Acta Scientiarum

http://periodicos.uem.br/ojs/acta
ISSN on-line: 1807-8664
Doi: 10.4025/actascitechnol.v41i1.38954

STATISTICS

Using Sample Entropy to assess complexity of wind speed dynamics

Gutenberg Ferreira da Silva[1] Antonio Samuel Alves da Silva[2*] and Tatijana Stosic[2]

[1]Universidade Federal Rural de Pernambuco, Recife, Pernambuco, Brazil. [2]Programa de Pós-graduação em Biometria e Estatística Aplicada, Departamento de Estatística e Informática, Universidade Federal Rural de Pernambuco, Rua Dom Manuel de Medeiros, s/n., 52171-900, Recife, Pernambuco, Brazil. *Author for correspondence. E-mail: samuelmatematico@yahoo.com.br

ABSTRACT. In this paper we analyzed the complexity of the time series of wind speed in Petrolina, Brazil at heights of 25 and 50 m. We applied the sample entropy (*SampEn*) method on wind speed temporal series for each month of 2010, in order to analyze the intra-annual variability of complexity for wind dynamics and its relation with the wind potential. The results showed that wind speed fluctuations are more regular when comparing consecutive data sequences over one hour periods. Although the average wind speed was higher at 50m, indicating higher wind potential, the values of *SampEn* were also higher between June and September indicating a greater complexity of wind dynamics and consequently lower efficiency in wind energy capture. In the months of February, October, and November, when the average wind speed exceeded the minimum value for generation of wind power (3.5 m s^{-1}) at both heights, the average wind speed increased and the entropy decreased with height. This indicates a higher wind potential and more regularity of wind dynamics at a height of 50 m, both favorable for wind energy production. For the rest of the year the wind power generation was possible only at height of 50 m.

Keywords: wind speed; sample entropy; wind energy.

Received on August 4, 2017.
Accepted on March 28, 2018

Introduction

Wind power is one of the most rapidly growing renewable energy sources because of its high efficiency and low pollution. The participation of wind power in US electricity production is projected to be 20% until 2030 (Lindenberg, 2009) and 12% until 2020 in Europe (Van Hulle et al., 2009). Wind energy production in Brazil increased from 22 MW in 2003 to 602 MW in 2009 as the result of the government Program for Incentive of Alternative Electric Energy Sources (*Programa de Incentivo às Fontes Alternativas de Energia Elétrica* - Proinfa), which was created in 2002 to stimulate the electricity generation from wind power, biomass, and small hydroelectric plants (Dutra & Szklo, 2008). It will contribute significantly to the electricity supply, especially during the dry season in the Northeast of Brazil, where the temporal variation of the wind potential shows complementarity with the flows of the São Francisco River. However, large scale integration of wind power into electricity grid is still challenging due to intermittency and high spatio-temporal variability of wind speed, and there is a constant effort in the development of new more accurate models for wind speed and wind power forecasting (Jung & Broadwater, 2014).

The evaluation of the wind potential at certain locations requires a detailed statistical analysis of the wind speed and its frequency distribution. Although various probability models for description of wind speed were proposed and evaluated (Jung & Broadwater, 2014), in most cases Weibull and Rayleigh distributions were shown to be suitable for wind energy analysis (Lima & Bezerra Filho, 2010; Rocha, Sousa, Andrade, & Silva, 2012; Pishgar-Komleh, Keyhani, & Sefeedpari, 2015; Ozay & Celiktas, 2016; Wais, 2017). In addition to this classical approach, the analysis of the dynamical structure (complexity) of wind speed can provide valuable information about stochastic processes that generate its temporal and spatial variability, which is important for the planning of wind energy production and for development and evaluation of predictive theoretical and computational models for wind speed and wind power.

During the last decades, various complexity measures such as fractal dimension, multifractal spectra, Lyapunov exponents, and entropies were developed to quantify complexity in real-world time series. These have been widely used to study the complexity of atmospheric phenomena, including wind speed (Millán, Rodríguez, Ghanbarian-Alavijeh, Biondi, & Llerena, 2011; Chang et al., 2012; Li & Zuntao, 2014; Telesca, Lovallo & Kanevski, 2016; Xavier et al., 2018). Among these methods, approaches that rely on entropy are particularly interesting due to their simplicity, lack of intensive computations, and robustness for short non-stationary and noisy data.

