Applied Energy 173 (2016) 283–295

Contents lists available at ScienceDirect

Applied Energy

journal homepage: www.elsevier.com/locate/apenergy

Stochastic modeling to represent wind power generation and demand in electric power system based on real data

Humberto Verdejo [a,*], Almendra Awerkin [b], Eugenio Saavedra [b], Wolfgang Kliemann [c], Luis Vargas [d]

[a] *Department of Electrical Engineering, University of Santiago, Santiago, Chile*
[b] *Department of Mathematics, University of Santiago, Santiago, Chile*
[c] *Department of Mathematics, Iowa State University, IA, USA*
[d] *Department of Electrical Engineering, University of Chile, Santiago, Chile*

HIGHLIGHTS

- The Ornstein–Uhlenbeck process is used to model stochastic random perturbations.
- Multi-correlated model that represents windmills interactions.
- Validation technique for continuous unidimensional stochastic models.
- Short-time multidimensional wind generation forecasting model (hours).
- Medium-time one dimensional residential power demand forecasting hybrid model.

ARTICLE INFO

Article history:
Received 8 October 2015
Received in revised form 17 March 2016
Accepted 6 April 2016
Available online 16 April 2016

Keywords:
Stochastic systems
Power systems
Estimation

ABSTRACT

A methodology to model two types of random perturbation that affect the operation of electric power systems (EPS) are presented. The first uncertainty is wind power generation and is represented by a one-dimensional and by a multidimensional continuous stochastic process. The second one is power demand, and is modeled by using an hybrid structure based on harmonic regression and the Ornstein–Uhlenbeck (O–U) process. The stochastic models are applied to a real Chilean case, using real data for parametric estimation and validation models.

1. Introduction

Dynamic and permanent regime studies that deal with electric power systems (EPS) are of vital importance to the electrical industry, because they make it possible to determine the adequate operating conditions for supplying the electric power required by society in an economic, reliable and safe manner. In this context, the most important approaches of the EPS studies are oriented at their planning and operation.

One of the main problems that concern planning and operation consists on keeping the system operating in a steady state, i.e., that the system does not lose its balance when it is subjected to perturbations that affect its behavior. The most frequent perturbations found in EPS are fault occurrence, load level variation, changes in the network topology, and the presence of random components caused by generation sources based on unconventional renewable energy.

In particular, wind generation and power demand will be the focus of this paper. Due to the stochastic nature governing these two types of disturbances, it is appropriate to consider statistical models to represent their behaviors and then, perform more realistically studies about their impact over the EPS [1].

Development of more precise and accurate models is of high importance for improving the results of their subsequent application [2–7]. For instance, paper [2] uses Kalman filtering in the context of short-term prediction wind speed, reflecting in turn into a better planning and usage of the power resource.

In [3], statistical regime-switching models are applied in order to modeling the fluctuations of offshore wind generation. The research is oriented to obtain models dedicated to enhance the existing control and energy management strategies at offshore wind parks.

Ref. [4] applied a Markov-switching model to perform point and interval forecasting of wind speed. It Emphasize that an accurate

* Corresponding author.
E-mail address: humberto.verdejo@usach.cl (H. Verdejo).

http://dx.doi.org/10.1016/j.apenergy.2016.04.004
0306-2619/

wind speed interval forecasting is beneficial to the robust optimization in wind farm operational planning.

A statistical and dynamical modeling are proposed in [5] for various classes of wind speed fluctuations distributions. A generated wind speed sequence in time is made solving the Langevin equation with different turbulence conditions. Ref. [6] uses Stochastic Differential Equations (SDE), based on Ornstein–Uhlenbeck (O–U) process, to develop methods to model wind speed.

Paper [7] presents models based on SDE and O–U process of wind power production, besides, base production and base consumption which can be used to evaluate the impact on power systems balancing. In addition, estimation methods for parameters are proposed and a case study is presented.

As can be seen, accurate wind speed and wind power simulations and forecasting have an important influence on studies and decision-making of EPS [8,9]. For this reason, a large number of techniques have been developed. Papers related to comparisons between different methods can be found in [10–14].

In [11] a comparison between time series and Artificial Neural Networks (ANN) models is presented by a long-therm prediction of production of wind power station in Mexico. Another comparison is made in [12], where Autoregressive – Moving Average (ARMA), 5 kinds of ANN and the Adaptive neuro fuzzy inference system (ANFIS) models are compare in different time horizons.

In [13], it is proposed a new method based on multiple architecture system (MAS).

Conversely, load uncertainty models can be found in [15–20], where they use continuous stochastic process in order to model the behavior of the loads for stability studies.

While continuous models are widely used for application to stability studies, other kinds of models are applied to modeling demand.

In [21–25], regression models to study residential energy demand are presented. Refs. [21,22] use multilinear regression with the aim of predicting future values of energy demand. On the other hand, logistic regression method is used in [23], where it is applied to analyse the domestic electric consumption types. In [24,25] harmonic regression is used taking advantage of their ability to describe processes with marked seasonality.

Time series analysis is also largely used in this context [26–30]. For example, in [29] residential demand is modeled and predicted with a Seasonal Autoregressive Integrated Moving Average with Exogenous Variables model (SARIMAX), considering as a data real residential demand measures.

The models and applications are vast, so in order to sum up all the techniques used to represent wind generation and power demand, Tables 1 and 2 classify the principal methods reviewed and expose their benefits and principal cases of application.

Classification of Table 1 is based on Refs. [31,32], and Table 2 maintains the structure classification but considering Ref. [33] as a guideline. It is important to say that Table 2 is focused on residential power demand because this type of load is the most commonly analysed.

As the literature shows, there is a wide variety of tools to study the impact of stochastic perturbation, as wind generation, and different approaches can be made. However the main purpose of this paper is to model stochastically in continuous time the behavior of wind farm power generation and the power demand due to residential consumption by means of a model that accounts for the random and self-sustained over time dynamics.

The novelty of the models proposed is the development of a multidimensional correlated model for a wind farm representation. The interesting thing in the equations obtained is the presence of a correlation matrix, which can provide potential

Table 1
Statistical methods to model and forecast wind speed and wind power generation.

Category	Subclass	Example	Advantage	Study focus
Conventional statistics	Recursive filter	Kalman filters	• Ability to provide the quality of the estimate • Relatively low complexity	• Used to predict the future wind speed. It is suitable for online forecasting of wind speed • Prediction to improve the planning and usage of power sources
	Stochastic process	Discrete time continuous state (Time series analysis: ARMA, ARIMA, ARIMA) Discrete time discrete space (Markov chains)	• Well established methodology • Well implemented estimation and validation techniques • Flexibility in assignment variables	• Long, medium and short term studies, where predicting future values are needed • Influence analysis of factors over wind power generation
		Continuous time continuous space (Brownian motion, O–U process, etc.)	• More general description for phenomena under study • Wider range of applicability of possible phenomena modeling	• Continuous time analysis, ideal for time depending studies, as analytic stability studies • Power wind impact studies
Artificial intelligence and new methods	Artificial neural networks	Feed-foreward Elman Radial Basis Function Multilayer perceptron	• More general than linear methods, which gives a major flexibility at the moment to fit a data series	• Accuracy improvement forecasting wind speed and wind power generation, applied to planning and control studies
	Fuzzy Logic Wavelets Entropy based training Spatial correlation		• Less condition about the data should be assumed, which is better in situations when the truly distribution is unknown or cannot be approximated easily	• Applied to non stationary wind speed prediction in wind power systems • Used where a system is difficult to model exactly
Hybrid methods	Mixtures of any method mentioned above	Adaptation neuro fuzzy inference system (ANFIS)	• Advanced ones and less error than others • Improvement of pure methods	• Depending on the hybrid mixture. In general used to improve the accuracy of forecasting methods • Short and medium term wind speed and wind power prediction

Table 2
Statistical methods to model and forecast load uncertainties.

Category	Subclass	Example	Advantage	Study focus
Conventional statistics	Regression analysis	Multilinear regression Harmonic Regression Conditional Demand Analysis (CDA) Analysis of Variance (ANOVA)	• Easy implementation and interpretation. Well established estimation techniques • Intuitive description of seasonal changes • Scientific acceptance and a wide range of applicability	• Long term studies of demand behavior, with great time intervals, varying to hour to years • Influence analysis of factors over electric consumption behavior
	Stochastic process	Discrete time continuous state (Time series analysis: ARMA, ARIMA, SARIMA, etc.) Discrete time discrete space (Markov chains)	• Well established methodology • Well implemented estimation and validation techniques • Flexibility in assignment variables	• Long and short term studies, where predict future values are needed
		Continuous time continuous space (Brownian motion, O–U process, etc.)	• More general description for phenomena under study • Wider range of applicability of possible phenomena modeling	• Continuous time analysis, ideal for time depending studies, like transient changes
Artificial intelligence and new methods	Artificial neural networks	Feed-foreward Elman NN Radial Basis Function	• More general than linear methods, which gives a major flexibility at the moment of fitting a data series	• Accuracy forecasting power demand, applied to planning and control studies
	Fuzzy logic		• Less condition about the data should be assumed. Which is better in situations when the truly distribution is unknown or cannot be approximated easily	
	Wavelets			
Hybrid methods	Mixtures of any method mentioned above	ANFIS	• Improvement of pure methods	• Depending on the hybrid mixture. In general used to improve the accuracy of forecasting methods

information about the effect of a windmill over another and allowing future studies on wind farm geometry.

Summarizing, the novelty of the present paper is:

- Multi-correlated model that represents windmills interactions.
- Validation technique for continuous unidimensional stochastic models.
- Short-time multidimensional wind generation forecasting model (hours).
- Medium-time one dimensional residential power demand forecasting hybrid model, based on harmonic regression and O–U process

The rest of the paper is structured as follows: Section 2 establishes the necessary knowledge in statistics and mathematics to propound models applicable to real cases. Section 3 presents the characterization of the perturbations that affect the operation of the EPS and the equations of the models are presented. Section 4 applies the proposed models to Chilean cases with real data. Section 5 provides the results obtained for the analysed cases, and finally Section 6 includes the conclusions and future work.

