Manuscript submitted to: Volume 2, Issue 4, 342-358.
AIMS Energy DOI: 10.3934/energy.2014.4.342
Received date 1 July 2014, Accepted date 4 September 2014, Published date 29 September 2014

Research article

Polynomial approximations of the Normal to Weibull Distribution transformation

Andrés Feijóo*, Daniel Villanueva

Departamento de Enxeñería Eléctrica, Universidade de Vigo, 36310 Vigo, Spain

*** Correspondence:** E-mail: afeijoo@uvigo.es; Tel: +34-986-812055

Abstract: Some of the tools that are generally employed in power system analysis need to use approaches based on statistical distributions for simulating the cumulative behavior of the different system devices. For example, the probabilistic load flow. The presence of wind farms in power systems has increased the use of Weibull and Rayleigh distributions among them. Not only the distributions themselves, but also satisfying certain constraints such as correlation between series of data or even autocorrelation can be of importance in the simulation. Correlated Weibull or Rayleigh distributions can be obtained by transforming correlated Normal distributions, and it can be observed that certain statistical values such as the means and the standard deviations tend to be retained when operating such transformations, although why this happens is not evident. The objective of this paper is to analyse the consequences of using such transformations. The methodology consists of comparing the results obtained by means of a direct transformation and those obtained by means of approximations based on the use of first and second degree polynomials. Simulations have been carried out with series of data which can be interpreted as wind speeds. The use of polynomial approximations gives accurate results in comparison with direct transformations and provides an approach that helps explain why the statistical values are retained during the transformations.

Keywords: Normal, Weibull, Rayleigh distributions, correlation, autocorrelation, polynomial approximation.

1. Introduction.

The increasing presence of wind farms (WF) in electrical power networks has made it important to simulate correlated wind speeds. The use of wind speed series in combination with the wind turbine (WT) power curves is common for the resolution of more than one typical problem in electrical power network analysis.

In order to attain a solution for some of these problems it is necessary to deal with wind speed series satisfying features regarding the frequency distribution of wind speeds and the correlation between series at different sites. There is wide agreement on considering the Weibull distribution as the best continuous approximation for the frequency distribution of wind speed in a site [1] and Kavasseri presents a study of the phenomena associated to the correlation [2]. Spatial correlation is explicity mentioned by Damousis *et al.* [3] and autocorrelation has been studied by Brown *et al.* [4], and also by Song and Hsiao [5].

Other solutions to the problem stated above have been proposed in different papers, and a review of them has been presented by Feijóo *et al.* [6], especially for the case when no additional chronological constraints are imposed.

Correia and Ferreira de Jesus [7] use the process of obtaining a Weibull distribution as a combination of two Normal distributions while Segura *et al.* [8] obtain correlated Weibull distributions from Uniform distributions.

A method based on the conditional probability theorem was presented by Karaki *et al.* [9] and an approach based on the inverse discrete Fourier transform was presented by Young and Beaulieu [10].

Huang and Chalabi [11] present a method based on chronological series and Shamshad *et al.* [12] a different one based on Markovian models.

Villanueva *et al.* [13] propose a solution for an application to the economic dispatch problem in networks with penetration of wind farms. It consists of the simulation of wind speed series satisfying statistical constraints such as those mentioned above together with an additional one regarding autocorrelation, which added a chronological nuance to the proposed method. As a result, series of correlated wind speeds are obtained with a very high degree of accuracy regarding correlations, Weibull parameters and autocorrelation of the series, although it is not so clear why these features are retained through the proposed transformation, which is a transformation from Normal to Weibull distributions.

The main objective of this paper is to return to this subject and offer an approach which helps understand and discuss why those features are retained through such a transformation. Polynomial approximations of first and second degree will be used to achieve this.

In the rest of the paper, a Normal distribution with mean μ and standard deviation σ is denoted as $N(\mu, \sigma)$, and a Weibull distribution with scale parameter c and shape parameter k is denoted as $W(c, k)$. Uniform distributions are also mentioned, and $U(0, 1)$ denotes one of these distributions in the interval $[0, 1]$. Subindices such as in μ_x or in μ_u are used for distinguishing between different data series, for example, for denoting the mean of the series $\{x_i\}_{i\in\{1,2,\ldots,M\}}$, or of the series $\{u_j\}_{j\in\{1,2,\ldots,M\}}$ respectively.

2. Normal to Weibull transformation.

A $N(0, 1)$ distribution can be converted into a $W(c, k)$ one by means of the following transformation [13]:

$$u = c\left(-log\left(\frac{1 - erf\left(\frac{x}{\sqrt{2}}\right)}{2}\right)\right)^{\frac{1}{k}} \quad (1)$$

where x is a value belonging to a normally distributed series, u to a Weibull one, log represents the natural logarithm and erf the error function defined as follows:

$$erf(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}dt \quad (2)$$

For such a transformation the fact must be taken into account that the cumulative distribution function (CDF) of normally distributed data with mean value μ_x and standard deviation σ_x can be expressed as $F_x(x) = \frac{1}{2}\left(1 + erf\left(\frac{x-\mu_x}{\sigma_x\sqrt{2}}\right)\right)$. For the $N(0, 1)$ distribution the transformation is $F_x(x) = \frac{1}{2}\left(1 + erf\left(\frac{x}{\sqrt{2}}\right)\right)$. In

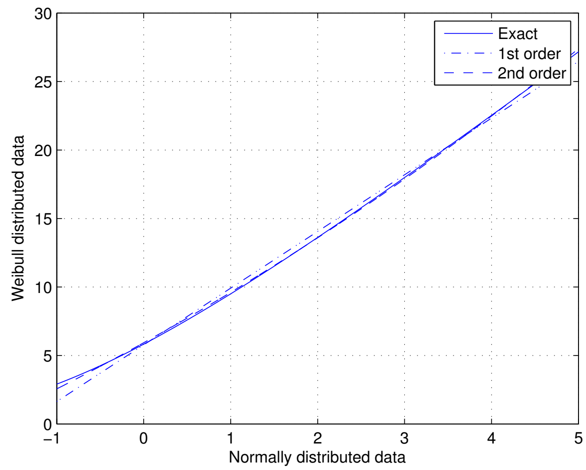

FIGURE 1. **Normal against Weibull data.**

the case of data belonging to a Weibull distribution, with scale parameter c and shape parameter k, the transformation is $F_u(u) = 1 - e^{-\left(\frac{u}{c}\right)^k}$. According to this notation, (1) is obtained by matching both distribution functions and clearing up the variate u, as $u = F_u^{-1}(F_x(x))$, which is distributed as a Weibull one. Both functions can be equalled due to the fact that any distribution function provides a Uniform distributed variate, i.e., $F_x(x) \sim U(0,1)$ and $F_u(u) \sim U(0,1)$.

