC), which is characterized by a sustained change in wind direction modeled typically with a cosine-shaped curve with a given return period; (iii) the extreme coherent gust (ECG), characterized by a sustained change in wind speed which ends, differently from EOG, on a constant high speed; and (iv) the extreme wind shear (EWS), characterized by a transient variation in the horizontal and vertical wind gradient across the rotor. These types of extreme wind events are not mutually exclusive. For example, so-called extreme coherent gusts with direction change can occur, which result from a simultaneous occurrence of EDC and ECG. Here we focus on EOG, which are standardly taken as typical extreme wind events. While attention is often centered on positive gusts -real Mexican-hats- since they are considered more dangerous, we will also consider "negative" gust with the symmetrical shape, resembling what we call a "Mexican-hole".

We start in Sec. 2 by describing the data and the surrogate data preprocessing. In Sec. 3 we characterize an extreme event and describe how to detect those events in data series. In Sec. 4 we present the statistical results and compare real and surrogate cases and Sec. 5 concludes the paper.

2. WIND MEASUREMENTS AND SURROGATES PROCESSING

The available data for extreme events detection consists of twelve data sets recorded at FINO1. Each of these files contains the wind speed during one month at different heights. Since all sets have long segments with missing data or segments at different sampling rates, we choose a fragment of them, namely segments of approximately 36 hours long ($\approx 10^5$ data points) were selected for each recording. Each of these segments was selected so that most of gaps last at most one second. Long segments with missing values (over 5 seconds) were eliminated. The remaining missing data was then added using linear interpolation. For the sake of simplicity, we focus on a single height, namely 80 m. *Figure 34*a shows illustrative data series of the wind which was then normalized, *Figure 34*b, according to

$$v_n(t) = \frac{v(t) - \langle v \rangle_{10m}(t)}{\sigma_{10m}(t)} \qquad (1)$$

where $\langle v \rangle_{10m}(t)$and $\sigma_{10m}(t)$ are, respectively, the mean and standard deviation computed within a window of 10 minutes centered at time-step t.

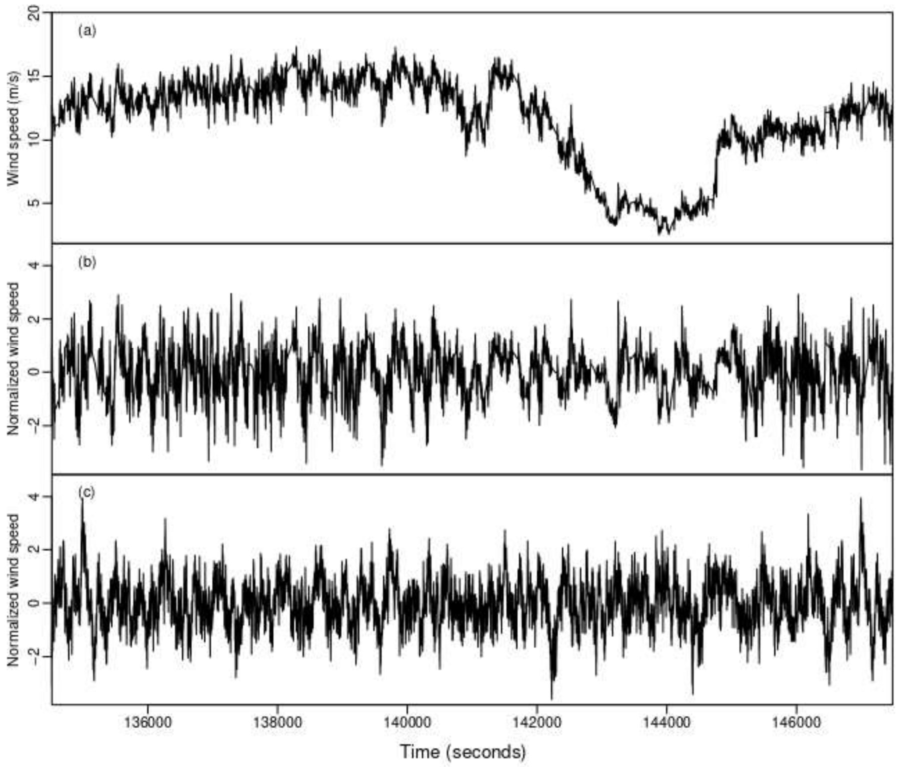

Figure 34. Illustration of wind data series: (a) real wind velocity $v(t)$ in FINO1 of Alpha Ventus at 80 m height, (b) the associated normalized data $v_n(t)$ (see Eq. [1]) and (c) the surrogate wind velocity $v_s(t)$ generated for the wind velocity (see text).

