



HAL
open science

Marginal Weibull diffusion model for wind speed modeling and short-term forecasting

Alain Bensoussan, Alexandre Brouste

► **To cite this version:**

Alain Bensoussan, Alexandre Brouste. Marginal Weibull diffusion model for wind speed modeling and short-term forecasting. 2017. hal-01590587

HAL Id: hal-01590587

<https://hal.science/hal-01590587v1>

Preprint submitted on 19 Sep 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Marginal Weibull diffusion model for wind speed modeling and short-term forecasting

Alain Bensoussan and Alexandre Brouste

Abstract We propose a dynamical model for the wind speed which is a Markov diffusion process with Weibull marginal distribution. It presents several advantages, namely nice modeling features both in terms of marginal probability density function and temporal correlation. The characteristics can be interpreted in terms of shape and scale parameters of a Weibull law which is convenient for practitioners to analyze the results. We calibrate the parameters with the maximum quasi-likelihood method and use the model to generate and forecast the wind speed process. We have tested the model with wind speed dataset provided by the National Renewable Energy Laboratory. The model fits very well with the data. Besides, we obtain a very good performance in point and probabilistic forecasting in the short-term in comparison to benchmarks.

1 Introduction

The two parameter Weibull probability density function has become widely used to fit wind speed datasets in the literature of wind energy (see [5, 9] and the reference therein). It has been included in regulations concerning wind energy and in most popular software on wind modeling like HOMER and WAsP¹.

The energy production of a wind farm is directly related to the wind speed on the site through the power transfer function of wind turbines. Operating a wind farm re-

Alain Bensoussan

International Center for Decision and Risk Analysis, School of Management, The University of Texas at Dallas, e-mail: alain.bensoussan@utdallas.edu

Alexandre Brouste

Institut du Risque et de l'Assurance du Mans, Laboratoire Manceau de Mathématiques, Le Mans Université Maine, e-mail: alexandre.brouste@univ-lemans.fr

¹ One can find more informations on this software respectively at <https://analysis.nrel.gov/homer/> and <http://www.wasp.dk/>.

quires forecasting production in the incoming hours. From seconds to minutes, this forecast allows an accurate indication to achieve smooth production on the electricity grid. From hours to half a day, it can be used to decide whether or not to store for efficient trading on electricity markets. Finally daily production forecasting helps in scheduling maintenance operations. These problems belong to the class of stochastic optimization problems due to the intermittent nature of the wind and the solutions depend mainly on the underlying model used for the wind speed.

Several dynamical models as classical time series (ARMA, FARIMA, ...), Markov chains [6, 13, 18, 20], semi-Markov chains [7, 8] and neural networks [17] have been considered in the literature for both modeling and short-term forecasting. We have considered the Cox-Ingersoll-Ross (CIR) diffusion in [2] to model the square of the wind speed. Diffusion processes in general provide efficient point and probabilistic forecasts obtained from transition probability density functions. For instance, the CIR model overperforms the persistence benchmark (*i.e.* last measured value as forecast value) in terms of one-step forecast mean square error (MSE). Moreover, stochastic optimization problems with such continuous Markov processes are handled conveniently.

In this paper, we propose a diffusion process for the wind speed whose marginal law is Weibull. Temporal correlation structure of the stochastic process is also parametrized. We show that the model is interesting as a wind generator model. Indeed, the fact that characteristics are understandable in terms of shape and scale parameters of a Weibull law helps practitioners analyzing the results. This model also shows good performances in short-term forecasting for a large class of datasets provided by the National Renewable Energy Laboratory (NREL). The model is compared to the basic persistence benchmark and the Ornstein-Uhlenbeck diffusion process both in terms of MSE (for point forecasts) and continuous ranked probability score (CRPS, for probabilistic forecasts) for short-term lead times (see also [12] for definitions of other score and other benchmark for one-step ahead forecasting). Conversely to the CIR process, the transition probability density function is no longer available anymore in closed form. Consequently, new calibration and forecast methods are presented that differ from our previous work.

In section 2, the dynamic marginal Weibull diffusion model is presented. Point forecast and probabilistic forecast methods are developed in section 3. In section 4, the estimation method of the parameters of the aforementioned model is presented and the dynamic model is calibrated on NREL dataset. Forecasting performances of the different models are also summarized.

2 Dynamic model for the wind speed

Let $\vartheta = (\vartheta_1, \vartheta_2, \vartheta_3)$ be a 3 dimensional parameter in $(\mathbb{R}_*^+)^3$. Practitioners consider the wind speed dataset as a Weibull law which probability density function is

$$f(z, \vartheta) = \frac{\vartheta_3}{\vartheta_2} \left(\frac{z}{\vartheta_2} \right)^{\vartheta_3-1} \exp \left(- \left(\frac{z}{\vartheta_2} \right)^{\vartheta_3} \right), \quad z \geq 0. \quad (1)$$

For this distribution, the mean is directly given by

$$\mu_W = \vartheta_2 \Gamma \left(1 + \frac{1}{\vartheta_3} \right)$$

and the standard deviation by

$$\sigma_W = \sqrt{\vartheta_2^2 \Gamma \left(1 + \frac{2}{\vartheta_3} \right) - \mu_W^2}.$$

In this section, we describe the model mentioned in the introduction. The 3-parameters marginal Weibull diffusion process $(Z_t, t \geq 0)$ is the solution of the stochastic differential equation (sde)

$$Z_t = z_0 + \int_0^t v_0(Z_s, \vartheta) ds + \int_0^t v_1(Z_s, \vartheta) d\beta_s, \quad t \geq 0, \quad (2)$$

where $(\beta_t, t \geq 0)$ is a standard Wiener process. Moreover the drift coefficient $v_0(\cdot, \cdot)$ and the diffusion coefficient $v_1(\cdot, \cdot)$ are known functions that are described below. The stochastic process $(Z_t, t \geq 0)$ is an homogeneous Markov diffusion process which can be characterized by its transition probability density function $p(t, \cdot; x, \vartheta)$ representing the conditional probability density function of Z_{s+t} given $Z_s = x$ (for more information on Markov diffusion processes, see [1] or [11]). The long term law of Z_t (as $t \rightarrow \infty$) is called the stationary distribution or invariant distribution. It is denoted $p(\cdot; \vartheta)$. If the initial condition z_0 is random and follows the invariant distribution, then the law of Z_t is still the invariant distribution for all $t \geq 0$. In this setting, the stationary distribution of our model is Weibull with scale parameter ϑ_2 and shape parameter ϑ_3 (see equation (1)).

