A Methodology to Generate Statistically Dependent Wind Speed Scenarios

J. M. Morales [☆a], R. Mínguez [☆b], A. J. Conejo [☆*,a]

[a] *Department of Electrical Engineering, Univ. Castilla – La Mancha, Campus Universitario s/n, 13071 Ciudad Real, Spain*

[b] *Environmental Hydraulics Institute "IH Cantabria", Univ. Cantabria, Avenida de los Castros s/n, 39005 Santander, Spain*

Abstract

Wind power - a renewable energy source increasingly attractive from an economic viewpoint - constitutes an electricity production alternative of growing relevance in current electric energy systems. However, wind power is an intermittent source that cannot be dispatched at the will of the producer. Modeling wind power production requires characterizing wind speed at the sites where the wind farms are located. The wind speed at a particular location can be described through a stochastic process that is spatially correlated with the stochastic processes describing wind speeds at other locations. This paper provides a methodology to characterize the stochastic processes pertaining to wind speed at different geographical locations via scenarios. Each one of these scenarios embodies time dependencies and is spatially dependent of the scenarios describing other wind stochastic processes. The scenarios generated by the proposed methodology are intended to be used within stochastic programming decision models to make informed decisions pertaining to

[☆] J. M. Morales, R. Mínguez, and A. J. Conejo are partly supported by Junta de Comunidades de Castilla – La Mancha through project PCI-08-0102 and by the Ministry of Education and Science of Spain through CICYT Project DPI2006-08001.

[*] Corresponding author. Tel.: +34 926 295433; fax: +34 926 295361.

Email addresses: JuanMiguel.Morales@uclm.es (J. M. Morales [☆]), roberto.minguez@unican.es (R. Mínguez [☆]), Antonio.Conejo@uclm.es (A. J. Conejo [☆])

wind power production. The methodology proposed is accurate in reproducing wind speed historical series as well as computationally efficient. A comprehensive case study is used to illustrate the capabilities of the proposed methodology. Appropriate conclusions are finally drawn.
Key words: Wind speed correlation, scenario generation, stochastic programming, stochastic processes, time series.

---

1. Introduction

1.1. Motivation

Wind power constitutes an electricity production alternative of growing relevance in current electric energy systems across the world. Moreover, wind power is a renewable source that is increasingly attractive from an economic viewpoint. However, wind power is an intermittent source that cannot be dispatched at the will of the wind producer. Its production is conditioned to the availability of an appropriate level of wind.

The integration of wind energy resources into power system networks is currently posing unique challenges for system operators and planners, which should take preventive and corrective actions (e.g., scheduling and deploying additional reserves) in order to maintain system security and reliability, potentially threatened by the unpredictable and variable nature of these resources [1, 2, 3, 4, 5, 6]. Likewise, wind producers should gather as much accurate information as possible on the wind speed characteristics at their wind farms to make informed decisions concerning electricity trading and risk management [7, 8, 9, 10, 11].

The wind speed at a particular location can be described through a stochastic process that is spatially correlated with the stochastic processes describing wind speeds at other locations. An appropriate description of these stochastic processes requires recognizing

the time dependent nature of each of these processes as well as the spatial correlation among them.

Note that characterizing the stochastic process describing wind speed at a particular location is a problem of a different nature that predicting wind speed for that particular site. Prediction entails forecasting the most likely wind speed time series for a particular future horizon, e.g., wind speed values for the 24 hours of the next day; while describing the involved stochastic process entails producing a set of possible wind speed time series with their associated probabilities, e.g., 1000 series of 24 wind speed values describing plausible wind speed outcomes for next day and their associated probabilities.

1.2. Procedure

This paper provides a methodology to characterize the stochastic processes pertaining to wind speed at different geographical locations. Each wind stochastic process is characterized by a set of scenarios. Each one of these scenarios embodies time dependencies and is spatially dependent of the scenarios describing other wind stochastic processes geographically close.

The uncertainty characterization achieved through this methodology is such that the generated scenario sets corresponding to different geographical locations retain the main statistical properties of the wind speed stochastic processes involved, namely:

1. The marginal distribution associated with the stochastic process corresponding to the wind speed at each geographical location considered.
2. The temporal correlations characterizing the stochastic process corresponding to the wind speed at each geographical location.
3. The predominant spatial correlations among the stochastic processes representing wind speed at the different considered locations.

The proposed methodology involves two procedures, namely, model characterization and scenario generation. Schematically, the model characterization involves the steps below:

1. Fitting of a known distribution function (e.g., a Weibull) to the available data (collapsed in time) of each wind site. This step allows retaining the characteristics of the marginal distribution that distinguishes each wind site.
2. Using the marginal distribution function obtained in 1 above, transform the time series of historical values of each site into a normalized Gaussian time series. This step allows using consistently time series models.
3. Fitting of a time series model (e.g., an ARMA process) to the transformed historical values obtained in 2 above. The obtained model allows taking into account temporal correlations while generating wind speed scenarios for a wind site.
4. Estimation of a stationary variance-covariance matrix describing spatial correlations among the considered wind speed sites. While producing scenarios, this matrix allows enforcing spatial correlations.

The procedure to generate one plausible wind speed scenario per wind site involves the steps below:

5. Generate a time series of white noise values of appropriate length for each wind site.
6. Using the variance-covariance matrix describing spatial correlations among sites (obtained in 4), transform the set of white noise series (one per wind site) to make these series spatially correlated.
7. Use the time series model of each wind site (obtained in 3) and the corresponding cross-correlated white noise series (obtained in 6) to generate a normalized wind speed series for each wind site.

8. Use the inverse of the transform performed in step 2 above for each wind site to un-transform the corresponding normalized wind speed series (obtained in 7), thus producing a plausible wind speed series for each site.

The above procedure (5 to 8) is repeated as many times as the number of required scenarios.

The scenarios generated by the proposed methodology are intended to be used within stochastic programming decision models to make informed decisions pertaining to wind power production. The methodology proposed is accurate in reproducing wind speed historical series as well as computationally efficient.

The validation of the proposed model is carefully carried out by comparing the statistical properties of (a) historical data and (b) scenarios generated by the proposed algorithm. Most relevant statistical properties of both historical data and synthetic scenarios are equal within a reasonable numerical threshold.

1.3. Literature review

Procedures for scenario generation are required to model the stochastic processes involved in decision-making based on stochastic programming [12]. A number of works addressing the problem of how to build appropriate scenario sets can be found in the technical literature [13, 14, 15, 16, 17]. In the context of electricity markets, several methods have been suggested to generate scenarios modeling the uncertainty of spot prices, reserve prices, and demand [18, 19, 20, 15]. However, there is a lack of scenario generation procedures capable of recognizing and managing the intricacies of wind stochastic processes, a lack that the proposed methodology is intended to fill. Specifically, the referred intricacies have to do with the non-normality of the associated marginal distribution and the likely spatial correlations existing among different wind sites.

Even though correlation at multiple wind power sites has a significant impact on the reliability of power systems, it has not yet received sufficient attention. In this respect, references [21] and [22] analyze the autocorrelation function of wind speeds at several sites and conclude that the autocorrelogram tends to zero as the time lag increases. Karaki et al. [23] present a general probabilistic model of a two-site wind-energy conversion system, which assumes that wind speeds at the two sites are not independent and convolutions can not be directly applied. The work developed in [24] constitutes a methodology for capacity adequacy evaluation of power systems including wind energy based on a Monte Carlo simulation approach of the hourly wind speeds. For this purpose, an autoregressive moving average time-series model is used. Wangdee and Billinton [25] present a generation adequacy assessment including wind energy at multiple locations and show that the degree of wind speed correlation between two wind farms has a considerable impact on the resulting reliability indices. In this case, an autoregressive moving average time series model is also used to simulate hourly wind speeds. Miranda and Dunn [26] introduce a spatially correlated wind speed model using a multivariate time-series approach, which models the correlation between various sites in the UK. Reference [27] proposes a model for power system reliability assessment that preserves some of the statistical characteristics of the wind speed distributions at two different locations, namely, their means and standard deviations, and allows simulating a given contemporaneous wind speed cross-correlation between these locations through a time-shifting technique. The optimal shifting time at the two sites is determined using a linear interpolation algorithm.

Note that the method proposed in this paper has the following advantages with respect to the existing ones:

1. It reproduces the autocorrelation function of each considered wind site as in [21, 22].
2. It is an approximate multivariate time-series approach which allows decoupling the

model into different univariate site components, whose parameters can be estimated separately, thus avoiding the complexity of multivariate parameter estimation methods.

3. It replicates the main cross-correlations coefficients characterizing wind speeds at different sites, and not just the contemporaneous one as in [27].
4. It preserves the statistical properties of wind speeds, not only in terms of mean and standard deviation as in [27], but the complete marginal distribution.

1.4. Contributions

The contributions of this paper are fourfold:

1. A precise statistical characterization of the stochastic processes describing wind speed at several geographical locations, considering both temporal and spatial correlations.
2. The development of a procedure to generate scenarios of wind speed values at a particular geographical site and spanning a pre-specified time horizon. These scenarios reproduce both the time dependency that characterizes wind speed time values and the marginal distribution that characterizes any wind site if time is collapsed.
3. The extension of the above procedure to generate wind speed scenarios that in addition of embodying the attributes in 2, incorporates the spatial dependencies among different wind sites.
4. Reporting the analysis and discussion of the results obtained from of a comprehensive case study involving five spatially dependent wind locations.

