Applied Energy 101 (2013) 541–550

Contents lists available at SciVerse ScienceDirect

Applied Energy

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

A methodology for the synthetic generation of hourly wind speed time series based on some known aggregate input data

Roberto Carapellucci *, Lorena Giordano

Dipartimento di Ingegneria Meccanica, Energetica e Gestionale, University of L'Aquila, Via Giovanni Gronchi 18, 67100 L'Aquila, Italy

HIGHLIGHTS

- ► We proposed a methodology for synthetic generation of hourly wind speed time series.
- ► The generation model is able to adapt to a different number and type of input data.
- ► Model validation is carried out examining two Italian localities.
- ► Model performances are assessed comparing features of measured and generated data.
- ► Best results are obtained when mean and maximum measured wind speeds are known.

ARTICLE INFO

Article history:
Received 9 February 2012
Received in revised form 8 June 2012
Accepted 9 June 2012
Available online 21 July 2012

Keywords:
Wind speed
Synthetic data generation
Diurnal pattern
Optimization algorithm
Autocorrelation function

ABSTRACT

The availability of hourly wind speed data is becoming increasingly important for ensuring the proper design of wind energy conversion systems. For many sites, measured series of such high resolution are incomplete or entirely lacking; hence the need for a model for synthesizing wind speed data.

The objective of this paper is to construct a model for synthetically generating hourly wind speed data, adopting a physical–statistical approach. This generation model defines four parameters for characterizing the wind speed time series in terms of probability distribution and autocorrelation functions. As opposed to the numerous methodologies reported in literature, the proposed approach can be adapted to a different number and type of available input data.

Model validation has been carried out by examining two Italian sites, having different characteristics in terms of mean monthly wind speeds and autocorrelation function. To demonstrate its flexibility, in both sites wind speed time series have been synthesized for three different cases, increasing the amount of known input data.

1. Introduction

The availability of hourly wind speed data is very important for accurately estimating the energy output of a single wind turbine or a wind farm. These data are also indispensable for designing stand-alone power systems, since their sizing is strongly influenced by the mismatch between energy production and electrical load profiles.

The production of synthetic series of wind speed data has traditionally adopted a variety of alternative methods, that can be classified as physical or statistical, depending on their input data. Physical models take into consideration several factors, including shelter from obstacles, pressure, temperature, local surface roughness, and orography effects. Statistical methods are based on models that establish a relationship between statistical parameters and wind speed data. Compared to physical methods, statistical methods are usually simpler and provide accurate results with reduced computational effort [1].

Statistical methods are typically based on probabilistic and stochastic models for the synthetic generation of wind speed time series. In the probabilistic approach suitable distribution functions (i.e. Rayleigh, Weibull, Gumbel) are used to generate independent and identically distributed random numbers [2], while in the stochastic approach various models can be used, including Autoregressive Moving Average (ARMA), Markov chain and the Wavelet-based approach.

ARMA models are a group of linear stochastic models, classified into three categories: purely autoregressive models (AR), moving average models (MA), and mixed models (ARMA), that combine autoregressive and moving average processes. As opposed to the probabilistic approach, ARMA models take into account the fact that hourly wind speeds are independent of each other, but with a degree of persistence defined by a typical autocorrelation

* Corresponding author. Tel.: +39 0862 434348; fax: +39 0862 434403.
E-mail address: roberto.carapellucci@univaq.it (R. Carapellucci).

0306-2619/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.apenergy.2012.06.044

Nomenclature

a	autoregressive coefficient
c	scale factor (m/s)
D	wind turbine rotor diameter (m)
E_{kin}	available kinetic energy (MWh)
e	percentage error (%)
f	Weibull distribution function
f_i	ith objective function
f_i^o	ith utopian point
g	random number from normal distribution
h	hour (h)
i_d	diurnal pattern strength
k	shape factor
n	hours of the year
n_d	daily noise
p	probability density
P	cumulative probability
r	random wind speed component
r_l	autocorrelation function at lag l
v	wind speed (m/s)
$\bar{v}$	mean wind speed (m/s)
w	weight
x	statistical parameter value
z	height (m)

Greek letters

α	wind shear coefficient
Γ	gamma function
ρ_{air}	air density (kg/m^3)
σ	standard deviation

Superscripts

gen	generated
meas	measured

Subscripts

d	daily
det	deterministic
m	monthly
max	maximum
n	normal
p	peak
rand	random
W	Weibull

Acronyms

erf	error function
RMSE	root mean square error

function. These models have been applied successfully in several studies [3–7]; their main disadvantage is the high complexity for estimating coefficients, order and parameters.

The application of the Markov chain model involves the discretization of the stochastic process in a number of states and the definition of the probabilities for the inter-state transition. In the first-order Markov chain approach, the current state of wind speed can only be determined from the known previous state, whereas two previous states are used for determining the current state of wind speed in the second-order Markov chain. The main disadvantage of this method is the possible loss of information with respect to recorded data, related with the discretization procedure, and the large number of parameters, i.e. the inter-state transition probabilities [8]. Hourly wind speed generation based on Markov chain models has been widely studied, the main focus being to assess potential improvements achievable varying the transition matrix order and state size [9–11].

The Wavelet-based approach is proposed as a non-parametric data generation tool. The idea behind this method is the decomposition of the data sample into its signals which is then reconstructed by randomly adding the signals to generate new data [12]. This technique was first applied to wind speed forecasting by Hunt and Nason [13], who attempted to model the wind time-series in terms of a multi-scale wavelet decomposition of data collected at a reference site.

Recently, with the development of artificial intelligence techniques and other forecasting methods, a number of alternative models are becoming popular for predicting wind speed data. These include Artificial Neural Network (ANN), fuzzy logic methods, support vector machine and some hybrid methods [14–18].

The common feature of all the methodologies discussed so far is the need for a series of historical data for calculating their parameters (ARMA models), constructing the transition matrices (Markov chain model) and defining the rules (ANN methods). However, most wind speed databases only provide average and maximum wind speeds on a monthly or yearly basis and lack information on the autocorrelation properties of wind speed data.

The purpose of this paper is to develop a methodology for synthesizing an hourly wind speed time series, using readily available statistical input data, such as average and maximum wind speed on a monthly or yearly basis. With the proposed model deterministic and random components of wind speed data are determined simply on the basis of four parameters influencing the persistence and autocorrelation properties of the generated time series. Parameter values are determined using a stochastic optimization algorithm, that minimizes a multi-objective function dependent on the number and type of input data.

To validate this methodology two Italian locations, representative of on- and off-shore sites, have been considered. To highlight the flexibility of the proposed model, i.e. its ability to operate regardless of the amount of known input data, wind speeds are generated for three cases, for increasing amounts of input data. Generated and measured wind speed data are compared in terms of probability distribution, general statistical parameters and autocorrelation function of the time series, and the effects on kinetic energy available for a wind turbine estimated.