In this work we applied Sample Entropy method (Richman & Moorman, 2000) to evaluate the complexity of wind speed time series in Petrolina, which is considered one. of most promising locations for wind energy production in Pernambuco, Brazil (Silva, Alves, Cavalcanti, & Dantas, 2002). We analyzed intra-annual variability of wind dynamical complexity by calculating entropy values for each month during the year 2010, and investigated the relation between complexity and wind power potential.

Material and methods

Data

The data used in this work are wind speed temporal series recorded at Sonda (*Sistema Nacional de Organização de Dados Ambientais*) in Petrolina, Brazil for 09° 04' 08" S latitude, 40° 19' 11 O longitude, and altitude of 387 m. The Sonda station belongs to National System for Environmental Data and was implemented by the National Institute of Space Research (Inpe- *Instituto Nacional de Estudos Espaciais*) to provide a physical infrastructure and the human resources to raise and improve the database of solar and wind energy resources in Brazil. The data was obtained from Inpe and are available at the electronic address <http://sonda.ccst.inpe.br/basedados/petrolina.html>. As one year long measurement of data from an observation station is sufficient in order to determine the wind power potential and project feasibility (Köse, 2004), we chose the most complete dataset: series of 10 min. observations recorded at heights 25 and 50 m in the year 2010 (roughly 52000 data points) which are shown in Figure 1.

Sample entropy

Sample entropy (*SampEn*) was introduced by Richman and Moorman (2000), as a modification of the Approximate entropy (ApEn) method (Pincus, 1991). Both methods evaluate the complexity of short non-stationary signals by examining time series for similar epochs, where more frequent and similar epochs (i.e. increased regularity in the time series) lead to lower values of sample entropy. *SampEn* (m, r, N) is defined as the negative natural logarithm of the conditional probability that two sequences (within the time series) which are similar for m points remain similar at the next point, where self-matches are not included (in contrary with ApEn) in calculating the probability.

Sample entropy algorithm can be described as follows (Richman & Moorman, 2000):

i) For a time series of length N , $u(j), j = 1,\dots,N$, we form $N - m + 1$ vectors of length *m,* $x_m(i), i = 1,\dots, N - m + 1$ where $x_m(i) = \{u(i+k) : k = 0,\dots, m-1\}$;

ii) The distance between vectors $x_m(i)$ and $x_m(j)$ is defined as $d[x_m(i), x_m(j)] = \max\{|u(i+k) - u(j+k)| : k = 0,\dots, m-1\}$;

iii) Next we count the number B_i of vectors $x_m(j)$ such that $d[x_m(i), x_m(j)] \le r$ where r is a tolerance level ($r \equiv r\sigma$, σ - standard deviation of $u(i), i = 1,\dots, N$), $j = 1,\dots, N - m$ and $j \ne i$ to exclude self-matches;

iv) We then define $B_i^m(r) = \dfrac{B_i}{N - m - 1}$ and $B^m(r) = \dfrac{\sum_{i=1}^{N-m} B_i^m(r)}{N - m}$ where $B^m(r)$ is the probability that two vectors will match for m points;

v) We repeat steps (i)-(iv) for vectors of length $m+1$ and count the number A_i of vectors $x_{m+1}(j)$ which are within tolerance level r of $x_{m+1}(i)$ and again we exclude self-matches. We define $A_i^m(r) = \dfrac{A_i}{N - m - 1}$ and $A^m(r) = \dfrac{\sum_{i=1}^{N-m} A_i^m(r)}{N - m}$; where $A^m(r)$ is the probability that two vectors will match for $m+1$ points;

