Comparing wind generation profiles: A circular data approach

Ana Martins
ISEL; BRU-UNIDE
Lisbon, Portugal
anamartins@deea.isel.ipl.pt

Alda Carvalho
ISEL; CEMAPRE
Lisbon, Portugal
acarvalho@adm.isel.pt

Jorge A. M. Sousa
ISEL/ADEEEA; INESC-ID
Lisbon, Portugal
jsousa@deea.isel.ipl.pt

***Abstract*— The importance of wind power energy for energy and environmental policies has been growing in past recent years. However, because of its random nature over time, the wind generation cannot be reliable dispatched and perfectly forecasted, becoming a challenge when integrating this production in power systems. In addition the wind energy has to cope with the diversity of production resulting from alternative wind power profiles located in different regions. In 2012, Portugal presented a cumulative installed capacity distributed over 223 wind farms [1].**

In this work the circular data statistical methods are used to analyze and compare alternative spatial wind generation profiles. Variables indicating extreme situations are analyzed. The hour(s) of the day where the farm production attains its maximum daily production is considered. This variable was converted into circular variable, and the use of circular statistics enables to identify the daily hour distribution for different wind production profiles.

This methodology was applied to a real case, considering data from the Portuguese power system regarding the year 2012 with a 15-minutes interval. Six geographical locations were considered, representing different wind generation profiles in the Portuguese system.

***Index Terms*-- Renewable generation, wind power generation, circular statistics.**

I. Introduction

Due to the increasing importance of wind power generation, many attention has been paid to the study of the wind characteristics – wind speed and wind direction. In addition, to make more profitable the performance of wind turbines, is important to understand the behavior of wind power generation. In particular the characterization of the daily extreme values is very relevant. In [2] the extreme power generation events are characterized and estimated using meteorological records.

In this work we propose to use the methods of circular statistics to study the daily temporal patterns of extreme wind generation events. The maximum daily production is considered, taking into account the spatial diversity of wind generation. This proposed a novel approach to wind power analysis that enables to identify times of day when important events occur, which may be of special interest for the system operator in order to better prepare the maintenance and operation of the grid.

Circular statistics is a set of techniques for dealing with directional data ([3], [4], [5], [6]). Typically, these data are expressed angular measurements, such as wind direction. The tools of circulars statistics have been used for modeling the random nature of this variable ([7], [8]).

Circular statistics can also be used to analyze any kind of data that are cyclic in nature, like time-of-day data measured on a 24h clock, with 0:00 corresponding to 0^o, 6:00 to 90^o, 18:00 to 180^o, and 24:00 to 360^o [3].

Due to the circular geometry of the sample space, standard statistical techniques cannot be used to model circular data. Consider a simple example: consider the times 23:30 and 00:30, using the usual mean between these two values we obtain after covert to decimal hours, (23.5+0.5)/2=12, which corresponds to 12:00, right opposite the circular mean which is 00:00.

This paper begins with an introduction of the circular statistics. Following this, a real case application is presented as well as the corresponding results. In the last section of the paper, final conclusions are presented.

II. Circular statistics

A circular observation can be seen as a point on an unit ratio circumference or as a unit vector. Once the initial direction and an orientation has been chosen, the circular observation, θ, represents the angle from the initial direction and it can be specified using polar coordinates $(1,\theta)$ or Cartesian coordinates $(\cos\theta, \sin\theta)$. The observation θ can be either measured in degrees or radians.

A. Descriptives Measures

Let $\theta_1, \theta_2, \ldots, \theta_n$ be a sample of circular data. The center of gravity in two-dimensional space has coordinates $(\bar{C}, \bar{S})$ where $\bar{C}$ and $\bar{S}$ are the mean of X and Y coordinates, respectively, that is:

$$\bar{C} = \frac{1}{n}\sum_{i=1}^{n} \cos\theta_i \qquad \bar{S} = \frac{1}{n}\sum_{i=1}^{n} \sin\theta_i \tag{1}$$

The mean direction, $\bar{\theta}$, is defined as the angle of the vector $(\bar{C}, \bar{S})$:

$$\bar{\theta} = \begin{cases} atan\left({}^{S}/_{C}\right), & if\ C > 0, S \geq 0 \\ atan\left({}^{S}/_{C}\right) + 2\pi, & if\ C > 0, S < 0 \\ atan\left({}^{S}/_{C}\right) + \pi, & if\ C < 0 \\ {}^{\pi}/_{2}, & if\ C = 0, S > 0 \\ {}^{3\pi}/_{2}, & if\ C = 0, S < 0 \\ undefined, & if\ C = 0, S = 0 \end{cases} \tag{2}$$

where the function $atan$ is the inverse tangent function.

