

CONFIDENCE INTERVALS FOR ANNUAL WIND POWER PRODUCTION *, **, ***

ALAIN BENSOUSSAN^{1,2}, PIERRE RAPHAËL BERTRAND³, ALEXANDRE BROUSTE⁴, NABIHA HAOUAS^{3,5}, MEHDI FHIMA⁶ AND DAOU DA KOU LIBALY⁷

Abstract. Wind power is an intermittent resource due to wind speed intermittency. However wind speed can be described as a stochastic process with short memory. This allows us to derive a central limit theorem for the annual or pluri-annual wind power production and then get quantiles of the wind power production for one, ten or twenty years future periods. On the one hand, the interquantile spread offers a measurement of the intrinsic uncertainties of wind power production. On the other hand, different quantiles with different periods of time are used by financial institutions to quantify the financial risk of the wind turbine. Our method is then applied to real datasets corresponding to a French wind turbine. Since confidence intervals can be enhanced by taking into account seasonality, we present some tools for change point analysis on wind series.

Key words: Intrinsic uncertainties of annual wind power production; Central Limit Theorem; Quantile of annual or pluri-annual wind power production; Seasonality; Intermittency ; Change point analysis.

* Research supported by a grant from *Électricité de France (EDF)* and *EDF Énergies Nouvelles (EEN)*.

** The first author acknowledges support from the Research Grants Council of the Hong Kong Special Administrative Region (*CityU 500111*).

*** The second author acknowledges support from grant ANR-12-BS01-0016-01 for developing change point analysis software used in Section 3.

¹ University of Texas at Dallas, School of Management.

² City University Hong Kong, Department SEEM.

³ Laboratoire de Mathématiques, UMR 6620 CNRS et Université Blaise Pascal (Clermont-Ferrand 2), France.

⁴ Laboratoire Manceau de Mathématiques, Université du Maine, France.

⁵ Computational Mathematics Laboratory, Monastir, Tunisia.

⁶ Altran Research, France.

⁷ EDF Énergies Nouvelles, France.

Résumé. L'énergie éolienne est une ressource intermittente à cause de l'intermittence du vent. Cependant la vitesse du vent peut être décrite comme un processus stochastique à mémoire courte. Nous pouvons alors obtenir un théorème central limite pour la production annuelle ou pluri-annuelle d'énergie éolienne, ce qui nous permet de calculer les différents quantiles de la production d'énergie éolienne avec des horizons de un, dix ou vingt ans. La différence des écarts interquantiles fournit, d'une part, une mesure de l'*incertitude intrinsèque* sur la production annuelle d'énergie éolienne. D'autre part, différents quantiles avec différents horizons de temps sont utilisés par les institutions financières afin de quantifier le risque lié à la construction d'une éolienne. Nous appliquons ensuite notre méthode à de vrais jeux de données correspondant à une turbine française. De plus, la prévision peut être améliorée par la prise en compte de la saisonnalité, nous présentons alors quelques outils pour la détection de rupture sur la moyenne de la vitesse du vent.

Mots clés : Incertitude intrinsèque sur la production annuelle d'énergie éolienne. Théorème central limite. Quantile de la production annuelle ou pluri-annuelle d'énergie éolienne. Saisonnalité. Intermittence. Détection de ruptures.

INTRODUCTION

To begin with, let us quote a well-known consultant on renewables and wind energy:

"Financiers want accurate future energy predictions for wind farms. It is normal to provide a Central Estimate energy prediction and an uncertainty analysis presenting P90, P99 or other probability levels. Reliable uncertainty analysis allows financiers to make better decisions. Reductions in uncertainty make a big difference to investor returns. Most owners are very sensitive to P90 and P99 results."

[18], Germanischer Lloyd, Garrad Hassan (2011).

In other words, evaluating wind resources and producible energy are key factors considered when deciding to build wind farms. The accuracy of the analysis made to evaluate producible energy beforehand is therefore vital to secure financing for the project and win over investors. Wind farm projects are subject to different types of risks, among which are production risks, critical to value wind-based projects. Being able to precisely determine the available wind resources therefore seems essential. It must be noted that most wind farms currently produce less than what was expected, which could reveal a certain bias in the theoretical producible energy estimation. This article offers an approach to reduce the uncertainty linked to producible energy estimation.