1.5. Paper organization

The rest of this paper is organized as follows. In Section 2, the statistical concepts, tools and hypotheses on which the suggested methodology is based are stated. In Sec-

tion 3, a detailed description of the algorithm to produce a plausible scenario set for multiple wind sites is provided. Section 4 includes numerical and graphical results testing the proposed technique through a comprehensive analysis comprising five different wind locations. Finally, in Section 5, some relevant conclusions are duly drawn.

2. Theoretical foundation

The wind speed at a given site A, $\boldsymbol{Y^A}$, can be studied as a time series, and then, mathematically modeled as a stochastic process. In the technical literature, a stochastic process is commonly defined as a collection of dependent random variables $\boldsymbol{Y^A} = \{y_t^A, t \in T\}$. That is, for each t in the index set T, y_t^A is a random variable. We often interpret t as time and call y_t^A the state of the process at time t. In the case we are dealing with, y_t^A is a random variable describing the wind speed at location A at time t.

The probabilistic structure of a stochastic process is determined by identifying the joint distribution of its random variables $\{y_t^A\}$. Roughly speaking, this distribution explains both the probabilistic behavior of each random variable y_t^A on its own (marginal distributions), and the interrelations existing among all of them (statistical dependencies).

In practice, the determination of the joint distribution is usually a complex and cumbersome endeavor. However, this estimation becomes much easier under the following two assumptions:

1. The joint distribution is a multivariate Gaussian distribution, and therefore, is determined by just specifying the mean vector and the variance-covariance matrix of the random variables constituting the stochastic process.
2. The stochastic process under analysis is *stationary*, which means that neither the mean vector nor the variance-covariance matrix depend on time t.

Note that the two assumptions above imply that the marginal distributions of the random variables involved are all identical univariate Gaussian distributions. Being so and for simplicity, hereafter we will just refer to the marginal distribution associated with the considered stochastic process.

With reference to assumption 1, the non-diagonal elements of a variance-covariance matrix are generally referred to as *autocovariances*. Through a simple standardization process consisting in just dividing the autocovariances by the corresponding product of standard deviations, the *autocorrelations* are obtained. These are also known as *temporal correlations* and quantify the statistical interdependencies among the random variables.

The time series theory based on *autoregressive moving average* (ARMA) models relies on these two premises. An ARMA(p, q) process $\boldsymbol{Y^A}$ is mathematically expressed as

$$y_t^A = \sum_{j=1}^{p} \phi_j y_{t-j}^A + \varepsilon_t^A - \sum_{j=1}^{q} \theta_j \varepsilon_{t-j}^A \tag{1}$$

with p autoregressive parameters ϕ_1, ϕ_2, ..., ϕ_p, and q moving average parameters θ_1, θ_2, ..., θ_q. The term ε_t^A in equation (1) stands for an uncorrelated normal stochastic process with mean zero and variance $\sigma^2_{\varepsilon^A}$, and is also uncorrelated with $y_{t-1}^A, y_{t-2}^A, \ldots, y_{t-p}^A$. Stochastic process ε_t^A is also referred to as *white noise*, *innovation term*, or *error term*.

Time series analysis based on ARMA models provides a comprehensive mathematical framework to represent an important group of stationary processes that result from imposing a linear dependence among the process variables and a series of white noises. In particular, the autoregressive part of the ARMA(p, q) model in (1), AR(p)=$y_t^A - \sum_{j=1}^{p} \phi_j y_{t-j}^A$, establishes that the present realization y_t^A of the stochastic process $\boldsymbol{Y^A}$ directly depends on its last p realizations. It can be proved that the inclusion of the moving-average component, MA(q)= $\varepsilon_t^A - \sum_{j=1}^{q} \theta_j \varepsilon_{t-j}^A$, in combination with the autoregressive one, allows obtaining simple and compact ARMA(p, q) models, i.e., with p and q sufficiently small [28].

Observe in (1) that y_t^A boils down to a linear combination of white noises, and as such,

the marginal distribution associated with the stochastic process $\boldsymbol{Y}^{\boldsymbol{A}}$ is necessarily normal, which is in accordance with the first assumption above.

When it comes to analyze wind speed processes, there are no reasons in advance to question the validity of the stationarity assumption, as the natural phenomena able to modify the wind characteristics at a given site change in general very slowly, exhibiting a large time constant. Only the existence of seasonal winds could wreck the stationarity, but in this case, seasonal ARMA models can be used to easily overcome this difficulty. Furthermore, the proposed methodology is intended to simulate short-time wind speed sequences, and consequently, the stationarity assumption is a reasonable premise.

On the contrary, the normality assumption does not hold in general. It is widely accepted that the marginal distribution characterizing the speed of local winds can be suitably approximated by a Weibull distribution. The technical literature is rich in references on this [29, 30, 31, 32, 33], including relevant works dealing with the problem of how to estimate the parameters of the Weibull distribution that best describes a given wind speed frequency distribution [34, 35, 36, 37]. In [38], Weibull distributions for different wind sites all over the world are provided.

In order to preserve the original marginal distribution associated with the wind speed stochastic process $\boldsymbol{Y}^{\boldsymbol{A}}$ while making use of the modeling capability of the ARMA models, a new stochastic process, $\boldsymbol{Z}^{\boldsymbol{A}}$, with a standard normal marginal distribution is defined through the following transformation:

$$\boldsymbol{Z}^{\boldsymbol{A}} = \Phi^{-1}\left[F_{Y^A}(\boldsymbol{Y}^{\boldsymbol{A}})\right], \tag{2}$$

where F_{Y^A} is the cumulative distribution function (CDF) of the marginal distribution associated with the original stochastic process $\boldsymbol{Y}^{\boldsymbol{A}}$ and $\Phi(\cdot)$ is the cumulative distribution function of the standard normal random variable.

Transformation (2) comes from [39] and is graphically illustrated in Fig. 1, where the

bold path in direction *d* (direct) represents the referred transformation. This transformation allows us to preserve the marginal distribution and the covariance structure of the random variables making up the stochastic process $\boldsymbol{Y^A}$ by assuming a multivariate Gaussian joint distribution of the transformed stochastic process $\boldsymbol{Z^A}$.

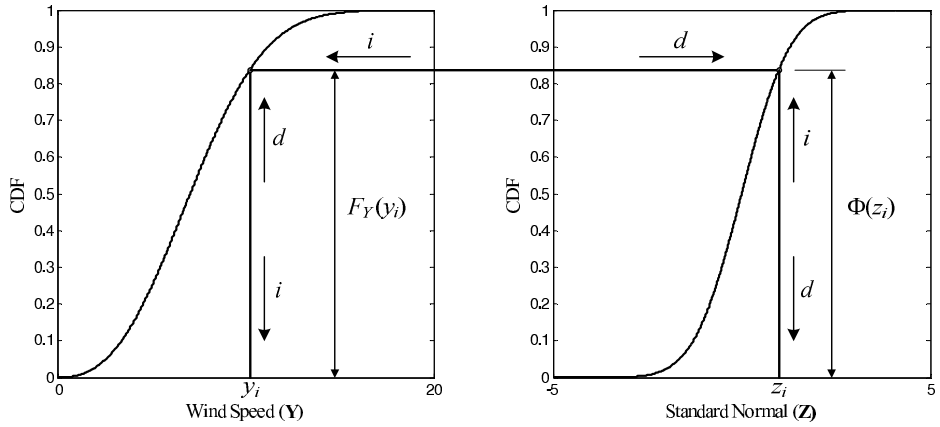


Figure 1: Graphic interpretation of the normal transformation expressed in equation (2).

Thus, the ARMA model (1) is adjusted to the transformed stochastic process $\boldsymbol{Z^A}$. From the fitted model, a *transformed* set of scenarios Ω_{Z^A} is first generated, and then *untransformed* in order to produce a set of scenarios, Ω_{Y^A}, characterizing the original wind speed stochastic process, $\boldsymbol{Y^A}$. The inverse transformation can be written as

$$\boldsymbol{Y^A} = F_{Y^A}^{-1}\left[\Phi(\boldsymbol{Z^A})\right], \tag{3}$$

and is graphically represented by the bold path in direction *i* (inverse) of Fig. 1.

The aim of the methodology introduced in this paper is to obtain a set of scenarios that allows us to properly characterize the uncertainty associated with multiple wind sites. This requires to account for both the marginal properties of each location and their interdependencies. In this respect, if an additional wind speed stochastic process at a different site B, $\boldsymbol{Y^B}$, comes into consideration, the time series analysis to be carried out becomes mul-

tivariate, and consequently, significantly thornier. In particular, if we define the following groups of vectors

$$\boldsymbol{y_{t-j}} = \begin{pmatrix} y^A_{t-j} \\ y^B_{t-j} \end{pmatrix}, \quad \boldsymbol{z_{t-j}} = \begin{pmatrix} z^A_{t-j} \\ z^B_{t-j} \end{pmatrix}, \quad j = 1, 2, \ldots, p; \tag{4}$$

$$\boldsymbol{\varepsilon_{t-j}} = \begin{pmatrix} \varepsilon^A_{t-j} \\ \varepsilon^B_{t-j} \end{pmatrix}, \quad j = 1, 2, \ldots, q; \tag{5}$$

scalars ϕ_j and θ_j in (1) turn into matrices, $\boldsymbol{\phi_j}$ and $\boldsymbol{\theta_j}$, respectively, in the following multivariate ARMA model:

$$\boldsymbol{z_t} = \sum_{j=1}^{p} \boldsymbol{\phi_j} \boldsymbol{z_{t-j}} + \boldsymbol{\varepsilon_t} - \sum_{j=1}^{q} \boldsymbol{\theta_j} \boldsymbol{\varepsilon_{t-j}}, \tag{6}$$

where the noise term $\boldsymbol{\varepsilon_t}$ has a zero mean vector and satisfies:

$$E[\boldsymbol{\varepsilon_t} \boldsymbol{\varepsilon_t^T}] = \begin{pmatrix} \sigma^2_{\varepsilon^A} & 0 \\ 0 & \sigma^2_{\varepsilon^B} \end{pmatrix}, \tag{7}$$

and $E[\boldsymbol{\varepsilon_t} \boldsymbol{\varepsilon^T_{t-j}}] = \mathbf{0}$ for $j \neq 0$. $E[\cdot]$ is the expectation operator and superscript T denotes the transpose of a matrix. In addition, it is assumed that $\boldsymbol{\varepsilon_t}$ is uncorrelated with $\boldsymbol{z_{t-1}}, \boldsymbol{z_{t-2}}, \ldots, \boldsymbol{z_{t-p}}$, and that $\boldsymbol{\varepsilon_t}$ is normally distributed.