In this model, we fix the drift term equal to

$$v_0(z, \vartheta) = \vartheta_1 \left(\vartheta_2 \Gamma \left(1 + \frac{1}{\vartheta_3} \right) - z \right). \quad (3)$$

Here, the parameter ϑ_1 is the temporal correlation parameter of the process. It can be shown, if the initial condition is the invariant distribution, that the correlation structure is given by

$$\text{corr}(Z_s, Z_t) = e^{-\vartheta_1(t-s)}. \quad (4)$$

For more information, see Appendix 1. Also, the computation of the diffusion coefficient $v_1(z, \vartheta)$ to obtain marginal Weibull probability density function is postponed to Appendix 2. Namely, the diffusion coefficient is given by

$$v_1^2(z, \vartheta) = \frac{2\vartheta_1\vartheta_2\Gamma\left(1 + \frac{1}{\vartheta_3}\right)}{f(z, \vartheta)} \left(\left(1 - e^{-\left(\frac{z}{\vartheta_2}\right)^{\vartheta_3}}\right) - \frac{1}{\Gamma\left(1 + \frac{1}{\vartheta_3}\right)} \int_0^{\left(\frac{z}{\vartheta_2}\right)^{\vartheta_3}} u^{\left(1 + \frac{1}{\vartheta_3}\right) - 1} e^{-u} du \right). \quad (5)$$

For the marginal Weibull diffusion model the transition probability density function $p(t, y; x, \vartheta)$ cannot be obtained in closed form. Fortunately, it is the solution of the Fokker-Planck (FP) equation

$$\frac{\partial}{\partial t} p(t, y; x, \vartheta) = -\frac{\partial}{\partial y} (v_0(y, \vartheta) p(t, y; x, \vartheta)) + \frac{1}{2} \frac{\partial^2}{\partial y^2} (v_1(y, \vartheta)^2 p(t, y; x, \vartheta)) \quad (6)$$

with initial condition $p(0, y; x, \vartheta) = \delta_x(y)$ where $\delta_x(y)$ represents a Dirac distribution at point x .

3 One-step forecasting

The operator of a wind farm is interested in forecasting the production in the incoming hours to provide the information to the entity in charge of the electric grid and compete on electricity markets. From the transfer function of the wind turbine, the production is directly related to the wind speed. So the problem boils down to the short-term forecasting of the wind speed.

3.1 Definitions

Suppose that we fix the present time at $t = 0$ and that the initial observed wind speed is \tilde{Z}_0 . Let us denote \tilde{Z}_t the true (random and unknown) value of the wind speed at time $t > 0$ and $\tilde{\psi}(t, \cdot)$ its (unknown) probability density function (see Figure 1).

In this paper, a point forecast is an estimator of the wind speed \tilde{Z}_t given the knowledge of \tilde{Z}_0 . It is denoted by $\pi(Z_t^{\tilde{Z}_0})$. We call persistence benchmark the current knowledge

$$\pi_{per}(Z_t^{\tilde{Z}_0}) = \tilde{Z}_0. \quad (7)$$

Following the diffusion model, the forecast value is defined by

$$\pi_d(Z_t^{\tilde{Z}_0}) = \mathbf{E}_{\vartheta}(Z_t)$$

where Z_t is the solution of (2) with initial condition $z_0 = \tilde{Z}_0$.

But, probabilistic forecast can also be proposed. It consists in defining an estimator of the probability density function $\tilde{\psi}(t, \cdot)$ of \tilde{Z}_t given the knowledge of \tilde{Z}_0 . It is denoted $\hat{\psi}(t, \cdot; \tilde{Z}_0)$. It is worth mentioning that no probabilistic forecasts can be specified in the basic persistence benchmark. In the diffusion model, it is natural to define the probabilistic forecast as the transition probability density function,

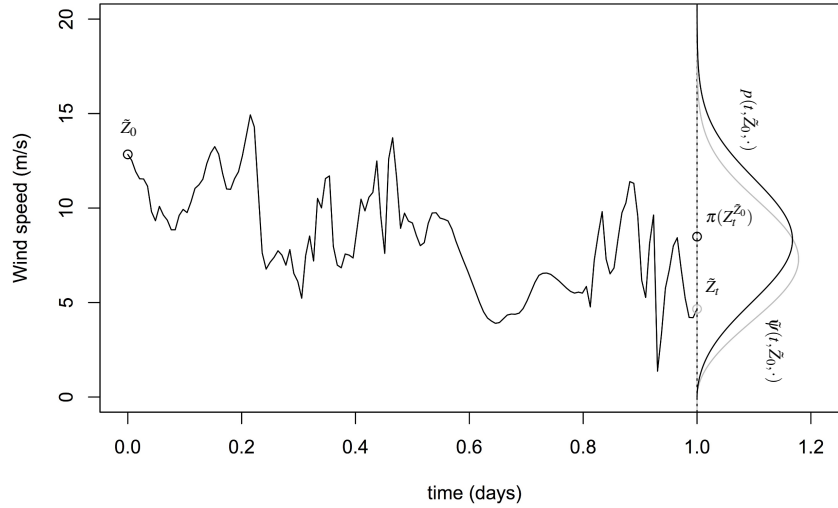


Fig. 1 One-step ahead (short-term) forecasting error

namely

$$\hat{\psi}(t, \cdot; \tilde{Z}_0) = p(t, \cdot; \tilde{Z}_0, \vartheta)$$

where p is the solution of (6).