The mean resultant length, $\bar{R}$, is the length of the vector $(\bar{C}, \bar{S})$ which value lies in $[0,1]$. Higher values of $\bar{R}$ are associated with less spread of the data and when $\bar{R} = 1$, all the data points are coincident [3].

The sample circular variance can be defined as [3]:

$$V = 1 - \bar{R} \qquad (0 \leq V \leq 1) \tag{3}$$

The greater the values of V, the more dispersed is the distribution.

Furthermore, the sample circular standard deviation is given by:

$$\vartheta = \sqrt{-2log\bar{R}} \tag{4}$$

Similarly to linear case, circular standard deviation takes only positive values and has no upper bound. Further details on this definition can be seen in [3] or [5].

The circular range is the length of the shortest arc on the circle containing all the sample data.

Another central location measure is the median, $\tilde{\theta}$. A median axis divides the circle in two slices in a way that the data is divided in two equal length groups. Then the median is the end of the median axis with more nearer data points. For a formal definition, see [3]. For quantiles determination the sample is treated as linear [3].

B. The Circular Uniforme and Von Mises Distributions

When dealing with circular statistics two fundamental distributions should be considered.

The first distribution is the circular uniform that reflects the case where all directions are equally spread all over the circle. The total probability is spread out uniformly and we get the circular uniform distribution:

$$f(\theta) = \frac{1}{2\pi}, \qquad 0 \leq \theta < 2\pi \tag{6}$$

When the distribution is not uniform, it is often to think on a distribution with some concentration on one (unimodal) or more (multimodal) preferred directions. The most common unimodal distribution is the circular normal distribution, also known as the von Mises distribution:

$$f(\theta; \mu, \vartheta) = \frac{1}{2\pi I_0(\vartheta)} e^{\vartheta\cos(\theta-\mu)}, \quad 0 \leq \theta < 2\pi, \tag{7}$$

where $0 \leq \mu < 2\pi$ and $\vartheta \geq 0$ are parameters and

$$I_0(\vartheta) = \frac{1}{2\pi}\int_0^{2\pi} e^{\vartheta\cos\theta} d\theta = \sum_{r=0}^{\infty}\left(\frac{\vartheta}{2}\right)^{2r}\left(\frac{1}{r!}\right)^2. \tag{8}$$

In circular statistics the Von Mises distribution takes the role of the normal distribution in standard linear statistics. In fact, it is shaped like the normal distribution, except that its tails are truncated.

C. Non-Parametric goodness-of-fit testes

When analyzing circular data it is important to check if the data follow a uniform distribution, which means that there is no modal region (or regions) in the distribution of the data. In addition, many parametric procedures in circular statistics require von Mises distribution. Thus, it is also important to check if the data follow this distribution.

In assessing whether von Mises or circular uniform distributions are suited to a data-set, the Watson test can be applied [4]. Like usual goodness of fit tests, it compares the empirical and a reference distribution, and the null hypothesis states that the data distribution is equal to the reference distribution. The alternative hypothesis is that the sample data distribution is not the stated reference distribution.

Furthermore when dealing with more than one sample, a keynote is the comparison of the empirical distributions. The Mardia-Watson-Wheeler (or uniform score) test is also a non-parametric test that enables to compare several samples concluding whether its populations´ distributions are identical or not. The null hypothesis is that the distributions of several populations are identical, against the alternative that the distributions are not the same. For further remarks on this test see [3] or [5].

III. Case Study

The previously described methodology was applied to a real case, considering data from the Portuguese power system of the year 2012 with a 15-minute interval [9]. Six geographical locations were considered, representing different wind generation profiles in the Portuguese system, as shown in Fig. 1. Although there are some missing values, for each location, the datasets are the farm production, corresponding to a time series with 33696 observations.