The use of the full multivariate ARMA model described above often leads to complex parameter estimation. Thus, a model simplification is suggested in this paper.

This simplification is based on assuming that $\boldsymbol{\phi_j}$ and $\boldsymbol{\theta_j}$ are diagonal matrices. Then, model (6) can be decoupled into the univariate ARMA models

$$\begin{aligned} z^A_t &= \sum_{j=1}^{p^A} \phi^A_j z^A_{t-j} + \varepsilon^A_t - \sum_{j=1}^{q^A} \theta^A_j \varepsilon^A_{t-j} \\ z^B_t &= \sum_{j=1}^{p^B} \phi^B_j z^B_{t-j} + \varepsilon^B_t - \sum_{j=1}^{q^B} \theta^B_j \varepsilon^B_{t-j} \, . \end{aligned} \tag{8}$$

Note the consequences of this assumption as the above model decoupling into univariate models allows us to avoid estimating model parameters jointly, and well-known and computationally efficient univariate modeling procedures can be employed instead. Likewise, note that one does not have to consider the same univariate ARMA(p, q) model for each site. Thus, the model components at each site are simply univariate ARMA(p, q) models where $\boldsymbol{\varepsilon^A}$ and $\boldsymbol{\varepsilon^B}$ are not *autocorrelated* (or temporarily correlated), i.e., $E[\varepsilon_t^A \varepsilon_{t-j}^A] = 0$ and $E[\varepsilon_t^B \varepsilon_{t-j}^B] = 0, \forall\, j \neq 0$; but, as a result of the assumption above, they may be indeed *cross-correlated* (or spatially correlated), i.e., $E[\varepsilon_t^A \varepsilon_{t-j}^B] \neq 0$ and/or $E[\varepsilon_t^B \varepsilon_{t-j}^A] \neq 0$ for some j.

Therefore, if the scenario generation process is designed in such a way that the cross-correlations between the error terms, $\boldsymbol{\varepsilon^A}$ and $\boldsymbol{\varepsilon^B}$, are reproduced, then the resulting wind speed scenario set will retain the spatial correlations between the original wind speed time series, $\boldsymbol{Y^A}$ and $\boldsymbol{Y^B}$.

Mathematically, the cross-correlations between $\boldsymbol{\varepsilon^A}$ and $\boldsymbol{\varepsilon^B}$ can be described by means of a variance-covariance matrix $\boldsymbol{G}$. This matrix is symmetric by definition, and therefore, can be always diagonalized. In other words, an orthogonal transformation can be used to model the $\boldsymbol{\varepsilon}$'s as

$$\boldsymbol{\varepsilon} = \begin{pmatrix} \boldsymbol{\varepsilon^A} \\ \boldsymbol{\varepsilon^B} \end{pmatrix} = \boldsymbol{B\xi} = \boldsymbol{B} \begin{pmatrix} \boldsymbol{\xi^A} \\ \boldsymbol{\xi^B} \end{pmatrix}, \tag{9}$$

in which $\boldsymbol{B}$ is the orthogonal matrix of the transformation, and $\boldsymbol{\xi}$ is normal and satisfies that $E[\boldsymbol{\xi}\boldsymbol{\xi}^T] = \boldsymbol{I}$, being $\boldsymbol{I}$ the identity matrix. That is, $\boldsymbol{\xi}$ is temporarily and spatially uncorrelated.

This way, white noises (independent standard normal errors) can be generated, and then, cross-correlated according to the variance-covariance matrix $\boldsymbol{G}$ by using the orthogonal transformation as expressed in (9).

In most engineering applications, matrix $\boldsymbol{G}$, besides being symmetric, is also positive definite, and as such, can be decomposed through the computationally advantageous

Cholesky decomposition, which avoids the calculation of eigenvalues and eigenvectors. The Cholesky decomposition can be stated as

$$\boldsymbol{G} = \boldsymbol{L}\boldsymbol{L}^T, \tag{10}$$

where $\boldsymbol{L}$ is an inferior triangular matrix that turns out to be the orthogonal matrix required for transformation (9), i.e., $\boldsymbol{B} = \boldsymbol{L}$, as shown below.

The variance-covariance matrix $\boldsymbol{G}$ can be computed as

$$\begin{aligned} \boldsymbol{G} &= cov(\boldsymbol{\varepsilon}, \boldsymbol{\varepsilon}^T) = cov(\boldsymbol{B}\boldsymbol{\xi}, \boldsymbol{\xi}^T\boldsymbol{B}^T) = \boldsymbol{B}\, cov(\boldsymbol{\xi}, \boldsymbol{\xi}^T)\boldsymbol{B}^T \\ &= \boldsymbol{B}\, E[\boldsymbol{\xi}\boldsymbol{\xi}^T]\boldsymbol{B}^T = \boldsymbol{B}\boldsymbol{I}\boldsymbol{B}^T = \boldsymbol{B}\boldsymbol{B}^T, \end{aligned} \tag{11}$$

where $cov(\cdot, \cdot)$ stands for the covariance operator.

By using the Cholesky decomposition (10), equation (11) yields:

$$\boldsymbol{G} = \boldsymbol{L}\boldsymbol{L}^T = \boldsymbol{B}\boldsymbol{B}^T \Rightarrow (\boldsymbol{B}^{-1}\boldsymbol{L})(\boldsymbol{L}^T(\boldsymbol{B}^T)^{-1}) = \boldsymbol{I} \Rightarrow \boldsymbol{B} = \boldsymbol{L}. \tag{12}$$

The simulation of the cross-correlated error $\boldsymbol{\varepsilon}$ becomes more costly as the number of significative cross-correlation coefficients increases. In this sense, the proposed methodology is efficacious provided that the statistical dependencies among errors can be explained by a finite set of cross-correlation coefficients, and its efficaciousness is dependent on the cardinality of this set.

In summary, the purpose of this theoretical development is the design of a procedure able to provide scenario sets that reproduce:

1. The marginal distribution associated with each stochastic process through the normal transformation (2).
2. The autocorrelations characterizing the dynamic behavior of each stochastic process over the time (temporal correlations) by employing univariate ARMA models.
3. The cross-correlations determining the interrelations among the stochastic processes (spatial correlations) by means of the orthogonal transformation (9).

3. Algorithm

In this section, the procedure to generate a *plausible* scenario set for multiple wind sites is described step by step. The algorithm is also illustrated through the flow chart depicted in Fig. 2.

Step 1. For each wind site, estimate the parameters of the probability distribution that best fits its wind speed frequency distribution. This is done using the available historical data.

Step 2. For each wind site, apply transformation (2) to the historical wind speed series, where the marginal cumulative distribution function $F(\cdot)$ to be used is the one estimated in Step 1 above. This way, a *transformed* series is obtained for each wind site with an associated standard normal marginal distribution.

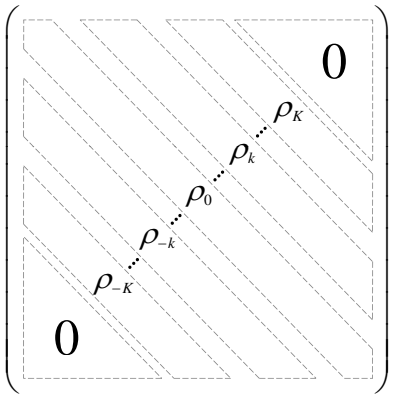
Step 3. For each wind site, adjust an univariate ARMA model to the corresponding *transformed* series (obtained in Step 2 above). The fitting process to be performed in this step is well known (see, e.g., [28]) and yields uncorrelated normal residuals (historical errors).

Step 4. Although the historical error series obtained in the preceding step are not *autocorrelated*, they should be *cross-correlated* if the wind sites under analysis are indeed interrelated. In this step, the cross-correlations are computed and, consequently, the variance-covariance matrix $\boldsymbol{G}$ is built. Considering for the sake of simplicity just two wind sites, A and B, the structure of matrix G is as follows:

$$\boldsymbol{G} = \begin{array}{c} \begin{array}{cc} \boldsymbol{\varepsilon}^A & \boldsymbol{\varepsilon}^B \end{array} \\ \left(\begin{array}{c|c} \boldsymbol{G}_{11} & \boldsymbol{G}_{12} \\ \hline \boldsymbol{G}_{21} & \boldsymbol{G}_{22} \end{array} \right) \end{array} \begin{array}{l} \boldsymbol{\varepsilon}^A \\ \boldsymbol{\varepsilon}^B \end{array} \tag{13}$$

where $\boldsymbol{G}_{11} = \sigma^2_{\varepsilon^A} \boldsymbol{I}_{K+N_T}$, $\boldsymbol{G}_{22} = \sigma^2_{\varepsilon^B} \boldsymbol{I}_{K+N_T}$, and

$$\boldsymbol{G}_{12} = \boldsymbol{G}_{21}^{T} = \sigma_{\varepsilon^A}\sigma_{\varepsilon^B}$$



.