When such a transformation is performed, a representation of the normally distributed values against the Weibull distributed ones gives Figure 1 as a result, where the Weibull values cover an interval including $[3, 25]$, i.e., the interval of wind speed values in which most WTs can run, which is the reason for being considered an important interval in this paper.

The graph represented in Figure 1 has been obtained for a Weibull distribution with parameters $c = 7$ and $k = 2$. A value of $k = 2$ corresponds to a particular case of the Weibull distribution known as Rayleigh distribution. Similar curves are obtained for other values of the pair (c, k), although they have not been represented here.

So far, in Figure 1 attention must be paid to the line with the legend Exact, obtained from (1).

3. Polynomial approximations.

A visual inspection of Figure 1 and the experience of having carried out many different simulations lead the authors to think that polynomial approximations to (1) could give accurate results.

If a least squares approximation is applied to the set of values of such a transformation, then the other curves of Figure 1 are obtained, i.e., those corresponding to the legends 1st order and 2nd order.

Both curves have been obtained by means of polynomial approximations of first and second order, respectively, such as:

$$f(x) = \sum_{n=0}^{D} a_n x^n \quad (3)$$

where $D \in \{1, 2\}$.

This means that u can be expressed as $u = f(x) + \epsilon(x)$, where $\epsilon(x)$ is a small error that, when neglected, involves accepting that $u \approx f(x)$.

The values of the constants obtained for these approximations are $a_0 = 5.7660$, $a_1 = 4.1384$ in the case of the first degree polynomial (i.e., $D = 1$) and $a_0 = 5.9301$, $a_1 = 3.5227$ and $a_2 = 0.1539$ for the second degree one (i.e., $D = 2$).

Summarizing, for the transformation proposed in (1) there are the following possible approximations.

$$f_1(x) = 5.7660 + 4.1384 \cdot x \quad (4)$$

$$f_2(x) = 5.9301 + 3.5227 \cdot x + 0.1539 \cdot x^2 \quad (5)$$

For each approximation, a measure of error can be defined as:

$$e_{f_k(x)} = \frac{\sqrt{\sum_{i=1}^{M} (u_i - f_k(x_i))^2}}{M} \quad (6)$$

where $k \in \{1, 2\}$, M is the number of samples, u_i are calculated from (1) and $f_k(x_i)$ are calculated from (4).

The error made in a set of 100,000 samples, according to (6), when obtaining the approximations of Figure 1 came to 0.0043 for the first order polynomial and 0.0011 for the second order one. In different simulations the values of these errors can be slightly different, but very close to the one given here.

Although the differences in error according to (6) are not so important, they are bigger when dealing with absolute errors. If an absolute error is measured according to:

$$e_{abs_{f_k(x)}} = max\{|u_i - f_k(x_i)|, i \in \{1, 2, ..., M\}\} \quad (7)$$

then in different simulations, values around 1.27 have been found for the first degree approximation and around 0.34 for the second degree one.

In Figure 2, the histogram obtained by means of a transformation based on (1) has been included with the notation Exact. In the same Figure, 1st deg. pol. and 2nd deg. pol. denote the histograms corresponding to both proposed approximations. This histogram has been obtained with 100,000 samples.

An observation must be made here. A first order polynomial does not convert a Normal distribution into a Weibull one. It is a linear conversion and the result has to be a Normal distribution, and this can be appreciated in Figure 2. However, a second order polynomial has the contribution of the second order term, with a strong trend to make the function asymmetric, and its similarity with a Weibull distribution is stronger.

Finally, in Table 1 some of the moments of u, $f_1(x)$ and $f_2(x)$ are given. As can be seen, both approximations have a lower mean and a higher standard deviation. It is intuitive that the three series are highly correlated. And indeed, correlations between them are $\rho_{uf_1(x)} = 0.9862$, $\rho_{uf_2(x)} = 0.9943$. If Spearman rank correlations, $\rho_{uf_i(x)_{Sp}}$, $i \in \{1, 2\}$, are calculated, then the result is $\rho_{uf_1(x)_{Sp}} = 1$ and $\rho_{uf_2(x)_{Sp}} = 1$.

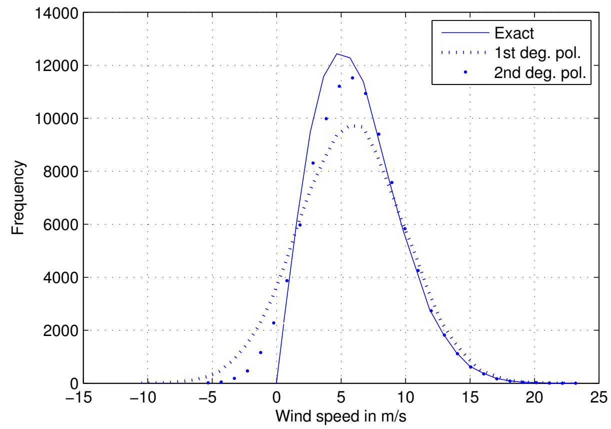

FIGURE 2. **Weibull distribution and approximations.**

TABLE 1. **Moments of u, $f_1(x)$ and $f_2(x)$.**

	mean μ	std. dev. σ
u	6.1975	3.2355
$f_1(x)$	5.7707	4.1289
$f_2(x)$	6.0787	3.5213

As $f_1(x)$ is a linear relationship, the calculation of $\rho_{xf_1(x)}$ should give 1 as result, and this is exactly what happens. And also, $\rho_{xu} = 0.9862$, just like $\rho_{uf_1(x)}$.

4. Influence of the constants c and k.

The next question to be answered is about the influence of constants c and k of the Weibull distribution on the transformation.

In this section the variation of constants a_i of the polynomial representation are studied as a function of c and k.

4.1. The effect of the variation of c.

In (1) it can be observed that the constant c is a factor of the equation, i.e., it multiplies the rest of the transformation. This is equivalent to saying that (1) can be written as an equation such as $u = c \cdot g(x)$ where $g(x)$ is a nonlinear function, or that u is proportional to $g(x)$ because c is a constant.

It is not difficult to deduce that variations of c should lead to proportional values of u for the same values of x. Consequently, this fact should be reflected in the polynomial approximation. And it is exactly what happens as can be seen in Figure 3, where these variations can be observed for a fixed value of $k = 2$, although similar results are obtained for different values of k.