2. Description of hourly wind speed data methodology

The proposed methodology is based on the assumption that wind speed comprises periodic deterministic and stochastic components [19]. Hence, the synthetic generation of hourly wind speed data consists of five main steps that can be summarized as follows.

(a) Generation of a time series of daily wind speed.

Based on statistical input data, daily mean wind speeds are generated adding a daily noise to mean yearly or monthly wind speeds (v_m). Referring to the latter case:

$$v_d = v_m(1 + n_d) \qquad m = 1, \ldots, 12, \qquad d = 1, \ldots, 365 \qquad (1)$$

where n_d is a daily noise extracted from a Gaussian distribution function.

(b) Generation of a time series of deterministic wind speed components.

Deterministic wind speed components are determined from the mean daily profile, that can be modeled using a cosine function [20]:

$$v_{det}(h) = v_d\{1 + i_d cos[2\pi/24(h - h_p)]\} \qquad h = 1, \ldots, 8760 \tag{2}$$

Eq. (2) is based on the assumption that wind speed has a repetitive pattern, exhibiting low speeds up to the early hours of the morning, increasing up to late afternoon, to decrease thereafter down to the minimum value in the evening [21]. The diurnal pattern strength, i_d, indicates how wind speed depends on the time of day, corresponding to the ratio of the amplitude of cosine function to the mean wind speed; typical values range from 0 to 0.3. The peak hour of wind speed, h_p, which depends on the time duration in which wind speed increases, ranges from 12 to 18.

(c) Generation of a time series of stochastic wind speed components.

Random wind speed components are estimated through an autoregressive model of the first order:

$$r(h) = ar(h-1) + g(h) \qquad h = 1, \ldots, 8760 \tag{3}$$

where a is the autoregressive coefficient and $g(h)$ is a random number extracted from a normal distribution function with zero mean and unit standard deviation.

(d) Summing the two series generated in the last two steps.

The time series $v_{det}(h)$ cannot be added to that of random components $r(h)$, as their distribution functions are different. This problem can be overcome by transforming the probabilities of the time series $v_{det}(h)$ so that it has the same distribution function of $r(h)$.
The fundamental transformation law of probabilities states that:

$$\int_{-\infty}^{r} p_r(r) = \int_{-\infty}^{v_{det}} p_{v_{det}}(v_{det}) \tag{4}$$

where $p_r(r)$ and $p_{v_{det}}(v_{det})$ are the distribution functions of $r(h)$ and $v_{det}(h)$ respectively. This is the fundamental equation for transforming random numbers from the distribution $p_{v_{det}}(v_{det})$ to a new distribution $p_r(r)$.
In terms of cumulative distribution function, Eq. (4) becomes:

$$P_r(r) = P_{v_{det}}(v_{det}) \tag{5}$$

Appling the inverse method, we get:

$$r(v_{det}) = P_r^{-1}[P_{v_{det}}(v_{det})] \tag{6}$$

The series $r(v_{det})$ is thus characterized by a normal distribution function with mean and standard deviation equal to those of the series $r(h)$. Denoting $r(v_{det})$ with v_{norm} and adding $r(h)$, a normalized time series of wind speed is obtained:

$$v'_{norm} = v_{norm} + r(h) \tag{7}$$

(e) Transition from a normal to a Weibull distribution function.

Generally, the probability density function that best fits the measured wind speed data is the *Weibull distribution* [22]:

$$f(v) = \frac{k}{c}\left(\frac{v}{c}\right)^{k-1} \exp\left[-\left(\frac{v}{c}\right)^{k}\right] \tag{8}$$

where k and c are shape and scale factors respectively.
These parameters are related to mean of wind speeds through:

$$\bar{v} = c\Gamma\left(\frac{1}{k} + 1\right) \tag{9}$$

where Γ is the gamma function.
Applying again the transformation law of probabilities, it is possible to obtain a series of wind speeds characterized by a Weibull distribution function, with the desired values of parameters k and c:

$$\int_{-\infty}^{v'_{norm}} p_{v'_{norm}}(v'_{norm}) = \int_{-\infty}^{v_W} p_{v_W}(v_W) \tag{10}$$

or

$$P_{v'_{norm}}(v'_{norm}) = P_{v_W}(v_W) \tag{11}$$

where v_W is the wind speed time series.
The Weibull distribution can be simple or mixed; in the latter case, the cumulative distribution function is obtained through a weighted average of several independent cumulative Weibull distributions:

$$P_{v_W}(v_W) = \sum_{i=1}^{n} w_i P_{v_W}(v_W)_i \qquad \text{with} \quad \sum_{i=1}^{n} w_i = 1 \tag{12}$$

where

$$P_{v_W}(v_W)_i = 1 - \exp\left[-\left(\frac{v_W}{c_i}\right)^{k_i}\right] \tag{13}$$

and $w_i > 0$ are the weight parameters.
In the case of a simple Weibull distribution, Eq. (11) becomes:

$$\int_{-\infty}^{v'_{norm}} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = 1 - \exp\left[-\left(\frac{v_W}{c}\right)^{k}\right] \tag{14}$$

from which:

$$\frac{1}{2} + \int_{0}^{v'_{norm}} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = 1 - \exp\left[-\left(\frac{v_W}{c}\right)^{k}\right] \tag{15}$$

The integral of the Gaussian distribution function is determined using the error function:

$$erf(y) = \frac{2}{\sqrt{\pi}} \int_0^y e^{-u^2} du \tag{16}$$

Hence, provided that:

$$u = \frac{x - \mu}{\sqrt{2\sigma^2}} \tag{17}$$

Eq. (15) becomes:

$$\frac{1}{2}\left[1 + erf\left(\frac{v'_{norm} - \mu}{\sqrt{2}\sigma}\right)\right] = 1 - \exp\left[-\left(\frac{v_W}{c}\right)^{k}\right] \tag{18}$$

and consequently:

$$v_W = c\left\{-ln\left[\frac{1}{2} - \frac{1}{2} erf\left(\frac{v'_{norm} - \mu}{\sqrt{2}\sigma}\right)\right]\right\}^{\frac{1}{k}} \tag{19}$$

3. Multi-objective optimization for parameter estimation

Referring to a simple Weibull distribution function, four variables need to be known to generate wind speed data: shape factor

(k), autoregressive coefficient (a) and mean daily profile parameters, i.e. diurnal pattern strength (i_d) and peak hour (h_p). These parameters are determined on the basis of available statistical wind speed data, using a genetic algorithm, that is a kind of stochastic optimization technique, first proposed by Holland [23].