ρ_k is the positive lag-k cross-correlation coefficient, K is the maximum lag (positive or negative) with a significant cross-correlation coefficient and N_T is the number of time periods to cover in the decision-making process.

The size of matrix $\boldsymbol{G}$ is $(K + N_T) \times N_S$, where N_S is the number of wind speed series (wind sites) under consideration. If $K \ll K + N_T$, then it is computationally advantageous to treat $\boldsymbol{G}$ as a sparse matrix.

Step 5. Apply Cholesky decomposition to $\boldsymbol{G}$ (i.e., compute $\boldsymbol{L}$ such that $\boldsymbol{G} = \boldsymbol{L}\boldsymbol{L}^{\boldsymbol{T}}$) so as to obtain the matrix $\boldsymbol{B}$ of the orthogonal transformation (9).

Step 6. Simulate $N_T \times N_S$ independent standard normal errors, $\boldsymbol{\xi}$.

Step 7. Cross-correlate the $N_T \times N_S$ standard normal errors generated in the previous step by using the orthogonal transformation (9). Specifically, for two wind sites A and B, this

cross-correlation process is performed as:

$$\begin{pmatrix} \hat{\varepsilon}^A_{n-K+1} \\ \vdots \\ \hat{\varepsilon}^A_n \\ \varepsilon^A_{n+1} \\ \vdots \\ \varepsilon^A_{n+N_T} \\ \hline \hat{\varepsilon}^B_{n-K+1} \\ \vdots \\ \hat{\varepsilon}^B_n \\ \varepsilon^B_{n+1} \\ \vdots \\ \varepsilon^B_{n+N_T} \end{pmatrix} = \boldsymbol{L} \begin{pmatrix} \varepsilon^A_{n-K+1} \\ \vdots \\ \varepsilon^A_n \\ \xi^A_{n+1} \\ \vdots \\ \xi^A_{n+N_T} \\ \hline \varepsilon^B_{n-K+1} \\ \vdots \\ \varepsilon^B_n \\ \xi^B_{n+1} \\ \vdots \\ \xi^B_{n+N_T} \end{pmatrix} \tag{14}$$

where n is the sample size of the available historical data set, $\varepsilon^A_{n-K+1}, \ldots, \varepsilon^A_n$, $\varepsilon^B_{n-K+1}, \ldots, \varepsilon^B_n$ are the last K error terms of the historical series, $\hat{\varepsilon}^A_{n-K+1}, \ldots, \hat{\varepsilon}^A_n$, $\hat{\varepsilon}^B_{n-K+1}, \ldots, \hat{\varepsilon}^B_n$ are their corresponding transformed values, and $\xi^A_{n+1}, \ldots, \xi^A_{n+N_T}$, $\xi^B_{n+1}, \ldots, \xi^B_{n+N_T}$ are the $N_T \times N_S$ ($N_S = 2$) independent standard normal errors simulated in Step 6 above.

Finally, the error series $\varepsilon^A_{n+1}, \ldots, \varepsilon^A_{n+N_T}$ and $\varepsilon^B_{n+1}, \ldots, \varepsilon^B_{n+N_T}$ are those to be introduced in the ARMA models fitted in Step 3.

Step 8. For each wind site, use the corresponding N_T cross-correlated normal errors to simulate N_T wind speed *transformed* values by means of the corresponding univariate ARMA model fitted in Step 3.

Step 9. At this point in the algorithm, N_S autocorrelated and cross-correlated series have been simulated covering N_T time periods and following a normal standard marginal distribution.

In this step, the inverse transformation (3) is applied to theses series in order to enforce the actual marginal distributions that have been estimated in Step 1.

Step 10. Repeat Steps 6-9 until the desired number (N_W) of wind speed scenarios is obtained.

4. Results and discussion

In this section, the proposed methodology for wind scenario generation is tested on four different case studies. Specifically, these case studies are aimed to validate the proposed scenario generation methodology by comparing the statistical properties of historical wind speed series with the scenarios generated by the designed algorithm.

4.1. Site description

The following case studies are based on the wind data collected at five sites in Massachusetts, USA. All the historical wind speed series employed are publicly available in [40]. The data sets comprise ten-minute averaged speed measurements that have been aggregated into hourly time intervals. An approximate geographical distribution of the sites as well as the distances among them are shown in Fig. 3.

Next, a brief description of each wind site is provided:

Site A: The wind monitoring station is located at Bishop and Clerks, on the top of a lighthouse placed within the three-mile state limit of Massachusetts's waters, at 41 34' 27.6" North and 70 14' 59.5" West. The anemometry is mounted at a height of 15 m relative to the mean low water level.

Site B: The measurement station is located at the town Water Treatment Plant in Falmouth. The location of the tower base is at 41 36' 21.6" North, 70 37' 15.60" West. The wind monitoring equipment includes anemometers at three different heights on

the tower: 10 m, 30 m and 39 m. Redundant anemometers exist at 30 m and 39 m. The wind data series used in this paper from this site corresponds to a height of 39 m.

Site C: The site is located at a parking lot very near the beach on Ocean View Drive in Wellfleet. The location of the tower base is at 41 56' 2.4" North, 69 58' 48" West. Wind monitoring equipment comprises anemometers situated at three heights on the tower: 50 m, 38 m, and 20 m. Redundant anemometers are positioned at 50 m and 38 m. The wind speed series considered here is the one obtained at a height of 38 m.

Site D: The monitoring station is located at the WBZ radio tower, near the salt marsh on the west coast of the Hull isthmus, on the southern portion of the Boston Harbor. The site coordinates are 42 16' 44.11" North by 70 52' 34.39" West, and the anemometer from which the wind data employed comes from is placed at a height of 61 m.

Site E: The monitoring station is located at the Yankee Network Tower on Mount Asnebumskit, southeast of the town of Paxton. Site coordinates are 42 18' 11.6" North, 71 53' 50.9" West. The wind data series collected from this site corresponds to a height of 78 m.

In each case study, the proposed methodology is used to simulate time series reproducing the main marginal and joint statistical properties of two wind sites from among those described above.

4.2. Case study 1

In this first case study, the proposed technique is applied to model wind speeds at sites A and B. The wind data considered for this analysis were collected between June 1 and

December 30, 2004. This case study also serves us to illustrate the algorithm described in Section 3 step by step.

Step 1. First, an appropriate probability distribution is adjusted to the wind speed frequency distribution built for each site from the available historical data. A Weibull distribution is selected for the fitting process. As an example, Fig. 4 plots both the wind speed frequency distribution of site A and the corresponding probability density function (pdf) of the fitted Weibull. The adjustments are carried out by using the distribution fitting tool provided in the statistics toolbox of Matlab 7.0.

Step 2. Next, the marginal transformation (2) is used in order to *normalize* the wind speed historical data. Fig. 5 is the normality plot of the transformed data corresponding to site A. Needless to say, the normal transformation is not perfect because neither is the Weibull fit. However, considering the purpose of this paper, it constitutes an accurate approximation.

Step 3. The next step of the algorithm consists in *eliminating* the temporal correlations from the historical data series. To this end, we appeal to the assumption that each *normalized* wind speed series can be statistically treated as a stationary stochastic process with a multivariate Gaussian joint distribution. This way, univariate ARMA(p, q) models can be used to explain the dynamic behavior exhibited by the historical wind speed series. The corresponding fitting process is well-known (see, e.g., [28]), and consequently, is not discussed here. Nonetheless, the type of ARMA model fitted for each wind site is provided in Table 1. As indicated in Section 2, ARMA models boil down to a linear dependence among the process variables and a sequence of white noises. The validation of the assumption above then involves the inspection and characterization of residuals derived from the fitting process. Since the ARMA fits summarized in Table 1 yield temporarily uncorrelated normal residuals with zero mean and constant variance, i.e., white noises, such fits can be deemed as consistent [28], and therefore, the previous assumption holds.

Table 1: Stochastic processes for wind sites

Site	Time series model
A	ARMA$(1, 2)$
B	ARMA$(1, 2)$
C	ARMA$(1, 3)$
D	ARMA$(1, 2)$
E	ARMA$(2, 2)$

Steps 4 and 5. The residual errors derived from the univariate time series analysis should not be autocorrelated, but they are cross-correlated if there exists an spatial correlation between wind sites A and B. In Fig. 6, the cross-correlogram of the residuals is depicted. It can be observed that the most significant correlation coefficients form a group around the lag zero, with the lag-0 correlation coefficient being precisely the largest one. This correlation coefficient is usually referred to as *contemporaneous correlation*. The triangular structure characterizing the cross-correlogram plotted in Fig. 6 is typical of stochastic processes with a common, but delayed, physical cause. The triangle base length gives an idea of the delay magnitude. The narrower and the more pointed this triangle is the more *contemporaneous* the stochastic processes are. In this case, the triangle-shaped cross-correlogram is justified by the proximity of sites A and B (around 31 km apart). Due to their closeness, it is no wonder that winds at both sites are highly related.