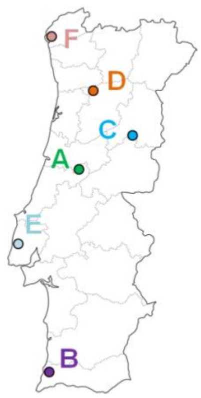

Figure 1. Location of wind farms under analysis.

A. Variables

In order to understand the different wind production profiles, the characterization of extreme conditions is very important.

Since the main goal of this work is to compare different wind generation profiles in the Portuguese system, a standardization procedure was performed by considering the production of farm *i* divided by the maximum production observed in all data set for that farm.

The hour(s) of the day where the farm attain its daily maximum production were analyzed. For this purpose, the study was developed using the variable ($HMaxD_i$,$PMaxD_i$), where $HMaxD_i$ is the hour of the day where there was the value of the maximum daily power and $PMaxD_i$ is the corresponding power in MW, $i \in \{A,B,C,D,E,F\}$ (see Fig. 1).

B. Circular data

The variables $HMaxD_i$ were converted into circular variables, enabling to identify the daily hour distribution of extreme values for different wind production profiles.

The hours of the day were the maximum occurs can be any hour from 00:00 to 23:45 (Fig. 2) with a 15-minute interval. Thus, the circle was divided into 96 equal portions and the circular value, in radians, is obtained by:

$$\mathrm{HMaxDr_i} = (k-1)\frac{2\pi}{96} \qquad (1)$$

For example, consider a situation where the farm production attains its maximum capacity at 00:45; this hour corresponds to the 4^{th} position (k=4) and $HMaxD_i$ = 0.19365.

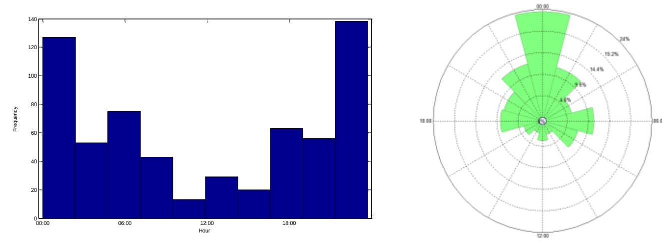

Figure 2. Histogram and Rose diagram of $HMaxD_A$.

The frequency distribution of the hours of the day where the maximum occurs can be seen in Fig. 2 (histogram on the left and rose diagram on the right). Usually, it is better to plot a circular variable as $HMaxD_A$ in a rose diagram, which is a circular histogram that displays directional data and the frequency of each class. In this case, it is possible to see a concentration around the preferred direction that corresponds to midnight.

IV. Results

Meteorologists usually use wind rose diagrams to give a succinct view of how wind speed (linear variable) and direction (circular variable) are typically distributed at a particular location. Making the analogy with our data, the hour were the maximum was obtained is the circular variable and the corresponding power is the linear variable.

In Fig. 3 one can see the wind rose of the pair ($HMaxD_i$,$PMaxDn_i$). For each location, one can see a different distribution of the hours of the day were the maximum occurs. Also, comparing the intensity - the value of the maximum power, is possible to see very different wind production profiles.

This information might be of special interest to the system operator since the load flow in the power grid will be influenced by the random generation of the several wind farms. In this regard, knowing the most likely hours of high generation of each wind location will enable the system operator to anticipate any constraints that the grid may face and prepare the maintenance and operational plans accordingly.