2. Mathematical foundations

This section presents the foundations of the processes and stochastic differential equations needed for the model, and the characterization of the random behavior of a wind farm and typical feeders in distribution networks. It also presents the methodology used for the parametric estimation of the models from the real data obtained from field measurements, and the numerical method for solving the stochastic differential equations.

2.1. Stochastic processes and stochastic differential equations

Definition 1. A one-dimensional stochastic process is a parametrized collection of random variables $\{X_t\}_{t\in T}$ defined in a probability space $(\Omega, \mathcal{F}, P)$, that assumes values in $\mathfrak{R}$. The parametric space $T \subset \mathfrak{R}$ is usually interpreted as time, see [34].

Definition 2. A Gaussian process is a stochastic process $\{X_t\}_{t\in T}$ such that every linear combination of the form $Z = \sum_{i=1}^{n} \alpha_i X_{t_i}$, with $n \geqslant 1, \alpha_i \in \mathfrak{R}, t_i \in T$, is a normal random variable.

Definition 3. A one-dimensional Brownian motion is a Gaussian process $\{W_t\}_{t\in T}$ that satisfies:

(i) $E(W_t) = 0, \forall t \in T$
(ii) $Cov(W_t, W_s) = min\{t, s\}$

Definition 4. A one-dimensional Ito process $\{X_t\}_{t\in T}$ is one that can be expressed in the form

$$X_t = X_0 + \int_0^t a(u, X_u)du + \int_0^t b(u, X_u)dW_u$$

where $a(u, X_u)$ and $b(u, X_u)$ are $\mathcal{F}_t$-adapted processes and $\int_0^t b(u, X_u)dW_u$ is Ito's integral (see [34]). This equation can be expressed as a stochastic differential equation:

$$dX_t = a(t, X_t)dt + b(t, X_t)dW_t \tag{1}$$

where W_t is a Brownian motion.Furthermore, if $\{X_t\}_{t\in T}$ is a continuous process, then it is said to be a diffusion process.

Theorem 1 (*one-dimensional Ito's formula*). *Let $\{X_t\}_{t\in T}$ be a diffusion process with (1) as an SDE. Let $f : [0, \infty[\times \mathfrak{R} \to \mathfrak{R}$ be a one time continuously differentiable function in t and twice continuously differentiable in x. Let $Y_t = f(t, X_t)$, then Y_t is a diffusion process that satisfies the following SDE:*

$$dY_t = \left(\frac{\partial f(t, X_t)}{\partial t} + a(t, X_t)\frac{\partial f(t, X_t)}{\partial x} + \frac{1}{2}b^2(t, X_t)\frac{\partial^2 f(t, X_t)}{\partial x^2}\right)dt + b(t, X_t)\frac{\partial f(t, X_t)}{\partial x}dW_t \tag{2}$$

Definition 5. The quadratic variation of a stochastic process $\{X_t\}_{t\in T}$, written as $[X]_t$, is defined as $[X]_t = \lim_{\|P\|\to 0}\sum_{k=1}^{n}(X_{t_k} - X_{t_{k-1}})^2$, where P corresponds to partitions over the interval $[0,t]$. This limit, if it exists, is defined using convergence in probability.

More generally, the cross variation of two processes $\{X_t\}_{t\in T}$ and $\{Y_t\}_{t\in T}$ is defined by $[X,Y]_t = \lim_{\|P\|\to 0}\sum_{k=1}^{n}(X_{t_k} - X_{t_{k-1}})(Y_{t_k} - Y_{t_{k-1}})$.

Definition 6. An m-dimensional standard Brownian motion $\{\mathbf{W}\}_{t\in T}$, is that with components $W_t^1, W_t^2, \ldots, W_t^m$, which are independent one-dimensional Brownian motions. The increments of the Brownian motion $W_t^j - W_s^j$ for $j \in \{1,2,\ldots,m\}, t \geqslant 0$ and $s \leqslant t$, are then independent random Gaussian variables with mean zero and variance equal to $t-s$. Therefore, $\mathbf{W}_t - \mathbf{W}_s \sim N_d(\mathbf{0}, (t-s)\mathbf{I})$, where $\mathbf{0}$ denotes the null vector and $\mathbf{I}$ the identity matrix [35].

Theorem 2 (*Multidimensional Ito's formula*). *Given the m-dimensional Brownian motion* $\{\mathbf{W}\}_{t\in T}$*, a d-dimensional drift coefficient vector function* $\mathbf{a} : [0,T] \times \Re^d \to \Re^d$ *and a* $d \times m$*-matrix diffusion coefficient function* $\mathbf{b} : [0:T] \times \Re^d \to \Re^{d\times m}$*, to form the following d-dimensional stochastic differential equation:*

$$d\mathbf{X}_t = \mathbf{a}(t,\mathbf{X}_t)dt + \mathbf{b}(t,\mathbf{X}_t)d\mathbf{W}_t \tag{3}$$

where for each component we have:

$$dX_t^i = a^i(t,X_t)dt + \sum_{j=1}^{m} b^{i,j}(t,X_t)dW_t^j \tag{4}$$

For sufficiently smooth vectorial functions $\mathbf{U} : [0,T] \times \Re^d \to \Re^k$ *of solution* $\mathbf{X}_t$ *of Eq. (3), we get a k-dimensional process* $\mathbf{Y}_t = \mathbf{U}(t,\mathbf{X}_t)$.

The expression for its pth-component, which results from applying Ito's formula to each component, satisfies the following differential equation:

$$dY_t^p = \left(\frac{\partial U^p}{\partial t} + \sum_{i=1}^{d} a^i \frac{\partial U^p}{\partial x_i} + \frac{1}{2}\sum_{i,j=1}^{d}\sum_{l=1}^{m} b^{i,l}b^{j,l}\frac{\partial^2 U^p}{\partial x_i \partial x_j}\right)dt + \sum_{l=1}^{m}\sum_{i=1}^{d} b^{i,l}\frac{\partial U^p}{\partial x_i}dW_t^l \tag{5}$$

for $p \in \{1,2,\ldots,k\}$*, where the terms on the right in (5) are evaluated in* $(t,\mathbf{X}_t), U^p, a^i$ *and* $b^{i,j}$ *are the corresponding coordinate functions of the vectorial functions* $\mathbf{U}, \mathbf{a}$ *and* $\mathbf{b}$ *[36].*

Definition 7. A correlated m-dimensional Brownian motion $\{\widetilde{\mathbf{W}}\}_{t\in T}$ is such that its components $\widetilde{\mathbf{W}}_t^1, \widetilde{\mathbf{W}}_t^2, \ldots, \widetilde{\mathbf{W}}_t^m$ are transformed scalar Brownian motions. In vectorial notation, the m-dimensional transformed Brownian motion can be expressed by the linear transformation:

$$\widetilde{\mathbf{W}}_t = \mathbf{a}t + \mathbf{B}\mathbf{W}_t \tag{6}$$

where $\mathbf{a} = (a_1, a_2, \ldots, a_d)^\top$ is a d-dimensional vector, $\mathbf{B}$ is a $d \times m$-matrix and $\mathbf{W} = \{\mathbf{W}_t, t \geqslant 0\}$ is a standard m-dimensional Brownian motion.

Applying Ito's formula we get the following for each component $d\widetilde{W}_t^k = a^k dt + \sum_{i=1}^{m} b^{k,i} dW_t^i$, for $k \in \{1,2,\ldots,d\}$.

In this case $\widetilde{\mathbf{W}}_t - \widetilde{\mathbf{W}}_s \sim N(\mathbf{0}, \Sigma)$, where $\Sigma = \mathbf{B}\mathbf{B}^\top$ is the correlation matrix [35].

Among the diffusion processes there is the Ornstein–Uhlenbeck process, widely used for the description of physical phenomena, which satisfies the following stochastic differential equation:

$$dU_t = \eta U_t dt + \nu dW_t \tag{7}$$

where $\eta < 0$ and $\nu > 0$.

This process is Gaussian and it is shown that its expectation is $U_0 e^{\eta t}$ and its variance is $\frac{\nu^2(e^{2\eta t}-1)}{2\eta}$. It is also verified that the variance is bounded and admits a stationary probability distribution. In other words for a given t_0, and for $\forall t > t_0$, we have that each random variable complies with $U_t \sim N(0, -\frac{\nu^2}{2\eta})$.

In the present work it will be the fundamental component of the models to be proposed, and its mathematical properties for the parameter fitting and later validation of the models will be used.

2.2. Parametric estimation

In the Ornstein–Uhlenbeck process (7), a technique for estimating parameter η is the use of quasi maximum likelihood [37], whose estimator turns out to be:

$$\tilde{\eta} = \frac{1}{\Delta}\log\frac{\sum_{i=1}^{n} U_{(i-1)\Delta}U_{i\Delta}}{\sum_{i=1}^{n} U_{(i-1)\Delta}^2} \tag{8}$$

where the random variable $U_{i\Delta}$ corresponds to an observation of the O–U process at instant $i\Delta$. For this fitting these observations will correspond to the actual data provided by the farm, under a certain transformation that will be indicated later; n is the number of available observations, and Δ is the time interval between observations.

Parameter ν, of Eq. (7) can be estimated by equating the quadratic variation of the O–U process with its discretization, from which the following estimator is obtained:

$$\tilde{\nu} = \sqrt{\frac{1}{t}\sum_{i=1}^{n}(U_{i\Delta} - U_{(i-1)\Delta})^2} \tag{9}$$

where t is the total time in which the process was observed, in the time scale considered for η.

For the estimation of possible correlations between two processes the cross variation will be used, in the same way as with the quadratic variation. In this case, for two processes $\{X_t\}_{t\in T}$ and $\{Y_t\}_{t\in T}$ that satisfy Eq. (1), with diffusion coefficients $b_x(t,X_t)$ and $b_y(t,Y_t)$, respectively. Its cross variation would be $\int_0^t b_x(s,X_s)b_y(s,Y_s)\rho ds$, where ρ is the correlation coefficient.

