SafeWind

Collaborative project funded by the European Commission
under the 7th Framework Program, Theme 2007-2.3.2: Energy

"Multi-scale data assimilation, advanced wind modelling & forecasting with emphasis to extreme weather situations for a safe large-scale wind power integration"

Grant Agreement N°: 213740

Deliverable Dp-6.1

"Probabilistic wind power forecasting - Novel approaches and their evaluation"

DOCUMENT TYPE	Deliverable
DOCUMENT NAME:	swind.deliverable_Dp-6.1.pdf
VERSION:	V1.0
DATE:	2012.09.15
CLASSIFICATION:	R0: General public
STATUS:	Approved

Abstract: This report compiles a large part of the work carried out in the Task 6.1 of the EU-FP7 project SafeWind, which placed emphasis on probabilistic forecasting of wind power generation (at lead times from a few minutes to a few days) with considerations for extremes. Various parametric and nonparametric approaches were considered, also relying on statistical and artificial-intelligence types of models. The probabilistic forecasts may take different forms since relating to the level of generated power itself, or to the occurrence of specific events i.e. ramps. All these works contributed to advancing the state of the art in probabilistic forecasting of wind power generation.

AUTHORS[1], REVIEWERS				
MAIN AUTHOR/EDITOR:	P. Pinson			
AFFILIATION:	Technical University of Denmark, DTU Informatics			
ADDRESS:	Asmussens Allee 305, 2800 Kgs Lyngby, Denmark			
TEL.:	+45 4525 3428			
EMAIL:	pp@imm.dtu.dk			
FURTHER AUTHORS:	Miguel Bermejo (UC3M), Ismael Sanchez (UC3M), George Sideratos (NTUA), Nikos Hatzygiriou (NTUA), Arthur Bossavy (Mines ParisTech), Robin Girard (Mines ParisTech), George Kariniotakis (Mines ParisTech), Joo Young Jeon (Uni. Oxford), James Taylor (Uni. Oxford)			
PEER REVIEWERS:	Robin Girard, Georges Kariniotakis (Mines ParisTech)			
REVIEW APPROVAL:	Approved :	X	Rejected (improve as indicated below) :	
SUGGESTED IMPROVEMENTS:	*For a long list of remarks make reference to another document*			
APPROVER:	Pierre Pinson (DTU)			

STATUS, CONFIDENTIALITY, ACCESSIBILITY									
STATUS:				CONFIDENTIALITY:				ACCESSIBILITY:	
S0	Approved/Released	**X**		**R0**	General public	**X**		Private web site	**X**
S1	Reviewed			**R1**	Restricted to project members			Public web site	**X**
S2	Pending for review			**R2**	Restricted to European **Commission**			Paper copy	
S3	Draft for comments			**R3**	Restricted to WP members + PL				
S4	Under preparation			**R4**	Restricted to Task members +WPL+PL				

PL: *Project leader* **WPL:** *Work package leader* **TL:** *Task leader*

[1] **The authors of this document are solely responsible for its content, which does not represent the opinion of the European Community and the European Community is not responsible for any use that might be made of data appearing therein.**

Contents

1 Acknowldegements

The various papers compiled in the Appendix were published in peer-reviewed international scientific journals and should be cited as such. It is their preprint version which is available here. For the final published version of these papers, one should access the relevant journal archive. These papers include:

Bermejo MA, Sánchez I (2012). Adaptive densitive estimation of wind power predictions. Technical Report, Universidad Carlos III Madrid, Dpt. of Statistics, 2012.

Bossavy A, Girard R, Kariniotakis G (2012). Forecasting ramps of wind power production with numerical weather prediction ensembles. *Wind Energy*, available online.

Jeon J, Taylor JW (2012). Using conditional kernel density estimation for wind power density forecasting. *Journal of the American Statistical Association* **107**(497): 66–79.

Pinson P (2012). Very short-term probabilistic forecasting of wind power with generalized logit-Normal distributions. *Journal of the Royal Statistical Society, Series C* **61**(4): 555-576.

Sideratos G, Hatziargyriou N (2012). Probabilistic wind Power forecasting using radial basis function neural networks. *IEEE Transactions on Sustainable Energy*, available online.

2 Overview

Massive deployment of renewable energy sources, with the leading role of wind energy historically, calls for the development of new integrated approaches for the optimal management of power systems operations, most often in a market environment. This in turn requires the use of forecasts as input to monitoring and decision-making, with forecasts that may be of different forms for different lead times and with varied space-time resolution (Jones 2011). Forecasts are especially valuable in a market environment like those existing in Europe or in the US (Botterud *et al.* 2010) among others. An extensive and recent review of the state of the art in wind power forecasting was compiled in Giebel *et al.* (2011).