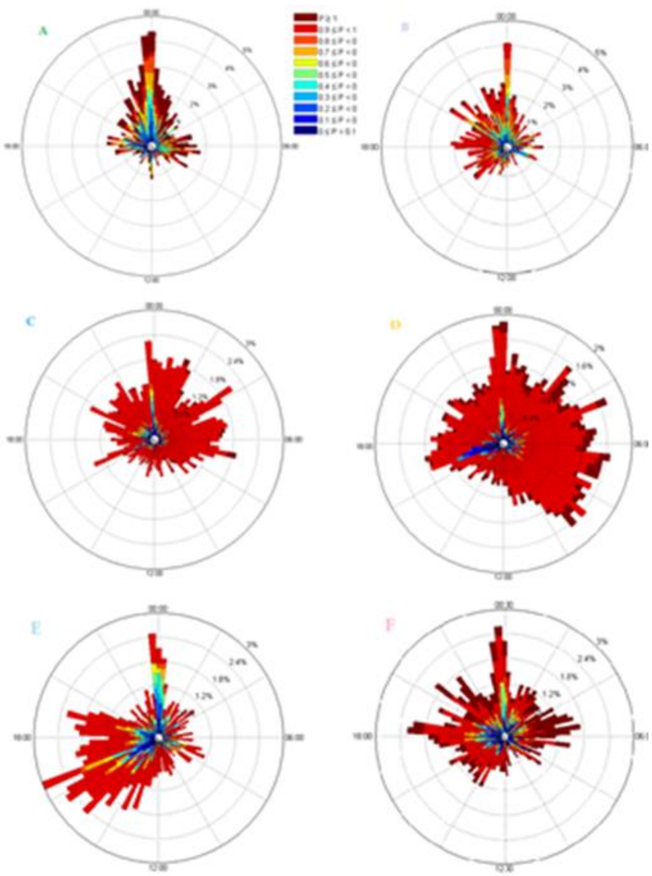

Figure 3. Histogram of HMaxD$_i$, $i \in \{A,B,C,D,E,F\}$.

For the six wind production profiles data sets, several descriptive statistics were computed, the results are presented in tables I and II. The data analysis results were obtained using the R software and "Circular" and "CircStats" packages.

The sample sizes are very different (Table I) reflecting the fact that the farm production attains its daily maximum production for several times in a single day. The mean direction is always obtained in night hours, mostly between 23:45 and 01:30. The profile D is the only one who presented a more distant mean value (04:00). The median values keep up with these results which are depictured in Fig. 4. In this figure a new proposal data representation is presented by considering circular boxplot representation. The comparison of multiple samples through boxplots is very usual with linear data. This new representation enables the circular comparison.

TABLE I. LOCATION MEASURES RESULTS

	Wind Profile					
	A	***B***	***C***	***D***	***E***	***F***
Sample Size	617	610	1376	1812	804	814
Mean **($\bar{\theta}$)**	00:15	23:45	01:30	04:00	23:45	23:45
Min	12:00	09:15	13:15	15:15	04:30	10:00
$Q_{0.25}$	04:00	00:30	05:45	08:00	23:45	03:00
Median **($\tilde{\theta}$)**	00:00	23:45	01:15	03:30	16:45	23:45
$Q_{0.75}$	23:45	17:30	23:45	23:45	13:45	17:45
Max	12:15	09:30	13:30	15:30	04:45	10:15

TABLE II. DISPERSION MEASURES RESULTS (IN RAD)

	Wind Profile					
	A	***B***	***C***	***D***	***E***	***F***
Range	6.09	6.02	6.22	6.22	6.22	6.15
Mean Resultant Length **($\bar{R}$)**	0.39	0.38	0.25	0.16	0.30	0.20
Variance (V)	0.61	0.62	0.75	0.84	0.70	0.80
Standart Deviation (ϑ)	1.37	1.40	1.67	1.91	1.56	1.79

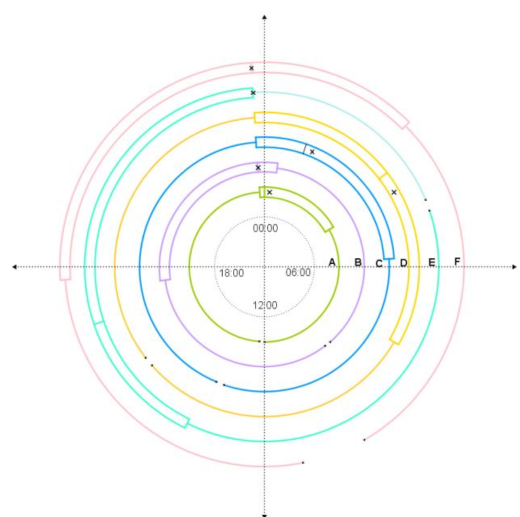

Figure 4. Circular boxplots

Regarding the dispersion of the data, several statistics results are presented in table II. The range of data is almost the entire circumference, meaning that the daily maximum production could be reached almost at any time of the day.

