Data-based continuous wind speed models with arbitrary probability distribution and autocorrelation

Guðrún Margrét Jónsdóttir and Federico Milano

School of Electrical & Electronic Engineering, University College Dublin, Ireland

Abstract

The paper presents a systematic method to build dynamic stochastic models from wind speed measurement data. The resulting models fit any probability distribution and any autocorrelation that can be approximated through a weighted sum of decaying exponential and/or damped sinusoidal functions. The proposed method is tested by means of real-world wind speed measurement data with sampling rates ranging from seconds to hours. The statistical properties of the wind speed time series and the synthetic stochastic processes generated with the Stochastic Differential Equation (SDE)-based models are compared. Results indicate that the proposed method is simple to implement, robust and can accurately capture simultaneously the autocorrelation and probability distribution of wind speed measurement data.

Key words: Stochastic differential equations, wind speed modeling, memoryless transformation, probability distribution, autocorrelation.

1 Introduction

1.1 Motivation

Wind power is the fastest growing among renewable energy sources worldwide [1]. For example, in Ireland the instantaneous wind generation can be up to 65% of the total demand and the system operator is planning to increase this limit to 75% by 2020. However, this growth comes with drawbacks. The power generated by a wind turbine depends on the weather conditions which makes it a highly volatile power source. In order to ensure a reliable and secure operation of the grid, it is essential to model the source of such a volatility, i.e., the wind speed. Existing literature does not provide a general method to synthesize continuous dynamic wind speed models that are adequate for the transient stability analysis of power systems. This research aims to address this issue and proposes a systematic method to develop continuous-time wind speed models that precisely reproduce both the autocorrelation and the probability distribution of actual measurement data.

1.2 Literature Review

Traditionally, wind speed is modeled as a stationary stochastic process. This process is characterized by its Probability Density Function (PDF) and Autocorrelation Function (ACF). The PDF describes all the possible values and likelihoods that the wind speed can take within a given range. The ACF is a measure of how the wind speed changes over time. That is, the ACF gives a measure of the relationship between the current wind speed value and past wind speed values. In some cases, e.g., steady-state analysis [2], it is enough to only consider the PDF. However, when a realistic synthetic wind speed time-series is required, the model must also capture the ACF.

A wide range of wind speed models are available in the literature. Most of the literature has focused on the forecasting of wind speed and the modeling of the forecast error. Discrete Autoregressive Moving Average (ARMA) models are the most common type found in the literature. These models have been widely used for wind speed forecast error modeling and short-term forecasting [3–8]. They are well-established and offer comprehensive tools to fit the model to data and reproduce both the PDF and ACF of wind speed data. Markov chains of first order and higher have also been widely used to model wind speed [9,10]. Additionally, physical models that use meteorological information have been widely used to predict the long-term wind speed [3].

Both ARMA models and Markov chains are discrete and have a fixed time step that must match the sampling interval of the available data. This constraint prevents using such models for transient stability analysis, for which continuous wind speed models are required. While methods to define a continuous-time equivalent of ARMA models have been proposed in the literature [11,12], the equivalencing procedure introduces numerical approximations and requires an involved modeling procedure.

In recent years, the use of SDEs for wind speed modeling has gained popularity. SDEs appear more suitable than classical discrete time series approaches as they are intrinsically continuous with respect to time and thus, they can better reproduce the transient fluctuations of the wind speed. More importantly, SDEs are not constrained to use the sampling time step of the original measurement data. Finally, power systems are typically formalized as a set of differential algebraic equations which allow readily incorporating SDEs.

In the literature, SDEs have been used for wind speed forecasting [13,14]; modeling the volatility of wind power generation [15–17]; and power system stability analysis [12, 18]. However, the SDE-based wind speed models that have been presented so far in the literature fail to capture either the PDF, the ACF, or both [16–19]. In [20–22], SDEs with an arbitrary PDF are presented but are limited to strictly exponentially decaying ACF.

1.3 Contributions

This paper proposes using SDE-based models with ACFs that are a weighted sum of decaying exponential and/or damped sinusoidal functions. This allows the fitting of a wider range of ACFs than what can be achieved by the techniques that are currently available in the literature. The proposed method is based on the superposition of stochastic processes that capture the desired ACF as presented in [23]. Then, to impose the desired PDF, a memoryless transformation is used as discussed in [24]. Either analytical PDFs that are fitted to the data or numerically estimated PDFs can be accommodated. The proposed method is both simple and flexible as it allows defining a SDE-based model from virtually any set of wind speed measurements.

The contributions of this paper are:

- A novel method to build SDE-based wind speed models in continuous-time that captures both the ACF and the PDF of measurement data. The proposed method can capture a wider range of measured wind speed ACFs than the SDE-based models proposed in the literature.
- The method can capture ACFs that can be modeled as a weighted sum of decaying exponentials and/or damped sinusoidal functions which in the authors experience can be used to model all wind speed data sets.
- The method can model any PDF as a numerical PDF is fitted to the probability distribution of the data set used. This makes it unnecessary to define an analytical PDF. However, the method can also be used to model any analytical PDF.
- The method is tested by building wind speed models based on seven measurement data sets with different sampling frequencies. It is shown that the method can capture the ACF and PDF of these data sets.

1.4 Organization

The remainder of this paper is organized as follows. Section 2 briefly outlines relevant definitions and concepts of SDEs and Section 3 presents the theoretical foundation of the proposed method to synthesize wind speed models. Section 4 shows how the parameters of the SDE models are identified from wind speed data sets. Section 5 discusses the generation of wind speed trajectories and illustrates their statistical properties through numerical simulations. Finally, Section 6 draws conclusions and outlines future work.

2 Outlines of Stochastic Differential Equations

Stochastic Differential Equations (SDEs) are a prominent mathematical modeling technique employed in areas such as finance for modeling stock prices or interest rates and physics to model particles in fluids.

A generic one-dimensional SDE has the form:

$$dX(t) = a(t, X(t))dt + b(t, X(t))dW(t), \quad X(t_0) = X_0, \tag{1}$$

where $a(t, X(t))$ and $b(t, X(t))$ are continuous functions and are referred to as the drift and diffusion term of the SDE, respectively. $W(t)$ represents the stochastic component driving the SDE. Commonly, this component is a Wiener process, $\{W(t), t > 0\}$, which is a random function characterized by the following properties:

(1) $W(0) = 0$, with probability 1.
(2) The function $t \mapsto W(t)$ is continuous in t.
(3) If $t_1 \neq t_2$, then $W(t_1)$ and $W(t_2)$ are independent.
(4) For $\forall t_i \geq 0$, all increments, $\Delta W_i = W(t_{i+1}) - W(t_i)$, are normally distributed, with mean 0 and variance $h = t_{i+1} - t_i$, i.e., $\Delta W_i \sim \mathcal{N}(0, h)$.

Wiener processes cannot be integrated in the conventional Riemann-Stieltjes sense as they are not bounded, i.e., the limit $\lim_{x\to 0}(W(t + \Delta t) - W(t))/\Delta t$ does not exist. A specific stochastic integral has to be defined to solve the SDE in (1). There are several different ways to interpret stochastic integrals. In this paper, the most widely used approach is used, namely, the Itô integral.