From this point, the algorithm focuses on simulating errors that reproduce the cross-correlogram of Fig. 6. Nevertheless, not all the cross-correlation coefficients should be

reproduced, but just those considered significant enough from a statistical point of view. The significance of a given cross-correlation coefficient is appraised by means of a *p-value* testing the hypothesis of no correlation. This p-value is the probability of obtaining a correlation as large as the observed one by random chance, when the true correlation is zero. In Fig. 7(a), only those coefficients with a p-value smaller than a *significance level* $\alpha = 0.05$ are represented. The variance-covariance matrix $\boldsymbol{G}$ is built (*Step 4*) from the cross-correlations coefficients of Fig. 7(a) as explained in Section 3, Step 4. In pursuit of computational efficiency, the Cholesky decomposition is then applied to matrix $\boldsymbol{G}$ (*Step 5*) in order to obtain the matrix $\boldsymbol{B}$ of the orthogonal transformation (9).

Steps 6 and 7. The simulation process starts from generating $2 \times N_T$ independent standard normal errors (*Step 6*), which are subsequently cross-correlated (*Step 7*) according to matrix $\boldsymbol{G}$ through the orthogonal transformation (9). For comparative purposes, the cross-correlogram of the simulated errors for $N_T = 50\ 000$ is depicted in Fig. 7(b). The time span to cover has been selected sufficiently long in order to properly observe and assess the stationary properties of the simulated processes. Note the remarkable fidelity of the reproduction, which highlights the effectiveness of the suggested cross-correlation procedure.

Step 8. Once the cross-correlated errors have been generated, they are introduced into the previously fitted ARMA models in order to obtain the simulated series of *normalized* wind speeds for sites A and B.

Step 9. Finally, by applying the inverse transformation (3), the simulated wind speed series preserving the main marginal and joint statistical properties of winds at both sites are built.

In Fig. 8, the cross-correlogram of the simulated wind speed series is compared with that obtained from the historical data. The 95% confidence limits are also shown. The

lower and upper confidence bounds of the population lag-k cross-correlation coefficient, r_k^{lo} and r_k^{up}, respectively, are computed by applying the Fisher Transformation [41] to the sample cross-correlation coefficient ρ_k, that is:

$$r_k^{\text{lo}} = \tanh\left[\frac{1}{2}\log\left(\frac{1+\rho_k}{1-\rho_k}\right) + \frac{\text{erfinv}\,(\beta-1)\,\sqrt{2}}{\sqrt{n-k-3}}\right] \tag{15}$$

$$r_k^{\text{up}} = \tanh\left[\frac{1}{2}\log\left(\frac{1+\rho_k}{1-\rho_k}\right) - \frac{\text{erfinv}\,(\beta-1)\,\sqrt{2}}{\sqrt{n-k-3}}\right] \tag{16}$$

where erfinv$(\cdot)$ is the value of the inverse error function, n is the sample size of the available historical data set, and β is the confidence level (equal to 0.05 in this case). Caution should be exercised with respect to these confidence limits as they are accurate for a multivariate normal distribution, which is not the case here. Consequently, these limits are approximate and used in this paper just for comparative purposes. Nevertheless, extensive numerical simulations show that the confidence bands obtained in the original space as indicated (with random variables following Weibull distributions) are very similar to those resulting in the *transformed* space (with normally distributed variates).

Note that the proposed scenario generation methodology succeeds in capturing the shape of the sample cross-correlogram with a remarkable degree of precision, albeit some cross-correlation coefficients fall out of the confidence limits. This is mostly due to the irremediable inaccuracies associated with the fitting processes (both ARMA and Weibull fits) and with the cross-correlogram reproduction task. It is also important to underline the significant spatial correlation existing between wind speeds at sites A and B, with a contemporaneous correlation coefficient equal to 0.8643, which, as pointed out above, is due to the geographical proximity of both sites.

This strong correlation can be graphically observed in Fig. 9(a), where a 48-h sample from the historical time series is depicted. Fig. 9(b) illustrates a 48-h extract from the

simulated wind speed series. By simple inspection, one can intuitively realize that the simulation retains the major cross-correlation properties.

Finally, in order to illustrate the ability of the proposed methodology to preserve the marginal properties of the wind stochastic processes, Fig. 10 shows, for wind site A, a comparison between the empirical cdf of the simulated series and the cdf of the Weibull distribution estimated from the historical data. Their similarity is clear from this figure. So as to complement this graphical contrast by giving simulation error values, Table 2 offers a numerical comparison between the 5, 25, 50, 75, and 95% percentiles computed from both the simulated and the historical wind series. The last row of this table provides the error of the simulation relative to the sample percentile. Note that this error keeps reasonably low.

Table 2: Percentile comparison. Wind site A

	Percentile				
	5	25	50	75	95
Sample	1.85	3.62	5.11	6.69	9.04
Simulated	1.91	3.68	5.06	6.61	9.09
Error (%)	3.31	1.63	0.85	1.22	0.56

4.3. Case studies 2, 3 and 4

In this subsection, the results obtained for case studies 2, 3 and 4 are presented, discussed and compared. These case studies are described below:

Case study 2 tackles the modeling of wind speeds at sites A and C (see Fig. 3). The

historical data used for this purpose are comprised between June 1 and December 31, 2007.

Case study 3 analyzes wind speeds at sites A and D (see Fig. 3). The data set employed in this analysis includes wind speed measurements from June 1 to December 31, 2007.

Case study 4 focuses on modeling wind speeds at sites A and E (see Fig. 3). The wind data series used in this study were collected in the time span between June 1 and December 31, 2004.

In Fig. 11, the cross-correlograms of both the residuals derived from the ARMA fits and the errors simulated by the proposed methodology are plotted for the three case studies. Only those correlation coefficients with a p-value smaller than a significance level $\alpha = 0.05$ are depicted. The rest of them are considered to be nil. In view of these cross-correlograms, the following observations are in order:

1. In terms of spatial correlation, case studies 1 and 2 are very similar. They are both characterized by cross-correlograms with a well-defined triangular shape in which the contemporaneous correlation coefficient stands out among the rest. This similarity is mainly due to the fact that distances A-B (31.05 km) and A-C (45.80 km) are of the same order of magnitude. However, the contemporaneous correlation between wind sites A and B is a slightly greater than that between sites A and C. In this sense, note that A is closer to B than to C.
2. As the distance between wind sites increases, the triangular structure characterizing the cross-correlograms flattens. In fact, the cross-correlogram of sites A and E is not a triangle, rather it can be described as a rectangle made up of a finite number of correlation coefficients with none particularly standing out from the others.
3. The suggested technique succeeds in reproducing by simulation the sample cross-correlograms. Moreover, in light of the results obtained from these case studies, its

accuracy (in absolute terms) is observed to be independent of the distance between wind sites.

Fig. 12 is analogous to Fig. 11, but considering a significance level α equal to 0.2. Consequently, a greater number of correlation coefficients are taken into account in the analysis. It should be noted that this increase in the significance level introduces important modifications into the cross-correlogram of wind sites A and E, to which a notable number of correlations coefficients, previously disregarded, are now added. Furthermore, these new correlations are not as *negligible* as they are in the other case studies if compared with the rest of coefficients composing such cross-correlogram.

In Fig. 13, the cross-correlograms of the sample and simulated wind speed series are compared for each case study. The 95% confidence limits obtained from the historical series are also drawn. The difference between the graphics labelled as "$\alpha = 0.05$" and those tagged as "$\alpha = 0.2$", stems from the significance level imposed on the residual cross-correlations coefficients to be reproduced in the simulation process. Note that the consideration of a higher significance level, or equivalently, a greater number of residual cross-correlations coefficients, does not have a significant effect on the simulated cross-correlograms in case studies 2 and 3 (A-C and A-D), but just slight improvements are observed, especially for high lags. The reason for this is that the residual cross-correlation coefficients making up the characteristic triangular structure referred to above explain almost completely the spatial interrelations between the wind sites defining these two case studies. In contrast, the impact of the considered significance level on the simulated cross-correlogram of case study 3 (A-E) is much more remarkable. In this respect, it should be noted that, for $\alpha = 0.05$, the simulated cross-correlogram is far from falling within the 95% confidence limits, as opposed to what happens if $\alpha = 0.2$. In this case, the residual cross correlations that are disregarded if the significance level is reduced from 0.2 to 0.05

are important in relative terms, that is, in comparison with the magnitude of the rest of coefficients.

Concerning computational issues, all the simulations have been carried out using Matlab 7.0 on an Intel Core Duo, 1.83 GHz and 1 GB of RAM. CPU time required to generate 1000 wind speed scenarios covering 1 day (24 hourly periods) is around 3 minutes for all the considered case studies and their variants ($\alpha = 0.05$ or $\alpha = 0.2$). If the number of hourly periods is increased up to 2200 (i.e., wind speed scenarios spanning three months approximately), the CPU time keeps below 7 minutes in all cases.

In short, the results presented in these four case studies allow us to conclude that the proposed wind scenario generation procedure performs satisfactorily in:

1. Reproducing the marginal distribution of the stochastic process modeling wind speed at each site (Fig. 10 and Table 2).
2. Capturing the dynamic relationships in time series data by means of the appropriate fit of ARMA models (Table 1).
3. Preserving the main cross-correlations among the stochastic processes representing wind speed at different locations (Figs. 7–9, 11–13).