The farms A and B are the ones with less dispersed data. On the other hand the D profile data present the higher spread. In fact, the A and B farms are the ones with fewer sample observation and profile D have the greater sample size.

For testing if any of the six wind profiles follow a circular uniform distribution, the Watson test was performed. The results of the six performed tests indicate that the circular uniform distribution is not suitable to describe the empirical distributions (in all cases, p-value<0.01). This result underline that in this samples not all directions are equally spread all over the circle which is already been suggested by the analyses of the circular histograms presented in Fig 3.

Furthermore, the Watson test was also performed using the von Mises distribution. The conclusion is that none of the six sample data sets have been drawn from a population with von Mises distribution (in all cases, p-value<0.01).

For comparing the distributions of the six wind profiles, the Mardia-Watson-Wheeler was performed. The test enables to conclude that the six distributions are not equal (p-value < 2.2E-16). Also, in order to compare each pair of distributions, the Watson test for two samples was applied ([3]). Fifteen tests were performed for comparing each pair of wind profiles. In all of them, the p-values were lower than 0.001, indicating the rejection of the null hypotheses that the distributions are equal. In conclusion, all the six wind profiles have different empirical hourly distributions for the daily maximum productions.

This is a useful result that reinforces the smoothing effect that is known from several studies about wind generation from aggregated wind farms located in different regions.

V. CONCLUSIONS

In this work a new approach based in circular statistics is used in order to characterize the daily occurrence of extreme wind power generation events. The hour(s) of the day where the farm production attains its maximum daily production is considered. In addition, the geographical diversity of wind production is taken into account.

Considering data from the Portuguese power system of the year 2012 with a 15-minute interval, six geographical locations were considered, representing different wind generation profiles in the Portuguese system.

For the six wind production profiles data sets, several descriptive statistics were computed. The results showed that the mean direction is always obtained in night hours, mostly between 23:45 and 01:30. A new proposal data representation is presented by considering circular boxplot representation, which enables the circular comparison. The range of data is almost the entire circumference, meaning that the daily maximum production could be reached almost at any time of the day.

Several non-parametric tests enable to conclude that all the six wind profiles have different empirical hourly distributions for the daily maximum productions. These results step up the well known smoothing effect due to the aggregation of different wind production profiles.

Finally, this work results bring same insight into the characterization of the occurrence of wind power generation events, which is important for system planning purposes.

VI. ACKNOWLEDGMENTS

The authors express their gratitude to Eng. Rui Pestana for his support in data collection.
This work was supported by national funds through Fundação para a Ciência e a Tecnologia (FCT) with reference UID/CEC/50021/2013.

REFERENCES

[1] 2010 IEA Wind Annual Report, IEA - International Energy Agency, Jul.2011. [Online]. Available: http://www.ieawind.org/.
[2] Cannon, D.J.; D.J. Brayshaw, D.J.; Methven, J.; Coker, P.J.; Lenaghan, D. (2015). Using reanalysis data to quantify extreme wind power generation statistics: A 33 year case study in Great Britain. Renewable Energy, 75, 767-778.
[3] Fisher, NI. (1993). Statistical Analysis of Circular Data. Cambridge University Press. ISBN 0-521-35018-2.
[4] Jammalamadaka, S. Rao; Sengupta, A. (2001). Topics in Circular Statistics. World Scientific Publishing Company. ISBN 978-981-02-3778-3.
[5] Mardia, KV. and Jupp P., Directional Statistics (2nd edition), John Wiley and Sons Ltd., 2000. ISBN 0-471-95333-4.
[6] Lee, A. (2010). Circular data. Wiley Interdisciplinary Reviews: Computational Statistics, Vol. 2, Issue 4, 477–486.
[7] Carta, J. A., Ramírez, P. & Bueno, C. (2008). A joint probability density function of wind speed and direction for wind energy analysis. Energy Conversion & Management, 49, 1309-1320.
[8] Qin, X., Zhang, J. and Yan, X. (2010). A New Circular Distribution and Its Application to Wind Data. Journal of Mathematics Research Vol. 2, No. 3, 12-17. ISSN 1916-9795 E-ISSN 1916-9809.
[9] REN, technical data 2012, [Online]. Available: [Accessed in Jan 10, 2014].