An in-depth discussion on SDEs is outside the scope of this paper. The interested reader is referred to [25–27] for details on SDEs theory and numerical methods.

2.1 2-dimensional Ornstein-Uhlenbeck Process

The Ornstein-Uhlenbeck (OU) process is a particular stochastic process with a Gaussian PDF which exhibits mean reversion, i.e., it drifts towards its mean value at an exponential rate. Moreover, the OU process has a bounded variance which makes it suitable to model physical processes such as wind fluctuations [28].

The following 2-dimensional OU is utilized as the building block of the proposed method to synthesize wind speed models:

$$\begin{pmatrix} dX(t) \\ dY(t) \end{pmatrix} = \begin{pmatrix} -\alpha & -\omega \\ \omega & -\alpha \end{pmatrix} \begin{pmatrix} X(t) \\ Y(t) \end{pmatrix} dt + \begin{pmatrix} \sigma \\ 0 \end{pmatrix} dW(t). \tag{2}$$

where $\alpha > 0$, $\sigma > 0$, $\omega \geq 0$ and $W(t)$ is a standard Wiener process. The correlation matrix of the SDE in (2) is:

$$\boldsymbol{R}(\tau) = \mathbb{E}\begin{pmatrix} X(t+\tau) \\ Y(t+\tau) \end{pmatrix}\big(X(t), Y(t)\big) = \exp(-\alpha\tau)\begin{pmatrix} \cos(\omega\tau) & -\sin(\omega\tau) \\ \sin(\omega\tau) & \cos(\omega\tau) \end{pmatrix}. \tag{3}$$

The process $X(t)$ is used throughout this paper for building the wind speed models. Its ACF is

$$R_X(\tau) = \exp(-\alpha\tau)\cos(\omega\tau), \tag{4}$$

In stationary conditions, $X(t)$ is Gaussian distributed with zero mean and variance $\sigma^2/(2\alpha)$.

For $\omega = 0$, $X(t)$ and $Y(t)$ are decoupled and $X(t)$ becomes a conventional 1-dimensional Ornstein-Uhlenbeck (OU) process:

$$dX(t) = -\alpha X(t)dt + \sigma dW(t), \tag{5}$$

with exponentially decaying autocorrelation:

$$R_X(\tau) = \exp(-\alpha\tau). \tag{6}$$

3 Proposed Method to Synthesize SDEs

The main idea behind the proposed approach is to use a summation of a set of SDEs of the form of (2). The resulting compound stochastic process is build in such a way that it has the same Probability Density Function (PDF) and Autocorrelation Function (ACF) as the given wind speed data. It is shown below that the proposed method can capture both a wide range of ACFs and PDFs, even those that do not fit any analytical PDF function.

The steps to build the desired compound stochastic process are twofold: (i) a superposition of SDEs that captures the desired ACF is defined (see Subsection 3.1); and (ii) an analytical or numerical memoryless transformation that imposes the desired PDF is applied to the SDEs determined in the previous step (see Subsection 3.2). This is shown in Fig. 1.

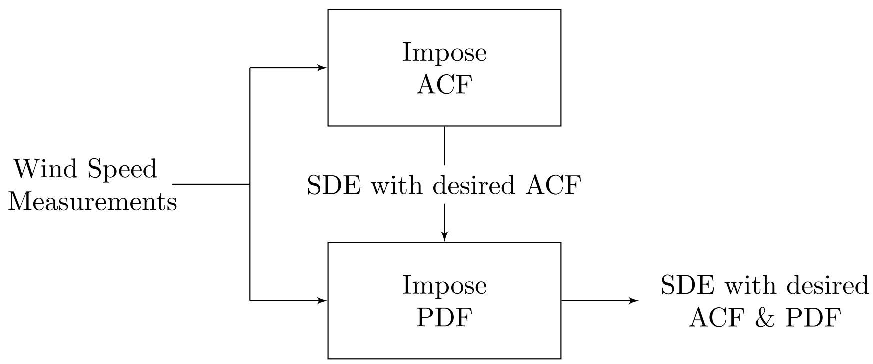


Fig. 1. The proposed method to capture the ACF and PDF of measurement data.

3.1 Impose the Autocorrelation

In [23], a superposition of OU processes is used for imposing an ACF represented with a weighted sum of Exponentially Decaying Functions (EDFs). We generalize the technique in [23] with the 2-dimensional OU in (2) instead of the conventional OU process. In this way, the proposed model is able to reproduce not only exponentially decaying ACFs but also periodical behaviors, such as those due to daily effects.

Let $Z(t)$ be a stochastic process obtained as the weighted sum of n SDE processes, as follows:

$$Z(t) = \sum_{i=1}^{n} \sqrt{w_i}\, X_i(t), \tag{7}$$

where $X_i(t)$, $i = 1, \dots n$, are SDE processes with ACFs $R_{X_i}(\tau)$, $w_i > 0$ and

$$\sum_{i=1}^{n} w_i = 1. \tag{8}$$

If all n processes have an identical Gaussian PDF $\mathcal{N}(\mu_X, \sigma_X)$, the stochastic process $Z(t)$ has the same Gaussian PDF, $\mathcal{N}(\mu_X, \sigma_X)$, and an ACF which is a weighted sum of the ACFs of the n SDE processes, that is:

$$R_Z(\tau) = \sum_{i=1}^{n} w_i R_{X_i}(\tau). \tag{9}$$

If the n SDE processes in (7) are $X(t)$ processes as in (2), the resulting ACF of $Z(t)$ is a weighted sum of damped sinusoidal and decaying exponential functions and (9) can be rewritten as:

$$R_Z(\tau) = \sum_{i=1}^{n} w_i \exp(-\alpha_i \tau) \cos(\omega_i \tau). \tag{10}$$

Hence, the superposition of SDE processes allows capturing any ACF that can be modeled as a weighted sum of exponential and/or sinusoidal ACFs. If the ACF does not show a periodic behavior than $\omega_i = 0$, $\forall i = 1, \dots, n$. In the authors experience, (10) is general enough to reproduce the autocorrelation of all wind speed time series. This conclusion is drawn based on all the wind speed measurements the authors have had access to and been able to analyse.

3.2 Impose the Probability Distribution

The sum of the SDE processes in (7) resulting in the compound process $Z(t)$ has a Gaussian probability distribution. However, the PDF of the wind speed is typically not Gaussian. To impose the target PDF of the wind speed, the memoryless transformation is used that consists of transforming a standard Gaussian stochastic process into another one with the desired distribution while retaining the ACF of the original process. This is achieved by applying

Table 1
Description of the wind speed data sets.