For testing if the extreme events are characteristic of wind data, we compare the existence of specific extreme events in real data and in surrogate data derived from it. Surrogate data testing is usually employed for detecting non-linear behavior in time series. This method requires specifying a null hypothesis describing a linear process sharing the same linear properties as the original time series. Then, several surrogate data sets according to the null hypothesis are generated using Monte Carlo methods. A discriminating statistic, such as nonlinear prediction error, time reversal symmetry or correlation dimension, is then calculated for the original time series and for the surrogate set. If the statistics are significantly different from each other, the null hypothesis is rejected and non-linearity assumed [10].

We can adopt this method for testing if the number of extreme events in wind data is significantly larger than in a linear process. The null hypothesis is that there is no significant difference in the number of extreme events observed in wind data when compared with a Gaussian linear process. Accepting the null hypothesis would suggest that such specific shapes of the "extreme events" in wind data are in fact statistical fluctuations present in every Gaussian process, and are not characteristic of wind data.

For generating the surrogate data, the Theiler's phase randomization method was applied, as described in Ref. [10]. *Figure 34*c shows a plot of the surrogate $v_s(t)$ series obtained from the original velocity series $v(t)$**.** To avoid performing heavy computations, the surrogate data was generated from segments of just 90 minutes (≈ 10^4 points).

3. METHOD FOR EXTREME EVENTS DETECTION

For detecting EOG events present in wind data we will use the same approach employed by M. Ahmann [11], where the extreme events are assumed to follow a Mexican-hat pattern defined as

$$F_{A,D}(t) = \frac{A}{\sqrt{2\pi}D^3}\left(\frac{t^2}{D^2} - 1\right) exp\left(\frac{-t^2}{2D^2}\right) \qquad (2)$$

where A is the amplitude of the event and D controls the duration of the Mexican shape. An example of a Mexican-hat pattern in wind data is shown in *Figure **35***a.