3.2 Point forecasts and mean-square error

3.2.1 Point forecasts

For diffusion models, the point forecast is

$$\pi_d(\tilde{Z}_t^0) = \mathbf{E}_\vartheta(Z_t) = \int_{\mathbb{R}} y p(t, y; \tilde{Z}_0, \vartheta) dy \quad (8)$$

where Z_t is the solution of (2) with initial condition $z_0 = \tilde{Z}_0$. In the particular case of diffusion processes with a linear drift $v_0(z, \vartheta) = \vartheta_1(\alpha - z)$, it can be shown (see Appendix 1) that the point forecast has the following closed-form

$$\pi_d(\tilde{Z}_t^0) = \alpha + (\tilde{Z}_0 - \alpha) e^{-\vartheta_1 t}. \quad (9)$$

The parameter α can be understood by practitioners as the mean value of the stationary distribution, namely

$$\alpha = \int_{\mathbb{R}} y p(y; \vartheta) dy.$$

We recall that the parameter ϑ_1 parametrizes the correlation structure and stands for the mean-reverting speed. For the marginal Weibull diffusion model

$$\alpha = \vartheta_2 \Gamma \left(1 + \frac{1}{\vartheta_3} \right).$$

It is worth mentioning that the Ornstein-Uhlenbeck process with

$$v_0(z, \vartheta) = \vartheta_1(\alpha - z) \quad \text{and} \quad v_1(z, \vartheta) = \sigma \quad (10)$$

and the Cox-Ingersoll-Ross process with

$$v_0(z, \vartheta) = \vartheta_1(\alpha - z) \quad \text{and} \quad v_1(z, \vartheta) = \sigma \sqrt{z} \quad (11)$$

have also linear drift (see also [2]) and consequently generate similar point forecast.

3.2.2 Mean-square error

Given a point forecast $\pi(Z_t^{\tilde{Z}_0})$, the mean-square forecasting error is defined by

$$MSE(t) = \mathbf{E} \left(\left(\pi(Z_t^{\tilde{Z}_0}) - \tilde{Z}_t \right)^2 \right). \quad (12)$$

This indicator makes it possible to compare persistence model and diffusion model point forecasts. For the persistence model defined in (7),

$$MSE_{per}(t) = \mathbf{E} \left(\left(\tilde{Z}_0 - \tilde{Z}_t \right)^2 \right).$$

Diffusion models (2) have the property that the well-specified MSE (namely MSE (12) where $\tilde{Z}_t = Z_t$) can be written as

$$MSE(t) = \mathbf{E}_{\vartheta} (Z_t^2) - (\mathbf{E}_{\vartheta} (Z_t))^2 = u(t, Z_0) - (\mathbf{E}_{\vartheta} (Z_t))^2$$

in which $u(t, x)$ solves the Feynman-Kac pde, *i.e.*

$$\frac{\partial u}{\partial t} = v_0(x, \vartheta) \frac{\partial u}{\partial x} + \frac{v_1^2(x, \vartheta)}{2} \frac{\partial^2 u}{\partial x^2}$$

with

$$u(0, x) = x^2.$$

For instance, the MSE can be obtained in closed form for the Ornstein-Uhlenbeck process and the Cox-Ingersoll-Ross process (see [2]). Direct computations lead to

$$MSE(t) = \frac{\sigma^2(1 - e^{-2\vartheta_1 t})}{2\vartheta_1}$$

for the Ornstein-Uhlenbeck process (10) with

$$\lim_{t \rightarrow \infty} MSE(t) = \frac{\sigma^2}{2\vartheta_1} \quad \text{and} \quad MSE(t) \sim_0 \sigma^2 t,$$

and, comparatively,

$$MSE(t) = \frac{Z_0 \sigma^2 (e^{-\vartheta_1 t} - e^{-2\vartheta_1 t})}{\vartheta_1} + \frac{\alpha \sigma^2 (1 - e^{-2\vartheta_1 t})^2}{2\vartheta_1}$$

for the Cox-Ingersoll-Ross process (11) with

$$\lim_{t \rightarrow \infty} MSE(t) = \frac{\alpha \sigma^2}{2\vartheta_1} \quad \text{and} \quad MSE(t) \sim_0 \sigma^2 Z_0 t.$$

In the marginal Weibull setting, no closed form are available but numerical computations can be performed for the MSE with $MSE(t) \sim_0 v_1^2(Z_0, \vartheta)t$.

3.3 Probabilistic forecasts and continuous ranked probability score

3.3.1 Probabilistic forecasts

For a diffusion model in general, the probabilistic forecast is given by its transition probability density function

$$\hat{\psi}(t, \cdot; \tilde{Z}_0) = p(t, \cdot; \tilde{Z}_0, \vartheta). \quad (13)$$

The corresponding forecasting empirical cumulative distribution function is defined by

$$\hat{F}(u) = \int_{-\infty}^u p(t, y; \tilde{Z}_0, \vartheta) dy.$$

3.3.2 Continuous ranked probability score

Different probabilistic forecasts can be compared in term of Continuous Ranked Probability Score (CRPS). Let $\hat{\psi}(t, \cdot; \tilde{Z}_0, \vartheta)$ be a probabilistic forecast and \hat{F} its corresponding cumulative distribution function. For any $x \in \mathbb{R}$, continuous ranked probability function can be defined by

$$\text{CRP}(\widehat{F}, x) = \int_{\mathbb{R}} \left(\widehat{F}(u) - \mathbb{1}_{\{u \geq x\}} \right)^2 du. \quad (14)$$

The continuous ranked probability score (CRPS, see [10]) is defined by

$$\text{CRPS}(t) = \mathbf{E} \left(\text{CRP}(\widehat{F}, \tilde{Z}_t) \right). \quad (15)$$

For the Gaussian distribution, the previous function is explicitly computable, namely

$$\text{CRP}(\mathcal{N}(\mu, \sigma^2), x) = -\sigma \left(\frac{1}{\sqrt{\pi}} - 2\varphi \left(\frac{x - \mu}{\sigma} \right) - \frac{x - \mu}{\sigma} \left(2\Phi \left(\frac{x - \mu}{\sigma} \right) - 1 \right) \right) \quad (16)$$

where φ and Φ are respectively the probability distribution function and the cumulative probability distribution function of a standard Gaussian random variable. For a deterministic forecast,

$$\text{CRP}(\delta_y, x) = \int_{\mathbb{R}} \left(\mathbb{1}_{\{u \geq y\}} - \mathbb{1}_{\{u \geq x\}} \right)^2 du = |y - x|. \quad (17)$$

It is worth mentioning that for deterministic forecast, thanks to (17) the CRPS is reduced to the mean absolute error (MAE) defined by

$$\text{MAE}(t) = \mathbf{E} \left(\left| \pi(Z_t^{\tilde{Z}_0}) - \tilde{Z}_t \right| \right). \quad (18)$$

Conversely to the CIR model [2], transition (forecast) probability density functions are not in closed form. Consequently, computations of the CRPS requires using the Fokker-Planck finite element numerical solving scheme that is sophisticated (see section 3.3.3). In the following, we propose probabilistic forecasts methods based on the approximation of the transition probability density function $p(t, \cdot; x, \vartheta)$.