#	**Sampling rate**	**Averaged**	**Duration**	**Location**
1	1 hour	Yes	3 years	Mace Head, Galway, Ireland
2	1 hour	Yes	3 years	Malin Head, Donegal, Ireland
3	10 minute	Yes	1 year	Ashburnham, Massachussetts
4	10 minute	Yes	1 year	Orleans, Massachussetts
5	1 minute	Yes	1 month	Johnstown, Wexford, Ireland
6	1 minute	Yes	1 month	Oak Park, Carlow, Ireland
7	1 second	No	1 month	Tracy, California

the Gaussian Cumulative Distribution Function (CDF) to the inverse of the target CDF, as follows:

$$Z_F(t) = F^{-1}(\Phi(Z(t)), \tag{11}$$

where $F^{-1}(\cdot)$ is the inverse CDF of the desired process and $\Phi(\cdot)$ is the CDF of the Gaussian distribution [24]. The resulting process is the target SDE with the desired PDF and ACF.

A relevant advantage of this approach based on the memoryless transformation, is that it can be used with any analytical or a numerical PDF that has been fitted to the probability distribution of the data. A thorough discussion on this point is provided in the next section.

4 Fitting Procedure

This section outlines the procedure used to identify the parameters of wind speed models from data using the method presented in Section 3. The parameters are identified from the statistical properties of the equivalent wind speed data sets shown in Table 1. The wind speed data sets have sampling rates ranging from 1 second to 1 hour and are collected in locations in Ireland and USA. Further details on the measued wind speed data can be found in [29–31] as all data sets used are available open source.

The PDF and ACF that best describe the wind speed variability depend on the location and the time frame [32–38]. Commonly, wind speed has a PDF that is shifted to the left and tails heavily to the right. This is because, in most areas, strong extreme winds are rare, while moderate winds are quite common. The most widely used PDF for wind speed is the two-parameter Weibull distribution. However, a wide range of distributions have been proposed in the literature to fit the PDF of the wind speed at a specific location [32–34].

The ACF of wind speed is characterized as an Exponentially Decaying Function (EDF) over the first 12 hours for hourly averaged data and then

Table 2
The chi-squared test results for fitting a weighted sum of 1, 2 and 3 EDFs to the ACF of Data Set 1.

EDF #	**p-value**
1	563.05113902
2	0.1422387958
3	0.1422342338

settles to zero or a value bigger than zero or shows damped periodic fluctuations due to daily variations [20, 37, 38]. However, if the data is sampled more frequently and/or not averaged, fast wind speed variations change the shape of the ACF. Such short-term wind speed variations, e.g., turbulence and gusts, typically occur within a 10 minute time frame and result in the ACF initially decreasing rapidly before settling to the same slope as the hourly data [11, 12]. This kind of autocorrelation can be well described as a weighted sum of decaying exponential functions.

The fitting procedure involves fitting the function in (10) to the ACF of the data as well as identifying a PDF that best captures the probability distribution of the data. Two methods to define the PDF are presented: (1) fit an analytical PDF to the data; and (2) find a numerical estimated PDF. The procedure is demonstrated below with Data Set 1. The same procedure is used to identify the parameters of the wind speed models for all the data sets of Table 1.

4.1 Find the ACF parameters

Figure 2 shows the autocorrelation for Data Set 1. To capture this ACF using the method presented in Section 3, (10) has to be fitted to the autocorrelation. This can be done with any typical curve fitting algorithm. In this work, a non-linear least squares method, included in the Python package SciPy [39], is utilized. The number of decaying exponential and/or damped sinusoidal functions used to fit the ACF can most often be estimated visually or, if not, by trial and error. In this case, three EDFs are considered. In order to determine the ideal number of EDFs the chi-squared test is used and the results are presented in Table 2. The chi-square results indicate that it is sufficient to use two components as adding the third component does not significantly improve the fit.

In Figs. 2-3, the fit of the 3-component models are compared to the ACF of Data Set 1. Component 1 EDF only manages to capture the decay over the first 24 hours. To capture the correlation after that more components are needed. In Fig. 2, the 2 and 3 component EDF processes are indistinguishable. To further examine the fit, the difference in the fitted processes from the actual ACF of Data Set 1 is shown in Fig. 3. Component 2 performs slightly worse but the error for both components 2 and 3 is less than 3 %. The difference between the 2- and 3-component models is minimal in this case. Therefore, the simpler model is chosen, i.e., the 2-component model.

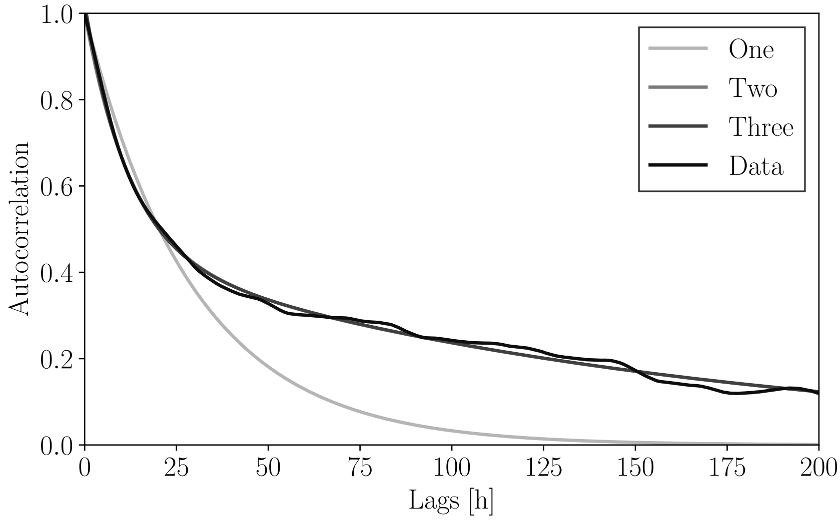


Fig. 2. The ACF of Data Set 1 and the fitted sum of $1-3$ component EDFs.

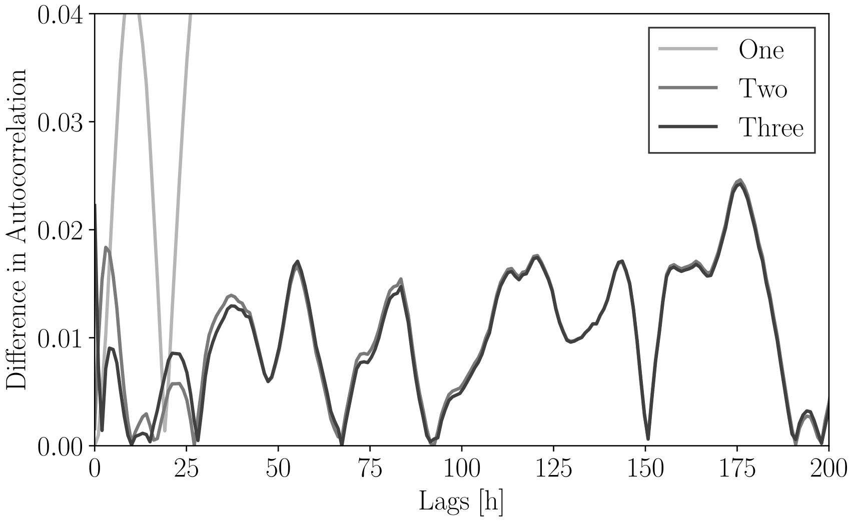


Fig. 3. The difference in the ACF of Data Set 1 and the fitted sum of $1-3$ component EDFs.