For this value of $k = 2$, the values of a_0, a_1 and a_2, of the polynomial approximation, which will be denoted as a_{0_c}, a_{1_c} and a_{2_c}, can be calculated directly as a function of c, according to the following equations:

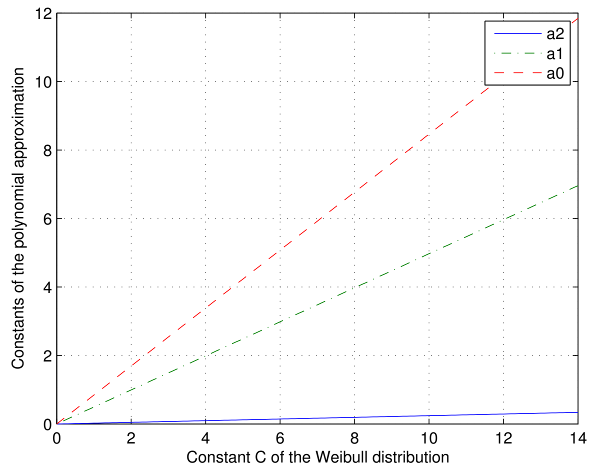

FIGURE 3. **Variation of a_0, a_1 and a_2 of the polynomial approximation with the constant c of the Weibull distribution.**

$$\begin{aligned} a_{0_c}(c) &= 0.8464 \cdot c \\ a_{1_c}(c) &= 0.4972 \cdot c \\ a_{2_c}(c) &= 0.0243 \cdot c \end{aligned} \tag{8}$$

As has been reflected in Figure 3 the simulation was made for values of c in the interval $[0, 14]$.

The following comment should be added. As mentioned before, sometimes the Rayleigh distribution has been recommended as a good continuous approximation of the frequency distribution of wind speeds at a given site. In a Weibull distribution the mean and standard deviation are calculated by means of the Gamma function first described by Euler, $\Gamma(p) = \int_0^\infty e^{-x}x^{p-1}dx$, according to $\mu_{Weibull} = c\Gamma\left(1+\frac{1}{k}\right)$ for the mean, and $\sigma^2_{Weibull} = c^2\left(\Gamma\left(1+\frac{2}{k}\right) - \Gamma^2\left(1+\frac{1}{k}\right)\right)$ for the standard deviation. However, one of the advantages of the Rayleigh distribution is the fact that the calculation of these statistical values is simpler, and they can be expressed such as $\mu_{Rayleigh} = c\frac{\sqrt{\pi}}{2}$ for the mean and $\sigma^2_{Rayleigh} = c^2\left(1-\frac{\pi}{4}\right)$ for the variance.

Taking the previous paragraph into account, the use of (8) can be of interest when using the Rayleigh distribution. In this case, $k = 2$ and the value c is calculated from the mean value as $c = \frac{2\mu_{Rayleigh}}{\sqrt{\pi}}$. And after that, the transformation of the Normal values to the Rayleigh ones can be performed with a high degree of accuracy by means of (8), which can be alternatively expressed as follows:

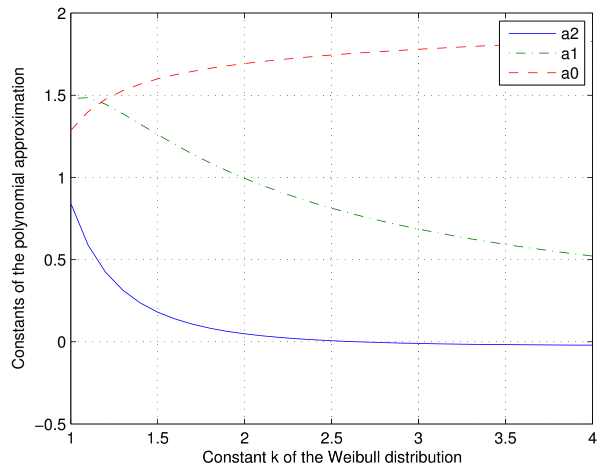

FIGURE 4. **Variation of a_0, a_1 and a_2 of the polynomial approximation with the constant k of the Weibull distribution.**

$$
\begin{aligned}
a_{0_c}(\mu) &= 0.8464 \cdot \frac{2}{\sqrt{\pi}} \cdot \mu = 0.9551 \cdot \mu \\
a_{1_c}(\mu) &= 0.4972 \cdot \frac{2}{\sqrt{\pi}} \cdot \mu = 0.5610 \cdot \mu \\
a_{2_c}(\mu) &= 0.0243 \cdot \frac{2}{\sqrt{\pi}} \cdot \mu = 0.0274 \cdot \mu
\end{aligned} \tag{9}
$$

where μ denotes the mean value of the Rayleigh distribution, that in the previous paragraph was denoted as $\mu_{Rayleigh}$.

4.2. The effect of the variation of k.

The presence of k in (1) is quite different from the case of the presence of c.

As can be expected, changes in the value of k produce nonlinear effects. This can be seen in Figure 4, where the value of c has been kept constant, allowing changes in k in the interval $[1, 4]$.

A first order approximation does not seem to be able to fit these curves, for which a second order one is here recommended, and it reveals that a_0, a_1 and a_2, now denoted as a_{0_k}, a_{1_k} and a_{2_k} can be expressed as follows:

$$
\begin{aligned}
a_{0_k}(k) &= -0.0715 \cdot k^2 + 0.4862 \cdot k + 0.9886 \\
a_{1_k}(k) &= 0.0967 \cdot k^2 - 0.8153 \cdot k + 2.2533 \\
a_{2_k}(k) &= 0.1503 \cdot k^2 - 0.9132 \cdot k + 1.3177
\end{aligned} \tag{10}
$$

Both effects of the variations of c and k can be studied in combination, and this is explained in next section.

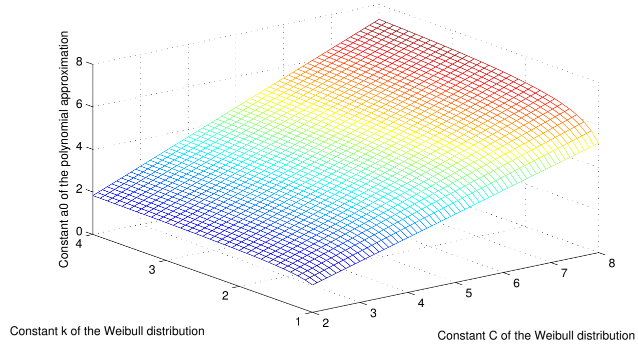

FIGURE 5. **Variation of a_0 in the polynomial approximation (degree 2) with the constants c and k of the Weibull distribution.**

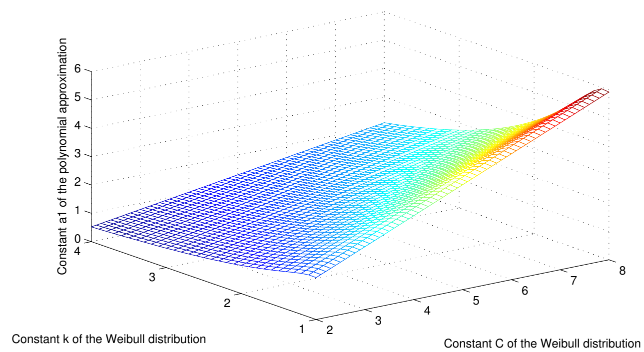

FIGURE 6. **Variation of a_1 in the polynomial approximation (degree 2) with the constants c and k of the Weibull distribution.**

4.3. **Combined effect of the variation of c and k.**

The results of the variation of both c and k values have been combined for the constants a_0, a_1 and a_2, and can be seen in Figures 5, 6 and 7.