5. Conclusions

This paper proposes a procedure to produce a set of *plausible* scenarios characterizing the uncertainty associated with wind speed at different geographic sites. This characterization constitutes an essential part within the decision-making processes faced by both power system operators and producers with a generation portfolio including wind plants at several locations.

The uncertainty characterization achieved through this methodology is such that the generated scenario sets retain the main statistical properties of the wind speed stochastic

processes involved, namely:

1. The marginal distribution associated with each wind speed stochastic process.
2. The autocorrelations characterizing each wind speed stochastic process as a time series (temporal correlations).
3. The predominant cross-correlations among the wind speed stochastic processes (spatial correlations).

Comprehensive simulations carried out for different case studies show the effectiveness of the proposed methodology in generating statistically consistent wind scenario sets to be used within a stochastic programming framework.

Scenarios generated by the proposed technique are most useful for both operations and planning studies involving wind power plants. Operations applications include offering algorithms for wind producers, market-clearing procedures for market operators in systems with a high penetration of wind power, etc.; planning applications include siting and timing of wind facilities, transmission expansion in systems with a high penetration of wind power, etc. The limitations of the technique proposed are mostly related to the size of the problems for which scenarios are generated, because the number of considered scenarios should not lead to problem intractability.

References

[1] P. A. Østergaard, Ancillary services and the integration of substantial quantities of wind power, Appl. Energy 83 (2006) 451–463.

[2] P.J. Luickx, E.D. Delarue, W.D. D'haeseleer, Considerations on the backup of wind power: Operational backup, Appl. Energy 85 (2008) 787–799.

[3] J.M. Morales, A.J. Conejo, J. Pérez-Ruiz, Economic Valuation of Reserves in Power Systems with High Penetration of Wind Power, IEEE Trans. Power Syst. 24 (2009) 900–910.

[4] F. Bouffard, F.D. Galiana, Stochastic Security for Operations Planning with Significant Wind Power Generation, IEEE Trans. Power Syst. 23 (2008) 306–316.

[5] Y.P. Cai, G.H. Huang, Z.F. Yang, Q. Tan, Identification of optimal strategies for energy management systems planning under multiple uncertainties, Appl. Energy 86 (2009) 480–495.

[6] J. Aghaei, H.A. Shayanfar, N. Amjady, Joint market clearing in a stochastic framework considering power system security, Appl. Energy 86 (2009) 1675–1682.

[7] G.N. Bathurst, J. Weatherill, G. Strbac, Trading Wind Generation in Short-Term Energy Markets, IEEE Trans. Power Syst. 17 (2002) 782–789.

[8] P. Pinson, C. Chevallier, G.N. Kariniotakis, Trading Wind Generation from Short-Term Probabilistic Forecasts of Wind Power, IEEE Trans. Power Syst. 22 (2007) 1148–1156.

[9] J. Matevosyan, L. Söder, Minimization of Imbalance Cost Trading Wind Power on the Short-Term Power Market, IEEE Trans. Power Syst. 21 (2006) 1396–1404.

[10] J.M. Morales, A.J. Conejo, J. Pérez-Ruiz, Short-Term Trading for a Wind Power Producer, IEEE Trans. Power Syst. (in press).

[11] M. Mohr, H. Unger, Economic reassessment of energy technologies with risk-management techniques, Appl. Energy 64 (1999) 165–173.

[12] J.R. Birge, F. Louveaux, Introduction to stochastic programming, Springer-Verlag, New York, 1997.

[13] J.L. Higle, S. Sen, Stochastic decomposition – an algorithm for 2-stage linear-programs with recourse, Math. Oper. Res. 16 (1991) 650–669.

[14] J. Dupačová, G. Consigli, S.W. Wallace, Scenarios for multistage stochastic programs, Ann. Oper. Res. 100 (2000) 25–53.

[15] J. Dupačová, N. Gröwe-Kuska, W. Römisch, Scenario reduction in stochastic programming: An approach using probability metrics, Math. Program., Ser. A, 95 (2003) 493–511.

[16] K. Høyland, S.W. Wallace, Generating Scenario Trees for Multistage Decision Problems, Manage. Sci. 47 (2001) 295–307.

[17] J.M. Morales, S. Pineda, A.J. Conejo, M. Carrión, Scenario Reduction for Futures Market Trading in Electricity Markets, IEEE Trans. Power Syst. (in press).

[18] M. Carrión, A.J. Conejo, J.M. Arroyo, Forward Contracting and Selling Price Determination for a Retailer, IEEE Trans. Power Syst. 22 (2007) 2105–2114.

[19] A.J. Conejo, F.J. Nogales, M. Carrión, J.M. Morales, Electricity pool prices: long-term uncertainty characterization for futures-market trading and risk management, J. Oper. Res. Soc. (in press).

[20] M. Olsson M, L. Söder. Generation of regulating power price scenarios. In: Proceedings of the 8th International Conference Probabilistic Methods Applied to Power Systems PMAPS, Ames, Iowa; 2004. p. 26–31.

[21] G.C. Thomann, M.J. Barfield, The time variation of wind speeds and wind farm power output in Kansas, IEEE Trans. Energy Convers. 3 (1988) 44–49.

[22] R. Billinton, H. Chen, R. Ghajar, Time-series models for reliability evaluation of power systems including wind energy, Microelectron. Reliab. 36 (1996) 1253–1261.

[23] S.H. Karaki, B.A. Salim, R.B. Chedid, Probabilistic model of a two-site wind energy conversion system, IEEE Trans. Energy Convers. 17 (2002) 530–536.

[24] R. Billinton, G. Bai, Generating capacity adequacy associated with wind energy, IEEE Trans. Energy Convers. 19 (2004) 641–646.

[25] W. Wangdee, R. Billinton, Considering load-carrying capability and wind speed correlation of WECS in generation adequacy assessment, IEEE Trans. Energy Convers. 21 (2006) 734–741.

[26] M.S. Miranda, R.W. Dunn, Spatially correlated wind speed modelling for generation adequacy studies in the UK. In: Proceedings of the IEEE Power Engineering Society General Meeting, Tampa, USA; 2007. p. 4783–4788.

[27] K. Xie, R. Billinton, Considering wind speed correlation of WECS in reliability evaluation using the time-shifting technique, Electr. Power Syst. Res. 79 (2009) 687-693.

[28] D. Peña, G.C. Tiao, R.S. Tsay (Eds), A Course in Time Series Analysis, John Wiley, New York, 2001.

[29] J.P. Hennessey, Some aspects of wind power statistics, J. Appl. Meteorol. 16 (1977) 119–128.

[30] DP. Lalas, H. Tselepidaki, G. Theoharatos, An analysis of wind power potential in Greece, Sol. Energy 30 (1983) 497–505.

[31] H. Nfaoui, J. Buret, A.A. Sayigh, Wind characteristics and wind energy potential in Morocco, Sol. Energy 63 (1998) 51–60.

[32] A. Garcia, J.L. Torres, E. Prieto, A. de Francisco, Fitting wind speed distributions: a case study, Sol. Energy 62 (1998) 139–144.

[33] D. Feretic, Z. Tomsic, N. Cavlina, Feasibility analysis of wind-energy utilization in Croatia, Energy 24 (1999) 239–246.

[34] Y.F.I. Lun, J.C. Lam, A study of Weibull parameters using long-term wind observations, Renew. Energy 20 (2000) 145–153.

[35] J.V. Seguro, T.W. Lambert, Modern estimation of the parameters of the Weibull wind speed distribution for wind energy analysis, J. Wind Eng. Ind. Aerodyn. 85 (2000) 75–84.

[36] A.N. Celik, Assessing the suitability of wind speed distribution functions based on wind power density, Renew. Energy 28 (2003) 1563–1574.

[37] P. Ramírez, J.A. Carta, Influence of the data sampling interval in the estimation of the parameters of the Weibull wind speed probability density distribution: a case study, Energy Conv. Manag. 46 (2005) 2419–2438.

[38] Danish Wind Industry Association, Denmark, 2009.
http://www.windpower.org/en/tour/wres/pow/index.htm

[39] P.-L. Liu, A. Der Kiureghian, Multivariate distribution models with prescribed marginals and covariances, Probab. Eng. Mech. 1 (1986) 105–112.

[40] RERL Wind Data. Renewable Energy Research Laboratory. Center for Energy

Efficiency & Renewable Energy. University of Massachusetts at Amherst, 2009. http://www.ceere.org/rerl/publications/resource_data/index.html.

[41] H. Hotelling, New Light on the correlation coefficient and its transforms, J. R. Stat. Soc. Ser. B-Stat. Methodol. 15 (1953) 193–232.