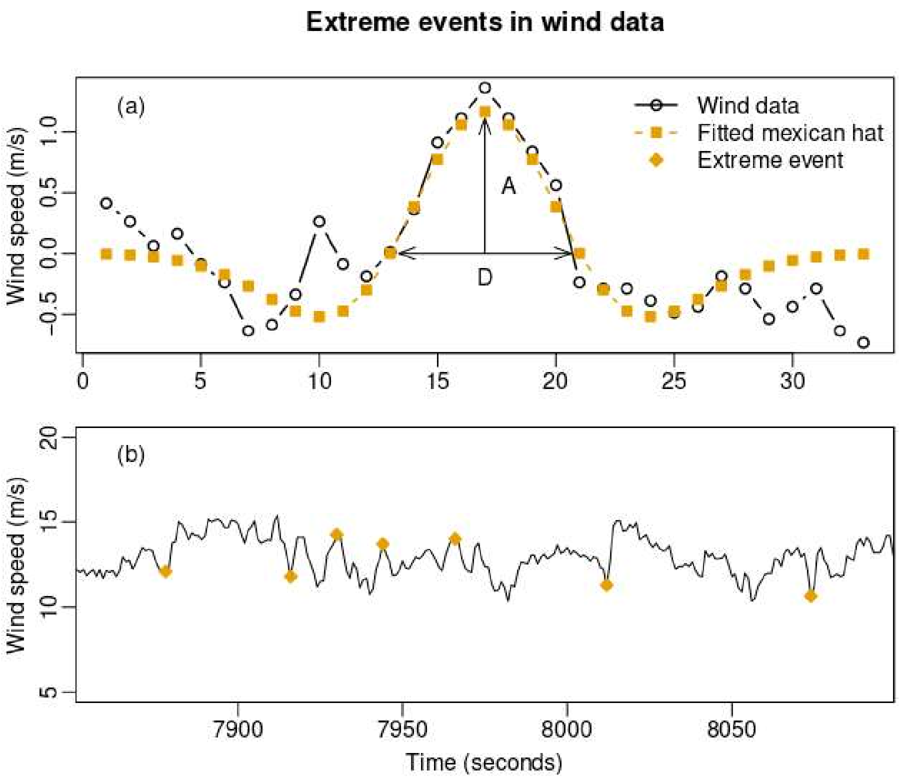

Figure 35. (a) Illustrative example of a "Mexican-hat" pattern in wind data (circles) with the corresponding fit of Eq. [2] with A = 1 and D = 9. In (b) a succession of detected extreme events are marked (diamonds) in an illustrative wind data series. Here, both "hats" and "holes" are considered (see text).

The extreme events are detected using a matched filter: the detection is performed by correlating the template of the Mexican-hat pattern with the wind data time series, through the following function

$$R_D(t) = \frac{\left\langle (v(t) - \overline{v}_D)\left(F_{A,D}(t) - \overline{F}_{A,D}(t)\right)\right\rangle}{\sigma_v \sigma_F} \tag{3}$$

where the mean $\langle \quad \rangle$ is performed from time t till $t + D$ and $\overline{v}_D$ (resp. $\overline{F}_{A,D}(t)$) and σ_D (resp. σ_F) are the mean and standard deviation of wind speed v (resp. of F). When a local extrema of the resulting correlation exceeds a certain threshold, $R_{th} = 0.9$, an extreme event is detected. The exact time t in Eq. [2] marking the occurrence of the extreme event is determined by exploring a small temporal window around the instant in which the correlation exceeds the threshold.

Extreme events are assumed to not occur simultaneously, i.e. during an extreme event with the time-duration D, no other extreme event can take place. The detection process is repeated for several values of the D and A parameters, till the maximum of the correlation between function and real data is found. An illustrative example of the performance of the algorithm is shown in *Figure **35***b, where one plots a short series of successive Mexican-hats and Mexican-holes.

4. ANALYSIS OF WIND EXTREME EVENTS

Statistical frequency

Sweeping our detection method through the wind data and surrogate series we first count the number of extreme events as a function of the duration D. *Figure **36***a and *Figure **36***b show the distribution of events depending on the D parameter.

The probability of an event follows a stretched exponential distribution $g(D) \sim exp(-D^{\alpha})$ with parameters $\alpha = 0.49 \pm 0.01$ for the wind data series and $\alpha = 0.47 \pm 0.01$for the corresponding surrogates. The exponential behavior showed in *Figure **36***a is a consequence from the fact that narrower Mexican-hats can occur more often than wider Mexican-hats.

In order to take into account the length of the events we calculate the probability for a Mexican-hat of duration D to occur as the quotient between the number of events $g(D)$ detected for a given D and the maximum number of possible events with duration D, given by T/D, being T the total length of the data set. *Figure **36**c* and *Figure **36**d* show the distributions $P(D)$ of the duration D for the real wind series and the surrogates. Both are well approximated by $P(D) \sim g(D)\frac{D}{T} \propto (1+\beta D)\cdot exp(-D^{\alpha})$.

