Wind Energy

WIND ENERGY
Wind Energ. 2016; **19**:1355–1365
Published online 27 August 2015 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/we.1896

SHORT COMMUNICATION

Cox–Ingersoll–Ross model for wind speed modeling and forecasting

Alain Bensoussan[1] and Alexandre Brouste[2]

[1] International Center for Decision and Risk Analysis, School of Management, The University of Texas, Dallas, Texas, USA
[2] Laboratoire Manceau de Mathématiques, Institut du Risque et de l'Assurance du Mans, Université du Maine, Le Mans, France

ABSTRACT

We propose a dynamic model for the squared norm of the wind speed which is a Markov diffusion process. It presents several advantages. Since the transition probability densities are in closed form, it can be calibrated with the maximum likelihood method. It presents nice modeling features both in terms of marginal probability density function and temporal correlation. We have tested the model with real wind speed data set provided by the National Renewable Energy Laboratory. The model fits very well with the data. Besides, we obtained a very good performance in forecasting wind speed at short term. This is an interesting perspective for operational use in industry. Copyright © 2015 John Wiley & Sons, Ltd.

KEYWORDS
wind speed modeling; wind speed forecasting; diffusion processes; Fokker–Planck equation; maximum likelihood estimation

Correspondence
Alexandre Brouste, Laboratoire Manceau de Mathématiques, Université du Maine, Le Mans, France.
E-mail: abrouste@univ-lemans.fr

Received 24 January 2015; Revised 12 May 2015; Accepted 20 July 2015

1. INTRODUCTION

Wind characteristics are essential for the evaluation of wind resources and performance of wind turbines. In the investment phase, they are used for the estimation of annual power production and valuation of a wind farm project. In the operational phase, they are important for selling on electricity markets or allocation on smart grids.

From 1950 to 1970, the horizontal wind $V = (V_1, V_2)^*$ at some fixed height (in the south–north/east–west coordinates) was usually described by a bivariate Gaussian vector with five parameters (see Crutcher and Baer[1] and the reference therein).

The increasing interest for renewable energies in the early 1970s elicited a variety of models for different applications (see Carta *et al.*[2] for a review). A particular interest has been shown towards simpler aggregate quantities such as the Euclidian norm of V or its square $Z = \|V\|^2$, namely,

$$Z = \|V\|^2 = V_1^2 + V_2^2$$

In that respect, the Rayleigh distribution was widely used[3,4] to model $\|V\|$. Its density is given by the formula

$$f(r) = \frac{r}{\sigma^2} \exp\left(-\frac{r^2}{2\sigma^2}\right), \quad r \geq 0 \tag{1}$$

Its origin stems from considering independent centered Gaussian components V_1 and V_2 of identical standard deviation $\sigma > 0$. It implies the $\chi^2(2)$ distribution for $\|V\|^2$. On the other hand, the two parameter Weibull probability density function (pdf)

$$g(r) = \frac{p}{\lambda}\left(\frac{r}{\lambda}\right)^{p-1} \exp\left(-\left(\frac{r}{\lambda}\right)^p\right), \quad r \geq 0 \tag{2}$$

has become widely used to fit wind speed data set in the literature of wind energy (see Carta *et al.*[2] and the reference therein). It has been included in regulations concerning wind energy and in the most popular softwares on wind modeling like HOMER and WAsP.*

The Rayleigh and the Weibull pdf are still unable to describe all the wind regimes encountered (high frequency of null wind speeds, bimodal distributions, *etc.*) and new models have also been proposed: bimodal, Gamma, mixtures, and hybrid distributions among the main ones.

It has been natural to move from static to dynamic models. We can mention, in particular, classical time series (ARMA, FARIMA, etc.), Markov chains, semi-Markov chains, and neural networks. In this paper, we propose a dynamic model for $\|V\|^2$ considered as a continuous-time Markov diffusion.

It is relevant in different wind contexts and improves substantially static models or persistence models. Consequently, it can be used either for wind speed short-term forecasting or for stochastic control problems considering wind speed as an entry. Moreover, point and probabilistic wind speed forecasts can be performed.

In Section 2, we give a brief recall on the usual distributions for $\|V\|^2$ and the dynamic model is presented. We mention, in Section 3, the estimation methods of the parameters for squared Weibull, non-central $\chi^2(2)$, and Gamma static models and for Cox–Ingersoll–Ross (CIR) diffusion process dynamic model. In Section 4, aforementioned models are calibrated on a real data set provided by the National Renewable Energy Laboratory. Applications of the dynamic model to wind speed short-term forecasting are also presented.

2. MODELS FOR $\|V\|^2$

In this section, we present the static and dynamic model for the random variable $Z = \|V\|^2$.