In the particular case of two O–U process, that satisfy (7), the estimator of its correlation coefficient, obtained by means of equating the theoretical cross variation with its discretization, is given by Eq. (10).

$$\tilde{\rho}_n = \frac{1}{\nu_1\nu_2 t}\sum_{k=1}^{n}\left(U_{t_k}^1 - U_{t_{k-1}}^1\right)\left(U_{t_k}^2 - U_{t_{k-1}}^2\right) \tag{10}$$

2.3. Numerical resolution schemes

The numerical schemes will be useful at the time of analysing the behavior of the stochastic process that is being studied and then simulating the trajectories for later applications.

For a stochastic differential equation of form (1), the Milstein numerical scheme would be:

$$X_{i+1} = X_i + a(t_i,X_i)(t_{i+1}-t_i) + b(t_i,X_i) + \frac{1}{2}b(t_i,X_i)b\prime(t_i,X_i) \times ((W_{i+1}-W_i)^2 - (t_{i+1}-t_i)) \tag{11}$$

In the multidimensional case, the kth-component of the Milstein scheme, for a process that satisfies (3) with diagonal matrix $\mathbf{b}$, would be:

$$X_{n+1}^k = X_n^1 + a^k\Delta + b^{k,k}\Delta W^k + \frac{1}{2}b^{k,k}\frac{\partial b^{k,k}}{\partial x^k}\{(\Delta W^k)^2 - \Delta\} \tag{12}$$

For further details of the above mentioned numerical schemes, see [36].

3. Perturbation models

This section presents the assumptions made for modeling the two random phenomena considered in the paper: wind generation and electric power demand. The procedure to obtain the stochastic equations that represent the uncertainties is detailed.

3.1. Characterization of the perturbations

3.1.1. Wind generation

Wind generation is subjected to various factors, among the most important of which there is the wind speed and the installed capacity of the wind farm. In general, the following aspects are distinguished:

- In short time intervals, as hours, the irregularity of the wind speed results in a generation that lacks a definite trend and seasonality. However, seasonalities can exist in longer time intervals, as months or days [38].
- It should be noted that the trajectories of the measurements of wind generation are bounded in the $[0, P]$ (where P corresponds to the total installed capacity of the farm). Moreover, the wind generation samples do not change discontinuously.
- Wind speed does not have symmetrical distribution Therefore, the representation of the power delivered by the wind farm should have similar characteristics in its distribution. Weibull and Rayleigh distributions are most commonly used in wind speed data analysis [39]. On the other hand, logarithmic normal distribution is used to characterize wind power generation, due to the possibility to incorporate the Brownian motion in the dynamic representation [40].
- The square of the mean wind generation cannot increase during many consecutive hours, and this is associated with the property of reversion to the mean. This implies that the trajectories of the process will be around the mean value.

3.1.2. Electric power demand

In general, the behavior of a load follows certain seasonal patterns associated with changes in the seasons of the year, day and night, weekends, etc. It is also subject of a certain trend due to the population growth and urbanization.

3.2. Wind generation model

Two stochastic models are developed that allow the dynamic representation of the power generation of the complete farm. The first corresponds to a one-dimensional representation of the total power output and the second to a multidimensional representation where each component symbolizes a windmill in the wind farm.

3.2.1. One-dimensional model

If it is desired to directly consider process (7) as a representation of the wind generation, it would have to be assumed that the wind generation follows a normal distribution, but as already stated, it is often considered that the wind power generation follows a normal-logarithmic distribution, suggesting that the process that describes the wind generation, which will be denoted by $\{Y_t\}_{t\in T}$, complies with:

$$\ln Y_t - h = U_t \tag{13}$$

$$\Rightarrow Y_t = e^{U_t+h} \tag{14}$$

where h is the mean of $\ln Y_t$. This ensures that $\ln Y_t - h$ has a zero mean and a normal distribution. So (14) is the candidate model and with transformation (13) one can work directly with the O–U process.

According to Ito's formula (3), Eq. (14) satisfies the following SDE:

$$dY_t = \left(\eta \ln Y_t - \eta h + \frac{\nu^2}{2}\right) Y_t dt + \nu Y_t dW(t) \tag{15}$$

3.2.2. Multidimensional model

As an experiment, the wind farm was modeled in a multidimensional manner, taking each component as a windmill. The objective was to determine if, at the time of describing the farm's power output, it is more appropriate to take into account the behavior of each windmill and its possible interaction, or if the overall view considered in the one-dimensional model is sufficient.

The same as in the case of the total output power, it will be proposed that each wind generator can be described by a process of the form (14), so that with transformation (13), each windmill can be fitted by an O–U process. Writing this in matrix form, following model is obtained:

$$d\mathbf{U}_t = \mathbf{A}\mathbf{U}_t dt + \mathbf{C} d\mathbf{W}_t \tag{16}$$

where vector $\mathbf{U}_t = (U_t^1, U_t^2, \ldots, U_t^n)$ has a one-dimensional O–U process in each component. The $d\mathbf{U}_t$ notation means that each component of the vector is differentiated, i.e., $d\mathbf{U}_t = (dU_t^1, dU_t^2, \ldots, dU_t^n)$. Matrices $\mathbf{A}$ and $\mathbf{C}$ are diagonal with components a_{ij} and c_{ij} respectively, given by:

$$a_{ij} = \begin{cases} \eta_i & i = j \\ 0 & i \neq j \end{cases} c_{ij} = \begin{cases} \nu_i & i = j \\ 0 & i \neq j \end{cases} \tag{17}$$

where vector $\mathbf{W}_t$ is an n-dimensional Brownian motion. Two cases will be considered in this stage: correlation and no correlation between the elements of the vectorial Brownian.

The case with no correlation is simply aimed at fitting an O–U to each transformed windmill, the same as in the one-dimensional case, i.e., every wind generator, transformed according to (13), is associated with the process $dU_t^i = \eta_i U_t^i dt + \nu_i dW_t^i$, where $i = \{1, 2, \ldots, n\}$. The W_t^i are independent from one another.

For the correlated case it will be assumed that the vectorial Brownian in (16) is correlated. This Brownian can be expressed as the linear transformation $\mathbf{W}_t = \mathbf{B}\widetilde{\mathbf{W}}_t$, where $\widetilde{\mathbf{W}}_t$ is an independent n-dimensional vectorial Brownian and $\mathbf{B}$ is the $n \times n$ square matrix, which is built in such a way that at the time of calculating the cross variation of the processes of each component the cross variation coefficient is obtained by some extra parameter of matrix $\mathbf{C}$.

In the case of the correlated model it is necessary to build and then estimate the correlation matrix $\Sigma = \mathbf{B}\mathbf{B}^t$. To show how to determine matrix $\mathbf{B}$, consider the three-dimensional case. Assume the following process:

$$\begin{pmatrix} dU_t^1 \\ dU_t^2 \\ dU_t^3 \end{pmatrix} = \begin{pmatrix} \eta_1 & 0 & 0 \\ 0 & \eta_2 & 0 \\ 0 & 0 & \eta_3 \end{pmatrix} \begin{pmatrix} U_t^1 \\ U_t^2 \\ U_t^3 \end{pmatrix} dt + \begin{pmatrix} \nu_1 & 0 & 0 \\ 0 & \nu_2 & 0 \\ 0 & 0 & \nu_3 \end{pmatrix} \begin{pmatrix} dW_t^1 \\ dW_t^2 \\ dW_t^3 \end{pmatrix} \tag{18}$$

where vector $(dW_t^1, dW_t^2, dW_t^3)^\top$ corresponds to a correlated three-dimensional Brownian motion that can be expressed as:

$$\begin{pmatrix} dW_t^1 \\ dW_t^2 \\ dW_t^3 \end{pmatrix} = \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{pmatrix} \begin{pmatrix} d\widetilde{W}_t^1 \\ d\widetilde{W}_t^2 \\ d\widetilde{W}_t^3 \end{pmatrix} \tag{19}$$

where vector $(d\widetilde{W}_t^1, d\widetilde{W}_t^2, d\widetilde{W}_t^3)^\top)$ is a standard three-dimensional Brownian motion.

Each component is written in the form:

$$dU_t^1 = \eta_1 U_t^1 dt + \nu_1(a_{11} d\widetilde{W}_t^1 + a_{12} d\widetilde{W}_t^2 + a_{13} d\widetilde{W}_t^3)$$
$$dU_t^2 = \eta_2 U_t^2 dt + \nu_2(a_{21} d\widetilde{W}_t^1 + a_{22} d\widetilde{W}_t^2 + a_{23} d\widetilde{W}_t^3)$$
$$dU_t^3 = \eta_3 U_t^3 dt + \nu_3(a_{31} d\widetilde{W}_t^1 + a_{32} d\widetilde{W}_t^2 + a_{33} d\widetilde{W}_t^3)$$

It is expected that when the quadratic and cross variations are calculated we will have that: $\left[U^1\right]_t = \nu_1^2 t$, $\left[U^2\right]_t = \nu_2^2 t$, $\left[U^3\right]_t = \nu_3^2 t$, $\left[U^1, U^2\right]_t = k_{1,2}\rho_{1,2} t$, $\left[U^1, U^3\right]_t = k_{1,3}\rho_{1,3} t$, $\left[U^2, U^3\right]_t = k_{2,3}\rho_{2,3} t$, where $k_{i,j}$ is some constant, and $\rho_{i,j}$ is a correlation parameter to be estimated.

To begin, it will be assumed that $b_{11} = 1, b_{12} = b_{13} = 0$, so the first coordinate will be: $dU_t^1 = \eta_1 U_t^1 dt + \nu_1 d\widetilde{W}_t^1$, and its quadratic variation is $\nu_1^2 t$.