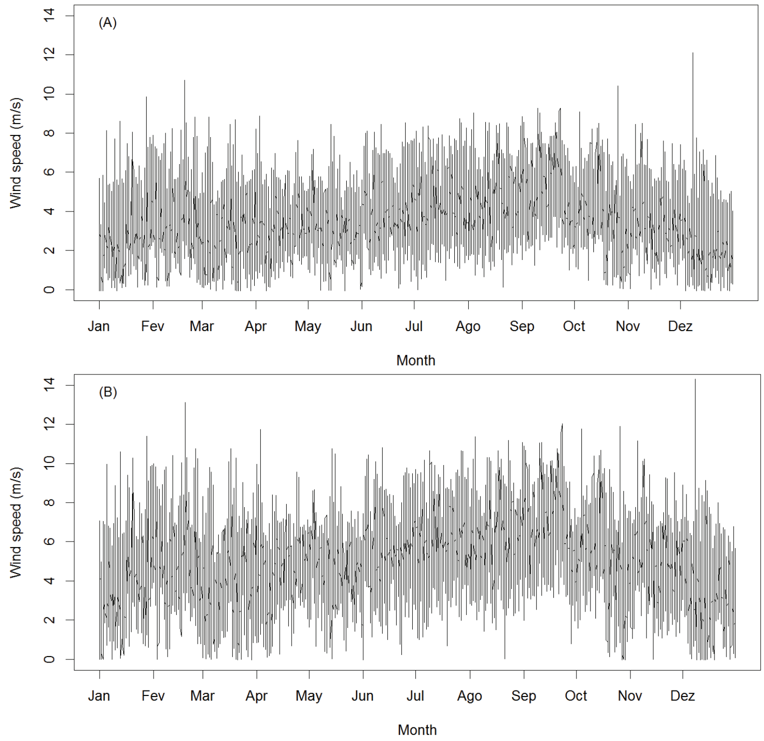

Figure 1. Original wind speed (m s^{-1}) series recorded at the Petrolina station at heights of 25 (A) and 50 m (B).

vi) Sample entropy (*SampEn*) is defined as $SampEn\,(m, r) = \lim_{N\to\infty}\left[-\ln\frac{A^m(r)}{B^m(r)}\right]$ which is estimated by the statistics $SampEn(m,r,N) = -\ln\frac{A^m(r)}{B^m(r)}$.

It can be shown that $A^m(r)$ / $B^m(r)$ = A/B where A is the number of forward matches of length m+1 and B is the number of forward matches of length m. The quantity A/B is precisely the conditional probability that two sequences within a tolerance r for m points remain within r of each other at the next point and *SampEn* (*m,r,N*) can be expressed as -1n(A/B) (Richman & Moorman, 2000). It is the negative natural logarithm of the conditional probability that two sequences similar for m points remain similar at the next point, where self-matches are not included in calculating the probability. SampEn statistics agrees much better with theory than ApEn statistics when applied on random numbers with known probabilistic character. The record length has less effects on SampEn than on ApEn, although both have residual bias for very short record lengths as a result of non-independence of the templates (Richman & Moorman, 2000).

Sample entropy method was used in analyzing physiological processes (Costa, Henriques, Munshi, Segal, & Goldberger, 2014), geophysical signals (Hernández-Pérez, Guzman-Vargas, & Ramírez-Rojas, 2010), hydrological (Chou, 2014) and financial time series (Alvarez-Ramirez, Rodriguez, & Alvarez, 2012).

Results and discussion

The results of the descriptive statistics are presented in Table 1. The power of the wind per unit area is given by $P(v)=1/2pv^3$, where P (Wm - 2) is the power per unit area, p kg m^{-3}) is the air density, and v (ms - 1) is the wind speed (Safari & Gasore, 2010). For most wind turbines, the range of cut-in wind speed (the speed at which the turbine starts producing the energy) is 3.5 – 423.5 m s^{-1}. It is seen from Table 1 that at 25 m, the average wind speed is above cut in level from June to November, and during the whole year at 50 m. At both heights, the average wind speed is the highest (with lowest variation) in September indicating the most favorable period of year for energy generation. Santos, Silva, and Moisés (2013) analyzed seasonal variability of wind speed in Northeast Brazil and also found that the highest wind speed were recorded in winter (June, July, August) and spring (September, October, November). Lima and Bezerra Filho (2010) have also demonstrated the existence of seasonality in wind speeds in the semiarid NEB region (Triunfo) with maximum values in the months of July to November and minimums in March and April.