This combined dependency can also be approximated by means of a polynomial transformation such as:

$$a_i(c,k) = \sum_{m=0}^{1} \sum_{n=0}^{2} p_{mn_i} c^m k^n \qquad i \in \{0,1,2\} \tag{11}$$

where the constants p_{mn_i} can be calculated for each a_i. Summarizing, (1) can be substituted by an equation like (3), where the constants a_i can be calculated as functions of the constants of the Weibull distribution, c and k, by means of (11).

Under the assumption of an interval for c equal to $[0, 14]$ and an interval for k equal to $[1, 4]$, the use of the above mentioned tool gives as a result the values for the constants given in Table 2 (the first subscript is for c and the second one for k).

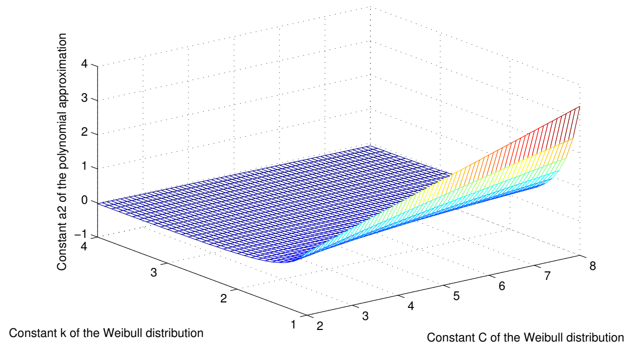

FIGURE 7. **Variation of a_2 in the polynomial approximation (degree 2) with the constants c and k of the Weibull distribution.**

TABLE 2. **Constants p_{mn} in the expresions of a_i, $i \in \{1, 2, 3\}$.**

	p_{00}	p_{10}	p_{01}	p_{11}	p_{02}
a_0	-0.9743	0.6891	0.8938	0.0644	-0.1788
a_1	1.3170	0.8632	-1.209	-0.1660	0.2417
a_2	2.0480	0.2492	-1.8790	-0.0808	0.3758

The coefficients p_{12} have not been included because they are very close to 0 in all cases (in fact, their values are lower than 10^{-15}).

5. Consequences of the proposed approximations.

Some consequences of the proposed approximations are given in this section. The conservation of autocorrelation when going from Normal distributions to Weibull distributions is not clearly easy to deduce directly from (1). But some operations with the statistical values based on the approximations can be of help for giving some explanation about why these values are retained. At the end of the section there are also some considerations about negative values in the simulations.

5.1. Conservation of the autocorrelation.

It has been mentioned that (1) was presented by Villanueva *et al.* in [13] as an option for the obtaining of Weibull distributions, with given mean, standard deviation and also lag 1 autocorrelation. To achieve such an objective, an autoregresive process known as AR(1) was used for randomly simulating Normal distributions with mean 0 and standard deviation 1 and then a transformation based on the Choleski decomposition of the covariance matrix was used to obtain the Weibull distributions.

Something that was observed when using this transformation process was the fact that lag 1 autocorrelation was apparently retained when converting data from the initial Normal distribution to the final Weibull one, i.e., the autocorrelation of each of the simulated Weibull series had a value very close to the value of the autocorrelation of the initial Normal series. In view of (1) it is not so evident why such statistical values are retained through the transformation.

The results obtained with the polynomial approximations presented in this paper can be used as an interesting approach to explain the fact given above.

In previous sections it has been shown that good approximations with polynomials of degrees 1 and 2 can be obtained for (1). The proposal is to look for relationships between the statistical values by means of these approximations.

5.1.1. *First degree.*

From this section on, some new notation will be used, and so $\mathbf{x} = [x_1, x_2, ..., x_M]$, $\mathbf{x_i} = [x_1, x_2, ..., x_{M-1}]$, and $\mathbf{x_{i+1}} = [x_2, x_3, ..., x_M]$. For the approximation, the notation will be $\mathbf{f_1}(\mathbf{x}) = [f_1(x_1), f_1(x_2), ..., f_1(x_M)]$, $\mathbf{f_1}(\mathbf{x})_\mathbf{i} = [f_1(x_1), f_1(x_2), ..., f_1(x_{M-1})]$ and in the same way to x_{i+1}, $\mathbf{f_1}(\mathbf{x})_\mathbf{i+1} = [f_1(x_2), ..., f_1(x_M)]$. Similar notation will be employed for $\mathbf{f_2}(\mathbf{x})$.

It is a well known fact that if the mean value of a Normal distribution is $\mu_\mathbf{x} = E[\mathbf{x}] = \frac{\sum_{k=1}^{M} x_k}{M}$, where E represents the expected value of x, then the mean value of the approximated Weibull distribution, if the approximation is by means of a first degree polynomial, i.e., $\mathbf{f_1}(\mathbf{x}) = a_0 + a_1\mathbf{x}$, can be estimated as $\mu_{\mathbf{f_1}(\mathbf{x})} = E[\mathbf{f_1}(\mathbf{x})] = E[a_0 + a_1\mathbf{x}] = a_0 + a_1\mu_\mathbf{x} = a_0$, taking into account that $\mu_\mathbf{x} = 0$, because $\mathbf{x} \sim N(0,1)$. It must not be forgotten that the transformed distribution is a Normal one because the transformation is linear, as the polynomial is of first degree.

If $\sigma_\mathbf{x}^2 = E[(x - \mu_\mathbf{x})^2] = \frac{\sum_{k=1}^{M}(x_k - \mu_\mathbf{x})^2}{M}$ is the variance of the Normal distribution, then the variance of the approximated Weibull one is $\sigma^2_{\mathbf{f_1}(\mathbf{x})} = E[(\mathbf{f_1}(\mathbf{x}) - \mu_\mathbf{u})^2] = E[(a_0 + a_1\mathbf{x} - a_0)^2] = a_1^2$, because $E[\mathbf{x}^2] = E[(\mathbf{x} - \mu_\mathbf{x})^2] = 1$, as $\mathbf{x} \sim N(0,1)$.