In the second component it will be assumed that $b_{21} = \rho_{1,2}, b_{22} = \sqrt{1-\rho_{1,2}}$ and $b_{23} = 0$, so dU_t^2 will become: $dU_t^2 = \eta_2 U_t^2 dt + \nu_2(\rho_{1,2} d\widetilde{W}_t^1 + \sqrt{1-\rho_{1,2}^2} d\widetilde{W}_t^2)$, and its quadratic variation is $[U^2]_t = (\nu_2^2\rho_{1,2}^2 + \nu_2^2(1-\rho_{1,2}^2))t = \nu_2^2 t$.The cross variation between U_t^1 and U_t^2 becomes: $[U^1, U^2]_t = \nu_1\nu_2\rho_{1,2} t$.

To determine the coefficients b_{31}, b_{32} and b_{33}we proceed as follows: It must be fulfilled that $[U^1, U^3]_t = k_{1,3}\rho_{1,3} t$, then $\nu_1\nu_3 b_{31} t = k_{1,3}\rho_{1,3} t$, so $k_{1,3} = \nu_1\nu_3$ and $b_{31} = \rho_{1,3}$. We must also have that: $[U^2, U^3]_t = k_{2,3}\rho_{2,3} t$, so $(\nu_2\nu_3\rho_{1,2}\rho_{1,3} + \nu_2\nu_3 b_{32}\sqrt{1-\rho_{1,2}^2})t$ $= k_{2,3}\rho_{1,3} t$, and therefore $k_{2,3} = \nu_2\nu_3$ and $b_{32} = \frac{(\rho_{2,3}-\rho_{1,2}\rho_{1,3})}{\sqrt{1-\rho_{1,2}^2}}$.

Finally, for factor b_{33}, we see the quadratic variation of U_t^3, where: $[U^3]_t = \nu_3^2 t$ must be fulfilled, therefore, $\left(\nu_3^2\left(\rho_{1,3}^2 + \frac{(\rho_{2,3}-\rho_{1,2}\rho_{1,3})^2}{1-\rho_{1,2}}\right.\right.$ $\left.\left.+ b_{33}^2\right)\right)t = \nu_3^2 t$, so $b_{33} = \sqrt{1-\rho_{1,3}^2 - \frac{(\rho_{2,3}-\rho_{1,2}\rho_{1,3})^2}{1-\rho_{1,2}}}$. From this we get the transformation matrix $\mathbf{B}$, given by:

$$\mathbf{B} = \begin{pmatrix} 1 & 0 & 0 \\ \rho_{1,2} & \sqrt{1-\rho_{1,2}^2} & 0 \\ \rho_{2,3} & \frac{(\rho_{2,3}-\rho_{1,2}\rho_{1,3})}{\sqrt{1-\rho_{1,2}^2}} & \sqrt{1-\rho_{1,3}^2 - \frac{(\rho_{2,3}-\rho_{1,2}\rho_{1,3})^2}{1-\rho_{1,2}^2}} \end{pmatrix}$$

Under this construction we will have that matrix $\mathbf{B}$ will be a triangular matrix where coefficients $b_{i,1}$ will correspond to the correlation between the first component and component i. The rest of the factors are calculated by equalizing the values that are expected in the cross and quadratic variations.

3.3. Electric power demand model

A general characterization of this type of behavior will be represented with the following one-dimensional process:

$$D_t = \alpha + \beta t + \sum_{i=1}^{n}(\gamma_i \cos(2\pi t f_i) + \delta_i \sin(2\pi t f_i)) + U_t \tag{20}$$

where D_t is the power demand, U_t is an O–U process, f_i are the frequencies associated with the seasonal changes, βt is associated with the electric consumption increase over time, and the rest of the coefficients are parameters that must be estimated.

Using Ito's formula (3), process (20) satisfies the following stochastic differential equation:

$$dD_t = (\beta + \sum_{i=1}^{n}((-2\pi f_i\gamma_i - \eta\delta_i)\sin(2\pi t f_i) + (2\pi f_i\delta_i - \eta\gamma_i)\cos(2\pi t f_i)) + \eta(D_t - \alpha - \beta t))dt + \nu dW_t \tag{21}$$

4. Application case

The models presented in the section above will be implemented in a real Chilean case, using real information data for parameter estimation.

4.1. Wind power generation

Observations measured in field from eleven windmills of Canela wind farm, located in the Fourth Region of Chile was used.

The treatment to which the data were subjected is presented below, to get the parameters of the stochastic processes using the data from the farm.

4.1.1. Presentation of data

The data delivered from the wind farm corresponds to measurements of active power in kW from 11 windmills, made every second during three hours. For this time interval under study, seasonal changes are not appreciable, so they will be not considered.

The set of observations will be denoted by $\{y_{i\Delta}^n\}_{i=0,\dots,10800}$, where $\Delta = 1$ s, 1 s represents the time between observations, n is the identifier for each mill, and i is the number of samples considered. The total sum of the power generation of the mill every second will be denoted by $\{\tilde{y}_{i\Delta}\}_{i=0,\dots,10800}$.

Fig. 1 shows the trajectory generated by the data of the total power output generated by the farm. It is seen that the power delivered by the farm varies in time, with bounded trajectories and non-deterministic characteristics.

Some statistical characteristics will be observed in order to justify the distribution assumptions and the use of the proposed models in the previous section, as an application with these real data. In this case, and for practical matters, only the characteristic of global wind generation data $\{\tilde{y}_{i\Delta}\}$ are presented. Moreover, the statistical information of each windmill separately shows similar attributes.

Table 3 shows four statistical measures of the observation sample, both the real and the transformed by (8). It is important to analyse the transformation data characteristics, because these will be used for the model calibration. Positive skewness of the raw data indicates that the tail on the right side of the probability density function is longer than the left side, as in the case of log-normal distribution. The skewness of transformed data is close to zero, as it could be expected for a normal distribution.

Kurtosis measure has a value proximal to three, in both cases (raw and transformed). This gives another beneficial point to Gaussian distribution consideration for the data that will be adjusted and modeled.

In relation to the variance, its value decreases with the logarithmic transformation, which in somehow improves the data quality.

On the other hand, Figs. 2(a) and (b) shows respectively the autocorrelation function (ACF) of the real and transformed data. The exponential decay of the functions exhibit a similar behavior of process that correlated the future state with the present, as O–U process [41]. This can suggest that the application of the mentioned process to represent the data will be appropriate.

Summarizing:

- Data need to be transformed to reduce its variance.
- Transformed data shows indicators with Gaussian characteristics.

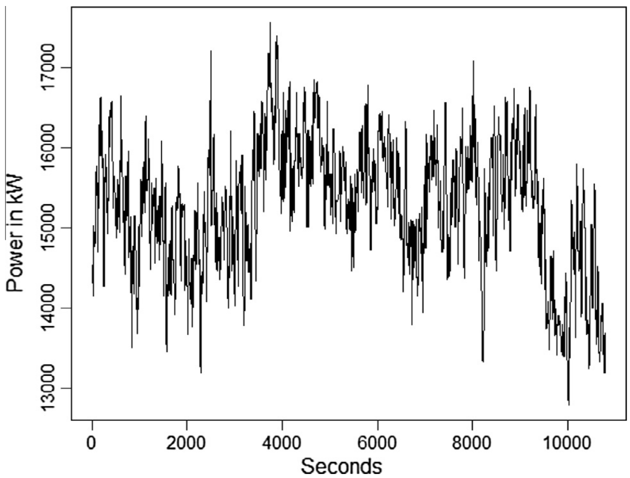


Fig. 1. Active power generation in kW.

Table 3
Statistical information about total wind generation data.

Measure	Raw data	Transformed data
Mean	15263.51	−5.857338e−16
Variance	558840.9	0.002373935
Skewness	0.1599429	0.03704418
Kurtosis	2.720979	2.67714

- Autocorrelation function exhibits an exponential decay, which suggest a process with correlation between the future state and the past state, as O–U process.

4.1.2. Parameter fitting of one-dimensional model

The parameters involved in Eq. (15) will be obtained under the following considerations:

As the estimators respond well under small pass sizes, a time scale greater than that of seconds (actual sampling time) will be considered, i.e., instead of considering that a time interval of 10,800 s was observed with a Δ of 1 s, it will be said that a time interval of 3 h was observed with $\Delta = 3/10800$. According to this:

- η is estimated from Eq. (8) considering as observations of the process the $\{\tilde{y}_{i\Delta}\}_{i=1,\dots 10800}$ data transformed according to (13).
- v is obtained from Eq. (9), with $t = 3$ h. and considering the same transformation for η.

Table 4
Parameters estimated from the sum of active power observations of the 11 generators.

Parameter	Estimated value
η	−8.707802
v	0.2309347
h	9.633223

Table 5
η_i, v_i and h_i parameters of the 11 windmills.

Parameter	Value	Parameter	Value	Parameter	Value
η_1	−10.8191	v_1	0.8225	h_1	7.2231
η_2	−12.9155	v_2	0.7515	h_2	7.2770
η_3	−11.3623	v_3	0.7466	h_3	7.2640
η_4	−8.6407	v_4	0.5866	h_4	7.3235
η_5	−12.4291	v_5	0.8842	h_5	7.2402
η_6	−10.2790	v_6	0.9005	h_6	7.2014
η_7	−13.3266	v_7	0.9541	h_7	7.1574
η_8	−12.0976	v_8	1.0594	h_8	7.1044
η_9	−9.3986	v_9	0.9136	h_9	7.1942
η_{10}	−9.0323	v_{10}	0.9626	h_{10}	7.1972
η_{11}	−13.2814	v_{11}	0.9852	h_{10}	7.1972

Table 6
Estimation of correlation parameters between windmills.