3.1. Genetic algorithm

In a genetic algorithm (GA) each possible solution is represented by a vector of the independent variables, called chromosome. At the first stage of a genetic algorithm, an initial population of potential solutions to the problem is randomly generated and evaluated in terms of their fitness, which generally corresponds to the optimization function. In the selection procedure solutions are ranked according to their fitness function, such that the best individuals will have a greater probability of being selected as parents. This process is followed by the recombination of genetic material of the individuals selected, by means of crossover and mutation processes. In crossover, stochastically selected subsections of two parents are swapped with each other to produce the offspring. In mutation, randomly selected genes in chromosomes are changed, at a low mutation rate, to maintain genetic diversity from one generation to the next. The process is repeated a number of times, until the acceptable fitness level or the maximum number of generations is reached.

Table 1 summarizes the main operating parameters of the adopted genetic algorithm, implemented in Visual Basic environment.

3.2. Objective function

The optimization problem is solved using a distance function based method [24]. This involve minimizing the distance between the set of the objective function values (f_i) and some chosen reference points (f_i^o), also called ideal or utopian points:

$$OF = \left[\left(\sum_{i=1}^{N}(f_i - f_i^o)^p\right)\right]^{1/p} \quad (20)$$

With p = 2, the objective function corresponds to the Euclidean distance of the different objective functions (f_i) from expected values (f_i^o):

$$OF = \sqrt{\sum_{i=1}^{N}(f_i - f_i^o)^2} \quad (21)$$

In this paper, f_i and f_i^o represent the statistical wind speed data, such as monthly or yearly mean and maximum wind speeds, determined using simulated and measured wind speed data respectively.

4. Validation of the proposed methodology

To validate the methodology proposed for generating hourly wind speed data, two locations have been selected. The first, *Colle Val d'Elsa*, is a municipality in the Tuscany region; the second is *Pianosa*, a small island (area about 1.5 km^2), belonging to the archipelago of the Tremiti Islands. Due to its position with respect to the coast and its morphology – fairly flat land and absence of tree cover –, the site can be classed as an off-shore site.

Table 1
Genetic algorithm parameters.

Parameter	Value
Population,	16
Crossover probability, p_c	0.9
Mutation probability, p_m	0.01
Maximum number of generations	10

4.1. Experimental wind data and modeling assumptions

In the two sites Colle Val d'Elsa and Pianosa Island, wind measurements campaign were conducted by the CESI Research Institute, in the period from November 2008 to February 2010 [25].

Focusing on wind speed data for the year 2009, the two sites have been characterized by calculating the wind speed distribution function, average and maximum wind speeds, as well as the standard deviations, on a monthly or yearly basis. These parameters are summarized in Table 2 for the two sites. Moreover, the trend of the autocorrelation function has been also determined.

Synthetic wind speed time series have been generated for three cases, each with different input data. In Case 1 only the mean yearly wind speed is known, in Case 2 also mean monthly wind speeds, and in Case 3 also maximum monthly values (Table 3). Different input data affect directly the definition of the objective function, whose complexity increases with the number of available data. Hence, in Case 1 the objective function is simply the Euclidean distance between generated and measured mean annual wind speeds, while in Case 3 it is the Euclidean distance between generated values and measured data for monthly mean and maximum wind speeds.

Table 4 summarizes the range of variability for the different optimization parameters. Diurnal pattern strength (i_d) ranges from 0 to 0.1 for Pianosa Island and from 0.1 to 0.3 for Colle Val d'Elsa, while other parameters vary in the same ranges for both sites. The reason for choosing i_d lies with the morphology of the sites [26]. As previously mentioned, Pianosa Island, regarded as an off-shore site, is presumably characterized by a flatter daily wind speed profile than Colle Val d'Elsa.

4.2. Model results

Table 5 summarizes the values of the parameters determined using the optimization algorithm as a function of input data, for the two sites.

These parameters depend on the input data and on the site itself. In fact, they affect the anemological characteristics of the site. As is well known, k mainly influences the shape of the probability distribution function, while the other parameters have a greater impact on the autocorrelation properties. In fact the first autocorrelation coefficient (r_1) is an increasing function of a, while the

Table 2
Main statistical parameters for Colle Val d'Elsa and Pianosa Island.

	Colle Val d'Elsa			Pianosa Island		
	$\bar{v}$	v_{max}	σ	$\bar{v}$	v_{max}	σ
January	2.36	13.0	1.59	6.64	15.38	3.11
February	2.92	11.6	1.87	7.84	16.97	3.30
March	3.44	14.0	2.37	7.34	17.43	3.84
April	2.80	11.8	1.97	5.35	17.35	3.17
May	2.70	11.5	1.57	4.61	15.30	2.41
June	2.93	9.5	1.72	4.76	15.35	2.79
July	3.05	10.8	1.88	4.27	13.37	2.09
August	2.60	8.5	1.45	4.46	13.80	2.31
September	2.73	9.6	1.78	4.53	12.87	2.62
October	2.29	9.7	1.53	6.10	20.78	3.34
November	2.39	7.7	1.57	4.83	14.80	2.89
December	2.74	13.4	1.97	8.16	20.25	3.86
Year	2.75	14.0	1.82	5.74	20.78	3.31

Table 3
Input data for the three cases investigated.

	Input data
Case 1	Yearly mean wind speed
Case 2	Monthly mean wind speeds
Case 3	Monthly mean and maximum wind speeds

Table 4
Range of optimization parameter variability for the two sites.

	Colle Val d'Elsa		Pianosa Island	
	min	max	min	max
i_d	0.1	0.3	0	0.1
h_p	12	18	12	18
a	0.6	0.9	0.6	0.9
k	1	2.5	1	2.5

Table 5
Values of optimization parameters as a function of input data.

	Colle Val d'Elsa			Pianosa Island		
	Case 1	Case 2	Case 3	Case 1	Case 2	Case 3
i_d	0.26	0.28	0.29	0.06	0.09	0.05
h_p	13.60	13.19	12.10	17.13	15.00	15.04
a	0.71	0.65	0.68	0.83	0.65	0.73
k	1.69	2.47	1.58	1.51	2.40	1.88

parameters i_d and h_p affect the daily wind speed profile and also the shape of the autocorrelation function, as will be discussed below.

The synthetic wind speed time series has been thoroughly analyzed to assess its ability to preserve the general statistical properties of the measured series, and thus the applicability of the proposed methodology for wind speed generation. More specifically, measured and generated wind speed data have been compared in terms of the probability distribution function, general statistical parameters (including mean and standard deviation) and autocorrelation function of the time series.