2.1. Static models

Different static models are used to model wind speeds encountered in real world. The square of a Weibull variable, the non-central $\chi^2(2)$ distribution, and the Gamma distribution are different generalizations of the (central) $\chi^2(2)$ distribution (or the exponential distribution) mentioned in the introduction and defined by the pdf

$$h(z) = \frac{1}{2\sigma^2} \exp\left(-\frac{z}{2\sigma^2}\right)$$

Indeed, the $\chi^2(2)$ distribution is the square of a Weibull distribution (3) for $p = 2$ and $\lambda = \sigma\sqrt{2}$, it is the non-central $\chi^2(2)$ distribution (4) for $\nu = 0$, and it is the Gamma distribution (6) for $\alpha = 1$ and $\beta = 2\sigma^2$.

2.1.1. On the square of a Weibull distribution.

The two-parameter Weibull model (2) for $\|V\|$ is the most widely used distribution to fit wind speed in the literature of wind energy. Both implicit moment estimators and maximum likelihood estimators for p and λ are available. We can use the Weibull model for $R = \|V\|$ to derive the distribution of $Z = R^2$. By a change of variable,

$$k(z) = \frac{p}{2\lambda^p} z^{\frac{p}{2}-1} \exp\left(-\frac{z^{\frac{p}{2}}}{\lambda^p}\right), \quad z > 0 \tag{3}$$

2.1.2. On the non-central $\chi^2(2)$.

Let us note $V = (V_1, V_2)^*$ the horizontal wind speed at some fixed height. Indeed, if we assume that V_1 and V_2 are independent Gaussian variables of same standard deviation $\sigma > 0$ and respective mean μ_1 and μ_2, then $U = \frac{Z}{\sigma^2}$ is a non-central $\chi^2(2)$ random variable (with 2 degrees of freedom) with density

$$\frac{1}{2} \exp\left(-\frac{u+\lambda}{2}\right) I_0\left(\sqrt{\lambda u}\right), \quad u \geq 0$$

where $\lambda = \frac{1}{\sigma^2}\left(\mu_1^2 + \mu_2^2\right)$. Consequently, the pdf of Z is given by

$$\ell(z) = \frac{1}{2\sigma^2} \exp\left(-\frac{z+\nu^2}{2\sigma^2}\right) I_0\left(\frac{\nu\sqrt{z}}{\sigma^2}\right), \quad z \geq 0 \tag{4}$$

*One can find more information on this softwares respectively at https://analysis.nrel.gov/homer/ and http://www.wasp.dk/.

Wind Energ. 2016; **19**:1355–1365 © 2015 John Wiley & Sons, Ltd.
DOI: 10.1002/we

where

$$\nu = \sqrt{\mu_1^2 + \mu_2^2}$$

and $I_q(.)$ is the Bessel function of the first kind of order q defined by

$$I_q(w) = \sum_{k=0}^{\infty} \left(\frac{w}{2}\right)^{2k+q} \frac{1}{k!\Gamma(k+q+1)}, \quad w \in \mathbb{R} \tag{5}$$

where $\Gamma(.)$ is the Gamma function

$$\Gamma(z) = \int_0^{\infty} x^{z-1} e^{-x} dx, \quad z \in \mathbb{R}^+$$

2.1.3. On the Gamma distribution.

The Gamma distribution model for $Z = \|V\|^2$ is defined by the pdf

$$r(z) = \frac{z^{\alpha-1}}{\beta^{\alpha}\Gamma(\alpha)} \exp\left(-\frac{z}{\beta}\right), \quad z > 0 \tag{6}$$

2.2. Dynamic models

The aforementioned static models cannot take into account parametrically the temporal correlation which appears in the real wind speed data sets. It is a major drawback for numerous applications as operational short-term forecasting or annual wind production estimation. As indicated in the introduction, dynamic models were developed to overcome this difficulty (Markov and semi-Markov chains, neural networks, etc.). We propose, in this paper, a natural parametric dynamic model for $Z = \|V\|^2$ that stems from basic considerations.

The model is a Markov diffusion process and is closely related to the squared Bessel process.[5] It has been intensively studied and applied in biology and finance, and it is known as the CIR model. It is also mentioned in Bibby *et al.*,[6] but it is not directly used for the wind application where another diffusion process is preferred.

To the best of our knowledge, it is the first time that it is used to model the wind speed squared norm $\|V\|^2$ and to wind speed short-term forecasting. It is relevant in different wind contexts and presents good forecast performances (see Section 4.2).