In the rest of this paper, our plan will be the following: In Section 2, we give the theoretical results for the forecasting of the annual wind power production and we apply these results to real datasets. In Section 3, we look at the segmentation of year into different seasons. Eventually, Section 4 sums our contribution and addresses some open questions.

1. FORECASTING ANNUAL WIND PRODUCTION

Wind power is an intermittent resource. However, electricity companies have to forecast at least the annual wind power production and have to give an estimation of uncertainties on forecasting. We basically have three categories of uncertainties. The first one is called the *intrinsic uncertainties* of wind power. Intrinsic uncertainties of wind power concern the physical uncertainties, mainly the uncertainties due to the intermittency of the wind or those due to the turbine. These uncertainties cannot be eliminated by acquiring more information. We also have the *measurement uncertainties*. Wind speed measurements or turbine performance data are obtained through sensors which carry uncertainties, which affect the data. Finally we have the *statistical estimation uncertainties*. Quantities of interest are estimated according to a model, which is calibrated with observed data (including measurement errors). The estimates, which can be biased or unbiased involve an

uncertainty. It is this one which is the best considered in literature, see eg. [8, 15, 17] and the references therein. It is important to note that all these uncertainties combine their negative effects to increase the forecasting error on the wind power production. In particular, we have shown in [2] that the intrinsic uncertainties of wind power can degrade the forecasting performance at least as much as the parameter estimation uncertainties. This is an important information for the practitioners, who are more aware of the statistical estimation uncertainties.

1.1. Intrinsic uncertainties and quantile of annual wind power production

Wind speed and wind power are intermittent resources. Wind speed intermittency mainly results from turbulence. The intermittency of wind speed can be modelled by stochastic processes. The intermittency is then explained by the stochastic behavior of wind speed. Let us stress on the one hand the loss of independency of wind speed and on the other hand the finite range of dependency of wind speed [6, 13]. In this framework, we can show that, despite the randomness of wind speed and wind power production, we can calculate the quantiles of the annual wind power production [2]. Indeed, we get the average wind speed and wind power production every ten minutes. Therefore, we can write the annual wind power production as the sum of all the productions in periods of length ten minutes:

$$\mathcal{P}_{annual} = \sum_{t=1}^T P_t, \quad (1)$$

where P_t is the wind power production during ten minutes, and $T = 52,416$ is the number of periods of ten minutes during one year. Under assumptions of stationarity and weak dependency of wind speed, which also imply stationarity and weak dependency of wind power production, we have noticed that a Central Limit Theorem (CLT) is in force and that the annual production \mathcal{P}_{annual} is almost a Gaussian random variable. More precisely, we make the following set of assumptions:

(A1): P_t is wide sense stationary.

(A2): The second order moment of P_t is finite, that is $\mathbb{E}(P_t^2) < \infty$, for each $t \geq 1$.

(A3): The family of random variables P_t is weakly dependent (see [11, 12]).

More to the point, the correlation coefficient ρ_P can be defined by

$$\rho_P(\Delta t) = \frac{\text{cov}(P_t, P_{t+\Delta t})}{\text{Var}(P_t)}. \quad (2)$$

since it does not depend on time t (Assumption A1). Then, straightforward calculations provide us the following CLT:

Proposition 1.1 (CLT on annual wind power production).

i) *If the family of random variables P_t are wide sense stationary (A1), we have*

$$\mathcal{P}_{annual} = T\mathcal{P} + T^{1/2}\mathcal{V}^{1/2}\Gamma_T \cdot \varepsilon_T \quad (3)$$

where ε_T is a zero mean random variable with variance 1; $\mathcal{P} = \mathbb{E}(P_t)$ and $\mathcal{V} = \text{Var}(P_t)$ are respectively the mean and the variance of the ten minutes wind power production and Γ_T is defined by

$$\Gamma_T = \left\{ 1 + 2 \sum_{k=1}^T \rho_P(k) \left(1 - \frac{k}{T} \right) \right\}^{1/2}, \quad (4)$$

(1) *Moreover if assumptions (A2) and (A3) are fulfilled, then ε_T converges in distribution towards a standard Gaussian law, that is $\varepsilon_T \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1)$ when $T \rightarrow \infty$, where $\xrightarrow{\mathcal{D}}$ denotes the convergence in distribution.*