3.3.3 Finite-Element-Method approximation of the transition probability densities

Finite-Element-Method (FEM) scheme (see Appendix 3) computes the numerical approximation of the transition probability density $p(t, \cdot; \tilde{Z}_0, \vartheta)$ on a regular grid on a compact $\Omega \subset \mathbb{R}$ (the distribution is supposed to be zero outside Ω).

3.3.4 Gaussian approximations of the transition probability densities

For a very short lead time $t > 0$, the Gaussian approximation for transition probability density function is valid. Consequently, we use the Gaussian distribution as the probabilistic forecast with the first order or second order Itô-Taylor expansion characteristics. For the first order expansion, we recall that the mean is given by

$$m_x = x + v_0(x, \vartheta)t$$

and variance by

$$\sigma_x^2 = v_1^2(x, \vartheta)t.$$

Here v_0 and v_1 are given respectively by (3) and (5). For the second order expansion, the mean is given by

$$m_x = x + v_0(x, \vartheta)t + \left(v_0(x, \vartheta) \frac{\partial}{\partial x} v_0(x, \vartheta) + \frac{1}{2} v_1^2(x, \vartheta) \frac{\partial^2}{\partial x^2} v_0(x, \vartheta) \right) \frac{t^2}{2}$$

and the variance (see [14]) by

$$\begin{aligned} \sigma_x^2 = & x^2 + (2v_0(x, \vartheta) + v_1^2(x, \vartheta))t + \left(2v_0(x, \vartheta) \left(x \frac{\partial}{\partial x} v_0(x, \vartheta) + v_0(x, \vartheta) \right. \right. \\ & \left. \left. v_1(x, \vartheta) \frac{\partial}{\partial x} v_1(x, \vartheta) \right) + v_1^2(x, \vartheta) \left(x \frac{\partial^2}{\partial x^2} v_0(x, \vartheta) + 2 \frac{\partial}{\partial x} v_0(x, \vartheta) \right. \right. \\ & \left. \left. + \frac{\partial}{\partial x} v_1^2(x, \vartheta) + v_1(x, \vartheta) \frac{\partial^2}{\partial x^2} v_1(x, \vartheta) \right) \right) \frac{t^2}{2} - m_x^2. \end{aligned}$$

For the marginal Weibull diffusion model and linear drift diffusion models,

$$\frac{\partial^2}{\partial x^2} v_0(x, \vartheta) = 0$$

and corresponding terms in the previous equation disappear.

4 Application to wind production for one turbine

A large class of datasets is provided by the National Renewable Energy Laboratory (NREL). It offers from <http://wind.nrel.gov> wind speeds datasets for several years for more than 25000 wind turbines.

The dataset is a time series that includes 52560 wind speeds measured at the turbine rotor height every 10 minutes over 365 days in 2011 in US (ID 24310) ranging from 0.08 m/s to 24.79 m/s. The dataset contains no null wind.

For comparison purposes, the dataset is divided into two subsets: the first 183 days (training dataset) to estimate the parameter ϑ and the remaining 182 days (testing dataset) to compute the empirical MSE. Testing dataset is denoted (z_0, \dots, z_{N-1}) with $N = 26208$.

Time is considered in days. We denote Δ the time mesh (here $\Delta = \frac{1}{6 \times 24}$). Let us fix the first measure of the testing dataset at time 0 and the horizon time $\tau = k\Delta$ with the integer k which represents the number of periods. In our dataset, this is the number of periods of ten-minutes ; for instance $k = 3$ ($\tau = 0.5$ h), $k = 6$ ($\tau = 1$ h), $k = 18$ ($\tau = 3$ h), $k = 36$ ($\tau = 6$ h), $k = 72$ ($\tau = 12$ h) and $k = 144$ ($\tau = 24$ h).

The calibration method of the three parameters is presented in section 4.1 and the effective computation on the training dataset is detailed in section 4.2. Forecasting performance of the method is illustrated in section 4.3.

4.1 Calibration method

Let Θ be a bounded open subset of $(\mathbb{R}_*^+)^3$. We present in this section the calibration method for the parameter $\vartheta = (\vartheta_1, \vartheta_2, \vartheta_3) \in \Theta$ in the marginal Weibull diffusion model. In order to obtain the best estimate we use the maximum quasi-likelihood estimator, that we describe now.

Let us consider an observation of the process $(Z_t, t \geq 0)$ on a (regular) discrete temporal grid

$$0 = t_0 < t_1 < \dots < t_n.$$

The mesh is denoted $\Delta_n = \frac{t_n}{n}$. In the following, we denote $Z^{(n)} = (Z_{t_1}, \dots, Z_{t_n})$ the observation sample. The sequence $Z^{(n)}$ is a Markov chain and the corresponding loglikelihood is given by

$$\mathcal{L}(\vartheta, Z^{(n)}) = \sum_{i=1}^n \log p(\Delta_n, Z_{t_i}; Z_{t_{i-1}}, \vartheta) \quad (19)$$

where the transition probability density $p(t, y; x, \vartheta)$ is given by the solution of the Fokker-Planck equation (6).

In the Ornstein-Uhlenbeck process or CIR model presented in [2], a closed form of the transition probability density is known and maximum loglikelihood estimator

$$\hat{\vartheta}_n = \max_{\vartheta \in \Theta} \mathcal{L}(\vartheta, Z^{(n)})$$

can be computed numerically. But, for the 3-parameters marginal Weibull diffusion process, this is not the case anymore.

Several methods are available in this context depending on the discretization time scheme. For a large observation horizon t_n and large mesh size Δ_n , one can use an approximation of the likelihood function (numerical approximation of the Fokker-Planck equation in [15], Monte-Carlo simulation approximation [16], ...). A recent review of possible methods in this setting is proposed in [19].

When the mesh is relatively small (below Δ_n will be equal to 10 minutes), which is the case we are considering, it is possible to use the quasi-likelihood approach. For small Δ_n , under proper assumptions, it is possible to approximate the transition probability density function (and consequently the likelihood) by a Gaussian probability density function with the same mean and variance.