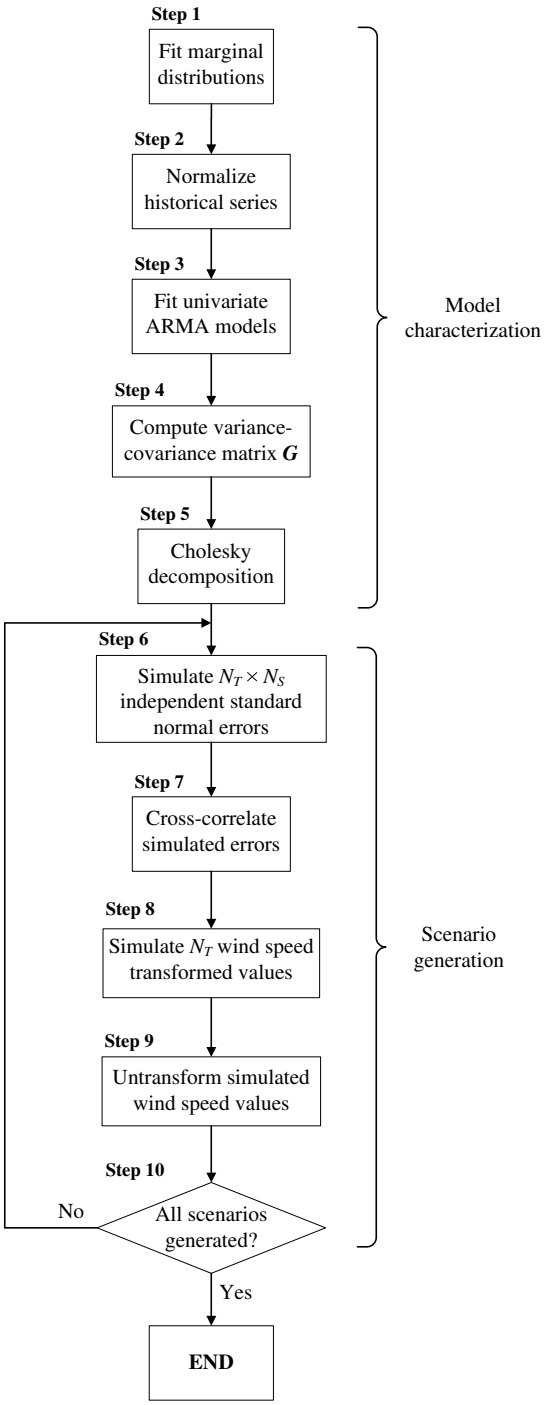


Figure 2: Flow diagram of the algorithm.

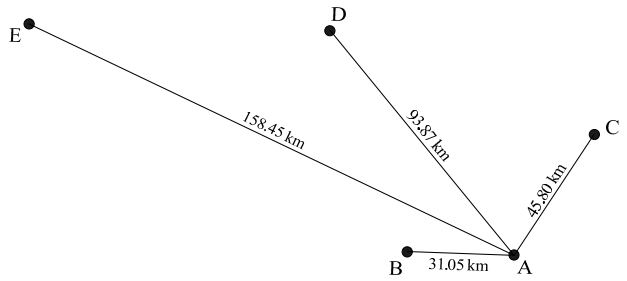


Figure 3: Schematic geographical distribution of wind sites and distances among them.

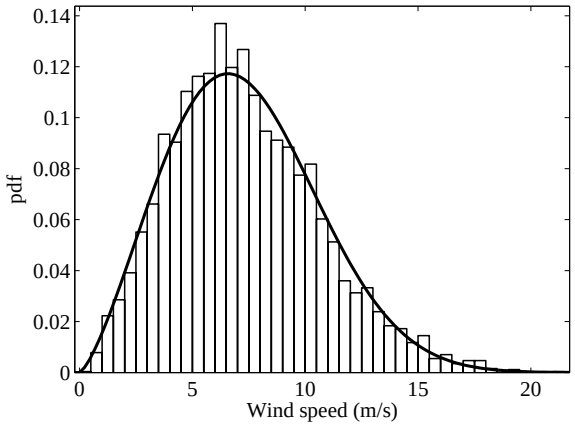


Figure 4: Wind speed frequency distribution for site A and the pdf of the Weibull fit.

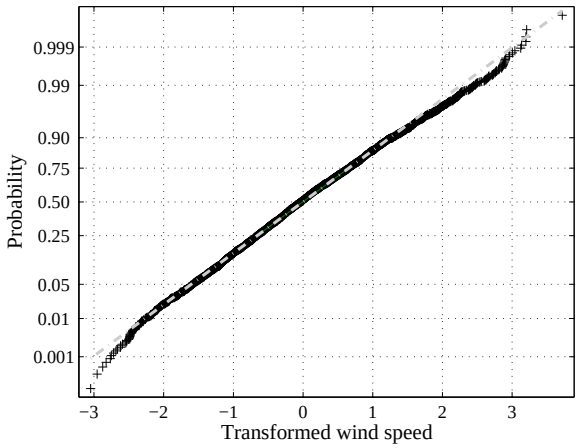


Figure 5: Normality plot of the transformed data for site A.

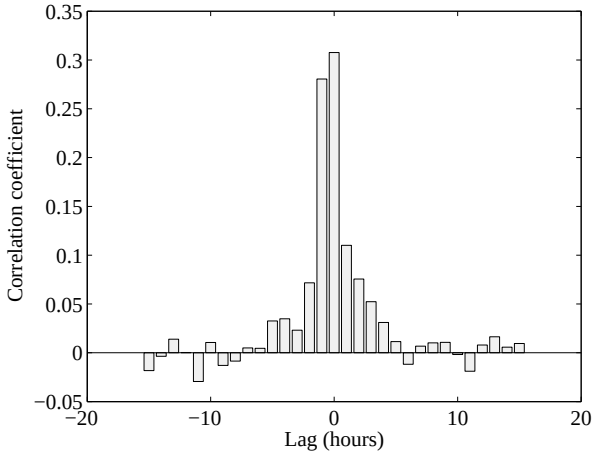
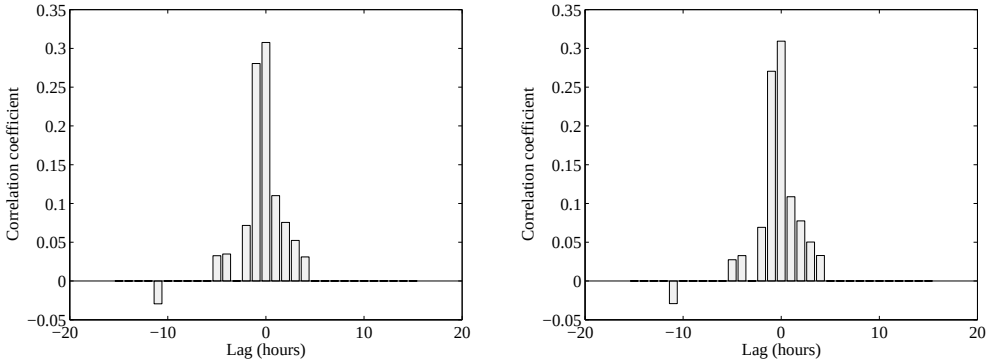


Figure 6: Cross-correlogram of the residuals obtained after the ARMA fit.



(a) Sample (b) Simulated

Figure 7: Cross-correlograms of errors with a significance level α= 0.05.

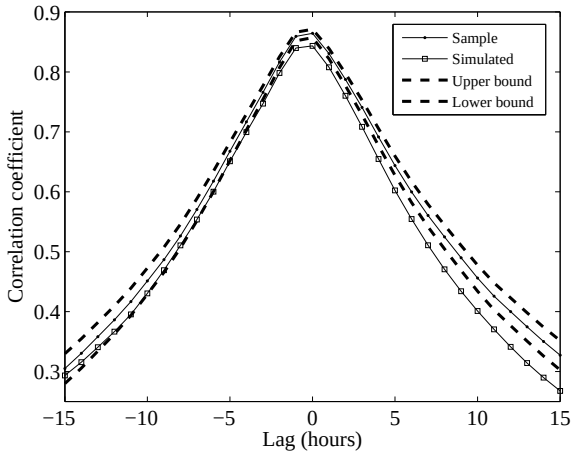
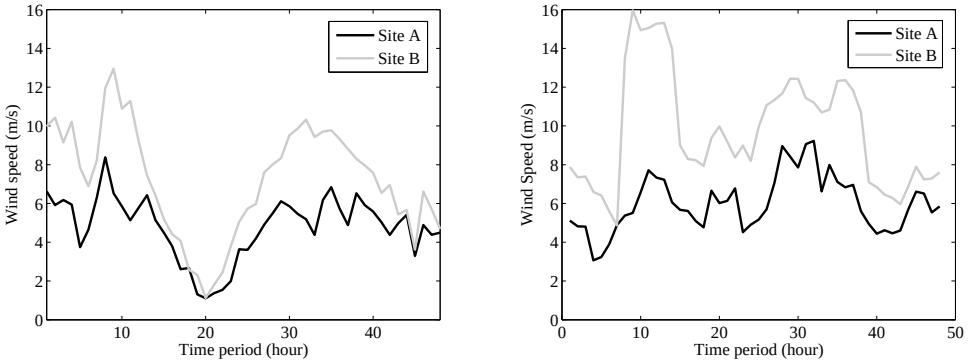


Figure 8: Comparison between sample and simulated cross-correlograms.



(a) Sample (b) Simulated

Figure 9: A 48-h extract from the historical and simulated wind speed series at sites A and B.