Parameter	Value	Parameter	Value	Parameter	Value
$\rho_{1,2}$	0.0356	$\rho_{3,4}$	−0.0758	$\rho_{5,9}$	0.0309
$\rho_{1,3}$	0.1028	$\rho_{3,5}$	−0.0400	$\rho_{5,10}$	0.0382
$\rho_{1,4}$	−0.0148	$\rho_{3,6}$	−0.0304	$\rho_{5,11}$	0.0160
$\rho_{1,5}$	0.0212	$\rho_{3,7}$	0.0499	$\rho_{6,7}$	−0.0398
$\rho_{1,6}$	0.0203	$\rho_{3,8}$	0.0802	$\rho_{6,8}$	0.0319
$\rho_{1,7}$	0.0158	$\rho_{3,9}$	−0.0581	$\rho_{6,9}$	0.0042
$\rho_{1,8}$	0.0171	$\rho_{3,10}$	0.0541	$\rho_{6,10}$	−0.0155
$\rho_{1,9}$	0.0532	$\rho_{3,11}$	0.0251	$\rho_{6,11}$	0.0115
$\rho_{1,10}$	0.0184	$\rho_{4,5}$	−0.0188	$\rho_{7,8}$	−0.0844
$\rho_{1,11}$	−0.0196	$\rho_{4,6}$	0.0086	$\rho_{7,9}$	−0.0325
$\rho_{2,3}$	−0.0096	$\rho_{4,7}$	−0.0062	$\rho_{7,10}$	−0.0053
$\rho_{2,4}$	0.0016	$\rho_{4,8}$	0.0227	$\rho_{7,11}$	0.0158
$\rho_{2,5}$	−0.0208	$\rho_{4,9}$	0.0196	$\rho_{8,9}$	−0.0181
$\rho_{2,6}$	−0.0162	$\rho_{4,10}$	−0.0669	$\rho_{8,10}$	−0.0081
$\rho_{2,7}$	0.0124	$\rho_{4,11}$	−0.0228	$\rho_{8,11}$	0.0209
$\rho_{2,8}$	0.0222	$\rho_{5,6}$	−0.0821	$\rho_{9,10}$	−0.1052
$\rho_{2,9}$	−0.0901	$\rho_{5,7}$	−0.0015	$\rho_{9,11}$	0.0523
$\rho_{2,10}$	0.0574	$\rho_{5,8}$	0.0023	$\rho_{10,11}$	−0.1451
$\rho_{2,11}$	−0.0237				

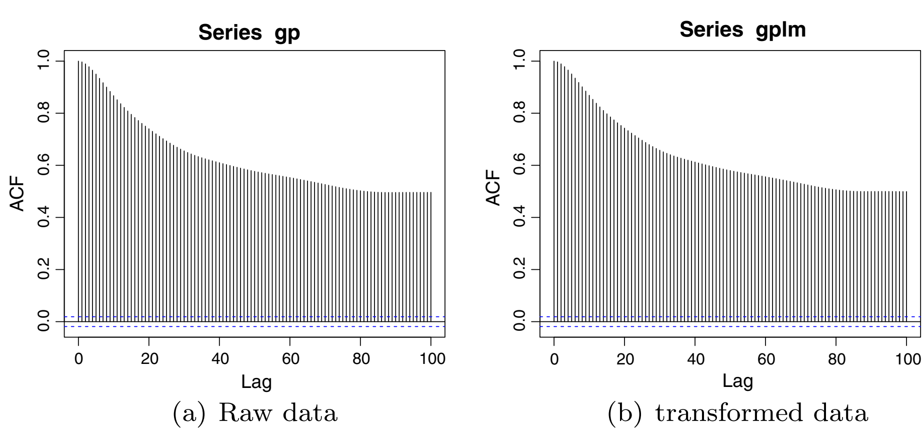


Fig. 2. Autocorrelation functions (Generation).

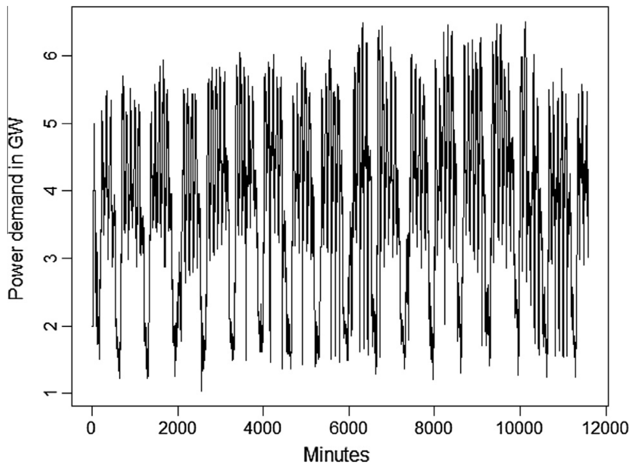


Fig. 3. Electric power demand.

Table 7
Statistical information about power demand data.

Measure	Raw data	Transformed data
Mean	3.6501	0.007082501
Variance	0.4881835	0.031888
Skewness	−0.2297685	−0.05376565
Kurtosis	1.940585	3.573246

- h is estimated using the average of the natural logarithm of the actual observations of the farm, i.e., the following estimator is used $\tilde{h} = \frac{1}{n}\sum_{i=0}^{n} \ln(Y_{i\Delta})$

Table 4 shows the results obtained from the process that represents the random component of the sum of the 11 generators of the wind farm. As expected, parameter η is negative, accounting for the bounded trajectories. As to parameter ν, the value is small, and this means that the amplitudes of the random variations do not undergo abrupt changes.

4.1.3. Parametric estimation of multidimensional model

The Eq. (16) will be considered as the representation of the transformation data by Eq. (13). Dimensions of matrices **A** and **C** in this case is $n = 11$, due to the available measure information of the 11 windmills of the farm denoted by $\{y^n_{i\Delta}\}_{i=0,\ldots,10800}$.

The estimation methods for the 22 parameters of matrices **A** and **C** are the same as those used in the one-dimensional case. The η_i are fitted by the quasi-maximum likelihood technique, and the coefficients ν_i are estimated by means of the quadratic variation of the process associated with its corresponding windmill.

Table 8
Frequencies in Hertz, associated with residential consumption.

Frequency	Value	Frequency	Value	Frequency	Value
f_1	0.00146	f_6	0.02109	f_{10}	0.009
f_2	0.00154	f_7	0.021	f_{11}	0.006
f_3	0.003	f_8	0.012	f_{12}	0.00454
f_4	0.01055	f_9	0.01046	f_{13}	0.00163
f_5	0.03155				

Table 9
Estimated harmonic regression parameters for residential consumption.

Parameter	Value	Parameter	Value
α	3.664	γ_8	−2.868e−02
β	2.610e−05	δ_8	−1.478e−01
γ_1	2.956e−01	γ_9	2.123e−01
δ_1	5.207e−01	δ_9	6.856e−02
γ_2	−6.073e−01	γ_{10}	1.602e−01
δ_2	−5.442e−01	δ_{10}	5.722e−02
γ_3	−1.400e−01	γ_{11}	7.104e−02
δ_3	6.625e−01	δ_{11}	1.295e−01
γ_4	−5.464e−01	γ_{12}	−2.202e−01
δ_4	−5.166e−01	δ_{12}	1.127e−01
γ_5	1.769e−01	γ_{13}	−1.508e−01
δ_5	−3.201e−02	δ_{13}	−1.594e−01
γ_6	−2.135e−01	γ_7	7.866e−02
δ_6	1.249e−01	δ_7	−1.752e−01

Since the data of each machine were transformed, it is necessary to estimate parameter h_i, which corresponds to the corresponding arithmetic mean of each generator.

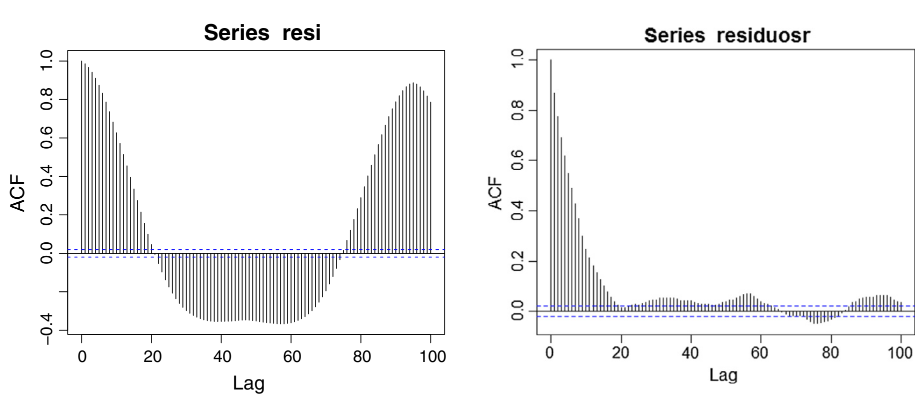
Table 5 show the results of the parametric estimation.

The same as in the one-dimensional case, it is seen that the values for parameters η_i are all negative, ensuring the stationary of the estimated process.

In the correlated case the vectorial Brownian motion involve in Eq. (16) is correlated, so it is necessary to estimate the correlation coefficients of the transformation matrix **B** to obtain the correlated matrix Σ. Using expression (10) the values of the estimated correlation coefficients between the eleven mills are obtained. The results are given in Table 6.

4.2. Electric power demand

To implement the model demand (20), a residential type consumption was considered. Loads with other characteristics, such



(a) Power demand data (b) adjusted power demand data

Fig. 4. Autocorrelation functions (Power Demand).

as industrial or commercial, can be represented in a similar way. The differences will lie in the periods and variations of the power consumption, but the general scheme is the same.

To analyse this behavior, actual observations provided by an electric power distributing company were used.

4.2.1. Presentation of data

The data provided by the company correspond to measurements every 15 min during 120 days and 20 h of electric power in GW, of a feeder in Santiago, Chile. The set of observations on the feeder will be denoted by $\{d_{i\Delta}\}_{i=0,\dots,11600}$, where $\Delta = 15$ min represents the time between observations and i is the number of samples considered.

Fig. 3 shows a trajectory drawn for the actual data.