2.2.1. First toy model.

Let $(X_t, t \geq 0)$ the $\mathbb{R}^2$-valued stochastic process which is the solution of

$$X_t = x + \sigma W_t$$

where $x \in \mathbb{R}^2$ and $(W_t, t \geq 0)$ is an $\mathbb{R}^2$-valued Wiener process. With Itò's formula, we get that $\left(Z_t = f(X_t) = \|X_t\|^2, t \geq 0\right)$ follows

$$Z_t = z_0 + 2\sigma^2 \sum_{i=1}^{2} \int_0^t W_s^i dW_s^i + 2\sigma^2 t$$

Fixing,

$$\beta_t = \sigma \sum_{i=1}^{2} \int_0^t \left(\frac{W_s^i}{\sqrt{Z_s}}\right) dW_s^i$$

We get that $(Z_t, t \geq 0)$ is the solution of a stochastic differential equation (for more information on this topic, see Ikeda and Watanabe[7])

$$Z_t = z_0 + 2\sigma \int_0^t \sqrt{Z_s} d\beta_s + 2\sigma^2 t, \quad t \geq 0 \tag{7}$$

where $(\beta_t, t \geq 0)$ is an $\mathbb{R}$-valued Wiener process. The process $(Z_t, t \geq 0)$ is known as squared Bessel process.[5] It is worth mentioning that this process does not reach zero.

2.2.2. CIR model.

The proposal in this paper is to consider the more general Markov diffusion process, solution of the stochastic differential equation (sde)

$$Z_t = z_0 + \vartheta_1 t - \vartheta_2 \int_0^t Z_s ds + \vartheta_3 \int_0^t \sqrt{Z_s} d\beta_s, \quad t \geq 0 \tag{8}$$

where $\vartheta_1, \vartheta_2 > 0$ and $(\beta_t, t \geq 0)$ is an $\mathbb{R}$-valued Wiener process. The transition probability densities $p^{\vartheta}(t,x,y)$ solve the Fokker–Planck equation

$$\frac{\partial}{\partial t} p(t,x,y) = -\frac{\partial}{\partial y} \left((\vartheta_1 - \vartheta_2 y) p(t,x,y)\right) + \frac{1}{2} \frac{\partial^2}{\partial y^2} \left(\vartheta_3^2 y\, p(t,x,y)\right)$$

with initial condition $p(0,x,y) = \delta_x(y)$ is a Dirac distribution at point x. Closed form of $p^{\vartheta}(t,x,y)$ has been obtained in Feller.[8] Precisely, they are related to non-central χ^2 distributions and given by

$$p^{\vartheta}(t,x,y) = c e^{-(u+v)} \left(\frac{v}{u}\right)^{\frac{q}{2}} I_q\left(2\sqrt{uv}\right) \tag{9}$$

with

$$c = \frac{2\vartheta_2}{\vartheta_3^2 \left(1 - e^{-\vartheta_2 t}\right)}, \quad u = cxe^{-\vartheta_2 t}, \quad v = cy, \quad q = \frac{2\vartheta_1}{\vartheta_3^2} - 1$$

Here, I_q is defined by Equation (5).

From some combination of the parameters, the state zero cannot be reached, namely, for $2\vartheta_1 > \vartheta_3^2$. For $0 < 2\vartheta_1 \leq \vartheta_3^2$, the state zero is reached with probability one but the process cannot become negative. The correlation structure of the process $(Z_t, t \geq 0)$ can be deduced from the transition probability densities (see Appendix B) and is given by

$$\operatorname{cov}(Z_s, Z_t) = \frac{z_0 \vartheta_3^2}{\vartheta_2} \left(e^{-\vartheta_2 t} - e^{-\vartheta_2 (t+s)}\right) + \frac{\vartheta_1 \vartheta_3^2}{2\vartheta_2^2} \left(e^{-\vartheta_2 (t-s)} - e^{-\vartheta_2 (t+s)}\right) \tag{10}$$

Moreover, for $\vartheta_2 < 0$, the process $(Z_t, t \geq 0)$ is ergodic. Its invariant measure (or stationary distribution) is Gamma distributed (Equation (6)) with specific shape and scale parameters, respectively,

$$\alpha = \frac{2\vartheta_1}{\vartheta_3^2} \quad \text{and} \quad \beta = \frac{\vartheta_3^2}{2\vartheta_2} \tag{11}$$

If the initial condition is the stationary probability distribution, the covariance structure is reduced to

$$\frac{\vartheta_1 \vartheta_3^2}{2\vartheta_2^2} e^{-\vartheta_2 (t-s)}$$

2.2.3. Comparison with other sde model.

The CIR model can be compared with more general diffusion processes, which are solutions of the stochastic differential equation

$$Z_t = z_0 + \vartheta_1 t - \vartheta_2 \int_0^t Z_s ds + \vartheta_3 \int_0^t b(Z_t, \vartheta_3) d\beta_s$$

In Bibby *et al.*,[6] one shows how to find the diffusion coefficient b in order to obtain a given invariant distribution. In this setting, the covariance structure if we start with the invariant measure is given by

$$ce^{-\vartheta_2 (t-s)}$$

It is worth mentioning that CIR model is obtained when the invariant measure is a Gamma distribution.