The fitted ACF function of Data Set 1 is thus approximated as:

$$R_1(\tau) = w_1\exp(-\alpha_1\tau) + w_2\exp(-\alpha_2\tau), \tag{12}$$

where $w_1 = 0.55$, $w_2 = 0.45$, $\alpha_1 = 0.0811$ and $\alpha_2 = 0.00648$. Since no periodical behavior is present in this data set, $\omega_1 = \omega_2 = 0$. In Table 4 the parameters for the fitted ACFs of the remaining datasets are shown. In the table, it is implied that $\omega_i = 0$ if not provided.

The σ_1 and σ_2 parameters are defined from the α_1 and α_2 parameters respectively and the set standard deviation, σ_X of the two OU processes. The standard deviation can be set to any value as long as this value is used to define the Gaussian CDF $\Phi(\cdot)$ in (11). In this case σ_X is set to be 1 and thus, $\sigma_1 = \sqrt{2\sigma_X^2\alpha_1} = 0.4027$ and $\sigma_2 = \sqrt{2\sigma_X^2\alpha_2} = 0.1138$.

4.2 Find the PDF parameters

Figure 4 shows the probability distribution of Data Set 1. The PDF of the wind speed is imposed using a memoryless transformation as discussed in Section 3. The PDF can be defined in two ways, analytically or numerically. The latter approach is to be preferred if the wind speed distribution is irregular or has two peaks.

For the sake of illustration, both analytical and numerical fitting of the PDF for Data Set 1 are discussed below.

4.2.1 Analytical PDF

Table 3 shows the results of the Kolmogorov-Smirnov tests for six PDFs applied to Data Set 1. The Kolmogorov-Smirnov test is used to compare the analytical Cumulative Distribution Function (CDF) to the Empirical Cumulative Distribution Function (ECDF) of the data and allows deciding which analytical PDF best captures the probability distribution of the data set. The six PDFs considered in Table 3 have all been used in the literature to model the probability distribution of wind speed data. These PDFs are given in Appendix A.

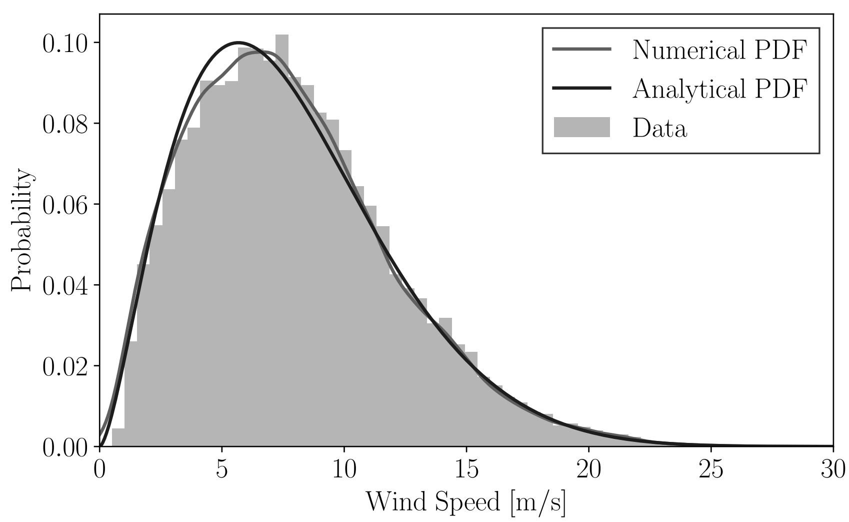


Fig. 4. The histogram for Data Set 1 and the fitted numerical and analytical PDF.

Table 3
The Kolmogorov-Smirnov test results for 6 analytical PDFs fitted to Data Set 1.

PDF	**p-value**
3-parameter Beta	$1.0688 \cdot 10^{-13}$
3-parameter Gamma	$1.3787 \cdot 10^{-77}$
2-parameter Inverse Gaussian	$3.5201 \cdot 10^{-166}$
2-parameter Lognormal	No fit
1-parameter Rayleigh	$4.7572 \cdot 10^{-14}$
2-parameter Weibull	$1.0849 \cdot 10^{-84}$

Based on Table 3, the 3-parameter Beta PDF has the highest p-value and thus, provides the best fit. The 3-parameter Beta function is defined as:

$$p_{\mathrm{B}}(x) = \begin{cases} \dfrac{1}{\lambda_3 \, B\left[\lambda_1, \lambda_2\right]} \left(\dfrac{x}{\lambda_3}\right)^{\lambda_1 - 1} \left(\dfrac{\lambda_3 - x}{\lambda_3}\right)^{\lambda_2 - 1} & \text{if } x > 0 \\ 0 & \text{if } x \leq 0 \end{cases} \tag{13}$$

where $B[\cdot, \cdot]$ is the Beta function, the shape parameters are $\lambda_1 = 2.6006$ and $\lambda_2 = 10.1357$ and the noncentrality parameter is $\lambda_3 = 38.0838$, for Data Set 1. The fitted analytical PDF is shown in Fig. 4.

With a similar procedure, the parameters of the best fitting PDF of all data sets in Table 1 can be determined. The results of such a fitting are shown in Table 4.

The PDFs in Table 3 are used to demonstrate that the method can be used with analytically defined PDFs. The authors are aware that more complex analytical PDFs have been proposed in the literature to model the probability distribution of wind speed, such as mixed distributions, Kappa etc. This method can be used to capture all more complex analytical PDFs if the equivalent inverse CDF can be defined. If the inverse CDF can not be defined analytically, it can be defined numerically using the method presented in Section 4.2.2.

4.2.2 Numerical PDF

Among the many possible numerical techniques to approximate the probability distribution of a set of measurements, the Empirical Cumulative Distribution Function (ECDF) is considered here. It is a non-parametric estimator of the underlying Cumulative Distribution Function (CDF) of a stochastic process [40]. The ECDF is found by first sorting the N data points from smallest to largest. Each data point is assigned a probability of $1/N$. The result is a step function that increases by $1/N$ for each data point. The ECDF is a discrete numerical approximation of the CDF of the data set. An estimation of the underlying continuous CDF function can be found from the ECDF by using interpolation.

Table 4
The ACF and PDF parameters of Data Sets $1-7$.