The yearly statistical parameters, such as yearly mean, standard deviation and maximum values of the generated and measured wind speed data are summarized in Tables 6 and 7; the last three columns of both tables show the percentage error:

$$e(\%) = \frac{x^{gen} - x^{meas}}{x^{meas}} \cdot 100 \tag{22}$$

where x^{gen} and x^{meas} are generated and measured values of a generic statistical parameter.

Considering the Colle Val d'Elsa site, the best results are obtained for Case 3, where monthly mean and maximum wind speeds were used as input data. In fact, the percentage error on mean wind speed is negligible (0.1%), while errors related to maximum value and standard deviation are −7.2% and −1.2% respectively. On the contrary, for Case 2, though the agreement remains very good for mean wind speed, there are significant errors on other statistical parameters, even greater than for Case 1. For Pianosa Island, the best results are again obtained for Case 3. However, for this site the percentage error on the maximum yearly wind speed in Case 2 (−20.4%) is lower than Case 1 (43.4%), while the opposite is true of the standard deviation.

In terms of probability distribution, synthetically generated data have been compared qualitatively and quantitatively with those of observed values.

Table 6
Comparison of mean, standard deviation and maximum values of measured and generated data for Colle Val d'Elsa.

	$\bar{v}$	v_{max}	σ	$e(\bar{v})$	$e(v_{max})$	$e(\sigma)$
Measured	2.75	14	1.82	–	–	–
Case 1	2.75	12.3	1.69	0.03	−12.4	−6.9
Case 2	2.75	8.0	1.19	0.04	−42.7	−34
Case 3	2.75	13.0	1.79	0.1	−7.2	−1.2

Table 7
Comparison of mean, standard deviation and maximum values of measured and generated data for Pianosa Island.

	$\bar{v}$	v_{max}	σ	$e(\bar{v})$	$e(v_{max})$	$e(\sigma)$
Measured	5.74	20.8	3.31	–	–	–
Case 1	5.74	29.8	3.82	−0.07	43.4	15.3
Case 2	5.75	16.5	2.55	0.12	−20.4	−23.2
Case 3	5.75	21.9	3.17	0.14	5.1	−4.4

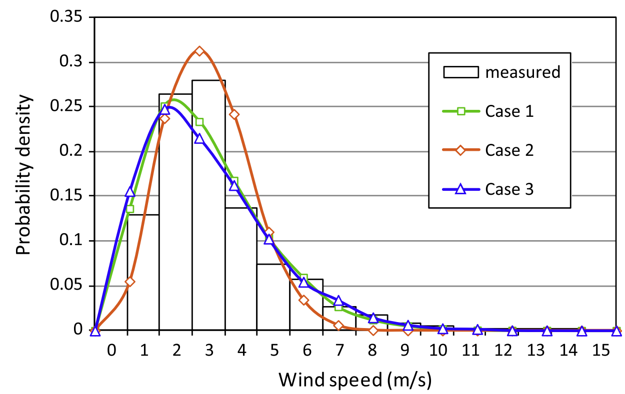

Fig. 1. Probability distribution of measured and synthetically generated wind speed data at Colle Val d'Elsa.

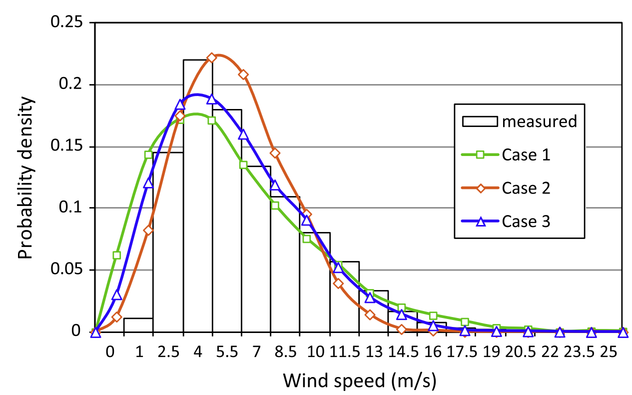

Fig. 2. Probability distribution of measured and synthetically generated wind speed data at Pianosa Island.

Table 8
Weibull parameters obtained for measured and generated data.

	Colle Val d'Elsa				Pianosa Island			
	Measured	Case 1	Case 2	Case 3	Measured	Case 1	Case 2	Case 3
k	1.57	1.69	2.47	1.58	1.81	1.51	2.40	1.87
c	3.06	3.07	3.09	3.06	6.45	6.36	6.48	6.47

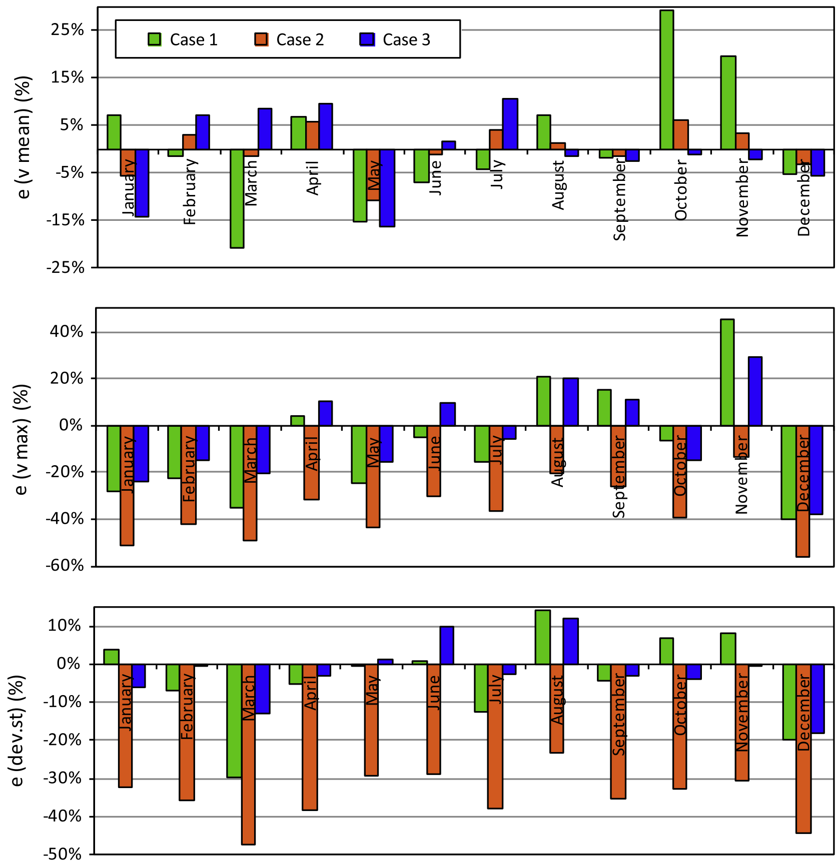

Fig. 3. Percentage errors on main statistical parameters for Colle Val d'Elsa.