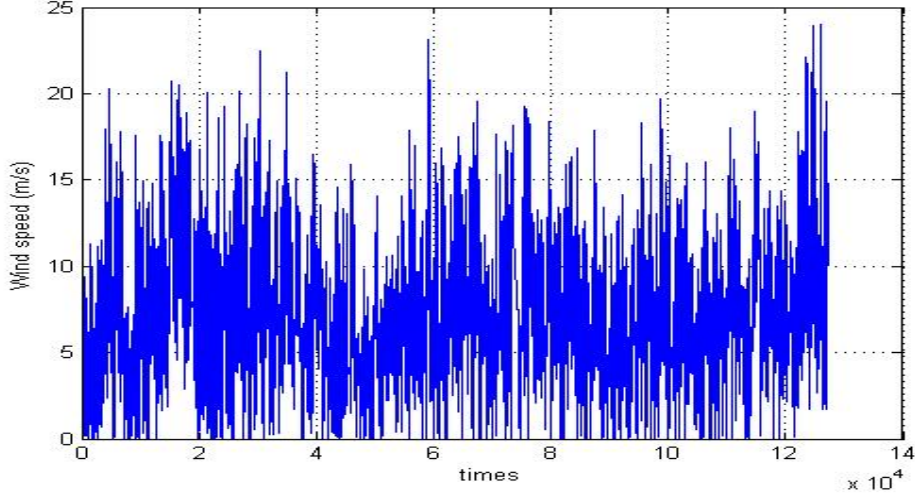


FIGURE 1. Ten minutes Wind speed as a time series, during 13 months

As required by financial institutions [18], Prop. 1.1 allows us to deduce the mean but also the quantiles Q50, Q90, Q10, or Q99 of the wind power production for 1, 10 or 20 years future periods T :

$$Q_{0.50} = \mathbb{E}(\mathcal{P}_{annual}) = T\mathcal{P}, \tag{5}$$

$$Q_{0.90} = T\mathcal{P} - 1.282 \cdot T^{1/2}\mathcal{V}^{1/2}\Gamma_T, \tag{6}$$

$$Q_{0.10} = T\mathcal{P} + 1.282 \cdot T^{1/2}\mathcal{V}^{1/2}\Gamma_T, \tag{7}$$

$$Q_{0.99} = T\mathcal{P} - 2.326 \cdot T^{1/2}\mathcal{V}^{1/2}\Gamma_T. \tag{8}$$

Both Assumption (A2) and (A3) are satisfied by the data and seem to fit physical description of wind power. But, assumption (A1) is less natural since it implies that the ten-minute average wind speed and the ten-minute wind power production do not depend on the period of the year or on the alternation of day and night. However, we can imagine that wind speed and wind power depend on the seasonality, see e.g. [9, 10]. This will be the subject of our forthcoming works.

1.2. A real Dataset

Our study is illustrated on real datasets furnished by EEN (*EDF Énergies Nouvelles*), see [2]. To sum up, we have a time series of average wind speed every ten-minute V_t for 13 months and the corresponding wind power series P_t for each ten-minute period, as illustrated by Fig. 1 and Fig. 2.

We can also plot the wind power during ten-minute as a function of the ten-minute average wind speed, see Fig. 3. This is called the power law by the electricity companies and the wind turbine manufacturers, and this map will be denoted $v \mapsto f(v)$ in the sequel. The manufacturer provides the theoretical power law whereas electricity companies can measure the empirical power law after the wind turbine building. Theoretically, fluid mechanics predict that the energy captured by a wind turbine is given by the formula $P = \frac{1}{2}\rho \mathcal{A} v^3$ where \mathcal{A} denotes the area of wind turbine, ρ the air density, and v the wind speed, see eg [16]. However, Betz' law asserts that the maximum power that can be extracted from the wind is given by Betz' Formula: $P = \frac{16}{27} \times \frac{1}{2}\rho \mathcal{A} v^3$. For a real turbine, there exist cut-in and cut-off wind speeds, namely $V_{cut.in}$ and $V_{cut.off}$, this means that the wind turbine is cut off for wind speed smaller than $V_{cut.in}$ or larger than $V_{cut.off}$. However, the energy