Data Set	ACF Parameters		PDF type & parameters
1	$w_1 = 0.62$		3-parameter Beta PDF
	$w_2 = 0.38$		$\lambda_1 = 2.601$
	$\alpha_1 = -0.0713$		$\lambda_2 = 10.136$
	$\alpha_2 = -0.0045$		$\lambda_3 = 38.084$
2	$w_1 = 0.72$	$\alpha_1 = -0.0961$	1-parameter Rayleigh PDF
	$w_2 = 0.24$	$\alpha_2 = -0.0109$	$\lambda_1 = 6.065$
	$w_3 = 0.04$	$\alpha_3 = 6.33 \cdot 10^{-4}$	
		$\omega_3 = 0.26$	
3	$w_1 = 0.05$		3-parameter Gamma PDF
	$w_2 = 0.95$		$\lambda_1 = 1.423$
	$\alpha_1 = -0.5592$		$\lambda_2 = 4.659$
	$\alpha_2 = -0.0096$		$\lambda_3 = 1.826$
4	$w_1 = 0.04$		3-parameter Gamma PDF
	$w_2 = 0.96$		$\lambda_1 = 1.1511$
	$\alpha_1 = -0.9232$		$\lambda_2 = 5.9695$
	$\alpha_2 = -0.0119$		$\lambda_3 = 2.2975$
5	$w_1 = 0.06$		3-parameter Beta PDF
	$w_2 = 0.94$		$\lambda_1 = 2.2422$
	$\alpha_1 = -0.9800$		$\lambda_2 = 4.6258$
	$\alpha_2 = -0.0014$		$\lambda_3 = 17.3688$
6	$w_1 = 0.045$	$\alpha_1 = -0.9800$	3-parameter Gamma PDF
	$w_2 = 0.925$	$\alpha_2 = -0.0015$	$\lambda_1 = 1.0155$
	$w_3 = 0.03$	$\alpha_3 = -0.1066$	$\lambda_2 = 5.2482$
			$\lambda_3 = 1.6766$
7	$w_1 = 0.035$	$\alpha_1 = -0.0242$	3-parameter Beta PDF
	$w_2 = 0.045$	$\alpha_2 = -0.1362$	$\lambda_1 = 2.1381$
	$w_3 = 0.830$	$\alpha_3 = -4.3 \cdot 10^{-5}$	$\lambda_2 = 8.5281$
	$w_4 = 0.084$	$\alpha_4 = -0.0841$	$\lambda_3 = 31.569$

The equivalent inverse CDF for the fitted numerical and/or analytical PDF is then used for imposing the probability distribution of the data set using the memoryless transformation, as discussed in Section 3. Figure 4 shows the PDF defined through the ECDF for Data Set 1. Similar results can be obtained for the other data sets in Table 1.

5 Simulation Results

In this section, the modeling method outlined in Section 3 is coupled with the data-fitted parameters from Section 4. Combined together, they make wind speed models for the data sets described in Table 1. These models are used for generating synthetic wind speed trajectories whose statistical properties accurately reproduce those of the actual wind speed data sets.

With this aim, (7) needs to be integrated and subsequently the transformation in (11) applied. To solve the SDE the Euler-Maruyama integration method is used. Other integration methods for SDEs can be found in [26] but, given the accuracy of the results discussed below, the Euler-Maruyama scheme works well and no higher order method is deemed to be required.

Synthetic models are simulated to produce N data points with the time step h (values for N and h shown in Table B.1). To illustrate the ability of the developed models to reproduce the statistical properties of the original data, the PDF and ACF of the synthetic processes are compared with those of the data sets. First the results of such comparisons for Data Set 1 are discussed in detail and then an overview is provided of the results for Data Sets 2 to 7.

The PDF and ACF results for Data Set 1 are shown in Figs. 5 and 6, respectively. In Fig. 5, the histogram of Data Set 1 (in gray) is compared to the results for the generated process using the analytical PDF (solid line) and numerical PDF (dashed line). Both the analytically and numerically defined PDF are well captured by the SDE-based models. In this case the analytical PDF is likely the best option.

Figure 7 compares analytical, numerical and data-based PDFs for the Data Sets 2 to 7. For Data Set 6, the analytical PDF gives a good fit to the probability distribution of the measurements. For the remaining PDFs, however, that is not the case. For Data Sets 2-5 and 7, in fact, the top of the PDF is uneven, i.e., it is wider or narrower than what can be reproduced through the analytical PDFs. In these cases, the numerically defined PDFs provide better approximations. These PDFs can, in principle, also be approximated through a combination of analytical PDFs, e.g., a superposition of Gaussian distributions. However, the numerical approach is simple, general and yet very accurate.

Figure 6 shows the ACF of Data Set 1, the fitted theoretical function and the ACF of the wind speed trajectory generated by the proposed SDE-based model. The latter consists of the weighted sum of two decaying EDFs and captures well the ACF of the data. It is important to note that this ACF

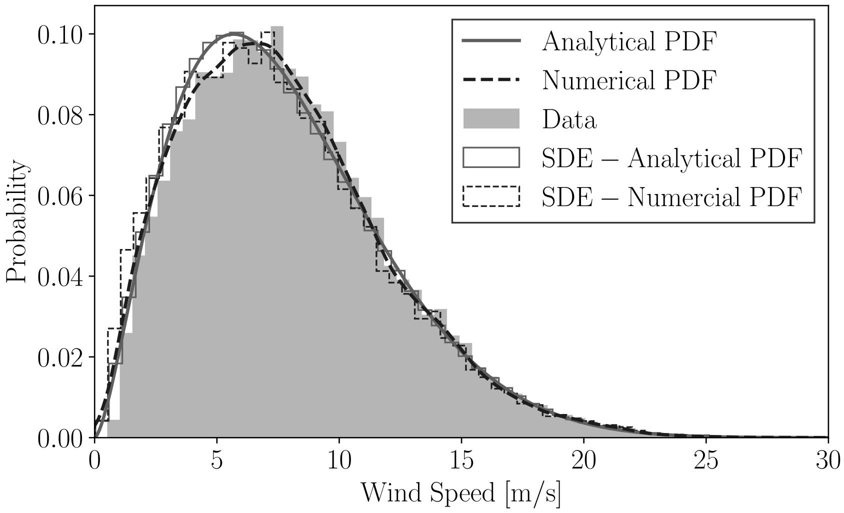


Fig. 5. Histogram of the data, fitted analytical and numerical PDFs and histogram of the simulated SDE with the analytical and numerical PDF for Data Set 1.

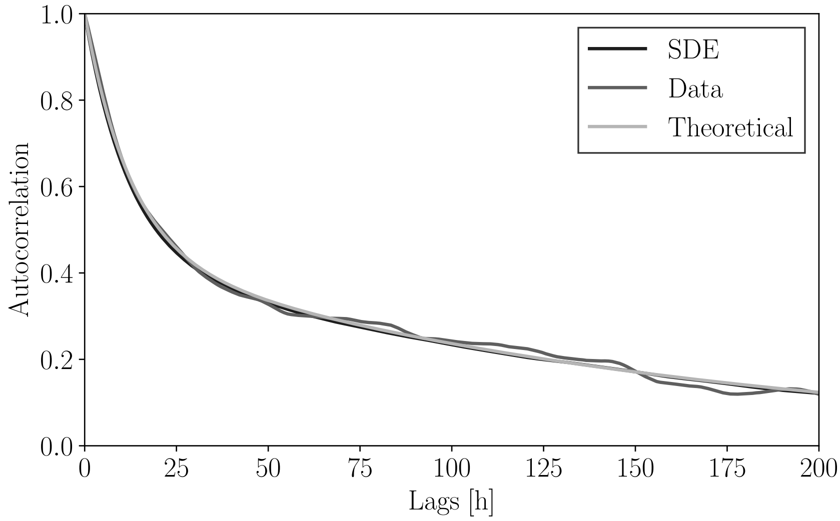
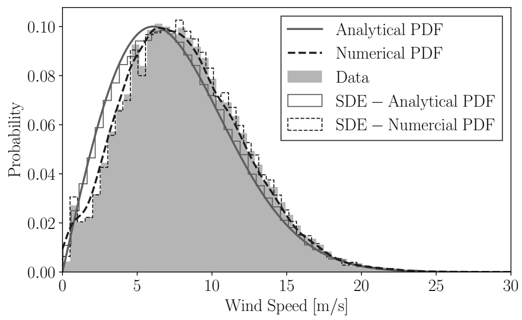


Fig. 6. ACF of the data, the fitted theoretical ACF and the ACF of the simulated SDE model for Data Set 1.