Table 2 presents the *SampEn* values for *m* = 2 to *m* = 6 (representing the sequences with 2 to 6 consecutive data - consecutive 10 min. observations for intervals of 20 min. up to 1 hour) of the wind speed at heights of 25 and 50 m for each month of 2010. The results provide additional information on wind dynamics that may be relevant for wind power production. The entropy values decrease with m, indicating a greater regularity (greater predictability) of the wind speed when the sequences of consecutive 10 min. values are considered during the intervals of 1 hour.

For June, July, August, and September the average wind speed is above cut-in level at both heights, however the entropy values are higher at 50 m indicating lower regularity in wind dynamics. This property can be considered as unfavorable for wind power generation at 50 m compared to 25 m, contrary to the average speed that is greater at 50 m indicating greater wind potential. For months February, October, and November the average wind speed is above cut-in level at both heights, and the entropy values are smaller at 50m, indicating greater regularity of the series. Both measures (average speed and entropy) indicate favorable conditions for wind power generation at 50 m. For other months (January, March, April, and December) the average wind speed at 25 m is below cut-in level meaning that here is no wind power production, while at 50 m, the average speed is above cut-in level, and the entropy values are lower both indicating favorable conditions for wind energy production. Ahmed and Mandic (2011) analyzed high frequency wind velocity data recorded using a 3D ultrasonic anemometer and found that for vertical component and east-west horizontal component the entropy increases with speed.

As a natural turbulent process wind is an extremely complex phenomenon and its spatial and temporal variability is less understood than for other atmospheric variables. Different factors contribute to surface wind speed variability such as the surface pressure gradient, vertical temperature gradient, surface friction which is directly related to surface roughness, synoptic-scale wind patterns (cyclones and anticyclones) and finally human activities such as urbanization (Klink, 1999). Better understanding of the nature of underlying stochastic process that generate complexity of wind variability requires the use of different methods, classical and emergent, as each provide the information about different aspects of such a process. Being a well established method in complex system science which can evaluate the level of regularity of temporal series, Sample Entropy analysis can contribute (along with classical statistical methods that are primarily based on the evaluation of mean wind speed and wind energy potential) to better planning of use of wind energy as it provides the information about the complexity and regularity of temporal fluctuations of wind speed.

Table 1. Descriptive statistics for wind speed (m s^{-1}) temporal series recorded in Petrolina during the year 2010.

Month	Minimum		Maximum		Average		Standard deviation		Coefficient of variation	
	25m	50m	25m	50m	25m	50m	25m	50m	25m	50m
January	0.00	0.00	9.85	11.40	3.24	4,26	1.31	1.56	0.40	0.37
February	0.12	0.00	10.71	13.13	3.76	4,88	1.41	1.65	0.38	0.34
March	0.00	0.00	8.83	10.77	2.86	3,82	1.33	1.67	0.46	0.44
April	0.00	0.09	8.87	11.73	3.20	4,38	1.14	1.37	0.36	0.31
May	0.00	0.01	8.45	10.76	3.32	4,55	1.09	1.36	0.33	0.30
June	0.07	0.00	8.54	10.82	4.00	5,35	1.53	1.73	0.38	0.32
July	0.00	0.67	8.74	10.65	4.22	5,64	1.42	1.50	0.34	0.26
August	0.13	0.01	9.03	11.38	4.53	5,96	1.38	1.51	0.30	0.25
September	0.99	0.78	9.27	12.04	4.87	6,27	1.33	1.52	0.27	0.24
October	0.44	0.00	10.40	11.91	3.83	4,96	1.42	1.74	0.37	0.35
November	0.00	0.00	8.49	11.15	3.81	4,94	1.23	1.44	0.32	0.29
December	0.00	0.00	12.11	14.30	2.79	3,62	1.34	1.64	0.48	0.45

Table 2. Sample Entropy (*m* = 2, ..., 6; *r* = 0, 20) for wind speed temporal series recorded in Petrolina during the year 2010.