Figs. 1 and 2 show the comparison between measured and generated wind speed distribution functions for Colle Val d'Elsa and Pianosa Island respectively. For both sites, the distribution functions for Cases 1 and 3 are very similar, notwithstanding the substantial differences in the amount of available input data, showing a good fit with the measured distribution function. For a quantitative assessment, Weibull distribution parameters have been compared. For the measured data, these parameters have been calculated using the experimental method [27]:

$$k = \left(\frac{\sigma}{\bar{v}}\right)^{-1.086} \tag{23}$$

$$c = \bar{v}/\Gamma\left(\frac{1}{k}+1\right) \tag{24}$$

The Weibull parameters at both stations for the observed and generated wind speed time series are compared in Table 8. The results show that in Case 3 the shape (k) and scale (c) parameters are very close to those calculated from experimental data. In Case 1 the agreement is still relatively good, whereas in Case 2 the optimization process always leads to an overestimation of the shape factor.

5. Comparing various features of measured and generated wind data

The extent to which the proposed methodology is able to preserve the features of the experimental distribution function depends on the amount and type of input data. However, the mere comparison in terms of distribution function is not sufficient to shed light on the actual differences between synthetic hourly wind speed time series and measured data. Hence, it is necessary to extend the investigation to other aspects, including some monthly statistical parameters (such as mean, maximum and standard deviations of wind speed), the autocorrelation function and the temporal pattern of time series.

The results of these investigations will be discussed in the following subsections, highlighting the differences between the cases examined.

5.1. Monthly statistical parameters

To highlight the differences among the three cases examined, percentage errors have also been calculated on a monthly basis for each statistical parameter. These are shown in Figs. 3 and 4 for Colle Val d'Elsa and Pianosa Island respectively.

Regarding the Colle Val d'Elsa site, errors on mean wind speeds for Case 2 are very low and range from −10.8% (May) to 6.1% (October), whereas errors on maximum wind speeds and standard deviations are certainly more significant. This can be explained by considering that monthly mean wind speeds do not differ substantially from the mean yearly value. Consequently, the optimization algorithm yields a relatively narrow distribution function around the mean value, that fails to reproduce the frequencies of the measured data at the tail.

On the contrary, for Case 1 there are no constraints on monthly mean wind speeds, so the distribution function determined using the optimization algorithm is broader. This leads to an increase of errors on the monthly mean wind speeds, but a reduction of those related to other statistical parameters.

Finally, for Case 3 errors on the mean wind speeds increase slightly compared to Case 2, ranging from −16.4% (May) to 10.6% (July), while errors on the maximum wind speeds and standard deviations decrease significantly. This is due to the need to find a compromise solution through the optimization algorithm, i.e. one which is able to combine often conflicting objectives. Referring to

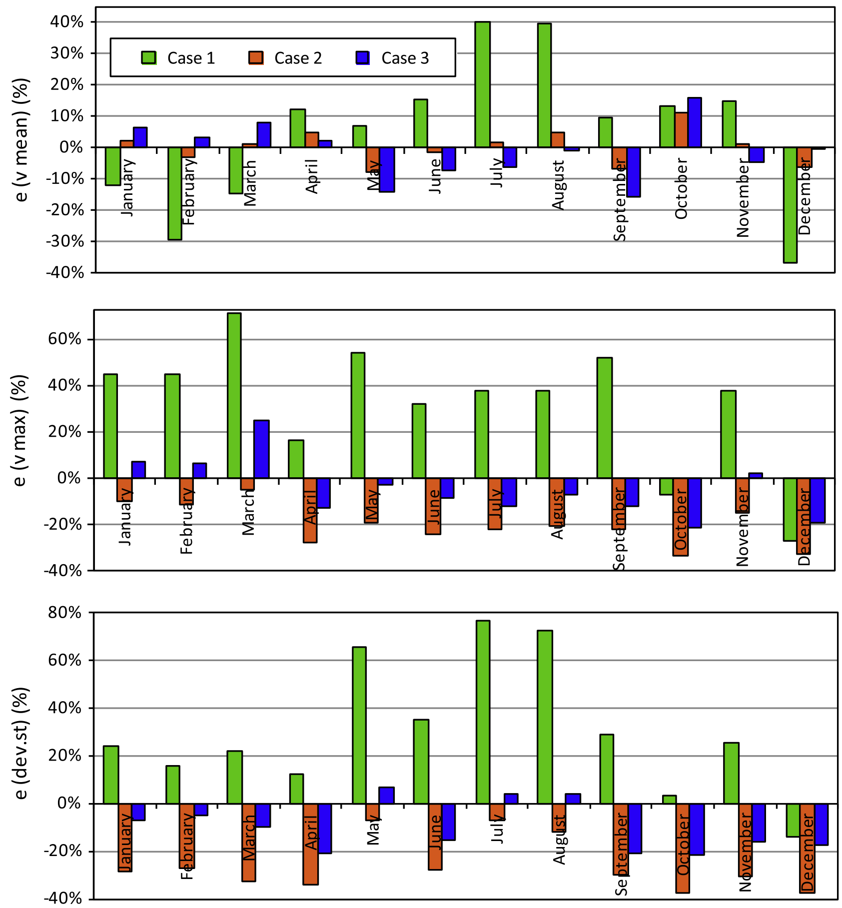

Fig. 4. Percentage errors on main statistical parameters for Pianosa Island.

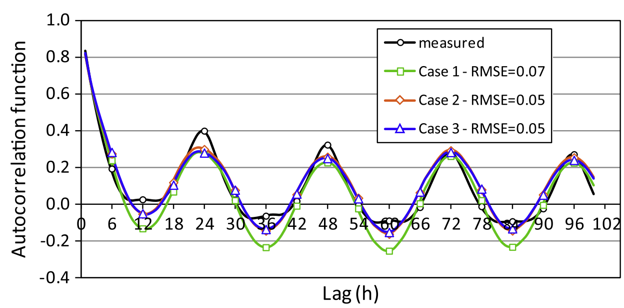

Fig. 5. Autocorrelation functions of measured and generated wind speed data for Colle Val d'Elsa.

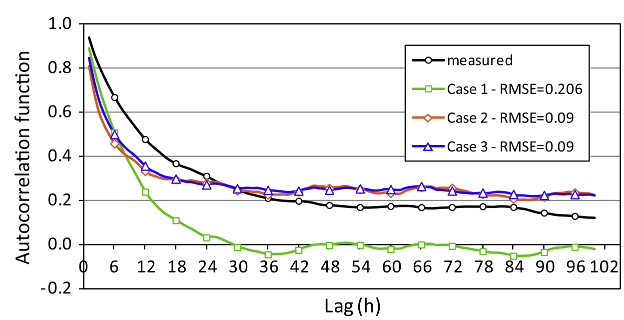

Fig. 6. Autocorrelation functions of measured and generated wind speed data for Pianosa Island.