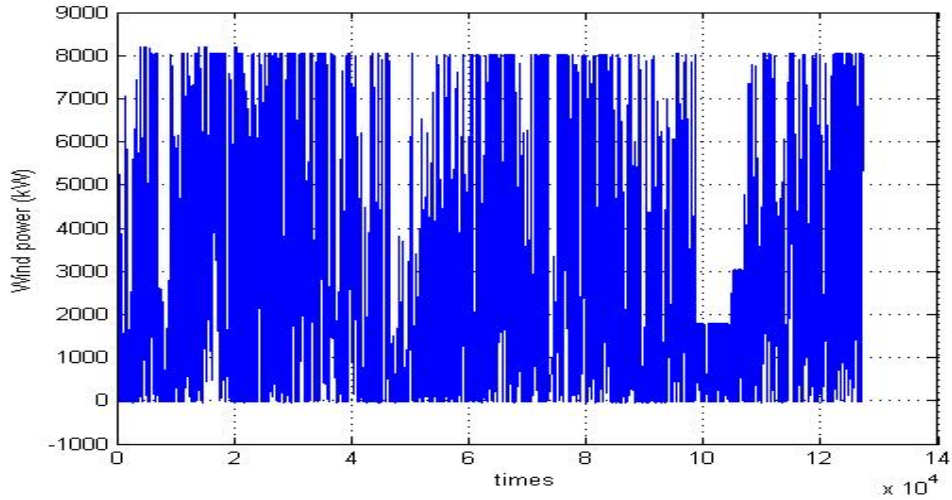


FIGURE 2. Ten minutes Wind Power as a time series, during 13 months

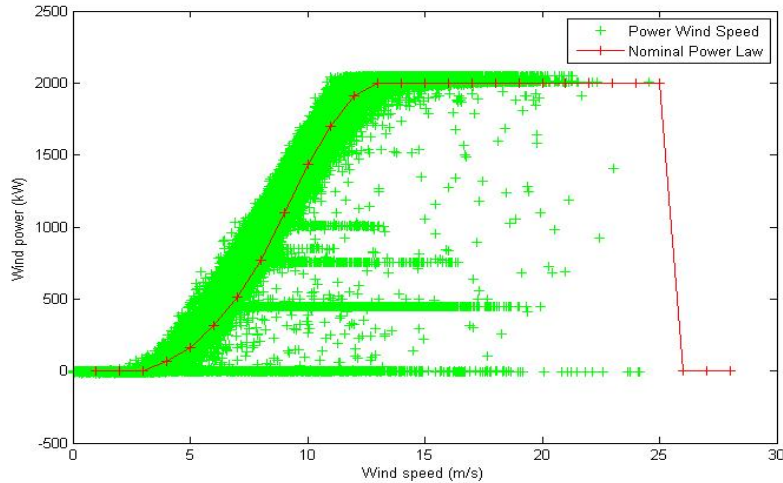


FIGURE 3. Theoretical and empirical power law

produced by the windmill is maintained at a fixed value long before the wind speed has reached the cut-off speed. This value is called the nominal value of the windmill and corresponds to a flat on the theoretical power law. Therefore, the nominal power $P_{nominal}$ is the energy produced by the windmill only for wind speed belonging to the interval $(V_{nominal}, V_{cut.off})$. Otherwise, the power law is approximately a polynomial map of order v^3 , but only for wind speeds ranging in $(V_{cut.in}, V_{nominal})$. Most often, the theoretical power law $v \mapsto f(v)$ is given by the manufacturer (at temperature 15 degrés Celsius, and air pressure 1.015 Pascal). On the other hand, we can plot the observed power versus observed average wind speed, that is (V_t, P_t) . The resulting landscape is a kind of cloud around the theoretical power law furnished by the manufacturer, as shown by figure below. These fluctuations around the theoretical power law can be explained both by measurement errors and by the variation of air density resulting in a variation of temperature or air pressure. In [2], we have investigated a

method of cancellation of aberrant data, based on the theoretical power law. However, a more systematic study is needed.

1.3. Forecasting based on real dataset

For simplicity reasons, we assume that the duration of each year is 365 days, thus the number of ten-minute periods becomes $T = 365 \times 24 \times 6 = 52,560$. In [2], we have proposed both a post-production and a preproduction method for forecasting annual wind power production. The postproduction method is based on the wind power production series P_t whereas the preproduction method is based on the wind speed series V_t . Let us point out that, in the preproduction method, we can derive a numerical wind power series by setting $P_t = f(V_t)$ where $v \mapsto f(v)$ is the power law furnished by the manufacturer. Therefore, we will only present in the following the results obtained by the post production method. Without seasonality, by using Formulae (5, 6, 7), we get

$$\begin{aligned} Q_{0.50} &= \mathbb{E}(\mathcal{P}^{annual}) &= 48.16 \text{ MWH}, \\ Q_{0.90} &= 45.99 \text{ MWH} \quad \text{and} \quad Q_{0.10} = 50.32 \text{ MWH}. \end{aligned}$$