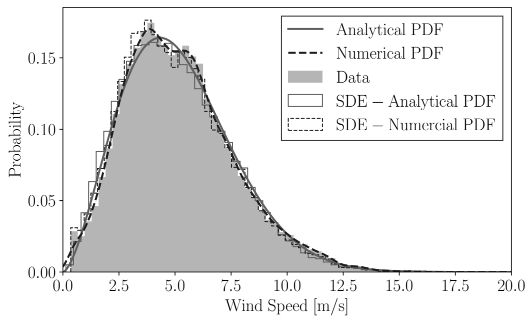
cannot be captured with the continuous-time wind speed models currently available in the literature. Such models can only model hourly averaged data with a single exponentially decaying ACF [20–22].

Figure 8 compares analytical and data-based ACFs for Data Sets 2 to 7. Results clearly show that the ACF is highly dependent on the sampling rate as the ACFs have different shapes for different sampling rates. For example, Data Sets 5 and 6 show two different sections, one in the time lags that ranges from 0 to 5 minutes, and another one for time lags larger than 5 minutes.

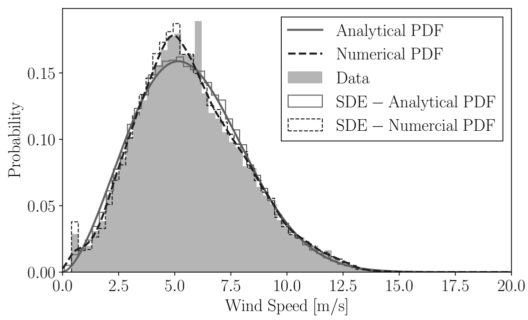
The autocorrelation of Data Set 2 shows a poorly damped sinusoidal mode with a period of 24 hours. The proposed superposition approach is able to reproduce all these different shapes of wind speed ACFs as a weighted sum of exponential and sinusoidal function.



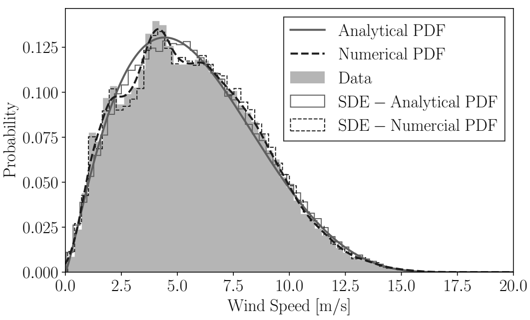
(a) Data Set 2



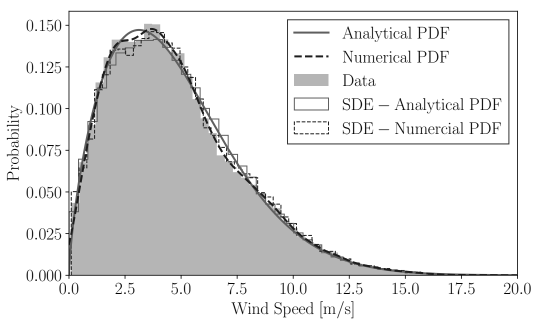
(b) Data Set 3



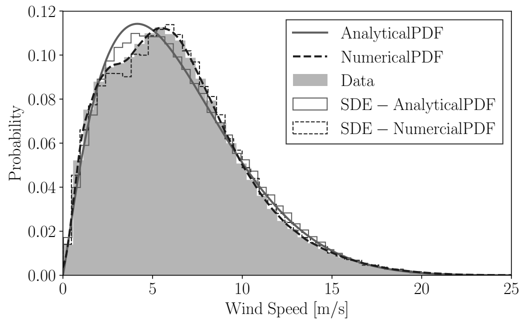
(c) Data Set 4



(d) Data Set 5

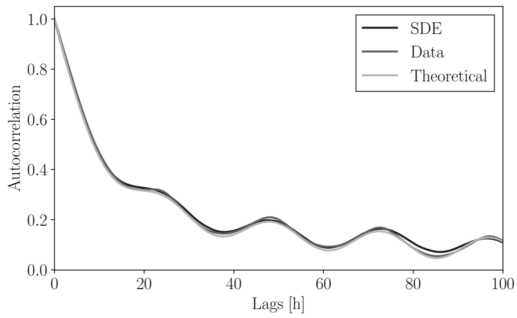


(e) Data Set 6

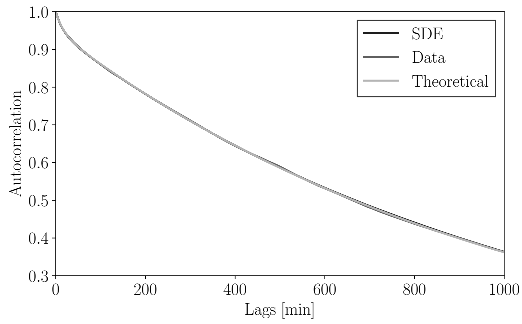


(f) Data Set 7

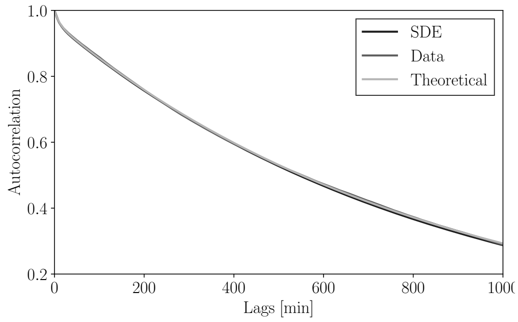
Fig. 7. The histograms of the data, fitted analytical and numerical PDFs and histograms of the simulated SDE with the analytical and numerical PDF for Data Sets 2-7.