Squared Weibull marginal sde could be interesting for practitioners. The major drawback of this model is the complexity of the corresponding diffusion coefficient b which implies that the transition probability densities cannot be obtained in a closed form. Consequently, more sophisticated methods than those presented in Section 3.2 have to be used in order

Wind Energ. 2016; **19**:1355–1365 © 2015 John Wiley & Sons, Ltd.
DOI: 10.1002/we

to calibrate the parameters. Moreover, it is necessary to solve numerically the corresponding Fokker–Planck equation to proceed with forecasting (see Section 4.2).

3. CALIBRATION METHODS

3.1. Static models

Since the density of the static models are known, the maximum likelihood estimation (MLE) method can be performed assuming independent and identically distributed observation sample.* Let φ_ϑ (k, ℓ, or r, respectively, given by Equations (3), (4), or (6)) be the static model pdf of $Z = \|V\|^2$ with multidimensional parameter $\vartheta \in \Theta \subset \mathbb{R}^m$. Let us denote $Z^{(n)} = (Z_1, \ldots, Z_n)$ the observation sample.

The loglikelihood is defined by

$$\mathcal{L}(\vartheta, Z^{(n)}) = \sum_{i=1}^{n} \log \varphi_\vartheta(Z_i)$$

and the maximum likelihood estimator satisfies

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

For these simple models, asymptotical properties of the estimators can be obtained. For instance, the maximum likelihood estimator is asymptotically normal, namely, as $n \to \infty$,

$$\sqrt{n}\left(\vartheta - \widehat{\vartheta}_n\right) \Longrightarrow \mathcal{N}\left(0, \mathcal{I}^{-1}(\vartheta)\right)$$

with Fisher information matrix

$$\mathcal{I}_{i,j}(\vartheta) = \mathbf{E}\left(\frac{\partial}{\partial \vartheta_i} \log \varphi_\vartheta(Z) \frac{\partial}{\partial \vartheta_j} \log \varphi_\vartheta(Z)\right) \tag{12}$$

Here, $\Longrightarrow$ stands for the convergence in law.

Unfortunately, the maximum likelihood estimator has no explicit form and has to be computed numerically. Moment estimators can be used as the starting point for numerical computation of the MLE. We present the moment estimators of the parameters for the three static models in Appendix A.

3.2. Dynamic model

The CIR model is a Markov process for which the transition probability densities are known. Consequently, the calibration of the parameters ϑ_1, ϑ_2, and ϑ_3 can be carried out with the maximum likelihood method from the observation of the process $(Z_t, t \geq 0)$ on a (regular) discrete temporal grid

$$0 = t_0 < t_1 < \cdots < t_n$$

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

$$\mathcal{L}(\vartheta, Z^{(n)}) = \sum_{i=1}^{n} \log p^\vartheta\left(\Delta, Z_{t_{i-1}}, Z_{t_i}\right) \tag{13}$$

where the transition probability density p is given by (9). The MLE is finally given by

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

It can be computed numerically (see Iacus[10] for more information).

*Practitioners usually consider that some wind speed are independent variables if there is a 2–3 days lag between the observations, see Bensoussan *et al.*[9] and the temporal correlation in Section 4.1.

4. APPLICATION TO WIND ENERGY

The data set is a time series that includes 52560 wind speeds at the turbine rotor height every 10 min over 365 days in 2005 in the USA (ID 24500 located 43.48N and 107.29W in Wyoming) ranging from 0.09 to 27.75 m/s. The data set presents no null wind and we consequently do not consider hybrid models. They are taken from http://wind.nrel.gov.

4.1. Calibration on the whole data set

The aforementioned static and dynamic models are calibrated on the whole data set. For the static models, the moment estimators presented in the sections A.1, A.2, and A.3 have been computed and used to deduce the maximum likelihood estimators. Results are summarized in Table I with the corresponding Fisher information matrix inverse (12).

We can remark that all probability functions are very similar (except in 0 for the non-central $\chi^2(2)$ distribution).

For the dynamic model, time is expressed in days and $\Delta = \frac{1}{6\times 24}$. MLE presented in Section 3.2 is

$$\widehat{\vartheta} = (79.43, 0.97, 11.17)$$

It is worth mentioning that for this wind speed data set, the estimations $2\widehat{\vartheta}_1 = 158.85 > 124.80 = \widehat{\vartheta}_3^2$. This model also calibrates the covariance structure by estimating the parameter ϑ_2 (Equation (10)). On this data set, the autocorrelation graphic is given in Figure 1.

It is interesting to compare the Gamma static model MLE in Table I with the Gamma stationary distribution estimated parameters computed with formulas (11), namely,

$$\tilde{\alpha}_{sta} = 1.27 \quad \text{and} \quad \tilde{\beta}_{sta} = 64.31.$$

4.2. Short-term forecasting

The operator of a wind farm needs to know what will be the production in the next hours to compete on electricity markets or provide the information to the entity in charge of the electric grid. 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.

Table I. Static calibration of the National Renewable Energy Laboratory data.