The simplest approximation is the Euler method², in which the transition probability density $p(\Delta_n, y; x, \vartheta)$ is approximated by a Gaussian with mean

$$m_x = x + v_0(x, \vartheta)\Delta_n \quad (20)$$

and variance

$$\sigma_x^2 = v_1^2(x, \vartheta)\Delta_n. \quad (21)$$

Consequently, the quasi-loglikelihood is given by

$$\mathcal{L}^*(\vartheta, Z^{(n)}) = -\frac{n}{2} \log(2\pi v_1^2(Z_{i-1}, \vartheta)\Delta_n) - \frac{1}{2} \sum_{i=1}^n \frac{(Z_{t_i} - (Z_{t_{i-1}} + v_0(Z_{t_{i-1}}, \vartheta)\Delta_n))^2}{v_1^2(Z_{t_{i-1}}, \vartheta)\Delta_n}. \quad (22)$$

The maximum quasi-likelihood (or quasi-loglikelihood) estimator is obtained by

$$\hat{\vartheta}_n^* = \max_{\vartheta \in \Theta} \mathcal{L}^*(\vartheta, Z^{(n)}).$$

Although the estimator is not in a closed form, it can be computed numerically. This approximation is valid for a "rapidly increasing experimental design" which means $\Delta_n \rightarrow 0$, $n\Delta_n \rightarrow \infty$ and $n\Delta_n^2 \rightarrow 0$ (see [14] for details).

4.2 Calibration of the different models

On the training set, the maximum quasi likelihood estimator presented in section 4.1 for the Weibull marginal diffusion model is

$$\hat{\vartheta}^* = (\hat{\vartheta}_1^*, \hat{\vartheta}_2^*, \hat{\vartheta}_3^*) = (1.19, 6.07, 2.06).$$

In term of marginal pdf, the scale parameter is $\hat{\vartheta}_2^* = 6.07$ and the shape parameter $\hat{\vartheta}_3^* = 2.06$. For the Weibull, the mean is

$$\hat{\mu}_{Wd} = \hat{\vartheta}_2^* \Gamma\left(1 + \frac{1}{\hat{\vartheta}_3^*}\right) = 5.38 m.s^{-1}$$

and the standard deviation

$$\hat{\sigma}_{Wd} = \sqrt{(\hat{\vartheta}_2^*)^2 \Gamma\left(1 + \frac{2}{\hat{\vartheta}_3^*}\right) - \hat{\mu}_{Wd}^2} = 2.73 m.s^{-1}.$$

² The Gaussian approximation of the conditional density function $p(\Delta, y; x, \vartheta)$ proposed in [14] uses the higher order Itô-Taylor expansion to approximate the mean and the variance. It is worth emphasizing that the Euler method is the one order Itô-Taylor expansion.

For the comparison, we choose to compare with persistence benchmark (which needs no calibration) and the Ornstein-Uhlenbeck process (10). For the Ornstein-Uhlenbeck process, maximum likelihood estimation gives

$$\hat{\vartheta} = \left(\hat{\vartheta}_1, \hat{\alpha}, \hat{\sigma} \right) = (0.69, 6.77, 4.41).$$

Ornstein-Uhlenbeck marginal probability distribution function is Gaussian of mean

$$\hat{\mu}_{OU} = \hat{\alpha} = 6.77 \text{ m.s}^{-1}$$

and standard deviation

$$\hat{\sigma}_{OU} = \frac{\hat{\sigma}}{2\hat{\vartheta}_1} = 3.75 \text{ m.s}^{-1}.$$

We can notice the difference of the mean and standard deviation in the two models. Stationary distribution for the marginal Weibull diffusion model and the Ornstein-Uhlenbeck process (respectively Weibull and Gaussian) are illustrated in Figure 2. Correlation structure for the marginal Weibull diffusion model is also illustrated in Figure 2. It shows that the marginal Weibull diffusion model clearly overperforms the (Gaussian) Ornstein-Uhlenbeck process in terms of wind speed generation.

4.3 Forecasting errors computations

For successive times $t_j = j\Delta$, $j = 0, \dots, N-k$, we want to compute the forecast of $\tilde{Z}_{t_j+\tau}$ at time $t_j + \tau$. We denote

$$\pi \left(Z_{t_j+\tau}^{t_j, z_j} \right)$$

the forecast value knowing that $\tilde{Z}_{t_j} = z_j$. The empirical mean square error (eMSE) is given by

$$eMSE(\tau) = \frac{\sum_{j=0}^{N-1-k} \left(\pi \left(Z_{t_j+\tau}^{t_j, z_j} \right) - z_{j+k} \right)^2}{N-k}.$$

We compute the empirical MSE for all models described above. Namely,

1. For the persistence benchmark, $\pi \left(Z_{t_j+\tau}^{t_j, z_j} \right) = z_j$;
2. For the dynamic 3 parameters marginal Weibull diffusion model with linear drift, the point forecast is given by (9) or

$$\pi \left(Z_{t_j+\tau}^{t_j, z_j} \right) = \vartheta_2 \Gamma \left(1 + \frac{1}{\vartheta_3} \right) + \left(z_j - \vartheta_2 \Gamma \left(1 + \frac{1}{\vartheta_3} \right) \right) e^{-\vartheta_1 \tau}$$

where, in practice, $\vartheta = \hat{\vartheta}^*$ is the maximum quasi-likelihood estimator of ϑ obtained as indicated in section 4.1.

We summarize the results on bias and RSME (root square of the MSE) in the Table 1 containing lead times of relevance ($\tau = 0.5$ h, 1 h, 3 h, 6 h, 12 h, 24 h).

	$\tau = 0.5$ h	$\tau = 1$ h	$\tau = 3$ h	$\tau = 6$ h	$\tau = 12$ h	$\tau = 24$ h
Weibull diffusion	-0.01 (0.79)	-0.02 (1.18)	-0.05 (2.00)	-0.09 (2.69)	-0.14 (3.29)	-0.22 (3.40)
Persistence benchmark	0.00 (0.80)	0.00 (1.20)	0.00 (2.08)	-0.01(2.93)	-0.01 (3.85)	-0.02 (4.14)
Ornstein–Uhlenbeck	0.01 (0.79)	0.03 (1.19)	0.09 (2.02)	0.16 (2.75)	0.31 (3.42)	0.53 (3.48)

Table 1 Bias and RMSE in parenthesis (all measures in m/s) for lead times of relevance

In term of RMSE, the results are comparable for forecast below one hour lead time. Marginal Weibull diffusion model overperforms the persistence benchmark for forecasting over 1h and the Ornstein–Uhlenbeck process for forecasting over 6h. Finally, on this data set, this bias is small representing 4% of the mean wind speed at 24h.