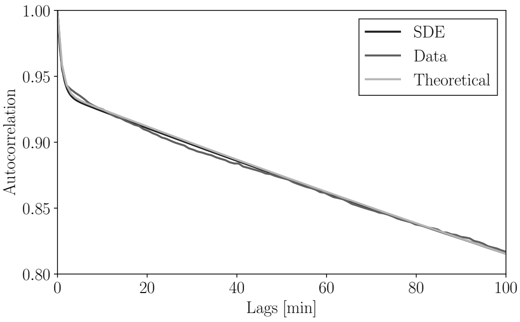
(a) Data Set 2



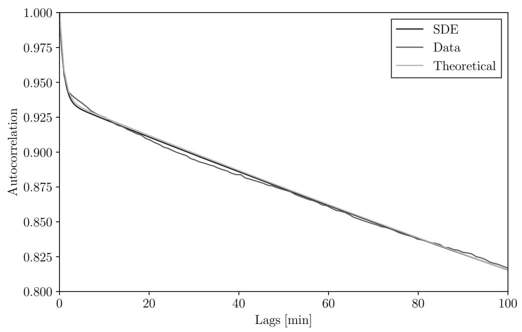
(b) Data Set 3



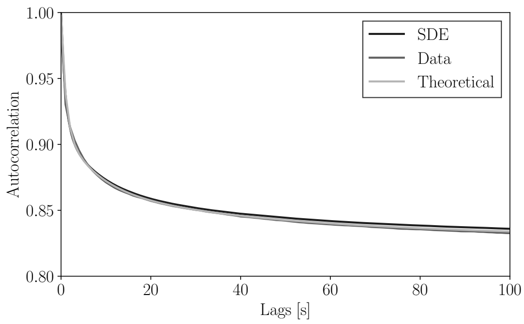
(c) Data Set 4



(d) Data Set 5



(e) Data Set 6



(f) Data Set 7

Fig. 8. The ACFs of the data, the fitted theoretical ACFs and the ACFs of the simulated SDE model for Data Sets 2-7.

6 Conclusions

This paper presents a method to construct wind speed models based on SDEs. The method consists of two steps. In the first step the superposition of SDE processes is used to capture the desired ACF. In the second step the memoryless transformation is used to capture an arbitrary PDF. The resulting process is a continuous-time wind speed time-series which captures the desired ACF and PDF simultaneously.

To demonstrate the flexibility of the proposed method, seven wind speed data sets with sampling rates from seconds to hours are modeled. The method is shown to accurately reproduce PDFs of any shape and ACFs that can be described as weighted sums of decaying exponential and/or damped sinusoidal functions. The comparison of simulation results with the real-world wind speed data sets shows that the proposed method is accurate and robust.

Future work will focus on extending the work presented in this paper to model other types of stochastic processes e.g., solar irradiance, in continuous-time based on measurement data. We are currently working on applications of the proposed method to model the forecast error of stochastic processes. Future work will also focus on extending the proposed method to build non-stationary stochastic processes.

Acknowledgments

This work is supported by the Science Foundation Ireland, by funding Federico Milano, under Investigator Programme, Grant No. SFI/15/IA/3074. The opinions findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the Science Foundation Ireland.

A Probability Density Functions

This appendix presents PDFs that have been proposed in the literature to model the PDF of wind speed data.

A.1 Three-parameter Beta distribution

The probability density function of the three-parameter Beta distribution, $p_{\mathrm{B}}(x)$, is

$$p_{\mathrm{B}}(x) = \begin{cases} \dfrac{1}{\lambda_3 \, B\left[\lambda_1, \lambda_2\right]} \left(\dfrac{x}{\lambda_3}\right)^{\lambda_1 - 1} \left(\dfrac{\lambda_3 - x}{\lambda_3}\right)^{\lambda_2 - 1} & \text{if } x > 0 \\ 0 & \text{if } x \leq 0 \end{cases}$$

where $B[\cdot, \cdot]$ is the Beta function, λ_1 and λ_2 are shape parameters, and λ_3 is a noncentrality parameter.

A.2 Three-parameter Generalized Gamma distribution

The probability density function of the three-parameter Generalized Gamma distribution, $p_{\mathrm{GG}}(x)$, is

$$p_{\mathrm{GG}}(x) = \begin{cases} \dfrac{1}{\lambda_2\,\Gamma\left[\lambda_1\right]}\,\lambda_3\,\left(\dfrac{x}{\lambda_2}\right)^{\lambda_1\,\lambda_3-1}\,\exp\left[-\left(\dfrac{x}{\lambda_2}\right)^{\lambda_3}\right] & \text{if } x>0 \\ 0 & \text{if } x\leq 0 \end{cases}$$

where $\Gamma[\cdot]$ is the Gamma function, λ_1 and λ_3 are shape parameters, and λ_2 is a scale parameter.

A.3 Two-parameter Inverse Gaussian distribution

The probability density function of the two-parameter Inverse Gaussian distribution, $p_{\mathrm{IG}}(x)$, is

$$p_{\mathrm{IG}}(x) = \begin{cases} \dfrac{1}{\sqrt{2\,\pi}}\,\sqrt{\dfrac{\lambda}{x^3}}\,\exp\left[-\dfrac{\lambda\,(x-m)^2}{2\,m^2\,x}\right] & \text{if } x>0 \\ 0 & \text{if } x\leq 0 \end{cases}$$

where m is the mean, and λ is a scale parameter.

A.4 Two-parameter Lognormal distribution

The probability density function of the two-parameter Lognormal distribution, $p_{\mathrm{LN}}(x)$, is

$$p_{\mathrm{LN}}(x) = \begin{cases} \dfrac{1}{\sqrt{2\,\pi}\,s\,x}\,\exp\left[-\dfrac{(\log\left[x\right]-m)^2}{2\,s^2}\right] & \text{if } x>0 \\ 0 & \text{if } x\leq 0 \end{cases}$$

where m and s are the mean and the standard deviation of the natural logarithm of variable x, respectively.

A.5 One-parameter Rayleigh distribution

The probability density function of the one-parameter Rayleigh distribution, $p_{\mathrm{R}}(x)$, is

$$p_{\mathrm{R}}(x) = \begin{cases} \dfrac{x}{\lambda^2}\,\exp\left[-\dfrac{x^2}{2\,\lambda^2}\right] & \text{if } x>0 \\ 0 & \text{if } x\leq 0 \end{cases}$$

where λ is a scale parameter.

A.6 Two-parameter Weibull distribution

The probability density function of the two-parameter Weibull distribution, $p_{\mathrm{W}}(x)$, is

$$p_{\mathrm{W}}(x) = \begin{cases} \dfrac{\lambda_1}{\lambda_2}\left(\dfrac{x}{\lambda_2}\right)^{\lambda_1-1} \exp\left[-\left(\dfrac{x}{\lambda_2}\right)^{\lambda_1}\right] & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}$$

where λ_1 is a shape parameter and λ_2 is a scale parameter.

B Simulation parameters

The simulation parameters used for the data sets in Table 1 are shown in Table B.1. N is the number of data points and h is the time step.

Table B.1
Simulation parameters for each data set.

Data Set	N	h
1	$1 \cdot 10^6$	1
2	$1 \cdot 10^5$	0.01
3	$5 \cdot 10^7$	0.1
4	$5 \cdot 10^7$	0.1
5	$5 \cdot 10^7$	0.1
6	$5 \cdot 10^7$	0.1
7	$1 \cdot 10^8$	0.1

References

[1] World Wind Energy Association (WWEA).Wind Power Capacity reaches 539 GW, 52.6 GW added in 2017 (Press Releases), http://www.wwindea.org/2017-statistics/ [accessed 10 July 2018].

[2] D. Villanueva, J. L. Pazos, A. Feijoo, Probabilistic load flow including wind power generation, IEEE Transactions on Power Systems 26 (3) (2011) 1659–1667.