Month	*m*=2		*m*=3		*m*=4		*m*=5		*m*=6	
	25m	50m	25m	50m	25m	50m	25m	50m	25m	50m
January	1.21	1.16	1.15	1.12	1.11	1.08	1.19	1.06	1.09	1.05
February	1.21	1.20	1.18	1.17	1.14	1.12	1.11	1.08	1.11	1.09
March	1.26	1.17	1.21	1.13	1.17	1.08	1.14	1.02	1.08	1.01
April	1.23	1.18	1.18	1.12	1.14	1.08	1.11	1.03	1.03	1.00
May	1.24	1.22	1.16	1.18	1.11	1.14	1.04	1.10	1.02	1.04
June	1.08	1.11	1.00	1.03	0.95	0.98	0.89	0.91	0.86	0.87
July	1.12	1.20	1.04	1.13	0.97	1.05	0.91	1.00	0.87	0.97
August	1.25	1.28	1.17	1.21	1.12	1.15	1.07	1.11	1.03	1.08
September	1.35	1.34	1.26	1.27	1.18	1.20	1.11	1.15	1.05	1.11
October	1.21	1.15	1.15	1.11	1.09	1.08	1.05	1.07	1.02	1.02
November	1.32	1.29	1.27	1.23	1.23	1.20	1.19	1.19	1.18	1.15
December	1.21	1.13	1.17	1.09	1.12	1.05	1.09	1.01	1.07	1.00

Conclusion

The results showed of SampEn analysis of wind speed in Petrolina that: i) Wind dynamics is more regular when considering consecutive 10 min. values at 1 hour intervals, indicating this temporal scale as better for forecasting models and ii) Although wind potential is higher at 50 m, in some periods, entropy values were also indicating less regular dynamics, which is unfavorable condition for the operation of wind turbines and consequently there is less efficiency in the capture of wind energy for wind electricity production.

References

Ahmed, M. U., & Mandic, D. P. (2011). Multivariate multiscale entropy: A tool for complexity analysis of multichannel data. *Physical Review E, 84*(6), 1-10. doi: 10.1103/PhysRevE.84.061918

Alvarez-Ramirez, J., Rodriguez, E., & Alvarez, J. (2012). A multiscale entropy approach for market efficiency. *International Review of Financial Analysis, 21*, 64-69. doi: 10.1016/j.irfa.2011.12.001

Chang, T.-P., Ko, H.-H., Liu, F.-J., Chen, P.-H., Chang, Y.-P., Liang, Y.-H., ... Chen, Y.-H. (2012). Fractal dimension of wind speed time series. *Applied Energy, 93*, 742-749. doi: 10.1016/j.apenergy.2011.08.014

Chou, C.-M. (2014). Complexity analysis of rainfall and runoff time series based on sample entropy in different temporal scales. *Stochastic Environmental Research and Risk Assessment, 28*(6), 1401-1408. doi: 10.1007/s00477-014-0859-6

Costa, M. D., Henriques, T., Munshi, M. N., Segal, A. R., & Goldberger, A. L. (2014). Dynamical glucometry: Use of multiscale entropy analysis in diabetes. *Chaos: An Interdisciplinary Journal of Nonlinear Science, 24*(3), 033139. doi: 10.1063/1.4894537

Dutra, R. M., & Szklo, A. S. (2008). Incentive policies for promoting wind power production in Brazil: Scenarios for the Alternative Energy Sources Incentive Program (PROINFA) under the New Brazilian electric power sector regulation. *Renewable Energy, 33*(1), 65-76. doi: 10.1016/j.renene.2007.01.013

Hernández-Pérez, R., Guzman-Vargas, L., Ramírez-Rojas, A., & Angulo-Brown, F. (2010). Pattern synchrony in electrical signals related to earthquake activity. *Physica A: Statistical Mechanics and its Applications, 389*(6), 1239-1252. doi: 10.1016/j.physa.2009.11.036.

Jung, J., & Broadwater, R. P. (2014). Current status and future advances for wind speed and power forecasting. *Renewable and Sustainable Energy Reviews, 31*, 762-777. doi: 10.1016/j.rser.2013.12.054

Klink, K. (1999). Trends in mean monthly maximum and minimum surface wind speeds in the coterminous United States, 1961 to 1990. *Climate Research, 13*(3), 193-205.