$\|V\|^2$	squared Weibull		Non-central $\chi^2(2)$		Gamma	
24500-2005	$\widehat{\lambda}$	$\widehat{p}$	$\widehat{\nu}$	$\widehat{\sigma}^2$	$\widehat{\alpha}$	$\widehat{\beta}$
Moment	9.17	2.15	5.49	25.77	1.16	70.56
Maximum likelihood	9.19	2.19	6.39	20.38	1.16	70.41
$\mathcal{I}^{-1}(\widehat{\vartheta}_n)$	$\begin{pmatrix} 19.60 & 2.36 \\ 2.36 & 2.87 \end{pmatrix}$		$\begin{pmatrix} 16.57 & 24.36 \\ 24.36 & 259.40 \end{pmatrix}$		$\begin{pmatrix} 2.14 & -129.96 \\ -129.96 & 12161.14 \end{pmatrix}$	

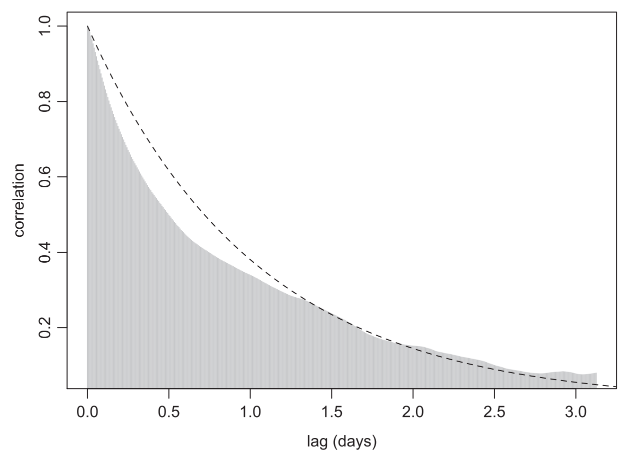

Figure 1. Autocorrelation calibration (dotted line) on the autocorrelation structure (plain gray) of the data set.

Wind Energ. 2016; **19**:1355–1365 © 2015 John Wiley & Sons, Ltd.
DOI: 10.1002/we

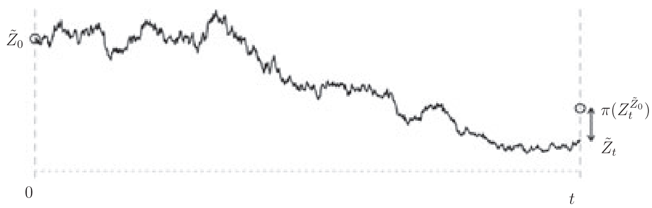

Figure 2. One-step ahead (short-term) forecasting error.

Suppose that we fix the present time at $t = 0$ and that the initial squared wind speed observed is $\tilde{Z}_0$. Let us denote $\tilde{Z}_t$ the true (random) value of the wind speed at time $t > 0$.

We consider the theoretical forecast value of $\tilde{Z}_t$. This forecast value is denoted by $\pi\left(Z_t^{\tilde{Z}_0}\right)$ (Figure 2). Then, the theoretical mean-square forecasting error is given by

$$MSE(t) = \mathbf{E}\left(\left(\pi\left(Z_t^{\tilde{Z}_0}\right) - \tilde{Z}_t\right)^2\right) \tag{14}$$

In the following, we compare three forecasting models: persistence benchmark, static models proposed in Section 3.1, and the dynamic CIR model proposed in Section 3.2.

1. For the persistence benchmark $\pi(Z_t^{\tilde{Z}_0}) = \tilde{Z}_0$. Consequently, bias is

$$b(t) = \mathbf{E}(\tilde{Z}_0) - \mathbf{E}(\tilde{Z}_t)$$

and mean square error (MSE) is defined by

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

2. For the static models, the forecast value is simply $\pi(Z_t^{\tilde{Z}_0}) = \mathbf{E}_\vartheta(Z)$ where Z is the static model of multidimensional parameter ϑ. Consequently, the bias is given by

$$b(t) = \mathbf{E}\left(\mathbf{E}_\vartheta(Z) - \tilde{Z}_t\right) = \mathbf{E}_\vartheta(Z) - \mathbf{E}(\tilde{Z}_t)$$

and the MSE is

$$MSE(t) = \mathbf{E}\left(\left(\mathbf{E}_\vartheta(Z) - \tilde{Z}_t\right)^2\right)$$