Case 3, the optimal solution must satisfy a greater number of objectives than Case 2; this leads to a greater penalty in the determination of mean wind speeds, in favor of a reduction of the error on the maximum values.

As far as Pianosa Island is concerned, the results obtained on a monthly basis confirm the trend observed in the investigation on yearly parameters. In fact, Fig. 4 shows that in Case 2 the errors on monthly mean wind speeds range from −7.9% to 11.4%. Moreover, differences in terms of maximum wind speed and standard deviation are lower than Case 1, where these parameters are always far overestimated. In fact, contrary to the Colle Val d'Elsa site, the measured monthly mean wind speeds are quite different from the mean yearly value. Consequently, the optimization algorithm yields to a broader distribution function, with a beneficial effect on the estimation of maximum monthly wind speeds.

Finally, in Case 3 the errors on mean wind speeds range from −14.3% to 15.8%, whereas those on the maximum values and standard deviations are, as expected, the lowest.

5.2. Autocorrelation function

The properties of wind speed persistence can be also assessed in terms of autocorrelation, not just by examining wind speed duration curves.

Autocorrelation can be defined as the degree of dependency of wind speed at a given time on preceding values. One measure of this dependency is given by the coefficient of autocorrelation, that

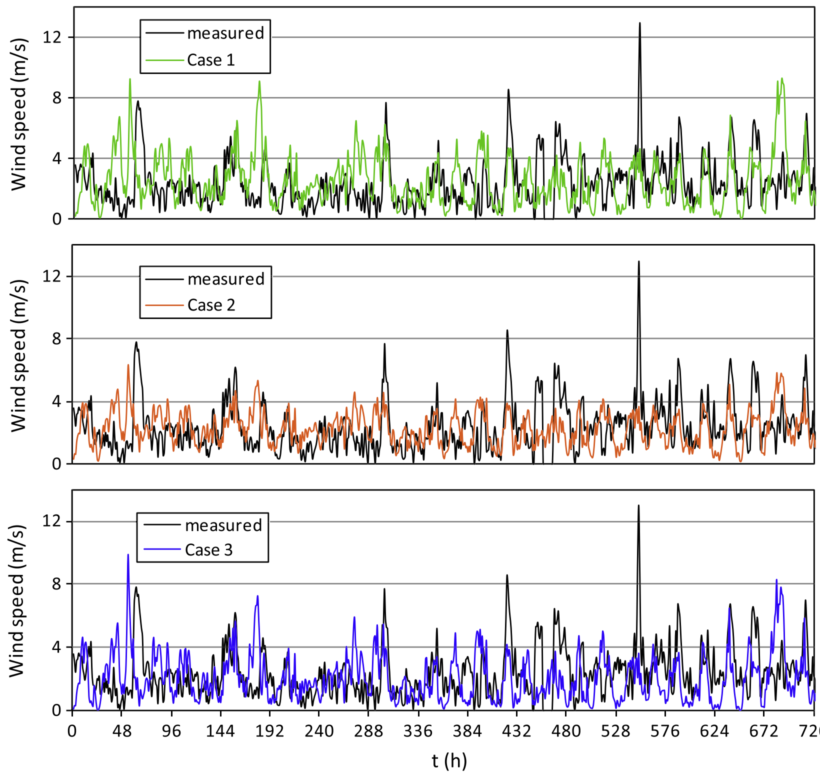

Fig. 7. Measured and generated wind speed data for Colle Val d'Elsa.

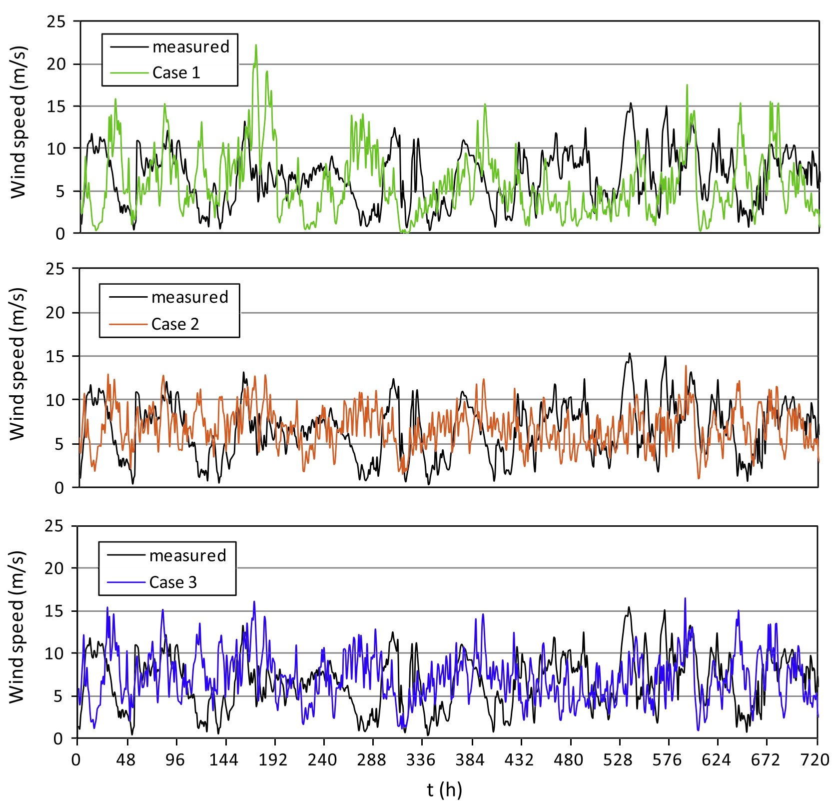

Fig. 8. Measured and generated wind speed data for Pianosa Island.

Table 9
Yearly kinetic energy (in MWh) available for a wind turbine with D = 12 m, calculated for measured and generated wind speed data.