Lag 1 autocorrelation is obtained as the correlation between a given series and the same series shifted by one position. For calculating covariances between two series, for example, $\mathbf{x} = [x_1, x_2, ..., x_M]$ and $\mathbf{y} = [y_1, y_2, ..., y_M]$, a formulation has to be used where terms like the following appear, $(x_i - \mu_\mathbf{x})(y_i - \mu_\mathbf{y})$. However, when calculating lag 1 autocorrelation, only one series is involved, $\mathbf{x} = [x_1, x_2, ..., x_M]$ and the terms are such as $(x_i - \mu_\mathbf{x})(x_{i+1} - \mu_\mathbf{x})$, i.e., the values are grouped as follows, (x_1, x_2), (x_2, x_3), ..., (x_{M-1}, x_M).

The covariance between a series of a Normal distribution and the shifted one will be denoted as $\sigma_{\mathbf{x_i}\mathbf{x_{i+1}}}$ and can be calculated as:

$$\sigma_{\mathbf{x_i}\mathbf{x_{i+1}}} = \frac{\sum_{i=1}^{M-1}(x_i - \mu_x)(x_{i+1} - \mu_x)}{M} = \frac{\sum_{i=1}^{M-1} x_i x_{i+1}}{M}$$

by taking into account that $\mu_\mathbf{x} = 0$ because $\mathbf{x} \sim N(0,1)$. One fact to bear in mind is that there are only $M - 1$ addends in the previous sum. The value of M in the denominator can remain, instead of being changed to $M - 1$, because this does not affect the rest of the reasoning, and also because it is generally accepted in the calculation of lag s autocorrelation, for a given value of s.

The correlation between both series can be defined as

$$\rho_{\mathbf{x_i}\mathbf{x_{i+1}}} = \frac{\sigma_{\mathbf{x_i}\mathbf{x_{i+1}}}}{\sigma_{\mathbf{x_i}}\sigma_{\mathbf{x_{i+1}}}} = \frac{\sigma_{\mathbf{x_i}\mathbf{x_{i+1}}}}{\sigma^2} = \sigma_{\mathbf{x_i}\mathbf{x_{i+1}}}$$

assuming that $\sigma = \sigma_{\mathbf{x_i}} = \sigma_{\mathbf{x_{i+1}}} = 1$.

This correlation between the two series is the lag 1 autocorrelation of the series.

What happens with the values of the approximated Weibull distribution is the following:

TABLE 3. **Moments of x and $f_(x)$.**

Variable	μ	σ
$\mathbf{x}$	0	1
$\mathbf{f_1}(\mathbf{x}) = a_0 + a_1\mathbf{x}$	a_0	a_1^2

$$\sigma_{\mathbf{f_1(x)_i f_1(x)_{i+1}}} = \frac{1}{M} \cdot \sum_{i=1}^{M-1} (f_1(x_i) - \mu_{\mathbf{f_1(x)}})(f_1(x_{i+1}) - \mu_{\mathbf{f_1(x)}}) = a_1^2 \sigma_{\mathbf{x_i x_{i+1}}}$$

taking into account that $f_1(x_i) - \mu_{\mathbf{f_1(x)}} = a_0 + a_1 x_i - a_0$, and $f_1(x_{i+1}) - \mu_{\mathbf{f_1(x)}} = a_0 + a_1 x_{i+1} - a_0$, the autocorrelation is, then, calculated, as a correlation between both series, i.e.

$$\rho_{\mathbf{f_1(x)_i f_1(x)_{i+1}}} = \frac{\sigma_{\mathbf{f_1(x)_i f_1(x)_{i+1}}}}{\sigma_{\mathbf{f_1(x)_i}} \sigma_{\mathbf{f_1(x)_{i+1}}}}$$

Previously it has been shown that $\sigma^2_{\mathbf{f_1(x)}} = a_1^2$, which leads to $\sigma^2_{\mathbf{f_1(x)_i}} = \sigma^2_{\mathbf{f_1(x)_{i+1}}} = a_1^2$, and $\sigma_{\mathbf{f_1(x)_i}} = \sigma_{\mathbf{f_1(x)_{i+1}}} = a_1$. Finally:

$$\rho_{\mathbf{f_1(x)_i f_1(x)_{i+1}}} = \frac{a_1^2 \sigma_{\mathbf{x_i x_{i+1}}}}{a_1 \sigma_{\mathbf{x_i}} a_1 \sigma_{\mathbf{x_{i+1}}}} = \frac{\sigma_{\mathbf{x_i x_{i+1}}}}{\sigma_{\mathbf{x_i}} \sigma_{\mathbf{x_{i+1}}}} = \rho_{\mathbf{x_i x_{i+1}}}$$

The conclusion is that substituting (1) by a first degree polynomial involves a transformation where lag 1 autocorrelation is retained.

In fact, if subindices $i+1$ are changed to subindices $i+s$, then s lag autocorrelation must be kept in such a transformation. Thus, if an AR(s) instead of an AR(1) process is used for generating the Normal distribution, the transformation to approximated Weibull series should keep s lag autocorrelations for all s.

A summary of all these moments can be read in Table 3.

5.1.2. *Second degree.*

The conclusions of the previous section are not surprising because the proposed transformation by means of f_1 is linear.

A better approximation to (1) consists of a second order polynomial, such as $\mathbf{f_2}(\mathbf{x}) = a_0 + a_1\mathbf{x} + a_2\mathbf{x}^2$. The term of second order contributes to confer a certain degree of asymmetry to the distribution, which is a feature that makes it more similar to a Weibull one.

The mean value of the given distribution is $\mu_{\mathbf{f_2(x)}} = E[a_0 + a_1\mathbf{x} + a_2\mathbf{x}^2] = a_0 + a_2$, again because $E[\mathbf{x}] = \mu_{\mathbf{x}} = 0$ and in addition because $E[\mathbf{x}^2] = \sigma^2_{\mathbf{x}} = 1$.

In the case of the variance the calculation is as follows, $\sigma^2_{\mathbf{f_2(x)}} = E[(a_0 + a_1\mathbf{x} + a_2\mathbf{x}^2 - \mu_{\mathbf{f_2(x)}})^2] = E[(a_1\mathbf{x} + a_2(\mathbf{x}^2 - 1))^2]$, by taking into account the fact that $\mu_{\mathbf{f_2(x)}} = a_0 + a_2$, shown in the previous paragraph. If some operations are performed, the following transformation can be obtained, $\sigma^2_{\mathbf{f_2(x)}} = E[a_2^2\mathbf{x}^4 + 2a_1a_2\mathbf{x}^3 + (a_1^2 - 2a_2^2)\mathbf{x}^2 - 2a_1a_2\mathbf{x} + a_2^2]$.

The values of the moments for a Normal distribution can be seen in appendix A, and substituting them in this transformation, the result is that $\sigma^2_{\mathbf{f_2(x)}} = a_1^2 + 2a_2^2$.

Autocorrelation can be defined with the help of $\sigma_{\mathbf{x_i x_{i+1}}} = \frac{\sum_{i=1}^{M-1} x_i x_{i+1}}{M}$, for the case of the original Normal series.