[3] M. Lei, L. Shiyan, J. Chuanwen, L. Hongling, Z. Yan, A review on the forecasting of wind speed and generated power, Renewable and Sustainable Energy Reviews 13 (4) (2009) 915–920.

[4] D. C. Hill, D. McMillan, K. R. W. Bell, D. Infield, Application of auto-regressive models to UK wind speed data for power system impact studies, IEEE Transactions on Sustainable Energy 3 (1) (2012) 134–141.

[5] K. Yunus, T. Thiringer, P. Chen, ARIMA-based frequency-decomposed modeling of wind speed time series, IEEE Transactions on Power Systems 31 (4) (2016) 2546–2556.

[6] E. Erdem, J. Shi, ARMA based approaches for forecasting the tuple of wind speed and direction, Applied Energy 88 (4) (2011) 1405–1414.

[7] A. Lojowska, Wind speed modeling, Master's thesis, Delft, the Netherlands (2009).

[8] L. Soder, Simulation of wind speed forecast errors for operation planning of multiarea power systems, in: Probabilistic Methods Applied to Power Systems, 2004 International Conference on, IEEE, 2004, pp. 723–728.

[9] G. D'Amico, F. Petroni, F. Prattico, First and second order semi-Markov chains for wind speed modeling, Physica A: Statistical Mechanics and its Applications 392 (5) (2013) 1194–1201.

[10] A. Shamshad, M. A. Bawadi, W. M. A. W. Hussin, T. A. Majid, S. A. M. Sanusi, First and second order Markov chain models for synthetic generation of wind speed time series, Energy 30 (5) (2005) 693–708.

[11] G. M. Jónsdóttir, B. Hayes, F. Milano, Optimum data sampling frequency for short-term analysis of power systems with wind, IEEE PES General Meeting, Portland, Oregon (2018) 1–5.

[12] G. M. Jónsdóttir, B. Hayes, F. Milano, Continuous-time ARMA models for data-based wind speed models, Power Systems Computational Conference (PSCC), Dublin, Ireland (2018) 1–7.

[13] E. B. Iversen, J. M. Morales, J. K. Møller, H. Madsen, Short-term probabilistic forecasting of wind speed using stochastic differential equations, International Journal of Forecasting 32 (3) (2016) 981–990.

[14] J. K. Møller, M. Zugno, H. Madsen, Probabilistic forecasts of wind power generation by stochastic differential equation models, Journal of Forecasting 35 (3) (2016) 189–205.

[15] H. Verdejo, A. Awerkin, E. Saavedra, W. Kliemann, L. Vargas, Stochastic modeling to represent wind power generation and demand in electric power system based on real data, Applied Energy 173 (2016) 283–295.

[16] W. Wu, K. Wang, G. Li, Y. Hu, A stochastic model for power system transient stability with wind power, IEEE PES General Meeting, Washington, DC (2014) 1–5.

[17] B. Yuan, M. Zhou, G. Li, Z. X-P, Stochastic small-signal stability of power systems with wind power generation, IEEE Transactions on Power Systems 30 (4) (2015) 1680–1689.

[18] X. Wang, H. D. Chiang, J. Wang, H. Liu, T. Wang, Long-term stability analysis of power systems with wind power based on stochastic differential equations: Model development and foundations, IEEE Transactions on Sustainable Energy 6 (4) (2015) 1534–1542.

[19] R. Calif, PDF models and synthetic model for the wind speed fluctuations based on the resolution of Langevin equation, Applied Energy 99 (2012) 173–182.

[20] R. Zárate-Miñano, F. Milano, Construction of SDE-based wind speed models with exponentially decaying autocorrelation, Renewable Energy 94 (2016) 186–196.

[21] R. Zárate-Miñano, M. Anghel, F. Milano, Continuous wind speed models based on stochastic differential equations, Applied Energy 104 (2013) 42–49.

[22] D. F. H. Larson, Modeling nonlinear stochastic ocean loads as diffusive stochastic differential equations to derive the dynamic responses of offshore wind turbines, Ph.D. thesis, Massachusetts Institute of Technology (2016).

[23] O. E. Barndorff-Nielsen, N. Shephard, Non-Gaussian OU based models and some of their uses in financial economics, Nuffield College Oxford, 1999.

[24] M. Grigoriu, Applied non-Gaussian processes: Examples, theory, simulation, linear random vibration, and MATLAB solutions, Prentice Hall, 1995.

[25] B. Oksendal, Stochastic differential equations: an introduction with applications, Springer Science & Business Media, 2013.

[26] P. E. Kloeden, E. Platen, H. Schurz, Numerical solution of SDE through computer experiments, Springer Science & Business Media, 2012.

[27] C. Gardiner, Stochastic methods, Vol. 4, Springer Berlin, 2009.

[28] M. Olsson, M. Perninge, L. Soder, Modeling real-time balancing power degmands in wind power systems using stochastic differential equations, Electric Power Systems Research 80 (8) (2010) 966 – 974.

[29] [Dataset] Met Éireann. The Irish meterological service online, historical data (2017), http://met.ie/climate-request/.

[30] [Dataset] RE<C: Renewable Electricity Less Than Coal. Surface Level Wind Data Collection (2017), https://code.google.com/archive/.

[31] [Dataset] University of Massachusetts. Wind Energy Center, Resource Data (2018), https://www.umass.edu/windenergy/resourcedata.

[32] V. L. Brano, A. Orioli, G. Ciulla, S. Culotta, Quality of wind speed fitting distributions for the urban area of Palermo, Italy, Renewable Energy 36 (3) (2011) 1026–1039.

[33] T. B. M. J. Ouarda, C. Charron, J.-Y. Shin, P. R. Marpu, A. H. Al-Mandoos, M. H. Al-Tamimi, H. Ghedira, H. T. N. Al, Probability distributions of wind speed in the UAE, Energy Conversion and Management 93 (2015) 414–434.

[34] J. A. Carta, P. Ramirez, S. Velazquez, A review of wind speed probability distributions used in wind energy analysis: Case studies in the Canary Islands, Renewable and Sustainable Energy Reviews 13 (5) (2009) 933–955.

[35] T. Burton, N. Jenkins, D. Sharpe, E. Bossanyi, Wind energy handbook, John Wiley & Sons, 2011.

[36] P. Milan, M. Wächter, J. Peinke, Turbulent character of wind energy, Physical Review Letters 110 (13) (2013) 138701.

[37] A. C. Brett, S. E. Tuller, The autocorrelation of hourly wind speed observations, Journal of Applied Meteorology 30 (6) (1991) 823–833.

[38] R. B. Corotis, A. B. Sigl, M. P. Cohen, Variance analysis of wind characteristics for energy conversion, Journal of Applied Meteorology 16 (11) (1977) 1149–1157.

[39] SciPy Reference Guide. Release 0.13.0, https://docs.scipy.org/doc/scipy-0.13.0/scipy-ref.pdf [accessed 23 August 2018].

[40] A. W. van der Vaart, Asymptotic Statistics, Cambridge University Press, 1998.