	Colle Val d'Elsa		Pianosa Island	
	E_{kin}	$e(E_{kin})$	E_{kin}	$e(E_{kin})$
Measured	44.8	–	292.0	–
Case 1	38.7	−13.8	355.5	+21.8
Case 2	26.3	−41.4	218.8	−25.1
Case 3	42.1	−6.2	273.2	−6.4

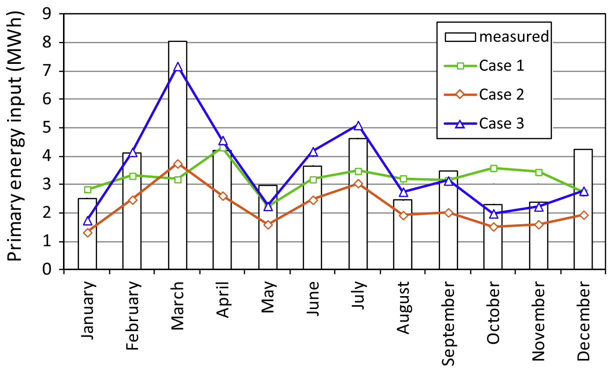

Fig. 9. Comparison of measured and generated primary energy input for Colle Val d'Elsa.

is a dimensionless parameter, defined as the ratio between the autocovariance at lag l and the variance of time series:

$$r_l = \frac{\frac{1}{n-l}\sum_{i=1}^{n-l}(v_i - \bar{v})(v_{i+l} - \bar{v})}{\frac{1}{n}\sum_{i=1}^{n}(v_i - \bar{v})^2} \quad (25)$$

where l is the lag between two wind speed values and $\bar{v}$ is the mean yearly wind speed of the time series. By definition r_0 = 1. The autocorrelation coefficient for l = 1,r_1, strongly depends on local topography; as a general rule, areas surrounded by different types of topography have lower values (0.7–0.8) than those with uniform topography (0.9–0.97) [20].

The autocorrelation functions calculated from the measured and generated wind speed data are given in Figs. 5 and 6 for Colle Val d'Elsa and Pianosa Island respectively. In particular, it can be seen that the Colle Val d'Elsa site is characterized by a fluctuating autocorrelation function, related to a diurnal pattern exhibiting higher wind speeds during the day than at night.

The proposed wind speed generation model is able to simulate the oscillating behavior of the autocorrelation function, regardless of the input data. The fit is quite good in Cases 2 and 3, i.e. when at least mean monthly wind speeds are known. Moreover, considering the first 24 lags, the minimum autocorrelation coefficient, both for measured and generated data, occurs for l = 12, which corresponds roughly to the time-lag between the maximum and minimum wind speed of the mean daily profile.

These general considerations can also be quantitatively supported, considering the root mean square error (RMSE):

$$\mathrm{RMSE} = \left[\frac{1}{n}\sum_{l=1}^{n}\left(r_l^{gen} - r_l^{meas}\right)^2\right]^{1/2} \quad (26)$$

and observing that it decreases from 0.07 (Case1) to 0.05 (Case 2 and Case 3), referring to the first 100 coefficients of autocorrelation.

Regarding Pianosa Island, the autocorrelation function is characterized by an asymptotic decreasing trend, due to a flatter mean daily profile. Discrepancies between measured and generated

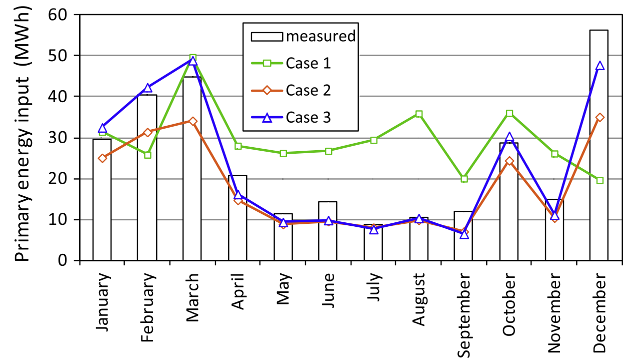

Fig. 10. Comparison of measured and generated primary energy input for Pianosa Island.

autocorrelation coefficients are greater, especially in Case 1 and with lags greater than 6 h. Consequently, RMSE takes higher values, ranging from 0.09 (Case 3) to 0.2 (Case 1).

5.3. Time series

The measured and generated wind speed data can be also compared in terms of time series. These are shown in Fig. 7 (Colle Val d'Elsa) and Fig. 8 (Pianosa Island), considering a time-window of the first 30 days.

As can be seen from the graphs, generated wind speed data do not match measured data perfectly. However, as measured wind speeds tend to increase, the wind speeds calculated with the model generally behave in a similar fashion. This can be explained considering that, unlike "curve fitting" models, the proposed methodology could not exploit the entire time series as input data.

Indeed, the purpose of this model, starting from general statistical parameters (mean and maximum wind speeds), is to generate a wind speed time series with very similar features to the measured data in terms of probability density function and autocorrelation, so as to provide an accurate estimate of energy production throughout the year.

With this aim in mind, the entire measured and generated time series can be compared in terms of kinetic energy available for a generic wind turbine, as discussed in the following subsection.

5.4. Effects on kinetic energy available for a wind turbine

The yearly kinetic energy available for a generic wind turbine can be estimated as follows:

$$E_{kin} = \sum_{i}^{8760}\frac{1}{8}\rho_{air}\pi D^2 v_i^3 \quad (27)$$

where D is the rotor diameter and v_i the wind speed at hub height. This value is calculated using the power law:

$$v(z) = v(z_{anem})\left(\frac{z}{z_{anem}}\right)^{\alpha} \quad (28)$$

where $v(z_{anem})$ is the mean wind speed at anemometer height (10 m, in this present study) and α is the wind shear coefficient, which depends on numerous factors, including atmospheric conditions, temperature, pressure, humidity, time of day, season of the year, mean wind speed, direction and nature of terrain [28].

Table 9 summarizes the results obtained, assuming a wind turbine diameter of 12 m and a wind shear coefficient of 0.143 (1/7) for Colle Val d'Elsa and 0.11 for Pianosa Island, regarded as an off-shore site.

The results in terms of kinetic energy available for a wind turbine confirm the good performance of the model when monthly mean and maximum wind speeds (Case 3) are used as input data, with a slight underestimation (close to 5%) for both the sites considered. On the contrary, in Case 1 the optimization process largely overestimates the yearly kinetic energy available for the Pianosa Island site (+21.8%), while it is significantly underestimated in Case 2 for both sites, the relative error being more than −40% for Colle Val d'Elsa and more than −25% for Pianosa Island.

Model performance for the three cases examined also emerges from analysis of the data on a monthly basis. The graphs of Figs. 9 and 10, for Colle Val d'Elsa and Pianosa Island respectively, show that the monthly kinetic energy calculated from generated wind speed data for Case 3 matches very closely the kinetic energy calculated using measured data. The same cannot be said of Cases 1 and 2 which are penalized by the less accurate estimation of maximum wind speeds in terms of amplitude and relative frequency, that significantly affect available kinetic energy with a cubic law.

6. Conclusions

In this study a novel approach for synthetically generating hourly wind speed data is proposed. It is based on a stochastic methodology that uses a number of statistical parameters, such as mean and maximum wind speed on a yearly or monthly basis.