In the next Table 2, we summarized the result of CRPS and MAE computed on the Weibull diffusion model with 1st-order Itô probabilistic forecast and FEM probabilistic forecast both presented in section 3.3.3 and 3.3.4.

	$\tau = 0.5$ h	$\tau = 1$ h	$\tau = 3$ h	$\tau = 6$ h	$\tau = 12$ h	$\tau = 24$ h
FEM	0.36 [0.46]	0.56 [0.73]	1.04 [1.39]	1.46 [2.00]	1.81 [2.47]	1.74 [2.33]
1st order TI Gaussian	0.36 [0.46]	0.56 [0.73]	1.03 [1.40]	1.44 [2.00]	1.76 [2.43]	2.11 [2.82]
Persistence benchmark	[0.46]	[0.74]	[1.44]	[2.16]	[2.91]	[2.65]
Ornstein–Uhlenbeck	0.36 [0.46]	0.57 [0.73]	1.05 [1.41]	1.50 [2.05]	1.89 [2.61]	1.86 [2.50]

Table 2 CRPS and MAE in brackets for lead times of relevance

Here again we can notice similar result for 3-parameters Weibull diffusion model and 3-parameters Ornstein-Uhlenbeck forecasts in term of CRPS and MAE at very-short term. The marginal Weibull diffusion model shows better result at medium range lead times and after 6h delay. It is worth emphasizing that first order approximation is valid at lead times less than 12 h.

5 Conclusion

Cox-ingersoll-Ross process and marginal Weibull diffusion model, presented in [2] and in this paper respectively, fit well the wind speed data and the wind speed datasets at short term. The stationary behavior is convenient to provide a wind generator. For instance, it is possible to consider a model

$$Y_t = f(t)Z_t + g(t)$$

where $f(\cdot)$ and $g(\cdot)$ are periodic functions (see for instance [3] for a related work.) The characteristics of the marginal Weibull diffusion model are understandable in terms of shape and scale parameter of a Weibull law that is appreciated by practitioners to analyze the results.

Properties of diffusion process are important to study the problem of optimal storage, optimal trading and optimal maintenance where wind speed and energy production are the stochastic entries.

Acknowledgements This research has been supported by a grant from EREN Groupe and partially by NSF grant DMS-1612880, Research Grants Council of HKSAR 11303316 and ANR project CAESARS (ANR-15-CE05-0024).

Appendix 1: Autocorrelation structure of homogeneous diffusion processus with linear drift

Let $(Z_s, s \geq 0)$ be the solution of equation (2). Let us denote $m(t, z_0) = \mathbf{E}(Z_s)$. Since $v_0(z, \vartheta) = \vartheta_1(\alpha - z)$ is linear in z with

$$\alpha = \vartheta_2 \Gamma \left(1 + \frac{1}{\vartheta_3} \right),$$

we have

$$dm(t, z_0) = \vartheta_1(\alpha - m(t, z_0)) dt$$

with $m(z_0, 0) = z_0$. Consequently,

$$m(z_0, t) = \alpha + (z_0 - \alpha) e^{-\vartheta_1 t}. \quad (23)$$

The joint distribution $f(y, z)$ of the pair (Z_s, Z_t) , $s \leq t$, is given by

$$f(y, z) = p(t - s, z; y, \vartheta) p(s, y; z_0, \vartheta)$$

due to the Markov property. Consequently,

$$\begin{aligned} \text{cov}(Z_s, Z_t) &= \int_{\mathbb{R}^2} (y - \mathbf{E}Z_s)(z - \mathbf{E}Z_t) f(y, z) dy dz \\ &= \int_{\mathbb{R}} (y - \mathbf{E}Z_s) p(s, z_0, y) \underbrace{\left(\int_{\mathbb{R}} (z - \mathbf{E}Z_t) p(t - s, y, z) dz \right)}_{e^{-\vartheta_1(t-s)}(y - \mathbf{E}Z_s)} dy \\ &= e^{-\vartheta_1(t-s)} \text{var}(Z_s) \end{aligned}$$

which proves the result. In the case of the stationary distribution,

$$\text{corr}(Z_s, Z_t) = e^{-\vartheta_1(t-s)}$$

is obtained.

Appendix 2: Marginal Weibull diffusion coefficient

Following [4], it is possible to construct a marginal Weibull diffusion process, solution of the stochastic differential equation

$$Z_t = z_0 + \int_0^t v_0(Z_s, \vartheta) ds + \int_0^t v_1(Z_s, \vartheta) d\beta_s, \quad t \geq 0, \quad (24)$$

where $(\beta_t, t \geq 0)$ is a Wiener process. Fixing the drift coefficient

$$v_0(z, \vartheta) = \vartheta_1 \left(\vartheta_2 \Gamma \left(1 + \frac{1}{\vartheta_3} \right) - z \right),$$

the solution of equation (24) has an invariant density equals to the Weibull probability density function

$$f(z, \vartheta) = \frac{\vartheta_3}{\vartheta_2} \left(\frac{z}{\vartheta_2} \right)^{\vartheta_3 - 1} \exp \left(- \left(\frac{z}{\vartheta_2} \right)^{\vartheta_3} \right), \quad z \geq 0.$$

If the diffusion coefficient satisfies

$$v_1^2(z, \vartheta) = \frac{2\vartheta_1}{f(z)} \left(\vartheta_2 \Gamma \left(1 + \frac{1}{\vartheta_3} \right) F(z) - \int_0^z y f(y) dy \right) \quad (25)$$

where F is the distribution function associated to the density f . Direct computations lead to

$$F(z) = 1 - \exp \left(- \left(\frac{z}{\vartheta_2} \right)^{\vartheta_3} \right)$$

and

$$\int_0^z y f(y) dy = \vartheta_2 \int_0^{\left(\frac{z}{\vartheta_2}\right)^{\vartheta_3}} u^{\frac{1}{\vartheta_3}} e^{-u} du \quad (\text{change of variable } u = \left(\frac{y}{\vartheta_2}\right)^{\vartheta_3}).$$