Köse, R. (2004). An evaluation of wind energy potential as a power generation source in Kütahya, Turkey. *Energy Conversion and Management, 45*(11-12), 1631-1641. doi: 10.1016/j.enconman.2003.10.016

Li, Q., & Zuntao, F. (2014). Permutation entropy and statistical complexity quantifier of nonstationarity effect in the vertical velocity records. *Physical Review E, 89*(1), 012905. doi: 10.1103/PhysRevE.89.012905

Lima, L. A., & Bezerra Filho, C. R. (2010). Wind energy assessment and wind farm simulation in Triunfo – Pernambuco, Brazil. *Renewable Energy, 35*(12), 2705-2713. doi: 10.1016/j.renene.2010.04.019

Lindenberg, S. (2009). *20% wind energy by 2030: increasing wind energy contribution to us electricity supply: diane publishing*. Retrieved from https://www.energy.gov/eere/wind/20-wind-energy-2030-increasing-wind-energys-contribution-us-electricity-supply

Millán, H., Rodríguez, J., Ghanbarian-Alavijeh, B., Biondi, R., & Llerena, G. (2011). Temporal complexity of daily precipitation records from different atmospheric environments: Chaotic and Lévy stable parameters. *Atmospheric Research, 101*(4), 879-892. doi: 10.1016/j.atmosres.2011.05.021

Ozay, C., & Celiktas, M. S. (2016). Statistical analysis of wind speed using two-parameter Weibull distribution in Alaçatı region. *Energy Conversion and Management, 121*, 49-54. doi: 10.1016/j.enconman.2016.05.026

Pishgar-Komleh, S. H., Keyhani, A., & Sefeedpari, P. (2015). Wind speed and power density analysis based on Weibull and Rayleigh distributions (a case study: Firouzkooh county of Iran). *Renewable and Sustainable Energy Reviews, 42*, 313-322. doi: 10.1016/j.rser.2014.10.028

Richman, J. S., & Moorman, J. R. (2000). Physiological time-series analysis using approximate entropy and sample entropy. *American Journal of Physiology-Heart and Circulatory Physiology, 278*(6), H2039-H2049. doi: 10.1103/PhysRevE.71.021906

Rocha, P. A. C., Sousa, R. C., Andrade, C. F., & Silva, M. E. V. (2012). Comparison of seven numerical methods for determining Weibull parameters for wind energy generation in the northeast region of Brazil. *Applied Energy, 89*(1), 395-400. doi: 10.1016/j.apenergy.2011.08.003

Safari, B., & Gasore, J. (2010). A statistical investigation of wind characteristics and wind energy potential based on the Weibull and Rayleigh models in Rwanda. *Renewable Energy, 35*(12), 2874-2880. doi: 10.1016/j.renene.2010.04.032

Silva, B. B. D., Alves, J. J. A., Cavalcanti, E. P., & Dantas, R. T. (2002). Potencial eólico na direção predominante do vento no Nordeste brasileiro. *Revista Brasileira de Engenharia Agrícola e Ambiental, 6*(3), 431-439. doi: 10.1590/S1415-43662002000300009

Santos, A. T. S., Silva, S., & Moisés, C. (2013). Seasonality, interannual variability, and linear tendency of wind speeds in the Northeast Brazil from 1986 to 2011. *The Scientific World Journal, 2013*, 1-10. doi: 10.1155/2013/490857

Telesca, L., Lovallo, M., & Kanevski, M. (2016). Power spectrum and multifractal detrended fluctuation analysis of high-frequency wind measurements in mountainous regions. *Applied energy, 162*, 1052-1061. doi: 10.1016/j.apenergy.2015.10.187

Van Hulle, F., Tande, J. O., Uhlen, K., Warland, L., Korpås, M., Meibom, P., ... Giebel, G. (2009). *Integrating wind: Developing Europe power market for the large-scale integration of wind power*. Bruxelas, BE: European Wind Energy Association (EWEA).

Wais, P. (2017). Two and three-parameter Weibull distribution in available wind power analysis. *Renewable Energy, 103*, 15-29. doi: 10.1016/j.renene.2016.10.041

Xavier, S. F. A., da Silva Jale, J., Stosic, T., dos Santos, C. A. C., & Singh, V. P. (2018). An application of sample entropy to precipitation in Paraíba State, Brazil. *Theoretical and Applied Climatology*, 1-12. doi: 10.1007/s00704-018-2496-3