For the transformed one:

$$\begin{aligned}
&\sigma_{\mathbf{f_2(x)_i f_2(x)_{i+1}}} = \\
&= \frac{1}{M} \cdot \sum_{i=1}^{M-1} (f_2(x_i) - \mu_{\mathbf{f_2(x)}})(f_2(x_{i+1}) - \mu_{\mathbf{f_2(x)}}) = \\
&= \frac{1}{M} \cdot \sum_{i=1}^{M-1} (a_1 x_i + a_2(x_i^2 - 1))(a_1 x_{i+1} + a_2(x_{i+1}^2 - 1))
\end{aligned}$$

For simplicity, M will not be taken into account now. By operating the previous equation:

$$\sum_{i=1}^{M-1} (a_1^2 x_i x_{i+1} + a_1 a_2 x_i x_{i+1}^2 - a_1 a_2 x_i + a_1 a_2 x_i^2 x_{i+1} \\ - a_1 a_2 x_{i+1} + a_2^2 x_i^2 x_{i+1}^2 - a_2^2 x_i^2 - a_2^2 x_{i+1}^2 + a_2^2)$$

This transformation can be simplified by taking into account that, due to properties of the $N(0,1)$ distribution, with the help of expressions given in appendix A, the final result is that $\sigma_{\mathbf{f(x)_i f(x_{i+1})}} = a_1^2 \sigma_{\mathbf{x_i x_{i+1}}} + 2a_2^2 \sigma^2_{\mathbf{x_i x_{i+1}}}$.

According to this:

$$\rho_{\mathbf{f_2(x)_i f_2(x)_{i+1}}} = \frac{\sigma_{\mathbf{f_2(x)_i f_2(x)_{i+1}}}}{\sigma_{\mathbf{f_2(x)_i}} \sigma_{\mathbf{f_2(x)_{i+1}}}} = \rho_{\mathbf{x_i x_{i+1}}} \frac{a_1^2 + 2a_2^2 \rho_{\mathbf{x_i x_{i+1}}}}{a_1^2 + 2a_2^2} \tag{12}$$

where $\rho_{\mathbf{x_i x_{i+1}}} = \sigma_{\mathbf{x_i x_{i+1}}}$.

It is interesting to point out the values of $\rho_{\mathbf{f_2(x)_i f_2(x)_{i+1}}}$ in some particular cases, which can be considered extreme cases:

$$\rho_{\mathbf{f_2(x)_i f_2(x)_{i+1}}} = \begin{cases} 1 & \text{if } \rho_{\mathbf{x_i x_{i+1}}} = 1 \\ 0 & \text{if } \rho_{\mathbf{x_i x_{i+1}}} = 0 \\ \frac{-a_1^2 + 2a_2^2}{a_1^2 + 2a_2^2} & \text{if } \rho_{\mathbf{x_i x_{i+1}}} = -1 \end{cases}$$

Anyway, as generally $a_1 \gg a_2$, it can be concluded that $\rho_{\mathbf{f_2(x)_i f_2(x)_{i+1}}} \approx \rho_{\mathbf{x_i x_{i+1}}}$. For example, by using the values obtained for a second degree approximation, i.e., $a_1 = 3.5227$ and $a_2 = 0.1539$, the result is that if $\rho_{\mathbf{x_i x_{i+1}}} = -1$, then $\rho_{\mathbf{f_2(x)_i f_2(x)_{i+1}}} = -0.9924$. The meaning of this is that the autocorrelation of the distribution obtained through the second degree polynomial is greater than 99% of the autocorrelation of the initial Normal distribution.

All that has been explained in this section can be summarized as follows: assuming a Normal distribution and the Weibull distribution obtained from this Normal one by means of (1), different approximations to the Weibull distribution can be run on a polynomial basis.

A degree one polynomial is a linear approximation which retains variance and autocorrelation without dependency of the lag.

A degree two polynomial includes a certain degree of asymmetry, which makes the transformed distribution be further from the Normal one and closer to the Weibull one. In this case, the autocorrelation is not retained and its degree of approximation to the values of the original distribution depends on the value itself.

As a second degree approximation is closer to the first degree approximation this allows the conclusion to be made that the autocorrelation is not retained through the exact transformation given by (1).

Table 4. **Moments of x and $f_(x)$.**

Variable	μ	σ
$\mathbf{x}$	0	1
$\mathbf{f_2}(\mathbf{x}) = a_0 + a_1\mathbf{x} + a_2\mathbf{x}^2$	$a_0 + a_2$	$a_1^2 + 2a_2^2$

5.2. **Conservation of the correlation.**
Correlations between series are retained through linear transformations, and this can be argued in a similar way to the assertion made for the lag 1 autocorrelations.

If $\mathbf{x} \sim N(0,1)$ and $\mathbf{y} \sim N(0,1)$, and $\mathbf{f_1}(\mathbf{x}) = a_0 + a_1\mathbf{x}$ and $\mathbf{g_1}(\mathbf{y}) = b_0 + b_1\mathbf{y}$, where $\mathbf{f_1}(\mathbf{x})$ and $\mathbf{g_1}(\mathbf{y})$ are now two new series transformed from $\mathbf{x}$ and $\mathbf{y}$, then the correlation between $\mathbf{x}$ and $\mathbf{y}$ is given by $\rho_{\mathbf{xy}} = \frac{\sigma_{\mathbf{xy}}}{\sigma_{\mathbf{x}}\sigma_{\mathbf{y}}}$. As $\sigma_{\mathbf{x}} = \sigma_{\mathbf{y}} = 1$, it can be expressed as $\rho_{\mathbf{xy}} = \sigma_{\mathbf{xy}}$. On the other side, $\sigma_{\mathbf{xy}} = E[(\mathbf{x} - \mu_{\mathbf{x}})(\mathbf{y} - \mu_{\mathbf{y}})]$, where $\mu_{\mathbf{x}} = \mu_{\mathbf{y}} = 0$, by which $\sigma_{\mathbf{xy}} = E[\mathbf{xy}]$.

This means that $\sigma_{\mathbf{f_1}(\mathbf{x})\mathbf{f_1}(\mathbf{y})} = E[(\mathbf{f_1}(\mathbf{x}) - \mu_{\mathbf{f_1}(\mathbf{x})})(\mathbf{f_1}(\mathbf{y}) - \mu_{\mathbf{f_1}(\mathbf{y})})] = E[(a_0 + a_1\mathbf{x} - a_0)(b_0 + b_1\mathbf{y} - b_0)] = a_1 b_1 \sigma_{\mathbf{xy}}$.