Forecasts arguably ought to be of probabilistic nature, even though the final prediction information to be conveyed to end-users may be single-valued. Their optimal format also depend upon the decision-making problem at hand. They may consist in quantile, interval or density forecasts when informing of marginal distributions of power production for each location and lead time, individually. Scenarios of short-term wind power generation may additionally inform on the space-time structure for prediction errors, as discussed by Pinson (2012) for instance.

In the frame of the EU-FP7 project SafeWind, a large emphasis was placed on the proposal and development of novel approaches to probabilistic forecasting (along with related forecast verification techniques), with a particular interest on the tails of distributions. Such tails are there to inform about likelihood of extreme events, which can be interpreted as expected large and rare deviations from the conditional expectation (that is, the commonly provided point forecast). This report is a collection of work that were published in scholarly journals or as technical reports, based on the research work carried out in the Work Package (WP) 6 of the EU-FP7 project SafeWind. Acknowledgements regarding publication information can be found in a previous section.

The works collected in the Appendix parts cover the following subjects. Bermejo and Sánchez (2012) concentrates on adaptive density estimation based on tracking conditional moments of forecast error distributions. It then compares rival approaches to density forecasting using these conditional moments, in a parametric and semi-parametric framework. The emphasis in Bossavy *et al.* (2012) is placed on the definition, characterization and probabilistic prediction of the so-called ramp events, with ensemble forecasts as input. Jeon and Taylor (2012) propose approaches for probbailistic forecasting of wind power generation based on the conversion of probabilistic prediction of wind speed and direction to be converted to power through appropriately designed and estimated power curves. Pinson (2012) introduces new types of predictive densities based on a discrete-continuous mixture embedding Generalized logit-Normal distributions (hence, in a parametric framework), also showing their relevance for the wind power application with focus on tails of forecast error distributions. Finally, Sideratos and Hatziargyriou (2012) describes an artificial-intelligence type of approaches for wind power probabilistic forecasting, with high skill overall and good calibration.

References

Bermejo MA, Sánchez I (2012). Adaptive densitive estimation of wind power predictions. Technical Report, Universidad Carlos III Madrid, Dpt. of Statistics, 2012.

Bossavy A, Girard R, Kariniotakis G (2012). Forecasting ramps of wind power production with numerical weather prediction ensembles. *Wind Energy*, available online.

Botterud A, Wang J, Miranda V, Bessa RJ (2010). Wind power forecasting in U.S. electricity markets. *Electricity Journal* **23**(3): 71–82.

Giebel G, Brownsword R, Kariniotakis G, Denhard M, Draxl C (2011). The state of the art in short-term prediction of wind power — A literature overview, 2nd Edition. Deliverable Report Dp-1.4, EU-FP7 project SafeWind [Available at:`www.safewind.eu`].

Jeon J, Taylor JW (2012). Using conditional kernel density estimation for wind power density forecasting. *Journal of the American Statistical Association* **107**(497): 66–79.

Jones LE (2011). Strategies and decision support systems for integrating variable energy resources in control centers for reliable grid operations. Technical Report, Alstom Grid, Washington DC (USA). [Available at uwig.org/doe_wind_integration_report.pdf]

Pinson P (2012). Very short-term probabilistic forecasting of wind power with generalized logit-Normal distributions. *Journal of the Royal Statistical Society, Series C* **61**(4): 555-576.

Pinson P, Girard R (2012). Evaluating the quality of scenarios of short-term wind power generation. *Applied Energy* **96**: 12–20.

Sideratos G, Hatziargyriou N (2012). Probabilistic wind Power forecasting using radial basis function neural networks. *IEEE Transactions on Sustainable Energy*, available online.

3 Appendix

Adaptive densitive estimation of wind power predictions

Bermejo, M.A. Sánchez, I.

Department of Statistics, UC3M

January 24, 2012

1 Introduction

The calculation of wind power point predictions is not enough when the problem we face requires to make decisions. To that end, the calculation of the uncertainty associated with the point prediction is necessary for making effectively tasks like trading in electricity markets or estimation of the stock of energy (Juban et al., 2007).

Usually, the calculation of wind power prediction through complex computation methods is usual. This methods only provide point predictions without their corresponding uncertainty. For this reason, the estimation of the conditional probability of the point prediction, called predictive density, is necessary.

The estimation of predictive density of wind power predictions can be studied from different perspectives:

- Using wind speed distributions (Lange, 2005; Carta et al., 2009). The predictive density is estimated by wind speed data, then the predictive density of wind power predictions is estimated through the power curve between wind speed and wind power.

- Using ensemble predictions (Nielsen, et al. 2004; Pinson and Madsen 2009; Taylor et al., 2009). The ensemble predictions of several meteorological variables are converted to wind power predictions by means of a power curve model. Then, the predictive density of wind power predictions is estimated.