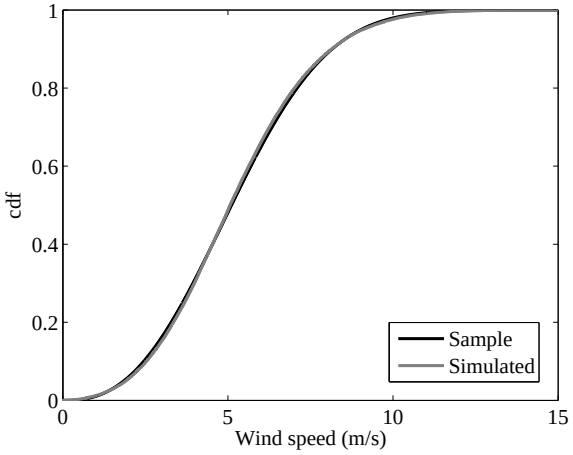
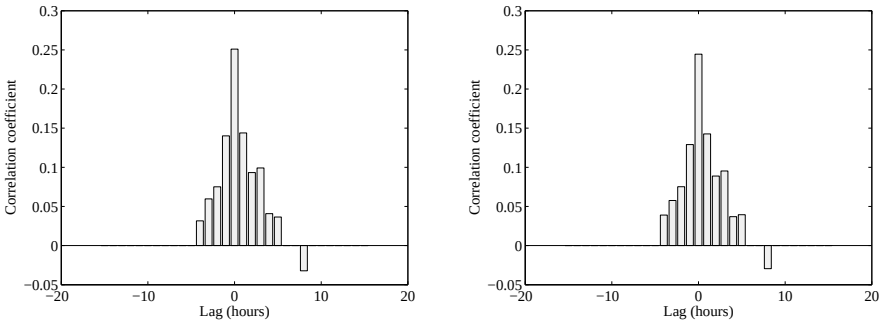


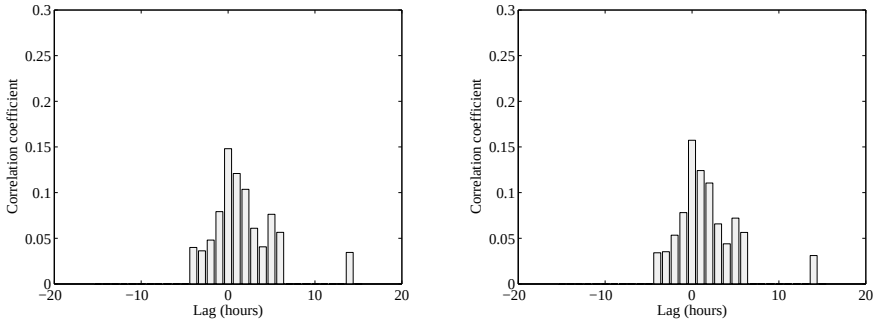
Figure 10: Comparison between fitted and simulated cumulative probability functions. Wind site A.

Sites A and C



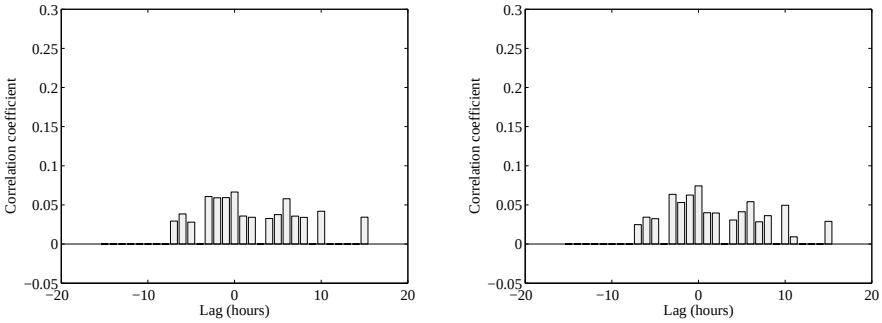
(a) Sample (b) Simulated

Sites A and D



(c) Sample (d) Simulated

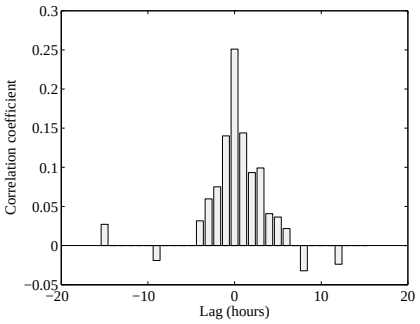
Sites A and E



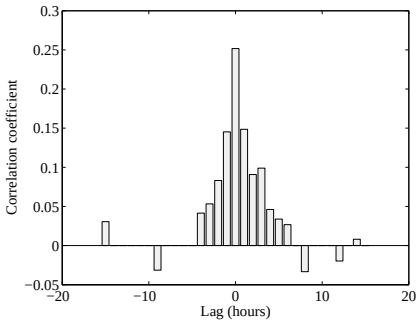
(e) Sample (f) Simulated

Figure 11: Cross-correlograms of errors with $\alpha = 0.05$

Sites A and C

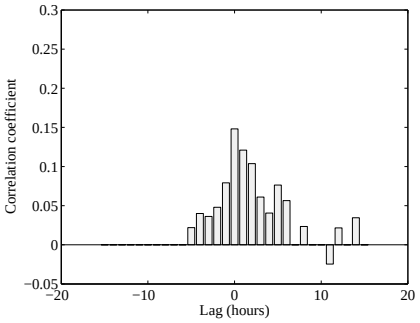


(a) Sample

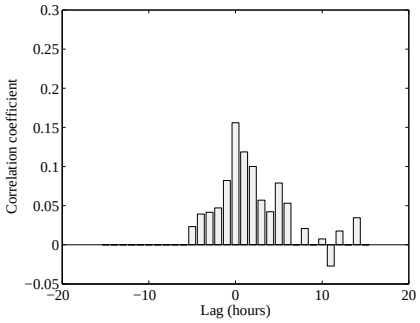


(b) Simulated

Sites A and D

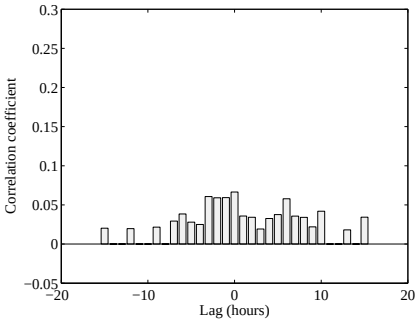


(c) Sample

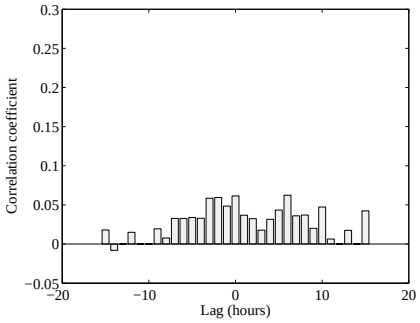


(d) Simulated

Sites A and E



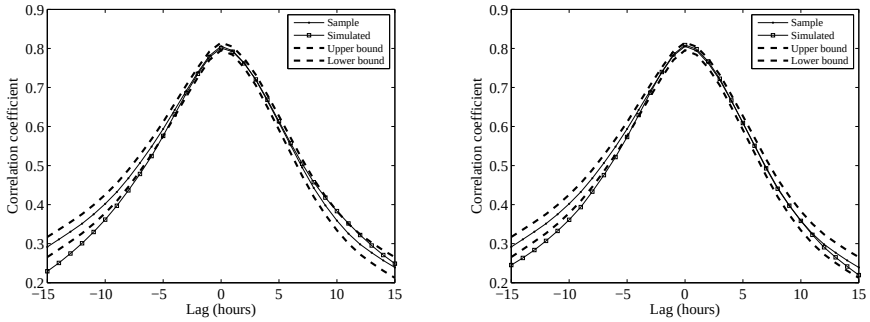
(e) Sample



(f) Simulated

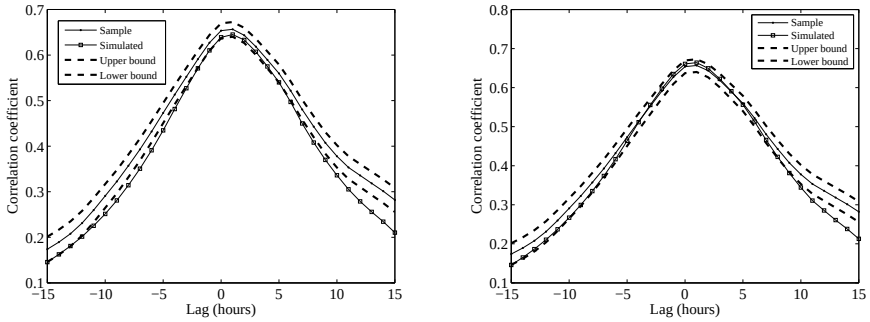
Figure 12: Cross-correlograms of errors with $\alpha = 0.2$.

Sites A and C



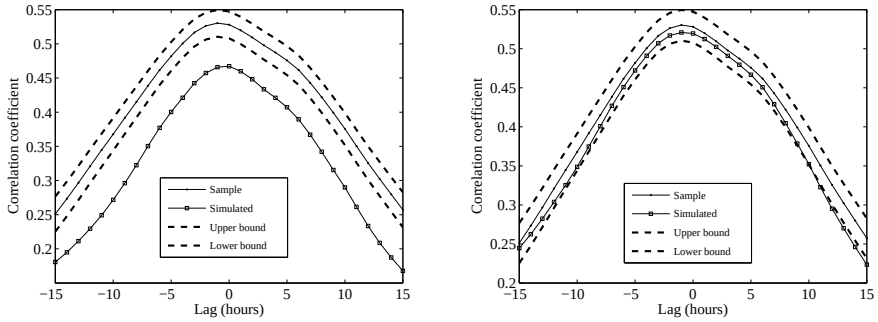
(a) $\alpha = 0.05$ (b) $\alpha = 0.2$

Sites A and D



(c) $\alpha = 0.05$ (d) $\alpha = 0.2$

Sites A and E



(e) $\alpha = 0.05$ (f) $\alpha = 0.2$

Figure 13: Comparison between sample and simulated cross-correlograms.