Fig. 3 shows a pronounced seasonal behavior, as was established in Section 3.1.2. On the other hand, a growing trend is not exhibit, due to the observation window is not wide enough.

Table 7 reveals information about four statistical measurements of the power demand data. Transformed data refers to real data without the seasonal and trend changes, expressed in (20) as a sine and cosine sum, and a time increment, respectively. Both cases are important to analyse, because real data will be adjusted by harmonic regression, and then, the residues of this fit will be adjusted by an O–U process, thus it is important to check that the data satisfy the requirements to use the proposed techniques.

In the case of transformed data, skewness and kurtosis show characteristic values for a normal distribution. In relation to the variance, its value decreases with a sinusoidal adjustment. It can be concluded, as a first approach, that the transformation is adequate.

On the other hand, Figs. 4(a) and (b) show respectively the autocorrelation function (ACF) of the real and transformed data. The sinusoidal behavior of the function in Fig. 4(a) exhibits the seasonal characteristic of the data. Furthermore, the exponential decay of the function in Fig. 4(b) reveals a similar behavior of process that correlated the future state with the present, as O–U process, but a little sinusoidal pattern can be seen. The above can suggest that the application of the mentioned process to represent the residues of harmonic regression will be appropriate and some seasonal changes are not been captured by the selected frequencies.

Summarizing:

- Raw data is described by a sinusoidal function.
- Transformed data show indicators with Gaussian characteristics.
- Autocorrelation function of adjusted data exhibits an exponential decay, which suggests a process with correlation between the future state and the past state, as O–U process.

4.2.2. Parametric estimation

The frequencies present in Eq. (20) are found by analysing the periodogram of the series of data, which corresponds to an estimation of the spectral density. Table 8 shows the most relevant frequencies considered.

With these frequencies it is possible to estimate the coefficients that go together with the trigonometric functions in (20) by means of harmonic regression. The results are shown in Table 9.

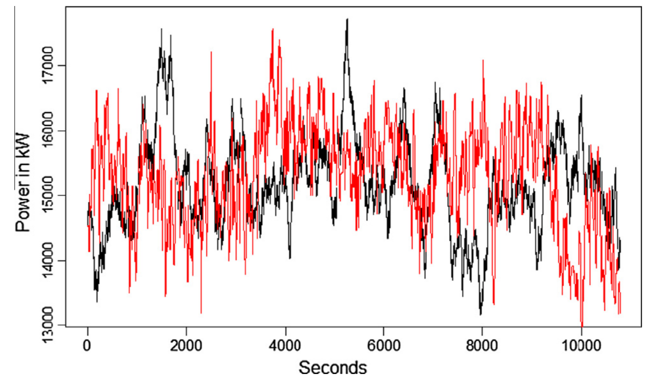


Fig. 5. Simulated trajectories of the active power output generation in kW.

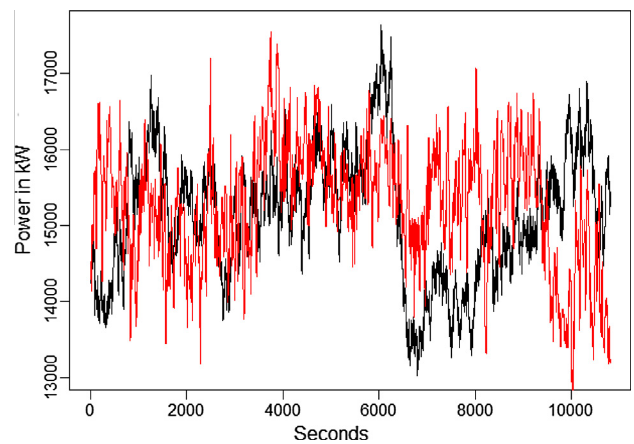


Fig. 6. Sum of simulated trajectories of the active power generation output in kW, uncorrelated.

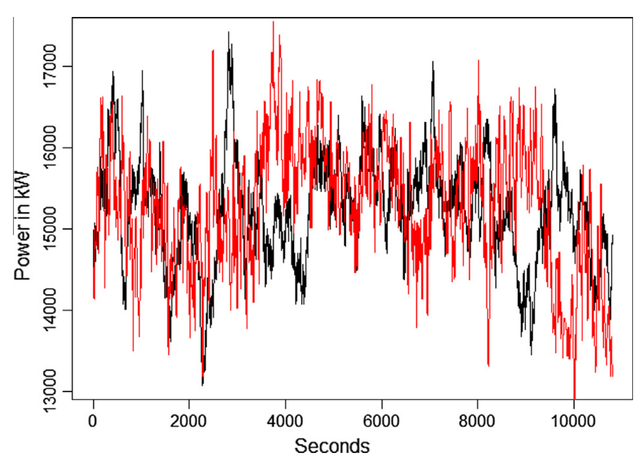


Fig. 7. Sum of simulated trajectories of the active power generation output in kW, correlated.

Table 10
Estimated parameters for the O–U process for residential consumption.

Parameter	Estimated value
η_c	−5.578228
ν_c	2.164175

To estimate the parameters associated with the O–U process, it was used the residues obtained from the fit through regression, carried out earlier. The estimation techniques and the estimators are the same that were used in the previous one-dimensional models. The results are shown in Table 10.

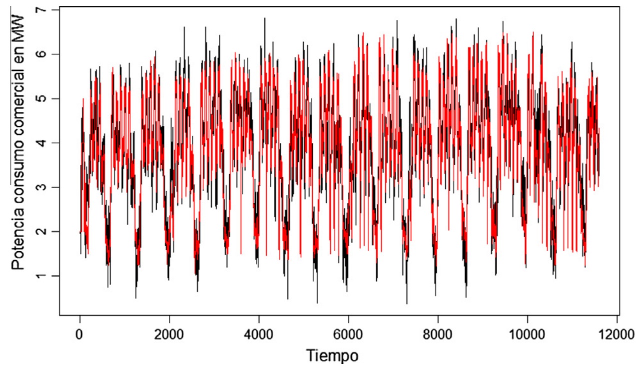


Fig. 8. Residential consumption simulation model.

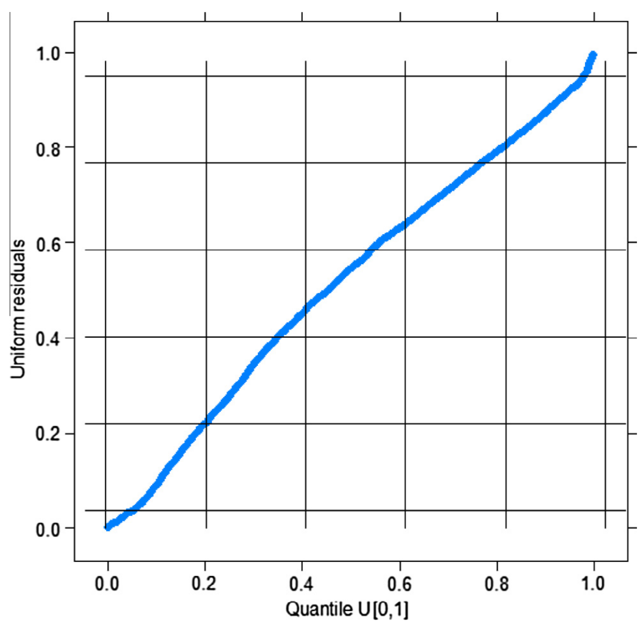


Fig. 9. Uniform residues quantile–quantile graph of the O–U model of the wind generation data.

5. Results

The models obtained from estimates of actual Chilean data will be evaluated and validated, with the aim of showing its potential usefulness for real studies.

5.1. Trajectory simulation

A first step to verify the good behavior of the obtained models is to observe them graphically. This requires using numerical schemes to simulate the solution of the equations proposed computationally. In this case Milstein scheme (12) or (13) will be used, depending on the dimension of the model. Each case is detailed below.

5.1.1. Wind generation

a. One-dimensional case

For process (15) we have the following discretization:

$$Y_{i+1} = Y_i + \left(\eta Y_i \ln(Y_i) - \eta Y_i h(t_i) + \frac{1}{2} Y_i \nu^2\right)(t_{i+1} - t_i) + \nu Y_i (W_{i+1} - W_i) + \frac{\nu^2 Y_i}{2}((W_{i+1} - W_i)^2 - (t_{i+1} - t_i)) \quad (22)$$

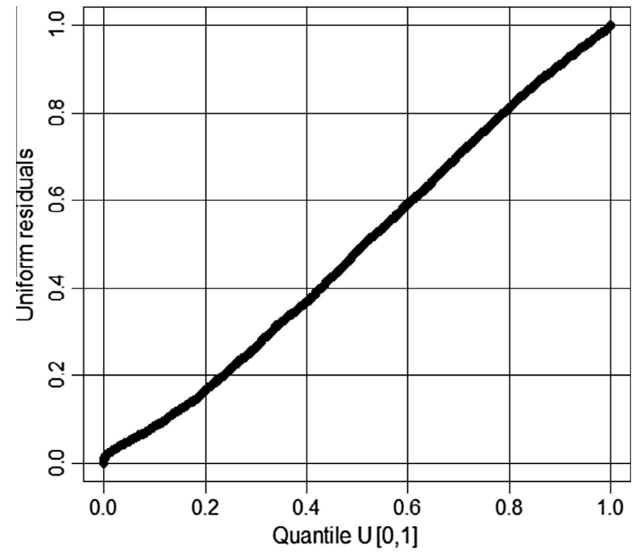


Fig. 10. Residential model residues quantile–quantile graph.

Fig. 5 shows a simulated trajectory in black, and the original data trajectory is shown in red for comparison. It shows that the simulated curve has a very similar behavior to the real one: both remain within the same rank and the shapes of their trajectories are very similar in the sense of how they change over time. This analysis shows, at first sight, a good representation of the power output of the wind farm obtained by the model.

b. Multidimensional case

In this case the discretization of each component satisfies the same scheme (22).