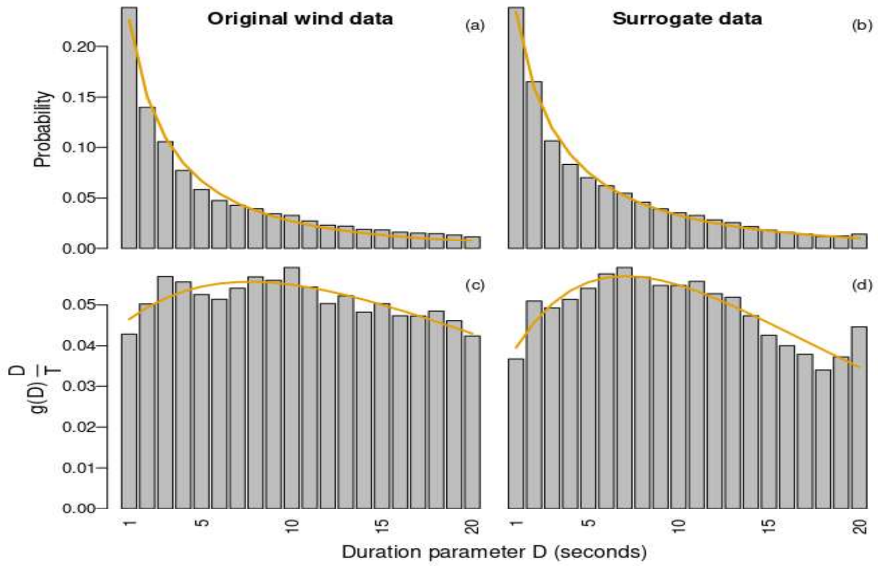

Figure 36. Figures (a) and (b) show the distribution of the Mexican-hat pattern for real wind data and the corresponding surrogates, depending on D. In (b) and (c) one shows the probability of occurrence of an extreme event for real wind data and the corresponding surrogates.

From *Figure **36*** one concludes that, from the perspective of its frequency, the occurrence of Mexican-hats in wind data is not characteristic of the data.

Amplitude and duration of extreme events

The fact that the number of events in wind data is similar to the number of events in Gaussian processes does not imply that their associated statistical features are also the same. To explore this question, we next study the distribution of the amplitudes of EOGs. Indeed, we may argue that the distributions of amplitudes for Gaussian processes and wind data are different. Due to the way they are generated, Gaussian processes are stationary and consequently, the amplitude of an extreme event will be concentrated around some specific value. On the other hand, wind data is non-stationary and thus, the amplitude of what we call an extreme event will change over time, generating broader distributions.

*Figure **37***a-d show the distribution of amplitudes for different values of the duration D for wind series in FINO1. From such plots one clearly sees that Mexican-holes (negative amplitudes) tend to be more probable than Mexican-hats (positive amplitudes). However the distribution of the positive amplitudes is similar to the distribution of negative amplitudes in wind data: both show a heavy tail for the extreme values (large negative or positive values) and another tail quickly decreasing towards zero when the amplitude tends to zero. As expected, the distribution is zero for $A = 0$, since there are no Mexican-hat with zero amplitude.

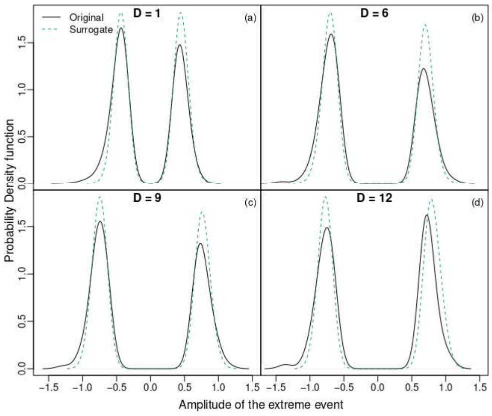

*Figure 37. (a-d) Distribution of amplitudes in both normalized wind and the corresponding surrogate data for different values of the **D** parameter, **D**=1, 6, 9 and 12. The probability density function was estimated using a Gaussian kernel. Note that wind data distribution has heavier tails than the distribution of the surrogate data.*

Altogether, for the distribution of amplitudes in wind differs from the one of surrogate data. However, we may hypothesize that these differences are only due to the non-stationary nature of wind data. To test this hypothesis, we can study the distribution of the normalized wind series $v_n(t)$ that feeds the detection algorithm, as defined in Eq. [1]. Note that, statistically, the standardization procedure should generate similar wind data series for all the different files. The statistics obtained for normalized data will be more robust than those from the original set: normalization allows one to average through all the months (different files) since we assume that these time series were generated from the same distribution.