Appendix 3: Finite Element Method for Fokker-Planck equation

Let us approximate the solution $(p(t, y), t \geq 0, y \in \mathbb{R})$ satisfying

$$\frac{\partial p}{\partial t} = - \frac{\partial}{\partial y} (a(y)p) + \frac{1}{2} \frac{\partial^2}{\partial y^2} (b(y)p) \quad (26)$$

with the initial condition $p(0, y) = \delta_x(y)$, $x \in \mathbb{R}$. For our problem $a = v_0$ and $b = v_1^2$. By integrating (26) with respect to y , we get

$$\begin{aligned} \frac{\partial}{\partial t} \int_{\mathbb{R}} p(t, y) dy &= \int_{\mathbb{R}} \left(-\frac{\partial}{\partial y} (a(y)p) + \frac{1}{2} \frac{\partial^2}{\partial y^2} (b(y)p) \right) dy \\ &= -a(y)p(t, y) + \frac{1}{2} \frac{\partial}{\partial y} (b(y)p) \Big|_{-\infty}^{\infty} \\ &= \left(\frac{1}{2} b'(y) - a(y) \right) p(t, y) \Big|_{-\infty}^{\infty} + \frac{1}{2} b(y) \frac{\partial}{\partial y} p(t, y) \Big|_{-\infty}^{\infty}. \end{aligned}$$

Consequently, using boundary conditions

$$\lim_{y \rightarrow \pm\infty} p(t, y) = 0 \quad \text{and} \quad \lim_{y \rightarrow \pm\infty} \frac{\partial}{\partial y} p(t, y) = 0$$

dominating possible growth of $a(\cdot)$ and $b(\cdot)$ we get that the solution is norm preserving, namely

$$\int_{\mathbb{R}} p(t, y) dy = 1, \quad t \geq 0.$$

Let us consider the approximation

$$\hat{u}(t, y) = \sum_{j=1}^{N-1} \lambda_j(t) \varphi_j(y)$$

where φ_i are the sequence of finite elements. This approximation satisfies (26) on the compact $\Omega = \text{supp}(\hat{u}) \subset \mathbb{R}$. Computing the variational formulation with respect to the test function $\varphi_i(y)$, $i = 1, \dots, N-1$, we obtain

$$\sum_{j=1}^{N-1} \lambda_j'(t) C_{i,j} - \lambda_j(t) K_{i,j} = 0 \tag{27}$$

with

$$C_{i,j} = \int_{\Omega} \varphi_i(y) \varphi_j(y) dy$$

and

$$\begin{aligned}
K_{i,j} &= \int_{\Omega} \left(-\frac{\partial}{\partial y} (a(y)\varphi_j(y)) + \frac{1}{2} \frac{\partial^2}{\partial y^2} (b(y)\varphi_j(y)) \right) \varphi_i(y) dy \\
&= -a(y)\varphi_j(y)\varphi_i(y) \Big|_{\partial\Omega} + \int_{\Omega} a(y)\varphi_i'(y)\varphi_j(y) dy \\
&\quad + \frac{1}{2} \frac{\partial}{\partial y} (b(y)\varphi_j(y)) \varphi_i(y) \Big|_{\partial\Omega} - \frac{1}{2} \int_{\mathbb{R}} \varphi_i'(y) \frac{\partial}{\partial y} (b(y)\varphi_j(y)) dy \\
&= \int_{\mathbb{R}} a(y)\varphi_i'(y)\varphi_j(y) dy - \frac{1}{2} \int_{\mathbb{R}} \varphi_i'(y) \frac{\partial}{\partial y} (b(y)\varphi_j(y)) dy \\
&= \int_{\mathbb{R}} \left(a(y) - \frac{1}{2} b'(y) \right) \varphi_i'(y)\varphi_j(y) dy - \frac{1}{2} \int_{\mathbb{R}} b(y)\varphi_i'(y)\varphi_j'(y) dy \quad (28)
\end{aligned}$$

considering that $\varphi_i(y) = 0$ for $y \in \partial\Omega$. It is worth mentioning that the first derivative of v_1 is needed to compute (28).

Consequently, equation (27) can be rewritten as a multidimensional first order linear EDO

$$C\lambda'(t) - K\lambda(t) = 0 \quad (29)$$

with the proper initial condition $\lambda(0)$. Namely, in order to mimic the true initial condition δ_x , $x \in \mathbb{R}$, we consider for $i = 1, \dots, N-1$

$$\int_{\Omega} \hat{u}(0, y)\varphi_i(y) dy = \varphi_i(x).$$

This can be written as

$$\sum_{j=1}^{N-1} \lambda_j(0) \int_{\Omega} \varphi_j(y)\varphi_i(y) dy = \varphi_i(x)$$

or

$$C\lambda(0) = \Phi_x \quad \text{where} \quad \Phi_x = \begin{pmatrix} \varphi_1(x) \\ \vdots \\ \varphi_{N-1}(x) \end{pmatrix}.$$

Finally, fixing $R \in \mathbb{N}$, we can apply an implicit Euler scheme to (29) on $[0, \tau]$, $\tau > 0$ with the mesh size $\frac{\tau}{R}$.

References

1. F. Baudoin. *Diffusion processes and stochastic calculus*. EMS textbooks in mathematics, 2014.
2. A. Bensoussan and A. Brouste. Cox-Ingersoll-Ross model for wind speed modeling and forecasting. *Wind Energy*, 19(7):1355–1365, 2016.
3. A. Bensoussan, A. Brouste, and P. Bertrand. A generalized linear model approach to seasonal aspects of wind speed modeling. *Journal of Applied Statistics*, 41(8):1694–1707, 2014.