In the uncorrelated case, the Brownian difference random vector follows a standard multivariate normal distribution. In the correlated case this vector follows a multivariate normal distribution with null vector mean and correlation matrix $\Sigma = \mathbf{B}\mathbf{B}^{\top}$.

Figs. 6 and 7 show in black the sum of the 11 simulated components for the uncorrelated and correlated cases. The trajectory of the original data is shown in red and in black the simulated one. As in the one-dimensional case, both trajectories (uncorrelated and correlated) move within the same rank and their variations over time are similar to the real trajectory. Nevertheless, the uncorrelated simulation shows more abrupt changes than the correlated case. This could indicate that the second model is more appropriate to describe the behavior of the power generated by each windmill simultaneously

5.1.2. Electric power demand

The discretization for Eq. (21) is the following:

$$D_{i+1} = D_i + \beta + \sum_{i=1}^{n}((-2\pi f_i \gamma_i - \eta\delta_i)\sin(2\pi t_i f_i) + (2\pi f_i \delta_i - \eta\gamma_i)\cos(2\pi t_i f_i)) + \eta(D_i - \alpha - \beta t_i)(t_{i+1} - t_i) + \nu(W_{i+1} - W_i) \quad (23)$$

Fig. 8 shows a simulated trajectory in black, and, for comparison, the trajectory of the original data is shown in red. In this case, the actual path is not behaving so randomly: it has certain changes, which are captured by the sinusoidal functions, behavior that is well represented in the simulated curve. However, there are slight variations in the amplitudes: the simulated process shows higher maximum and lower minimum than the actual curve, which

Table 11
Validation statistics.

Statistic	Result
Jarque Bera Test	$(p - value = 0.2745)$ Normality
Durbin Watson Test	$(p - value = 3.266e{-}05)$ Correlation
Breusch Pagan Test	$(p - value = 0.8075)$ Homoscedastic
R^2	0.7626
R^2 adjusted	0.7371

expands a little the rank. To determine whether the model overstates the real values another tool is required, which will be applied in the next point.

5.2. Residual analysis

An analytic and formal tool to verify and validate the goodness of a fit of a model based in SDE is the uniform residues study. This method consists basically in determining whether the actual observations obtained come from a process with the chosen characteristics of the model [42].

The procedure to use this technique, in the particular case of an O–U process will be explained. To begin, the expectation and conditional variance of the process must be known, and they can be calculated explicitly for the O–U process.

For the O–U process the expectation and conditional variance are respectively $E_{\eta,\nu}(U_\Delta|U_0 = U_{t_{i-1}}) = U_{t_{i-1}}e^{\eta\Delta}$ and $V_{\eta,\nu}(U_\Delta|U_0 = U_{t_{i-1}}) = \frac{\nu^2(e^{2\eta\Delta}-1)}{2\eta}$.

From this the following random variables are calculated:

$$R_{t_{i(\eta,\nu)}} = \frac{U_{t_i} - E_{\eta,\nu}(U_\Delta|U_0 = U_{t_{i-1}})}{\sqrt{V_{\eta,\nu}(U_\Delta|U_0 = U_{t_{i-1}})}} = \frac{U_{t_i} - U_{t_{i-1}}e^{\eta\Delta}}{\nu\sqrt{\frac{e^{2\eta\Delta}-1}{2\eta}}} \quad (24)$$

Eq. (24) is known as the standardized residues. For this process, as it is a Gaussian one, the standardized residues must follow a standardized normal distribution, which implies that if they are composed by the normal standard cumulative distribution function Φ, random variables with uniform distribution $U(0,1)$ must be obtained, i.e., the variable

$$\Psi_{t_i}(\eta,\nu) = \Phi\big(R_{t_i}(\eta,\nu)\big) \quad (25)$$

where $\Phi(z) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z} e^{-t^2/2}dt$, must comply with $\Psi_{t_i}(\eta,\nu) \sim U(0,1)$.

In order to formally validate the parametric one dimensional models obtained, the uniform residues associated with parameters estimated for the O–U process were analysed. For this propose, it will be verified if the data under the transformation (13) come from an O–U.

The multidimensional case is a slightly more difficult and it will be a future work.

5.3. One-dimensional wind power model

As stated, if transformation (13) of the data can be represented by an O–U, it must be true that taking these data as observations of the random variable U_{t_i} we must get that the variables (25) follow a uniform distribution $U(0,1)$. This can be verified by means of a quantile–quantile graph that compares the quantiles of the random variable (25)with the theoretical quantiles of a variable with uniform distribution $U(0,1)$.

For the numerical calculation of (24)the original time scale is considered, i.e., $\Delta = 1$ s. After getting $R_{t_i}(\eta,\nu)$ and $\Psi_{t_i}(\eta,\nu)$, we got the quantile–quantile graph of Fig. 9 for $\Psi_{t_i}(\eta,\nu)$:

Since the points follow approximately the identity line, it means that the observations obtained fit the stated type of model, and therefore there are no reasons to doubt the fit.

It can be concluded that this model seems to describe in a very good way the short-time interactions of the system and it is ready to be applied for real studies in short and continuous time in EPS.

5.3.1. Electric power demand model

Following the above mentioned procedure, the uniform residues for the O–U process associated with the residue of harmonic regression fit will be studied. According to the method described above, the following result was obtained and shown in Fig. 10:

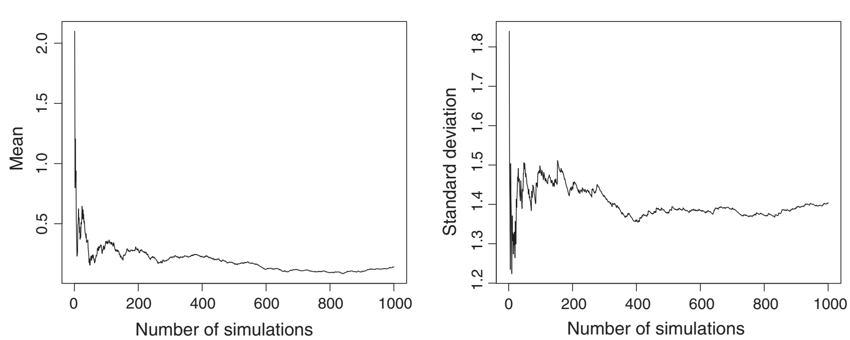
Again, in Fig. 10 the points follow approximately the identity line, so this quantile–quantile plot shows that the harmonic regression residue is well fitted by the chosen process. This states that the random behavior presented in the residential consumption is well represented by de the O–U process.

The results obtained by harmonic regression can be tested with the usual inference analysis. This is: homoscedasticity, correlation, normality of the residual and the determination coefficient R^2. Table 11 shows the results of statistical tests over the three residuals condition and the R^2 value

According to the result observed in Table 11, the non-correlation condition is not fulfilled. This might be because the chosen frequencies for description were not sufficient and more factors could be included to improve the model and obtain better fitting results.

Regarding the R^2 coefficient value, it indicates that the chosen parameters improve the electric energy demand variance in about 75%.

As can be seen, the selection of the variables accurately predicts the energy consumption behavior, however, the correlation and in the residuals reveals that some other elements could be explained by selecting another parameters.



(a) Mean errors behaviour (b) Standard deviation errors behaviour

Fig. 11. Behaviors for different number of simulations.

Table 12
Mean percentage error for the wind generation models.

Model	Error (%)
One-dimensional	0.115
Correlated multidimensional	0.206
Uncorrelated multidimensional	0.294

5.4. Wind generation comparison models

In this section the total power outputs delivered by the two models proposed earlier, correlated and uncorrelated, will be compared with the actual observations. In order to do this we will consider the mean percentage error between the sum of the actual observations and:

- The trajectories of the one-dimensional model.
- The sum of the trajectories of the 11 components of the uncorrelated model.
- The sum of the trajectories of the 11 components of the correlated model.

For the above, 700 trajectories were simulated for each model. The number of simulations was chosen for the following reasons:

- After 700 paths, mean and standard deviation of the calculated error are maintained around a constant value. This can be seen in the Figs. 11(a) and (b). The experiment was made just for the one dimensional case.
- The computational cost over 700 trajectories in multidimensional cases increases too much.

The mean percentage error of each trajectory was calculated, and then these 700 errors were averaged, respectively. Table 12 shows the results.

Although it could be expected that the correlated multidimensional model would be a better representation to the farm, the error of this model is greater than that of the one-dimensional due to the data transformation, since for the one-dimensional case the logarithm of the sum at the time of estimating is considered. However, for the multidimensional case each component must be transformed, turning into a sum of logarithms, which for the values of the provided observations is greater, increasing the error. It is for this reason, and because of its simplicity, that it is more convenient to work with the one-dimensional generation model of the farm. However, if an analysis of the distribution of the windmills wants to be carried out, the multidimensional approach would be appropriate.

6. Conclusions

The present work shows a methodology that allows characterizing, by means of a stochastic process, the random and self-sustained behavior in time of the output power of a wind farm and the demand for electric power.

For modeling wind generation, two short and continuous time representations, a one-dimensional and a multidimensional, were proposed. The one-dimensional case turned out to be a more accurate description of the output power of the farm. On the other hand, in the multidimensional case, the possible interaction between windmills was considered, including a correlated Brownian motion.

A correlation matrix was obtained to represent the effects of one windmill onto the others. The procedure to construct this matrix and the estimation method to obtain its components was explained in detail.

Future studies could be focused on this matrix, which can provide, for example, information on the geometric arrangement of the farm.

For power demand a hybrid model was proposed based on harmonic regression and O–U process.

The techniques for parameter estimation of all the models proposed are given. A validation method for one dimensional continuous model is explained. The latter confirms the good fit of the stated equations.

In the future it will be used the models obtained to evaluate the impact of wind penetration and demand on the operation of the systems in permanent regime.

Acknowledgment

This research was financed by Project FONDECYT 11130169 (Comisión Nacional de Investigación Científica y Tecnológica Chile)

References

[1] Hagspiel S, Papaemannouil A, Schmid M, Andersson G. Copula-based modeling of stochastic wind power in europe and implications for the swiss power grid. Appl Energy 2012;96:33–44. Smart grids. <http://www.sciencedirect.com/science/article/pii/S0306261911006933>.