As $\sigma_{\mathbf{f_1}(\mathbf{x})} = a_1$ and $\sigma_{\mathbf{g_1}(\mathbf{y})} = b_1$, then $\rho_{\mathbf{f_1}(\mathbf{x})\mathbf{g_1}(\mathbf{y})} = \frac{\sigma_{\mathbf{f_1}(\mathbf{x})\mathbf{g_1}(\mathbf{y})}}{\sigma_{\mathbf{f_1}(\mathbf{x})}\sigma_{\mathbf{g_1}(\mathbf{y})}} = \frac{a_1 b_1 \sigma_{\mathbf{f_1}(\mathbf{x})\mathbf{g_1}(\mathbf{y})}}{a_1 b_1} = \rho_{\mathbf{f_1}(\mathbf{x})\mathbf{g_1}(\mathbf{y})}$, and this proves that the correlation is also retained.

However, in nonlinear transformations, i.e., in the case of second degree approximations, things operate in a different manner. If new approximations are taken, such as $\mathbf{f_2}(\mathbf{x}) = a_0 + a_1\mathbf{x} + a_2\mathbf{x}^2$ and $\mathbf{g_2}(\mathbf{y}) = b_0 + b_1\mathbf{y} + b_2\mathbf{y}^2$, then both means are $\mu_{\mathbf{x}} = a_0 + a_2$ and $\mu_{\mathbf{y}} = b_0 + b_2$ and the standard deviations are $\sigma_{\mathbf{x}} = a_1 + 2a_2^2$ and $\sigma_{\mathbf{y}} = b_1 + 2b_2^2$.

The variance between $\mathbf{f_2}(\mathbf{x})$ and $\mathbf{g_2}(\mathbf{x})$ is calculated as $\sigma_{\mathbf{f_2}(\mathbf{x})\mathbf{g_2}(\mathbf{y})} = E[(\mathbf{f_2}(\mathbf{x}) - \mu_{\mathbf{f_2}(\mathbf{x})})(\mathbf{g_2}(\mathbf{y}) - \mu_{\mathbf{g_2}(\mathbf{y})})]$, which is $E[(a_1\mathbf{x} + a_2(\mathbf{x}^2 - 1))(b_1\mathbf{y} + b_2(\mathbf{y}^2 - 1))]$.

By rearranging the previous expression and by taking into account appendix A, it can be written as:

$$\sigma_{\mathbf{f_2}(\mathbf{x})\mathbf{g_2}(\mathbf{y})} = \sigma_{\mathbf{xy}}(a_1 b_1 + a_2 b_2 \sigma_{\mathbf{xy}})$$

Now, as $\sigma_{\mathbf{f_2}(\mathbf{x})} = (a_1^2 + 2a_2^2)\sigma_{\mathbf{x}}$ and $\sigma_{\mathbf{f_2}(\mathbf{y})} = (a_1^2 + 2a_2^2)\sigma_{\mathbf{y}}$, and $\sigma_{\mathbf{xy}} = \rho_{\mathbf{xy}}$ finally:

$$\rho_{\mathbf{f_2}(\mathbf{x})\mathbf{g_2}(\mathbf{y})} = \rho_{\mathbf{xy}} \frac{a_1 b_1 + 2a_2 b_2 \rho_{\mathbf{xy}}}{\sqrt{a_1^2 + 2a_2^2}\sqrt{b_1^2 + 2b_2^2}} \tag{13}$$

As can be deduced, things operate in a similar way to the case of lag 1 autocorrelation if both distributions coincide, i.e., if they have identical c and k parameters, because in this case $a_1 = b_1$, $a_2 = b_2$ and (13) is an equation similar to (12). In this case the conclusions achieved in (12) can be applied here.

However, when both distributions differ, then the four parameters a_1, a_2, b_1 and b_2 in (13) cannot be substituted by only two of them, and the dependency on them is more complex. But the conclusion is that correlation is not completely retained when using (1).

5.3. **Negative values.**
Another consequence of the approximation is the appearance of negative values in the conversion, a problem that was detected by Feijóo *et al.* [14], in a work where correlated Weibull and Rayleigh distributed series of wind speeds were simulated, and then avoided with new methods by Feijóo and Sobolewski [15], with the use of nonparametric correlations, i.e., Spearman rank correlations.

For representing wind speed values, the Weibull distribution can be treated just as it has so far, i.e., assuming that its minimum value is 0.

But more generally, the Weibull distribution CDF with an origin $\gamma \neq 0$ can be defined as $W : [\gamma, +\infty) \to [0, 1]$. As mentioned, in the case of wind speed distributions, $\gamma = 0$ because wind speeds are supposed to be positive, so there are no negative values in the distribution. The application of the explained approximations gives a certain number of negative values as a result. This can be observed in Figure 2. By multiple simulations, it has been estimated that no more than 0.5% of the data are negative.

According to this, it is natural that when data represent wind speeds, there is a tendency to reject negative values. However, other errors are accepted in all simulations and an interesting question that begs to be answered is how much of a problem it would be to accept negative values as wind speed values.

And the answer is that it does not necessary involve making a significant error in the calculations.

A typical situation consists of combining wind speed data with WT power curves with the aim of calculating either the power generated or the total energy produced during a certain period of time.

In order to check the error made when substituting the exact formulation given by (1) by a polynomial approximation of degree 2, a power curve for a WT has been combined with data corresponding to a site with a Weibull distribution of parameters $c = 7$ and $k = 2$.

The power curve has been proposed by Carta *et al.* [16] and described in appendix B. For the following values, $v_{CI} = 4\ ms^{-1}$ as cut-in wind speed, $v_R = 14\ ms^{-1}$ as rated wind speed, and $v_{CO} = 25\ ms^{-1}$ as cut-out wind speed, the error made in the calculation of energy generated by the WT rises to a 3%. The value of this error does not seem to depend on the maximum power. For WTs of rated powers 2, 3, ..., 7 MW the error is also around 3%.

As a conclusion, the acceptance of negative values as wind speeds for the calculation of power or energy values in a simulation is not so important, as they are filtered by the WT power curve, i.e., if the value of wind speed is negative, then the power generated by the WT will be 0. In many cases it will be just like when the value is positive but under 3 or 4 ms^{-1} or above 25 ms^{-1}.

An approximation with a polynomial of degree one is much less satisfactory, as the errors made when calculating energy rise to values close to 13% for all the rated powers given.

6. **Conclusions.**

In this paper two different polynomial approximations have been proposed for the transformation of sets of Normally distributed data to sets of Weibull distributed series, satisfying not only the parameters of Weibull distributions, but also their correlations and even autocorrelations.

The approximations have been used to provide an approach to a better understanding of why these features are retained when an exact transformation is carried out, with the following consequences:

1. The use of (1) for the Normal to Weibull transformation is very adequate and gives very good results. It has no disadvantages from a computational point of view. However, it is not easy to explain why certain statistical values, such as

correlation and autocorrelation are highly retained, which has been previously obtained by simulations.