*Figure **38***a-d show the first four moments of the amplitude distribution, namely the mean, the standard deviation, the skewness and the kurtosis, for both positive and negative amplitudes separately. As one sees, the distributions of the positive and negative amplitudes have similar statistical properties for the full range of D values, indicating that the distribution of the positive and negative amplitudes is the same, apart from their different absolute frequencies.

The mean, as well as the median and the mode (not shown), of surrogate and wind data are similar for small values of D, but diverge significantly from each other for larger values. The standard deviation of the surrogate data is smaller than the standard deviation of the original data, which explains why the density of surrogate data for the mode is greater than the density of the mode for the original data. From the skewness plot, we may conclude that wind data has still heavier tails than surrogate data. On the other hand, the skewness decreases with D: the longer the events last, the more symmetric their frequency distribution is. Kurtosis shows a tendency of also decreasing with the duration of the extreme event, showing a Gaussian shape (kurtosis=3) for the middle duration range.

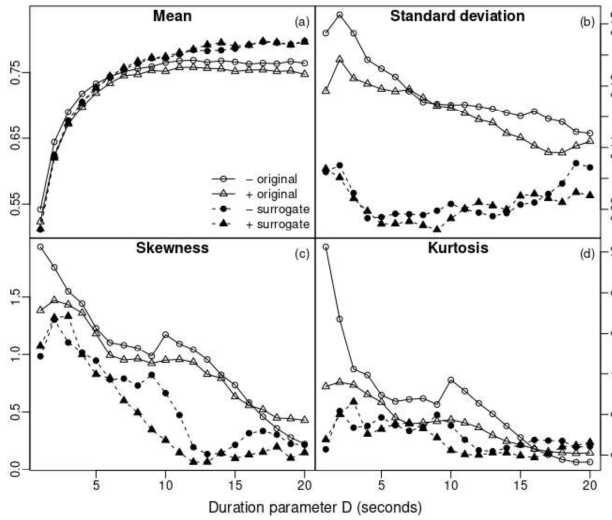

Figure 38. (a-d) Central metrics and moments of the amplitude distribution for both wind data and the surrogate data as a function of the event duration D.

Applying the Kolmogorov-Smirnov test we have ascertained that the positive amplitudes from surrogate data have a different distribution than the positive amplitudes in wind data, i.e. p-values closed to zero. The same applies for negative values. The difference between data and surrogates is mainly characterized by heavy tails, reflecting correlations in wind data that are not included in the surrogates.
We also compare the distributions of both positive and negative amplitudes in wind data, which seem to be equally distributed only for long enough events (see *Figure **39***). For extreme events occurring in short time-spans the amplitudes are no longer symmetrical and Mexican-holes tend to be significantly stronger than Mexican-hats, typically associated with gusts.

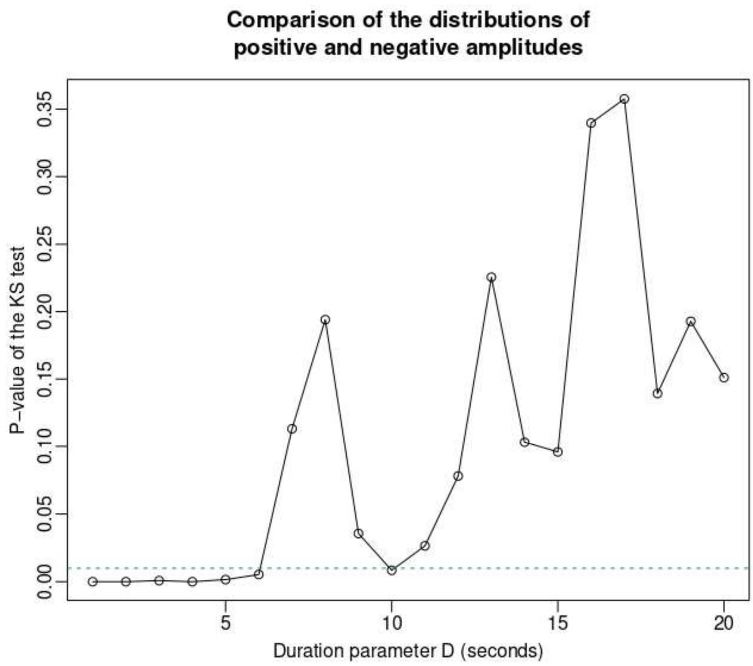

Figure 39. To test how close the two distributions are we calculate the p-values for the Kolmogorov-Smirnov test, taking positive and negative amplitudes in wind data separately. The dashed horizontal line marks a p-value of 0.01.