3. For the dynamic CIR model, the forecast value is the conditional expectation of Z_t defined by Equation (8) with $z_0 = \tilde{Z}_0$. For this Markov process, the transition probability densities are known and we can compute the mean and the variance of the random variable Z_t, respectively,

$$m_\vartheta(\tilde{Z}_0, t) = \frac{\vartheta_1}{\vartheta_2} + \left(\tilde{Z}_0 - \frac{\vartheta_1}{\vartheta_2}\right)e^{-\vartheta_2 t} \tag{15}$$

and

$$w_\vartheta(\tilde{Z}_0, t) = \frac{\tilde{Z}_0\vartheta_3^2\left(e^{-\vartheta_2 t} - e^{-2\vartheta_2 t}\right)}{\vartheta_2} + \frac{\vartheta_1\vartheta_3^2\left(1 - e^{-2\vartheta_2 t}\right)}{2\vartheta_2^2} \tag{16}$$

Consequently, $\pi(Z_t^{\tilde{Z}_0}) = m_\vartheta(\tilde{Z}_0, t)$. The bias satisfies

$$b(t) = \mathbf{E}(m_\vartheta(\tilde{Z}_0, t)) - \mathbf{E}(\tilde{Z}_t) = \frac{\vartheta_1}{\vartheta_2} + \left(\mathbf{E}(\tilde{Z}_0) - \frac{\vartheta_1}{\vartheta_2}\right)e^{-\vartheta_2 t} - \mathbf{E}(\tilde{Z}_t)$$

and the MSE is

$$MSE(t) = \mathbf{E}\left(\left(m_\vartheta(\tilde{Z}_0, t) - \tilde{Z}_t\right)^2\right)$$

In practice, the model $\tilde{Z}_t$ of the data is unknown and we need to replace the mean by its empirical version. For comparison purposes, the data set is divided into two subsets: the first 183 days (training data set) to estimate the parameter ϑ and the remaining 182 days (testing data set) to compute the empirical MSE. Testing data set is denoted $(v_0, \dots v_{N-1})$ 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 data set at time 0 and the horizon time $\tau = k\Delta$ with the integer k which represents the number of periods. In our data set, this is the number of periods of 10 min; for instance, $k = 18$ ($\tau = 3$ h), $k = 36$ ($\tau = 6$ h), $k = 72$ ($\tau = 12$ h), and $k = 144$ ($\tau = 1$ d).

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

$$\pi\left(Z_{t_j+\tau}^{t_j, v_j^2}\right)$$

this forecast value knowing that $\tilde{Z}_{t_j} = v_j^2$. Then the empirical mean square error (eMSE) is given by

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

For the previously mentioned model, eMSE can be computed. Namely,

1. For static models, we have

$$eMSE(\tau) = \frac{\sum_{j=0}^{N-1-k}\left(\mathbf{E}_{\widehat{\vartheta}}(Z) - v_{j+k}^2\right)^2}{N-k}$$

2. For the persistence benchmark,

$$eMSE(\tau) = \frac{\sum_{j=0}^{N-1-k}\left(v_j^2 - v_{j+k}^2\right)^2}{N-k}$$

3. For the dynamic CIR model

$$eMSE(\tau) = \frac{\sum_{j=0}^{N-1-k}\left(m_{\widehat{\vartheta}}\left(\tau, v_j^2\right) - v_{j+k}^2\right)^2}{N-k}$$

In item 1, $\widehat{\vartheta}$ is the maximum likelihood estimator or the moment estimator, and in item 3, $\widehat{\vartheta}$ is the maximum likelihood estimator of ϑ, all computed on the training set (Section 3).

Computation of the empirical MSE for the static and dynamic models are plotted in Figure 3. Practically, the empirical MSE is identical for the three static models if it is calibrated with moment estimators. We can remark in Figure 3 that the dynamic model overperforms the persistence benchmark in terms of mean square error. The dynamic model well behaves before half a day and is still better that the static models before 2 days.

We summarized this results in Table II containing lead times of relevance ($\tau = 3$ h, 6 h, 12 h, 1 d). It is worth mentioning that CIR model forecasts overperform other forecasts in terms of root mean square errors.

For the static models and the dynamic CIR model, the bias depends on the homogeneity of the distribution of the training and the testing sample. On this data set, this bias is relatively small. For this data set, bias of static models represents 3% of the mean of the testing set.

The CIR model allows to compute probabilistic forecast. As an example, we summarized the Continuous Ranked Probability Score (in short CPRS, for information of probabilistic forecasts scores, see[11]) for the lead times of relevance in the next Table III. Precisely, we compute

$$CPRS = \frac{1}{N-k}\sum_{j=0}^{N-1-k} \text{cprs}\left(F_{t_j+\tau}, v_{j+k}^2\right)$$