The proposed algorithm has been implemented for two Italian locations, representative of on- and off-shore sites. In order to highlight the model's flexibility, hourly wind speed data have been generated for three cases, each with different input data. In Case 1 only the mean yearly wind speed is known, in Case 2 only mean monthly wind speeds, and in Case 3 monthly mean and maximum wind speed values.

The generated wind speed time series have been compared with experimental ones, considering as comparison criteria the probability distribution function, the general statistical parameters, as well as the autocorrelation function. For both sites, the best results are obtained for Case 3, the proposed methodology being able to preserve the different characteristics of the measured wind speed data.

Referring to the distribution function, the Weibull parameters are very close to those calculated using the experimental data. Moreover there is good agreement in terms of mean and standard deviations of wind speeds on both a yearly and monthly basis. As regards the autocorrelation function, the method reproduces fairly closely the behavior in the two sites, exhibiting an oscillatory behavior for Colle Val d'Elsa, typical of sites with a highly variable daily profile and a decreasing asymptotic trend for Pianosa Island, typical of sites with a fairly flat daily profile.

In Case 2, the wind speed model yields a too narrow distribution function around the mean value, especially for Colle Val d'Elsa, characterized by monthly mean wind speed similar to the yearly mean; the autocorrelation function is quite close to that of Case 3. The opposite trend is observed for Case 1, where the generated wind speed time series is able to well reproduce the characteristics of the probability distribution function, but less accurately those of autocorrelation function of the measured data.

The good performance of the proposed model is confirmed by the results obtained for the kinetic energy available for a generic wind turbine. Compared to the values obtained with the measured data, in both sites the yearly kinetic energy is underestimated only by about 5% in Case 3, i.e. when the hourly wind speed time series is generated also using maximum monthly wind speeds.

References

[1] Ernst B, Oakleaf B, Ahlstrom ML, Lange M, Moehrlen C, Lange B, et al. Predicting the wind-models and methods of wind forecasting for utility operations planning. IEEE Power Energy Mag 2007;5(6):78–89.
[2] Aksoy H, Toprak ZF, Aytek A, Unal NE. Stochastic generation of hourly mean wind speed data. Renew Energy 2004;29:2111–31.
[3] Castellanos F, Ramesar VI. Characterization and estimation of wind energy resources using autoregressive modelling and probability density functions. Wind Eng 2006;30(1):1–14.
[4] Torres JL, Garcia A, De Blas M, De Francisco A. Forecast of hourly average wind speed with ARMA models in Navarre. Sol Energy 2005;79(1):65–77.
[5] Philippopoulos K, Deligiorgi D. Statistical simulation of wind speed in Athens, Greece based on Weibull and ARMA models. Int J Energy Environ 2009;3(4).
[6] Poggi P, Muselli M, Notton G, Cristofari C, Louche A. Forecasting and simulating wind speed in Corsica by using an autoregressive model. Energy Convers Manage 2003;44:3177–96.
[7] Erdem E, Shi J. ARMA based approaches for forecasting the tuple of wind speed and direction. Appl Energy 2011;88:1405–14.
[8] Papaefthymiou G, Klockl B. MCMC for wind power simulation. IEEE Trans Energy Convers 2008;23(1).
[9] Shamshad A, Bawadi MA, Wan Hussin WA, Majid TA, Sanusi SM. First and second order Markov chain models for synthetic generation of wind speed time series. Energy 2005;30:693–708.
[10] Nfaoui H, Buret J, Sayigh AM. Stochastic simulation of hourly average wind speed sequences in Tangiers (Morocco). Sol Energy 1996;56:301–14.
[11] Hocaoglu FO, Gerek ON, Kurban M. The effect of markov chain state size for synthetic wind speed generation. In: The tenth international conference on probabilistic methods applied to power systems (PMAPS2008), Rincón, Puerto Rico; May 25–29 2008.
[12] Bayazit M, Aksoy H. Using wavelets for data generation. J Appl Stat 2001;28(2):157–66.
[13] Hunt K, Nason GP. Wind speed modeling and short term prediction using wavelets. Wind Eng 2001;25(1):55–61.
[14] Gong L, Shi J. On comparing three artificial neural networks for wind speed forecasting. Appl Energy 2010;87:2313–20.
[15] Mabel MC, Fernandez E. Analysis of wind power generation and prediction using ANN: a case study. Renew Energy 2008;33(5):986–92.
[16] Fadare D. The application of artificial neural networks to mapping of wind speed profile for energy application in Nigeria. Appl Energy 2010;87:934–42.
[17] Damousis IG, Alexiadis MC, Theocharis JB, Dokopoulos PS. A fuzzy model for wind speed prediction and power generation in wind parks using spatial correlation. IEEE Trans Energy Convers 2004;19(2):352–61.
[18] Mohandes MA, Halawani TO, Rehman S, Hussain AA. Support vector machines for wind speed prediction. Renew Energy 2004;29:939–47.
[19] Burton T, Jenkins N, Sharpe D, Bossanyi E. Wind energy handbook. 2nd ed. 2011: John Wiley & Sons; 2011.
[20] Lilienthal P. The HOMER® micropower optimization model. 2004 DOE solar energy technologies program review meeting, Denver, Colorado; October 25–28 2004.
[21] Ephrath JE, Goudriaanb J, Marani A. Modelling diurnal patterns of air temperature, radiation wind speed and relative humidity by equations from daily characteristics. Agr Syst 1996;51(4):377–93.
[22] Usta I, Kantar Y. Analysis of some flexible families of distributions for estimation of wind speed distributions. Appl Energy 2012;89:355–67.
[23] Fouskakis D, Draper D. Stochastic optimization: a review. Int Stat Rev 2002;70:315e49.
[24] Wienke D, Lucasius C, Kateman G. Multicriteria target vector optimization of analytical procedures using a genetic algorithm part I. Theory, numerical simulations and application to atomic emission spectroscopy. Anal Chim Acta 1992;265:211–25.
[25] Lembo E, Rosito C. Misure di velocità e direzione del vento da stazioni anemometriche. CESI Research institute; 2008.
[26] Gibescu M, Ummels BC, Kling WL. Statistical wind speed interpolation for simulating aggregated wind energy production under system studies. In: 9th International conference on probabilistic methods applied to power systems KTH, Stockholm, Sweden; June 11–15 2006.
[27] Chang TP. Performance comparison of six numerical methods in estimating Weibull parameters for wind energy application. Appl Energy 2011;88:272–82.
[28] Firtin E, Güler O, Akdag SA. Investigation of wind shear coefficients and their effect on electrical energy generation. Appl Energy 2011;88:4097–105.