- Using wind power predictions (Bludszuweit et al., 2008; Al-Awami and El-Sharkawi, 2009; Lau

and McSharry, 2010; Pinson, 2010). The predictive density is directly estimated from point wind power predictions.

In this paper, the last option is used. The idea is the estimation of predictive densities by means of the estimation of the conditional moments of the wind power predictions. The data used in this paper is from different wind farms from Crete. The dataset contain the hourly wind power measures for several years. The point predictions are calculated by using different methods.

1.1 Notation

Wind power generation can be interpreted as a stochastic process, owing to the stochastic nature of wind itself. The realizations of the stochastic process will be denoted by p_t. To make easier the notation, we always use p_t for both stochastic process and its realizations. The $h-$step ahead point forecasts of wind power generation p_{t+h} will be denoted by $\widehat{p}_{t+h|t}$, and their predictive densities will be $f_{t+h|t}$. The $h-$step ahead forecast error will be $e_{t+h|t} = p_{t+h} - \widehat{p}_{t+h|t}$, and its first conditional moment will be $m^{(0)}_{t+h|t} = \mathrm{E}_t\left[e_{t+h|t}\right]$. Finally, the conditional central moments of the forecast error $e_{t+h|t}$ will be $m^{(j)}_{t+h|t} = \mathrm{E}_t\left[\left(e_{t+h|t} - m^{(0)}_{t+h|t}\right)^j\right]$, with $j = 2, 3, ..., M$.

2 Estimation of predictive density

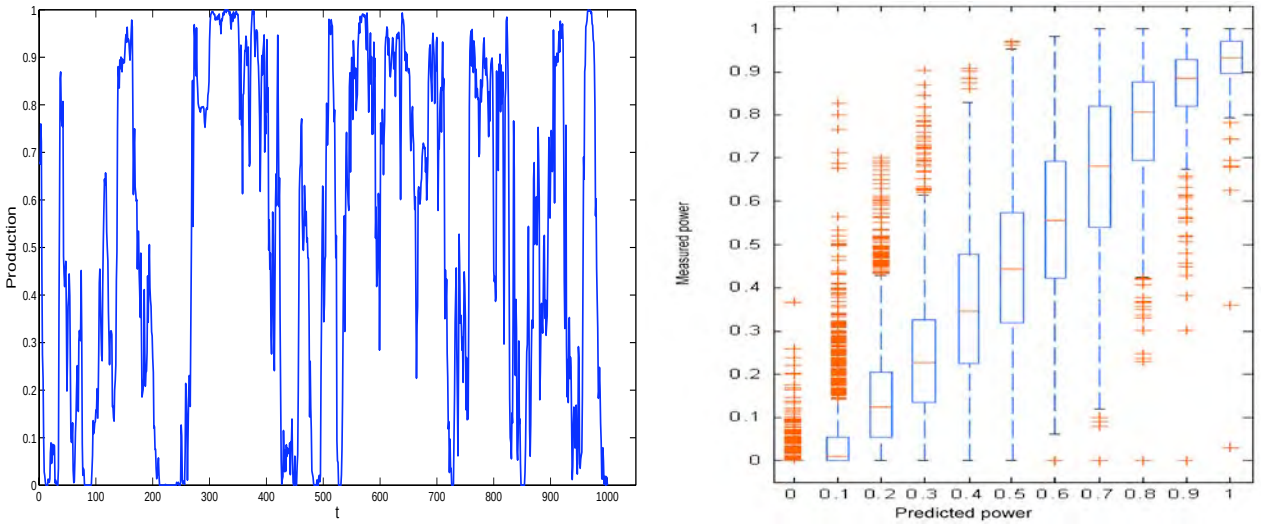
Wind power data have particular features which must be considered in the choice of probability density functions used. Figure 1a shows the main feature of wind power data, the observations are bounded, that is, probability density functions with bounded domain are necessary. Besides the box-plots of the wind power data for different levels of wind power predictions show the existence of different skewness and variance values, as it can be seen in Figure 1b.

This features lead us to propose the use of the maximum entropy distribution, because their main advantages are it is a bounded distribution and very flexible to adapt to a moment restrictions, that will be related to the prediction levels.

2.1 Maximum entropy distribution

Entropy can be interpreted as a measure of the uncertainty for any density function and can be defined by the following expression (Shannon, 1948)

$$S = -\int_R f_{t+h|t} \ln(f_{t+h|t}) dp. \tag{1}$$



(a) Wind power data

(b) Box-plot of wind power data for different levels of predictions

Figure 1: Main features of wind power data.