A good measurement of intrinsic uncertainties on wind power production is provided by the spread between the two quantiles, calculated as a percentile of mean wind power production. We get the following landscape:

- Without seasonality and without cleaning, the intrinsic spread uncertainties on annual wind power production is 9%:

$$Q_{0.10} - Q_{0.90} = 0.0899 \times \mathbb{E}(\mathcal{P}^{annual}) \tag{9}$$

On the one hand, we can emphasize that intrinsic uncertainties on annual wind power production are at least of the same order of magnitude as statistical uncertainties resulting from the parameter estimation. On the other hand, the statistical uncertainties decrease with the size of the sample of wind speed, whereas the intrinsic uncertainties of wind power decrease with the length of the forecast period: For instance, intrinsic uncertainties reduce to around 3% if we forecast the twenty-year production instead the one-year. However, it is still not negligible.

As pointed out by [18], the typical Debt Service Coverage is based in Europe on the P50 (with ratio 1.4) or ten-year P90 (with ratio 1.2), and in North America on one-year P99 (with ratio 1). For instance ratio 1.4 means that the loan can be $P50/1.4$. Formulas (8, 6) give us the one-year quantile $Q_{0.99} = 44.24 \text{ MWH}$, the ten-year quantiles $Q_{0.90} = 474.7 \text{ MWH}$, and $Q_{0.50} = 481.6 \text{ MWH}$. The quantiles on annual or ten-year wind power production are given in MegaWatt Hour (MWH), they have to be converted into monetary units, by multiplying them by the price of a MWH, which depends on the country. To sum up, we have the possible ten-year loans:

North America

$$Loan_{10years} = 10 \times Q_{0.99} \times price_{MWHin\$} = 442,4 \text{ MWH} \times price_{MWHin\$},$$

Europe

$$Loan_{10years} = \frac{Q_{0.50}}{1.4} \times price_{MWH in euro} = 344.0 \text{ MWH} \times price_{MWH in euro},$$

Europe

$$Loan_{10years} = \frac{Q_{0.90}}{1.2} \times price_{MWH in euro} = 395,6 \text{ MWH} \times price_{MWH in euro}.$$

Unfortunately, the dataset provided to us by EEN (*EDF Énergies Nouvelles*) is only 13months long, which is roughly one year. Elsewhere, if we were working with two years or more of data, it will be interesting to compute an interval of intrinsic uncertainties using one year (as presented in this paper) and then to evaluate how the theoretical interval behaves on the second year or the following ones.

2. CHANGE DETECTION ON THE MEAN

As pointed in Section 1.1, wind speed and wind power varies depending on the time of year. This addresses the issue of finding a natural division into seasons: We would like to detect change on the parameters of wind speed series (or wind power production), then test whether the division provides stationary series inside each season. This is a mid term objective. In this section, we will present some results obtained by the authors for change point analysis on the mean of a series, when data are independent. As pointed out in [4], statisticians have studied change point detection since the 1950s and there is abundant literature on this subject, see eg the textbooks [1, 7] and the papers [5, 14] for a more recent point of view. However, recent measurement methods allow to record large datasets: For instance a 10 minute wind speed series with one year duration leads to datasets of size $n = 52,560$. Actually, this phenomenon is general: time-dependent data are often recorded at high frequency, which, combined with size reduction of memory capacities, allows the recording of millions of data. This phenomenon is so-called a *numerical data deluge*. This data deluge has lead to revisit change point detection methods in the particular case of large or very large datasets and has addressed the question of time and memory complexity of the algorithms for change point analysis. Algorithms derived from the least square criterion (with or without penalization) lead to both time and memory complexity of order $\mathcal{O}(n^2)$. So, in [4] we have proposed a new method, called the Filtered Derivative with p-value, (FDpV), with both time and memory complexity of order $\mathcal{O}(n)$. The FDpV method is a two-step procedure: Step 1 is based on Filtered Derivative and selection of a set of potential change points, whereas Step 2 calculates the p-value associated with each potential change point, for distinguishing correct change points and false alarms. More precisely, the method is defined as follows:

- (1) **Step 1: Computation of the filtered derivative function**, which is defined as the difference between the estimators of the mean computed in two sliding windows respectively at the right and at the left of the time t , both of size A , that is as the function:

$$FD(t, A) = \widehat{\mu}(t+1, t+A) - \widehat{\mu}(t-A, t), \text{ for } A < t < N - A \quad (10)$$

where

$$\widehat{\mu}(t+1, t+A) := A^{-1} \sum_{j=t+1}^{t+A} X_j$$

denotes the empirical mean of the variables X_j in the box $(t+1, t+A)$. This method consists of filtering data by computing the estimators of the parameter μ before applying a discrete derivation, called the Filtered Derivative method [1]. Let us remark that quantities $A \times FD(t, A)$ can be iteratively calculated, thus, the computation of the whole function $t \mapsto FD(t)$ for $t \in [A, n - A]$ requires $\mathcal{O}(n)$ operations and the storage of n real numbers. Then, potential change points τ_k^* , for $k = 1, \dots, K^*$, are selected as local maxima of the absolute value of the filtered derivative $|FD(t, A)|$ where $|FD(\tau_k^*, A)|$ exceeds a given threshold C_1 , see [4].

- (2) **Step 2: Elimination of false alarms by p-value.** A potential change point τ_k^* can be an estimator of a correct change point or a false alarm. Next, we have to eliminate false detection in order to keep (as close as possible) only the right change points. In [4], we use as Step 2 single hypothesis tests: For each potential change point τ_k^* , we calculate the associated p-values $p_1^*, \dots, p_{K^*}^*$ and keep the change points with large enough p-value.

We have applied FDpV method to the same wind speed series X_t as in Section 1.2 for detecting change on the mean. We get Figure 4 below¹. We restrict ourself to the five first month of the period in order to get a more readable picture. We can see some changes on the mean wind speed. This should be take into account in order to enhance the forecasting of annual wind power production. Actually, a model with local stationarity of

¹Fig. 4 has been done by Guillaume Paugam (software developer), by using a Java version of the FDpV method. We develop fast change methods as a part of the ANR project "Do Well B.", grant ANR-12-BS01-0016-01.

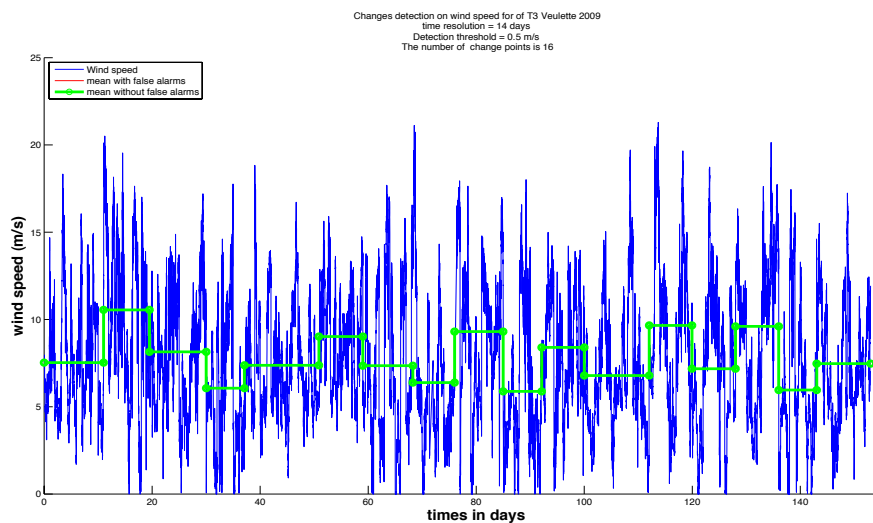


FIGURE 4. Change detection on wind speed during 5 months

the wind speed, and in consequence local stationarity of the wind power, seems to better fit the real dataset. Confidence intervals for annual or pluri-annual wind power production have been derived for this model in [3].