2. In order to explain such phenomena, polynomial approximations based on the least square method were used for the $\text{CDF}_{Weibull} = f(\text{CDF}_{Normal})$ curve.
3. A second degree approximation shows a high degree of accuracy, and also shows that correlations and autocorrelations are not exactly retained, but seem to explain why they are highly retained.
4. The appearance of negative values of the wind speed in simulations does not involve an important error. Although there are no negative values of wind speed, they can appear in the simulation, but even so, they are filtered by the power curves of the WTs when used for estimating power or energy captured from the wind.

Appendix A. Moments of the Normal distribution.

The following are the expressions of moments of the normal distribution that have been necessary in the paper:

$$\begin{aligned}
E[\mathbf{x}] &= \mu_{\mathbf{x}} = 0 \\
E[\mathbf{x}^2] &= \sigma_{\mathbf{x}}^2 = 1 \\
E[\mathbf{x}^3] &= \mu_{\mathbf{x}}^3 + 3\mu_{\mathbf{x}}\sigma_{\mathbf{x}}^2 = 0 \\
E[\mathbf{x}^4] &= \mu_{\mathbf{x}}^4 + 6\mu_{\mathbf{x}}^2\sigma_{\mathbf{x}}^2 + 3\sigma_{\mathbf{x}}^4 = 3 \\
E[\mathbf{x_i}] &= E[\mathbf{x_{i+1}}] = 0 \\
E[\mathbf{x_i^2}] &= E[\mathbf{x_{i+1}^2}] = 1 \\
E[\mathbf{x_i x_{i+1}}] &= \sigma_{\mathbf{x_i x_{i+1}}} \\
E[\mathbf{x_i x_{i+1}^2}] &= E[\mathbf{x_i^2 x_{i+1}}] = 0 \\
E[\mathbf{x_i^2 x_{i+1}^2}] &= \sigma_{\mathbf{x_i}}\sigma_{\mathbf{x_{i+1}}} + 2\sigma_{\mathbf{x_i x_{i+1}}}^2 = 1 + 2\sigma_{\mathbf{x_i x_{i+1}}}^2
\end{aligned}$$

Appendix B. Power curves description.

A WT power curve can be described by means of a function such as the following [17]:

$$P = \begin{cases} 0 & 0 \le v_w \le v_{CI} \\ h(v_w)P_R & v_{CI} \le v_w < V_R \\ P_R & v_R \le v_w < v_{CO} \\ 0 & v_w \ge v_{CO} \end{cases}$$

where v_w is the input variable, i.e., the wind speed, v_{CI} stands for cut-in wind speed, 4 ms^{-1} in the example proposed in the paper, v_{CO} for cut-out wind speed, 25 ms^{-1}, v_R is the rated wind speed, 14 ms^{-1}, and P_R the rated power, 1 MW.

The function $h(v_w)$ is calculated as $h(v_w) = A + Bv_w + Cv_w^2$, and

$$
\begin{aligned}
A &= a(v_{CI}b - 4v_{CI}v_{RC}) \\
B &= a(4bd - 3v_{CI}v_R) \\
C &= a(2 - 4d) \\
a &= \frac{1}{(v_{CI} - v_R)^2} \\
b &= v_{CI} + v_R \\
d &= \left(\frac{v_{CI} + v_R}{2v_R}\right)^3
\end{aligned}
$$

REFERENCES

[1] Freris, L. L. (1990) Wind energy conversion systems. Prentice Hall.

[2] Kavasseri, R. J. (2004) Evidence of crossover phenomena in wind-speed data. *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, Volume 51, 2255-2262.

[3] Damousis, I. G., Alexiadis, M. C., Theocharis, J. B., Dokopoulos, P. S. (2004) A fuzzy model for wind speed prediction and power generation in wind parks using spatial correlation. *IEEE Transactions on Energy Conversion*, Volume 19, 352-361.

[4] Brown, B. G., Katz, R. W., Murphy, A. H. (1984) Time series models to simulate and forecast wind speed and wind power. *Journal of Climate and Applied Meteorology*, Volume 23, 1184-1195.

[5] Song, W. T., Hsiao, L. C. (1993) Generation of autocorrelated random variables with a specified marginal distribution. *Proceedings of the Winter Simulation Conference*, 374-377.

[6] Feijóo, A., Villanueva, D., Pazos, J. L., Sobolewski, R. (2011) Simulation of correlated wind speeds. A review. *Renewable and Sustainable Energy Reviews*, Volume 15, 2826-2832.

[7] Correia, P. F., Ferreira de Jesus, J. M. (2010) Simulation of correlated wind speeds and power variates in wind parks. *Electric Power Systems Research*, Volume 80, 592-598.

[8] Segura-Heras, I., Escrivá-Escrivá, G., Alcázar-Ortega, M. (2011) Wind farm electrical power production model for load flow analysis. *Renewable Energy*, Volume 36, 1008-1013.

[9] Karaki, S. H., Salim, B. A., Chedid, R. B. (2002) Probabilistic model of a two site wind energy conversion system. *IEEE Transactions on Energy Conversion* Volume 17, 530-536.

[10] Young, D. J., Beaulieu, N. C. (2000) The generation of correlated Rayleigh random variates by inverse discrete Fourier transform. *IEEE Transactions on Communications*, Volume 48, 1114-1127.

[11] Huang, Z., Chalabi, Z. S. (1995) Use of time-series analysis to model and forecast wind speed. *Journal of Wind Engineering and Industrial Aerodynamics* Volume 56, 311-322.

[12] Shamshad, A., Bawadi, M. A., Wan Hussin, W. M. A., et al. (2005) First and second order Markov chain models for synthetic generation of wind speed time series. *Energy* Volume 30, 693-708.

[13] Villanueva, D., Feijóo, A., Pazos, J. L. (2012) Simulation of correlated wind speed data for Economic Dispatch Evaluation. *IEEE Transactions on Sustainable Energy*, Volume 3, 142-149.

[14] Feijóo, A., Cidrás, J., Dornelas, J. L. G. (1999) Wind speed simulation in wind farms for steady-state security assessment of electrical power systems. *IEEE Transactions on Energy Conversion*, Volume 3, 1582-1588.

[15] Feijóo, A., Sobolewski, R. (2009) Simulation of correlated wind speeds. *International Journal of Electrical Energy Systems*, Volume 1, 99-106.

[16] Carta, J. A., Ramírez, P., Velázquez, S. (2009) A review of wind speed probability distributions used in wind energy analysis. Case studies in the Canary Islands. *Renewable and Sustainable Energy Reviews* Volume 13, 933-955.

[17] Dialynas, E. N., Machias, A. V. (1989) Reliability modelling interactive techniques of power systems including wind generating units. *Archive für Elektrotechnik*, Volume 72, 33-41.