4. B. Bibby, I. Skovgaard, and M. Sorensen. Diffusion-type models with given marginal distribution and autocorrelation function. *Bernoulli*, 1(191-220), 2003.
5. J. Carta, P. Ramirez, and S. Velasquez. A review of wind speed probability distributions used in wind energy analysis, case studies in the Canary Islands. *Renewable and Sustainable Energy Reviews*, 13:933–955, 2009.
6. F. Castino, R. Festa, and C. Ratto. Stochastic modeling of wind velocities time series. *Journal of Wind Engineering and Industrial Aerodynamics*, 74:141–151, 1998.
7. G. D’Amico, F. Petroni, and F. Praticco. First and second order semi-Markov chains for wind speed modeling. *Physica a-Statistical Mechanics and Its Applications*, 392(5):1194–1201, 2013.
8. Guglielmo D’Amico, Filippo Petroni, and Flavio Praticco. Wind speed modeled as an indexed semi-Markov process. *Environmetrics*, 24(6):367–376, 2013.
9. P. Drobinski, C. Coulais, and Jourdir B. Surface wind-speed statistics modelling: Alternatives to the weibull distribution and performance evaluation. *Boundary-Layer Meteorol.*, 157:97–123, 2015.
10. T. Gneiting and A. Raftery. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378, 2007.
11. N. Ikeda and S. Watanabe. *Stochastic Differential Equations and Diffusion Processes*. North Holland, 1981.
12. T. Jonsson, P. Pinson, H. Madsen, and H. Nielsen. Predictive densities for day-ahead electricity prices using time adaptative quantile regression. *Energies*, 7:5523–5547, 2014.
13. H. Kantz, D. Holstein, M. Ragwitz, and N. K. Vitanov. Markov chain model for turbulent wind speed data. *Physica a-Statistical Mechanics and Its Applications*, 342(1-2):315–321, 2004.
14. M. Kessler. Estimation of an ergodic diffusion from discrete observations. *Scandinavian Journal of Statistics*, 24:211–229, 1997.
15. A. Lo. Maximum likelihood estimation of generalized Itô processes with discretely sampled data. *Econometric Theory*, 4:231–247, 1988.
16. A. Pedersen. A new approach to maximum-likelihood estimation for stochastic differential equations based on discrete observations. *Scandinavian Journal of Statistics*, 22:55–71, 1995.
17. K. Pourmousavi and M. Ardehali. Very short-term wind speed prediction: a new artificial neural network—Markov chain model. *Energy Conversion and Management*, 52(1):738–745, 2011.
18. A. D. Sahin and Z. Sen. First-order Markov chain approach to wind speed modeling. *Journal of Wind Engineering and Industrial Aerodynamics*, 89(3–4):263–269, 2001.
19. H. Sorensen. Parametric inference for diffusion processes observed at discrete points of time: a survey. *International Statistical Review*, 72(3):337–354, 2004.
20. J. Tang, A. Brouste, and K. Tsui. Some improvements of wind speed Markov chain modeling. *Renewable Energy*, 81:52–56, 2015.

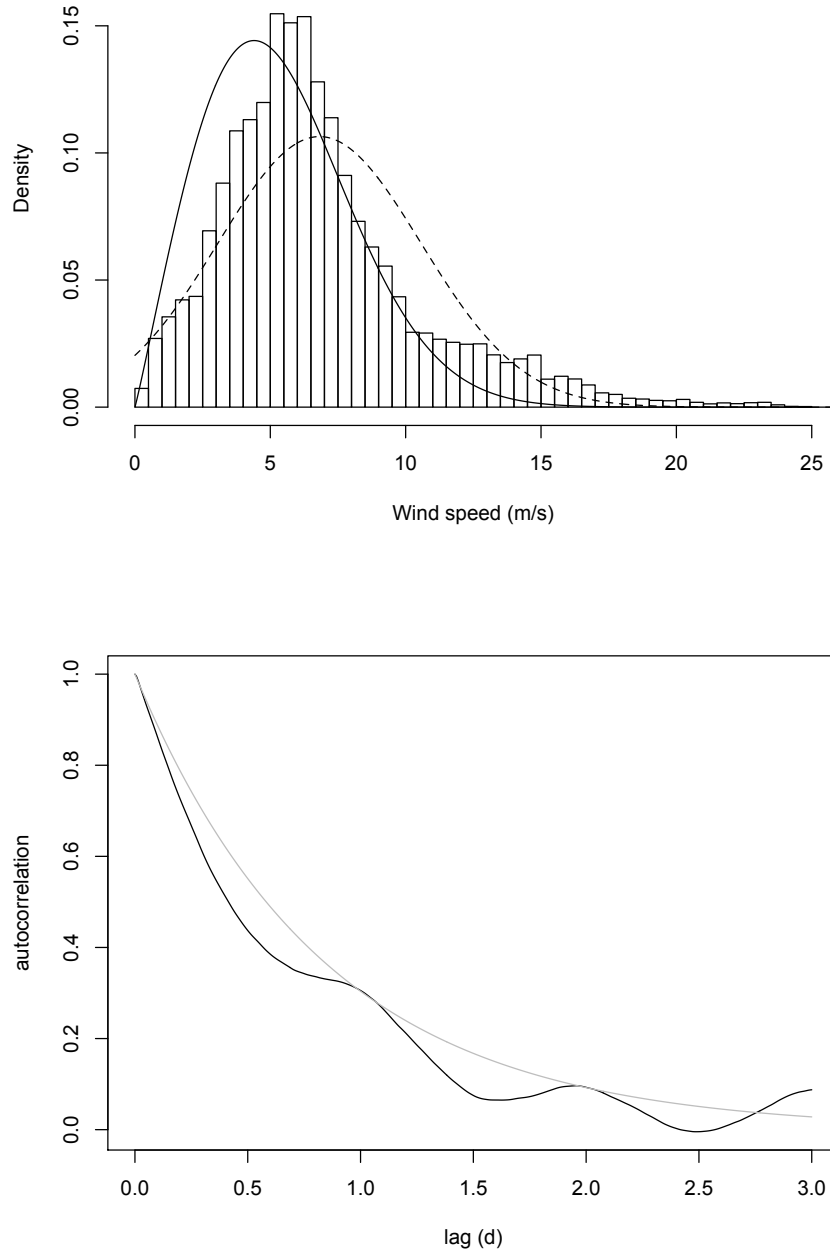


Fig. 2 On the left, histogram of the training NREL wind speed dataset (see section 4). Stationary distribution for the marginal Weibull diffusion model (plain line) and Gaussian Ornstein-Uhlenbeck process (dashed line) fitted on the training dataset are superposed. On the right, empirical autocorrelation of the training NREL wind speed dataset (plain black line) and calibrated correlation structure of the marginal Weibull diffusion model (plain gray line).