Wind Energ. 2016; 19:1355–1365 © 2015 John Wiley & Sons, Ltd.
DOI: 10.1002/we

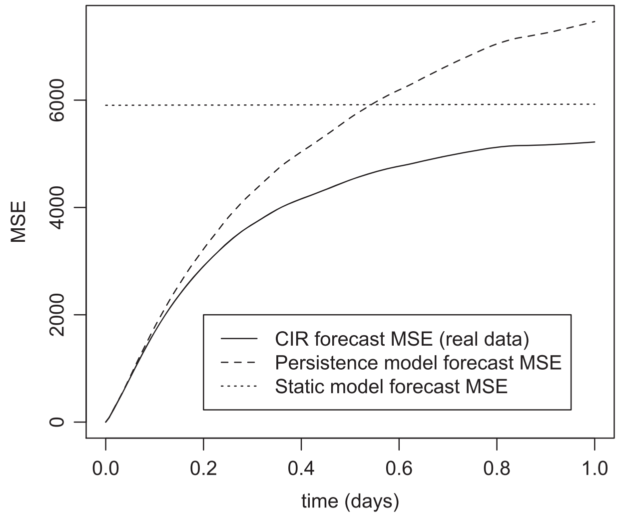

Figure 3. Empirical MSE for CIR model forecast (black plain line), persistence model forecast (black dashed line) and static models (black dotted line).

Table II. Bias, RMSE in parenthesis for lead times of relevance.

	$\tau = 3$ h	$\tau = 6$ h	$\tau = 12$ h	$\tau = 1$ d
CIR model forecast	0.3 (44.9)	0.7 (57.2)	1.2 (66.2)	1.9 (71.6)
Persistence benchmark	-0.1 (46.7)	-0.2 (61.7)	-0.3 (75.3)	-0.6 (86.4)
Static models	3.1 (76.9)	3.1 (76.9)	3.1 (76.9)	3.1 (77.0)

Table III. CPRS and MAE in brackets (all measures in m^2/s^2) for lead times of relevance.

	$\tau = 3$ h	$\tau = 6$ h	$\tau = 12$ h	$\tau = 1$ d
CIR model forecast	7.1 [29.3]	10.2 [38.8]	14.8 [46.2]	21.4 [52.0]
Gamma static model	38.7 [57.7]	38.7 [57.7]	38.6 [57.7]	38.5 [57.7]
Persistence benchmark	[30.1]	[41.1]	[50.6]	[59.1]

where $F_{t_j+\tau}$ is the forecast cumulative distribution function (cdf) and

$$\mathrm{cprs}(F,x) = \int_{\mathbb{R}} \left(F(u) - \mathbb{1}_{\{u \geq x\}}\right)^2 du$$

For the CIR model, $F_{t_j+\tau}$ is the cdf corresponding to (9). It can be shown that $F_{t_j+\tau}(u) = \Phi(2cu)$ where Φ is the cumulative distribution function of the non-central χ^2 distribution with non-central parameter $2cxe^{-\vartheta_2 t}$ (with $x = v_j^2$, $t = \Delta$ and $\vartheta = \widehat{\vartheta}$) and $r = \frac{4\vartheta_1}{\vartheta_3^2}$ degrees of freedom. In this relation, the constant $c = \frac{2\vartheta_2}{\vartheta_3^2(1-e^{-\vartheta_2 t})}$. It is worth mentioning that for deterministic forecast, the CPRS is reduced to the mean absolute error defined by

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

It is worth emphasizing that the CIR model also overperforms basic benchmarks in term of CPRS.

5. CONCLUSION AND COMMENTS

We propose in this paper a dynamic model which is a Markov diffusion process. From basic considerations, we show that the CIR process is a good model to describe the squared norm of the wind speed $Z = \|V\|^2$.

The CIR process has been intensively studied, particularly in finance. Nevertheless, it has not been applied to wind modeling or wind forecasting. It presents nice features and its transition probability densities are explicitly known.

Considering this model, the temporal correlation of a real data set can be estimated parametrically so as its marginal distribution. Finally, we exhibit the good performance of this process to forecast the wind speed at short-term horizon.

ACKNOWLEDGEMENTS

We would like to thank Pierre Pinson for the fruitful discussions and the anonymous referees for their valuable comments on this work.

This research has been supported by a grant from the EREN Groupe and partially by NSF grant DMS-1303775 and Research Grants Council of HKSAR (CityU 500113).

APPENDIX A: MOMENT ESTIMATORS OF STATIC MODELS

A.1. Square of a Weibull distribution

Non-central i-th moment of the square of a Weibull distribution (2), $i \geq 1$, are given by

$$m'_i = \mathbf{E}\left(Z^i\right) = \lambda^{2i}\Gamma\left(1 + \frac{2i}{p}\right) \tag{A.1}$$

where Γ stands for the Gamma function.

Replacing m'_1 and m'_2 by its respective empirical moments* in Equation (A.1) and solving numerically this system gives a natural moment estimator for p and λ.