[2] Zuluaga CD, Ivarez MA, Giraldo E. Short-term wind speed prediction based on robust Kalman filtering: an experimental comparison. Appl Energy 2015;156:321–30. <http://www.sciencedirect.com/science/article/pii/S0306261915008727>.

[3] Pinson P, Christensen L, Madsen H, Srensen P, Donovan M, Jensen L. Regime-switching modelling of the fluctuations of offshore wind generation. J Wind Eng Indust Aerodynam 2008;96(12):2327–47. <http://www.sciencedirect.com/science/article/pii/S0167610508001050>.

[4] Song Z, Jiang Y, Zhang Z. Short-term wind speed forecasting with markov-switching model. Appl Energy 2014;130:103–12. <http://www.sciencedirect.com/science/article/pii/S0306261914005212>.

[5] Calif R. Pdf models and synthetic model for the wind speed fluctuations based on the resolution of Langevin equation. Appl Energy 2012;99(C):173–82. <http://EconPapers.repec.org/RePEc:eee:appene:v:99:y:2012:i:c:p:173-182>.

[6] Zrate-Miano R, Anghel M, Milano F. Continuous wind speed models based on stochastic differential equations. Appl Energy 2013;104(C):42–9. <http://EconPapers.repec.org/RePEc:eee:appene:v:104:y:2013:i:c:p:42-49>.

[7] Olsson M, Perninge M, Soder L. Modeling real-time balancing power demands in wind power systems using stochastic differential equations. Electr Power Syst Res 2010;80(8):966–74.

[8] Liu H, qi Tian H, feng Liang X, fei Li Y. Wind speed forecasting approach using secondary decomposition algorithm and Elman neural networks. Appl Energy 2015;157:183–94. <http://www.sciencedirect.com/science/article/pii/S0306261915009393>.

[9] Wang J-Z, Wang Y, Jiang P. The study and application of a novel hybrid forecasting model a case study of wind speed forecasting in china. Appl Energy 2015;143:472–88. <http://www.sciencedirect.com/science/article/pii/S0306261915000446>.

[10] Kamal L, Jafri YZ. Time series models to simulate and forecast hourly averaged wind speed in Quetta, Pakistan. Solar Energy 1997;61(1):23–32. <http://www.sciencedirect.com/science/article/pii/S0038092X97000376>.

[11] Cadenas E, Rivera W. Wind speed forecasting in the south coast of Oaxaca, Mexico. Renew Energy 2007;32(12):2116–28. <http://www.sciencedirect.com/science/article/pii/S0960148106002801>.

[12] Giorgi MGD, Ficarella A, Tarantino M. Error analysis of short term wind power prediction models. Appl Energy 2011;88(4):1298–311. <http://www.sciencedirect.com/science/article/pii/S030626191000437X>.

[13] Bouzgou H, Benoudjit N. Multiple architecture system for wind speed prediction. Appl Energy 2011;88(7):2463–71. <http://www.sciencedirect.com/science/article/pii/S0306261911000559>.

[14] Liu H, qi Tian H, fei Li Y. Comparison of two new arima-ann and arima-kalman hybrid methods for wind speed prediction. Appl Energy 2012;98:415–24. <http://www.sciencedirect.com/science/article/pii/S0306261912002875>.

[15] Perninge L, Soder M. A stochastic control approach to manage operational risk in power systems. IEEE Trans Power Syst 2012;27(2):1021–31.

[16] Qiu Y, Zhao J, Chiang HD. Effects of the stochastic load model on power system voltage stability based on bifurcation theory. In: Transmission and distribution conference and exposition. IEEE/PES; 2008.

[17] Wang K, Crow ML. Numerical simulation of stochastic differential algebraic equations for power system transient stability with random loads. In: Power and energy society general meeting. IEEE; 2011.

[18] Zhao Yang Dong D, Hua Zhao Jun, Hill. Numerical simulation for stochastic transient stability assessment. IEEE Trans Power Syst 2012;27(4):1741–9.

[19] Milano R, Zarate-Minano F. A systematic method to model power systems as stochastic differential algebraic equations. IEEE Trans Power Syst 2013;28(4):4537–44.

[20] Verdejo H, Kliemann W, Vargas L, Becker C. Stability region and radius in electric power systems under sustained random perturbations. Int J Electr Power Energy Syst 2015;73:725–33. <http://www.sciencedirect.com/science/article/pii/S0142061515002689>.

[21] Kialashaki A, Reisel JR. Modeling of the energy demand of the residential sector in the united states using regression models and artificial neural networks. Appl Energy 2013;108:271–80. <http://www.sciencedirect.com/science/article/pii/S0306261913002304>.

[22] Hida Y, Yokoyama R, Iba K, Tanaka K, Yabe K. Load forecasting on demand side by multi-regression model for operation of battery energy storage system; September 2009. p. 1–5.

[23] Fuks M, Salazar E. Applying models for ordinal logistic regression to the analysis of household electricity consumption classes in rio de janeiro, brazil. Energy Econom 2008;30(4):1672–92. <http://www.sciencedirect.com/science/article/pii/S014098830700117X>.

[24] Bujosa M, Garca-Ferrer A, Young PC. Linear dynamic harmonic regression. Comput Stat ómeno son las series de tiempo Data Anal 2007;52(2):999–1024. <http://www.sciencedirect.com/science/article/pii/S0167947307002666>.

[25] Pedregal DJ, Trapero JR. Electricity prices forecasting by automatic dynamic harmonic regression models. Energy Convers Manage 2007;48(5):1710–9. <http://www.sciencedirect.com/science/article/pii/S0196890406003475>.

[26] Andersen F, Larsen H, Boomsma T. Long-term forecasting of hourly electricity load: identification of consumption profiles and segmentation of customers. Energy Convers Manage 2013;68:244–52. <http://www.sciencedirect.com/science/article/pii/S0196890413000381>.

[27] Dilaver Z, Hunt LC. Industrial electricity demand for turkey: a structural time series analysis. Energy Econom 2011;33(3):426–36. <http://www.sciencedirect.com/science/article/pii/S0140988310001702>.

[28] Erdogdu E. Electricity demand analysis using cointegration and {ARIMA} modelling: a case study of Turkey. Energy Policy 2007;35(2):1129–46. <http://www.sciencedirect.com/science/article/pii/S030142150600098X>.

[29] Felice MD, Alessandri A, Ruti PM. Electricity demand forecasting over Italy: potential benefits using numerical weather prediction models. Electric Power Syst Res 2013;104:71–9. <http://www.sciencedirect.com/science/article/pii/S0378779613001545>.

[30] Mohammadi S, Keivani H, Bakhshi M, Mohammadi A, Askari M, Kavehnia F. Demand forecasting using time series modelling and anfis estimator, vol. 1; September 2006. p. 333–7.

[31] Lei M, Shiyan L, Chuanwen J, Hongling L, Yan Z. A review on the forecasting of wind speed and generated power. Renew Sustain Energy Rev 2009;13(4):915–20. <http://www.sciencedirect.com/science/article/pii/S1364032108000282>.

[32] Soman SS, Zareipour H, Malik O, Mandal P. A review of wind power and wind speed forecasting methods with different time horizons. In: North American power symposium (NAPS) 2010. p. 1–8. September.

[33] McLoughlin F, Duffy A, Conlon M. Evaluation of time series techniques to characterise domestic electricity demand. Energy 2013;50:120–30. <http://www.sciencedirect.com/science/article/pii/S0360544212009164>.

[34] Øksendal B. Stochastic differential equations: an introduction with applications (universitext). 6th ed. Springer; 2014. January.

[35] Platen E, Rendek R. Exact scenario simulation for selected multi-dimensional stochastic processes. Quantitative Finance Research Centre, University of Technology, Sydney, Research Paper Series 259; 2009. <http://EconPapers.repec.org/RePEc:uts:rpaper:259>.

[36] Kloeden P, Platen E. Numerical solution of stochastic differential equations. ser. Applications of mathematics: stochastic modelling and applied probability. Springer; 1992. <http://books.google.cl/books?id=BCvtssom1CMC>.

[37] Martin Bibby B, Srensen M. Martingale estimation functions for discretely observed diffusion processes. Bernoulli 1995;1(1–2):17–39. 03. <http://projecteuclid.org/euclid.bj/1186078360>.

[38] Dimitriadis P, Koutsoyiannis D. Application of stochastic methods to double cyclostationary processes for hourly wind speed simulation. Energy Proc 2015;76:406–11. European Geosciences Union General Assembly 2015 – Division Energy, Resources and Environment, {EGU} 2015. <http://www.sciencedirect.com/science/article/pii/S1876610215016276>.

[39] Ouarda T, Charron C, Shin J-Y, Marpu P, Al-Mandoos A, Al-Tamimi M, Ghedira H, Hosary TA. Probability distributions of wind speed in the {UAE}. Energy Convers Manage 2015;93:414–34. <http://www.sciencedirect.com/science/article/pii/S0196890415000400>.

[40] Olsson M, Perninge M, Sder L. Modeling real-time balancing power demands in wind power systems using stochastic differential equations. Electric Power Syst Res 2010;80(8):966–74.

[41] Anishchenko VS, Astakhov V, Vadivasova T, Neiman A, Schimansky-Geier L. Nonlinear dynamics of chaotic and stochastic systems: tutorial and modern developments. ser. Springer series in synergetics. Berlin, Heidelberg: Springer; 2007. <https://cds.cern.ch/record/1338642>.

[42] Pedersen A. A.U., M.I.A. for Teoretisk Statistik, uniform residuals for discretely observed diffusion processes. Ser. Research reports – Department of Theoretical Statistics, Institute of Mathematics, University of Aarhus; 1994. <http://books.google.cl/books?id=LI7zMgEACAAJ>.