5. CONCLUSIONS AND DISCUSSIONS

Our findings can be summarized in two conclusions: (i) the frequency statistics of the number of extreme events in wind data is similar to the one in Gaussian processes (surrogate); (ii) however, the corresponding distributions of their parameters, namely duration and amplitude do not always match.
Such mismatch of the amplitudes and duration of extreme events can be explained by the removal of 2-point and higher n-point correlations in the linear Gaussian surrogates. Consequently, following the first conclusion we argue that, contrary to previous works, Mexican-hats are not necessarily characteristic of extreme wind patterns, namely gusts.
Since Mexican-hat is a pattern of at least five points, n-point correlations play a non negligible role in their occurrence and properties. Therefore, a comparison with surrogates that preserve the same n-point correlations as in wind data would be a step forward to investigate this problem. Recently [12, 13, 14], the n-point statistics of wave height in the ocean was recovered from empirical data, using a proper stochastic framework.

ACKNOWLEDGEMENTS
We acknowledge financial support by DFG (PE 478/18-1, WA 3793/2-1, LI 1599/3-1) as well as the financial support from the University San Pablo CEU and Santander Bank under the grant PCON10/2018.

REFERENCES
[1] Olabi A. 100% sustainable energy. Energy 77. 2017 1-5.
[2] Olabi A. State of the art on renewable and sustainable energy. Energy 61. 2013 2-7.
[3] Peinke J, Heinemann D, Kühn M. Windenergie-eine turbulente Sache? Physik Journal. 2014 Jul;13(7):36.
[4] Koçak K. Practical ways of evaluating wind speed persistence. Energy. 2008 Jan 1;33(1):65-70.

[5] Carapellucci R, Giordano L. A new approach for synthetically generating wind speeds: A comparison with the Markov chains method. Energy. 2013 Jan 1;49:298-305.
[6] Scholz T, Lopes VV, Estanqueiro A. A cyclic time-dependent Markov process to model daily patterns in wind turbine power production. Energy. 2014 Apr 1;67:557-68.
[7] Raischel F, Scholz T, Lopes VV, Lind PG. Uncovering wind turbine properties through two-dimensional stochastic modeling of wind dynamics. Physical Review E. 2013 Oct 29;88(4):042146.
[8] Shanmuganathan S, Sallis P. Data mining methods to generate severe wind gust models. Atmosphere. 2014 Mar;5(1):60-80.
[9] International Electrotechnical Commission. Wind turbine generator systems. Part 12: Wind turbine power performance testing, International Standard 61400-12-1. International Electrotechnical Commission, 2005.
[10] Theiler J, Eubank S, Longtin A, Galdrikian B, Farmer JD. Testing for nonlinearity in time series: the method of surrogate data. Physica D: Nonlinear Phenomena. 1992 Sep 15;58(1-4):77-94.
[11] Ahmann M. Untersuchungen über die struktur von extremeereignissen in windmessdaten (Ph.D. thesis). Carl von Ossietzky Universität Oldenburg, Germany, 2009.
[12] Hadjihosseini A, Peinke J, Hoffmann NP. Stochastic analysis of ocean wave states with and without rogue waves. New Journal of Physics. 2014 May 21;16(5):053037.
[13] Nawroth AP, Friedrich R, Peinke J. Multi-scale description and prediction of financial time series. New Journal of Physics. 2010 Aug 10;12(8):083021.
[14] Stresing R, Peinke J. Towards a stochastic multi-point description of turbulence. New Journal of Physics. 2010 Oct 29;12(10):103046.