3. CONCLUSION

This paper is a contribution to the modeling of structural uncertainties in wind energy. Here we have considered structural uncertainties arising from wind randomness, and identified speed correlations within 48 hours, as an important factor of variability for the annual power produced by turbines. Neglecting this element is a source of significant losses. Obviously, there are many more structural uncertainties. We have only considered the speed and not the direction of the wind. We have only considered one wind turbine, and not a wind farm. This paper does not consider the seasonality of the wind. It will be done in a forthcoming paper. The interesting discovery is that the annual production benefits from a central limit theorem effect. So the structural uncertainty that we have identified can be considered as Gaussian, whatever the underlying model for the stochastic process describing the wind speed might be. However, this model (at least the marginal laws) is needed to estimate the mean and the variance of the Gaussian. Besides, the model of correlation is not considered. We have relied purely on the empirical correlation. It would be interesting to further model the stochastic process representing the evolution of wind speed, in order to obtain the real correlation. It is important to keep in mind that this work is focused on the investment phase, and not on the operational phase. Forecasting the power production for the grid on a daily or weekly basis will be a different problem. We will not benefit from the central limit theorem effect. Describing the underlying process becomes a very important aspect in this context.

REFERENCES

- [1] M. Basseville and I. V. Nikiforov. *Detection of abrupt changes: theory and application*. Prentice Hall Information and System Sciences Series. Prentice Hall Inc., Englewood Cliffs, NJ, 1993.
- [2] A. Bensoussan, P.R. Bertrand, and A. Brouste. Forecasting the energy produced by a windmill on a yearly basis . *Stoch Environ Res Risk Assess*, 2012.
- [3] A. Bensoussan, P.R. Bertrand, A. Brouste, and N. Haouas. Impact of Seasonality on Forecasting the Energy produced by a Windmill . *submitted*.
- [4] P.R. Bertrand, Fhima M., and A. Guillin. Off-line detection of multiple change points with the filtered derivative with p-value method. *Sequential Analysis*, 30(2):172–206, 2011.

- [5] L. Birgé and P. Massart. Minimal penalties for Gaussian model selection. *Probab. Theory Related Fields*, 138(1-2):33–73, 2007.
- [6] F. Böttcher, St. Barth, and J. Peinke. Small and large scale fluctuations in atmospheric wind speeds. *Stochastic Environmental Research and Risk Assessment*, 21:299–308, 2007. 10.1007/s00477-006-0065-2.
- [7] B. E. Brodsky and B. S. Darkhovsky. *Nonparametric methods in change-point problems*, volume 243 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht, 1993.
- [8] J.A. Carta, P. Ramírez, and S. Velázquez. 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.
- [9] A. Charpentier, J.C. Bouette, J.F. Chassagneux, D. Sibai, and R. Terron. Wind in Ireland : long memory or seasonal effect ? *Stochastic Environmental Research and Risk Assessment*, 20:141–151, 2006.
- [10] F. Chellali, A. Khellaf, and A. Belouchrani. Application of time frequency representation in the study of the cyclical behavior of wind speed in Algeria: wavelet transform. *Stochastic Environmental Research and Risk Assessment*, 24:1233–1239, 2010.
- [11] J. Dedecker, P. Doukhan, G. Lang, J. R. León R., S. Louhichi, and C. Prieur. *Weak dependence: with examples and applications*, volume 190 of *Lecture Notes in Statistics*. Springer, New York, 2007.
- [12] P. Doukhan. *Mixing: Properties and examples*, volume 85 of *Lecture Notes in Statistics*. Springer-Verlag, New York, 1994.
- [13] M.C.H. Hui, Larsen A., and H.F. Xiang. Wind turbulence characteristics study at the Stonecutters Bridge site: Part II Wind power spectra, integral length scales and coherences. *Journal of Wind Engineering and Industrial Aerodynamics*, (97):48–59, 2009.
- [14] M. Hušková and C. Kirch. Bootstrapping confidence intervals for the change-point of time series. *J. Time Ser. Anal.*, 29(6):947–972, 2008.
- [15] P. Ramírez and J.A. Carta. The use of wind probability distributions derived from the maximum entropy principle in the analysis of wind energy. A case study. *Energy Conversion and Management*, (Issues 15-16):2564–2577, 2006.
- [16] V. Romero-Ternero. Influence of the fitted probability distribution type on the annual mean power generated by wind turbines: A case study at the Canary Islands. *Energy Conversion and Management*, (49):2047–2054, 2008.
- [17] J.V. Seguro and T.W. Lambert. Modern estimation of the parameters of the Weibull wind speed distribution for wind energy analysis. *Journal of Wind Engineering and Industrial Aerodynamics*, (85):75–84, 2000.
- [18] A. Tindal. Financing wind farms and the impacts of P90 and P50 yields. In *EWEA Wind Resource Assessment Workshop*, 11 May 2011.