A.2. On the non-central $\chi^2(2)$

In the following, the mean of a random variable ζ is denoted $m_1 = \mathbf{E}(\zeta)$. For $i \geq 2$, we denote m_i the i-th central moment of ζ, namely,

$$m_i = \mathbf{E}\left((\zeta - \mathbf{E}(\zeta))^i\right)$$

For the non-central $\chi^2(2)$ pdf (4), the mean and the second central moment are

$$m_1 = (2+\lambda)\sigma^2 \quad \text{and} \quad m_2 = 2\sigma^4(2+2\lambda)$$

Consequently, we obtain

$$\lambda = \frac{2\left(m_1^2 - m_2\right) + 2|m_1|\sqrt{m_1^2 - m_2}}{m_2}, \quad \sigma^2 = \frac{m_1}{2+\lambda} \quad \text{and} \quad \nu = \sigma\sqrt{\lambda}$$

In these formulas, the moments must satisfy the condition

$$m_1^2 - m_2 > 0 \tag{A.2}$$

Inserting empirical moment in the previous formulas provides the moment estimators for the parameters λ and ν. It is worth emphasizing, that the moment condition (A.2) is not always satisfied. In this case, the user discards this distribution.

A.3. On the Gamma distribution

Since for the Gamma distribution, the mean and the central second moment are given by

$$m_1 = \alpha\beta \quad \text{and} \quad m_2 = \alpha\beta^2$$

The moment estimators of α and β can be deduced from the following formulas

$$\alpha = \frac{m_1^2}{m_2} \quad \text{and} \quad \beta = \frac{m_2}{m_1} \tag{A.3}$$

where m_1 is replaced with the empirical mean and m_2 with the empirical variance.

*Another moment estimator using m'_1 and m'_3 can be found in Bensoussan *et al.*[9] or Carta *et al.*[2]

APPENDIX B: AUTOCORRELATION FUNCTION OF THE CIR PROCESS

Let $(Z_s, s \geq 0)$ be the solution of equation 8. The joint distribution $f(y,z)$ of the pair (Z_s, Z_t), $s \leq t$, is given by

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

due to the Markov property. Moreover, it is known that mean value and the variance of the distribution $p(t,x,y)$ (which is given by (9) and related to the non-central χ^2 distribution) are respectively

$$m(x,t) = \frac{\vartheta_1}{\vartheta_2} + \left(x - \frac{\vartheta_1}{\vartheta_2}\right) e^{-\vartheta_2 t}$$

and

$$v(x,t) = \frac{x\vartheta_3^2 \left(e^{-\vartheta_2 t} - e^{-2\vartheta_2 t}\right)}{\vartheta_2} + \frac{\vartheta_1 \vartheta_3^2 \left(1 - e^{-2\vartheta_2 t}\right)}{2\vartheta_2^2}$$

Consecutively,

$$\begin{aligned}
\operatorname{cov}(Z_s, Z_t) &= \int_{\mathbb{R}^2} (y - \mathbf{E}X_s)(z - \mathbf{E}X_t) f(y,z)\,dydz \\
&= \int_{\mathbb{R}} (y - \mathbf{E}X_s) p(s, z_0, y) \underbrace{\left(\int_{\mathbb{R}} (z - \mathbf{E}X_t) p(t-s, y, z)\,dz\right)}_{e^{-\vartheta_2(t-s)}(y - \mathbf{E}X_s)} dy \\
&= e^{-\vartheta_2(t-s)} \operatorname{var}(X_s) = e^{-\vartheta_2(t-s)} v(z_0, s)
\end{aligned}$$

that gives the result.

REFERENCES

1. Crutcher H, Baer L. Computations from elliptical wind distribution statistics. *Journal of Applied Meteorology* 1962; **1**: 522–530.
2. Carta J, Ramirez P, Velasquez S. A review of wind speed probability distributions used in wind energy analysis, case studies in the Canary Islands. *Renewable and Sustainable Energy Reviews* 2009; **13**: 933–955.
3. Celik A. A statistical analysis of wind power density based on the Weibull and Rayleigh models at the southern region of Turkey. *Renewable Energy* 2003; **29**: 593–604.
4. Haslett J, Kelledy E. The assessment of actual wind power availability in Ireland. *Energy Research* 1979; **3**: 333–348.
5. Revuz D, Yor M. *Continuous Martingales and Brownian Motion*. Springer-Verlag: London, 1991.
6. Bibby B, Skovgaard I, Sorensen M. Diffusion-type models with given marginal distribution and autocorrelation function. *Bernoulli* 2003; **1**(191–220).
7. Ikeda N, Watanabe S. *Stochastic Differential Equations and Diffusion Processes*. North Holland: New-York, 1981.
8. Feller W. Two singular diffusion problems. *Annals of Mathematics* 1951; **54**: 173–182.
9. Bensoussan A, Bertrand P, Brouste A. Forecasting the energy produced by a windmill on a yearly basis. *Stochastic Environmental Research and Risk Assessment* 2012; **26**: 1109–1122.
10. Iacus S. *Simulation and Inference for Stochastic Differential Equations with R Examples*. Springer-Verlag: New-York, 2008.
11. Gneiting T, Raftery A. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association* 2007; **102**: 359–378.