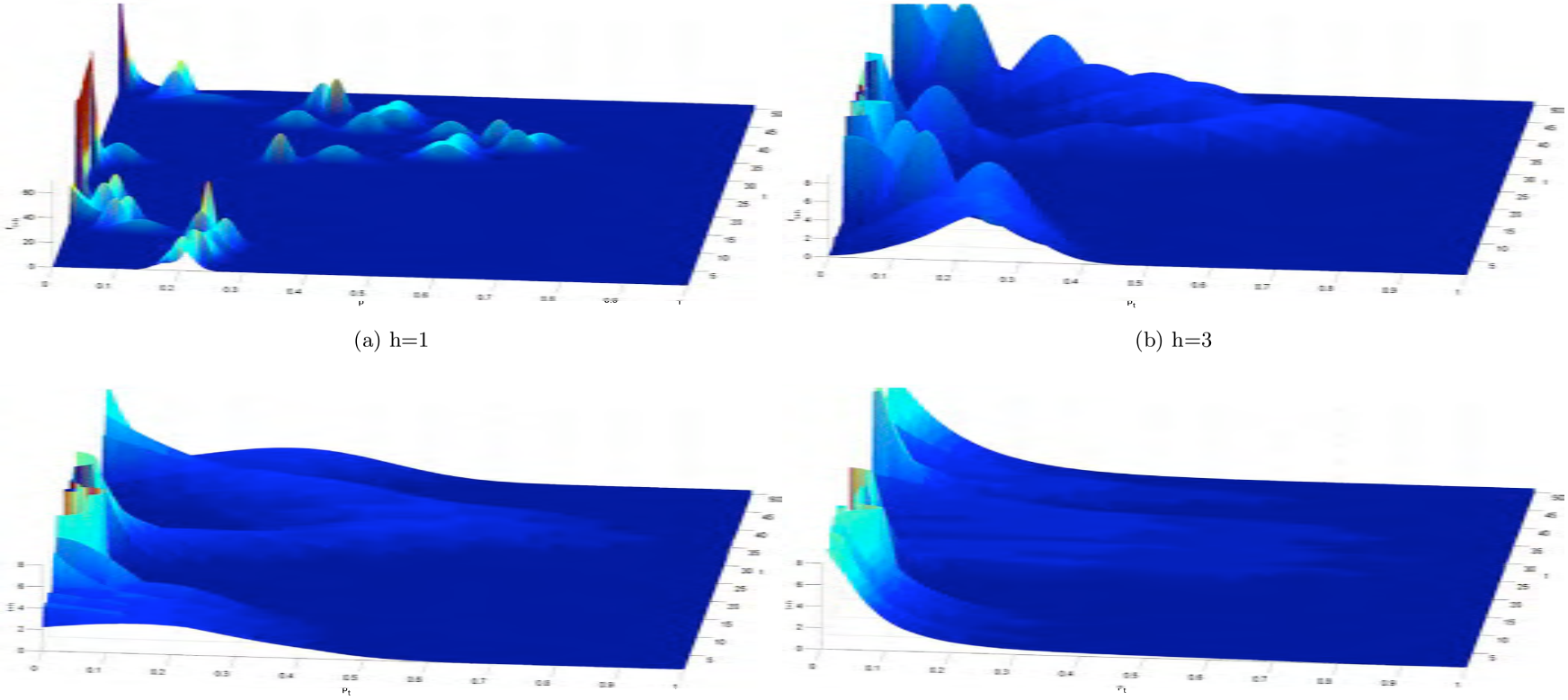
Jaynes (1957) extended this concept to formulate the maximum entropy principle (MEP). Jaynes proposed a method to determine the probability distribution by minimazing necessary prior information, which is obtained by maximizing Shannon's entropy. Then, the MEP probability density function $f_{t+h|t}$ which is defined in the interval $[a, b]$ is obtained by solving

$$\max - \int_a^b f_{t+h|t} \ln(f_{t+h|t}) dp, \tag{2}$$

subject to the following restrictions

$$\begin{array}{l} \int_a^b f_{t+h|t} dp = 1 \\ \int_a^b p^i f_{t+h|t} dp = m_{t+h|t}^{(j)}, \ j = 1, ..., M, \end{array} \tag{3}$$

where $m_{t+h|t}^{(j)}$ is jth conditional moment about the origin of the theoretical distribution, that is, the M low order statistical moment are determined numerically from the sample. Expression (2) has only analytical solution for $M = 1$, for this reason it has to be solved by numerical methods. The algorithm proposed by Siddall and Diab (1975) is used in this paper for this task. Figure 2 shows an example of the sequence of predictive densities $f_{t+1|t}^{(MEP)}$ estimated by using $M = 4$. The MEP distribution has been previously used in the estimation of wind speed distribution (see, e.g., Akpinar and Kavak Akpinar, 2007; Carta et al., 2009; Liu and Chang, 2010).



(a) h=1

(b) h=3

(c) h=6

(d) h=12

Figure 2: An example of the sequence of predictive density estimates for different horizons by using maximum entropy distribution.

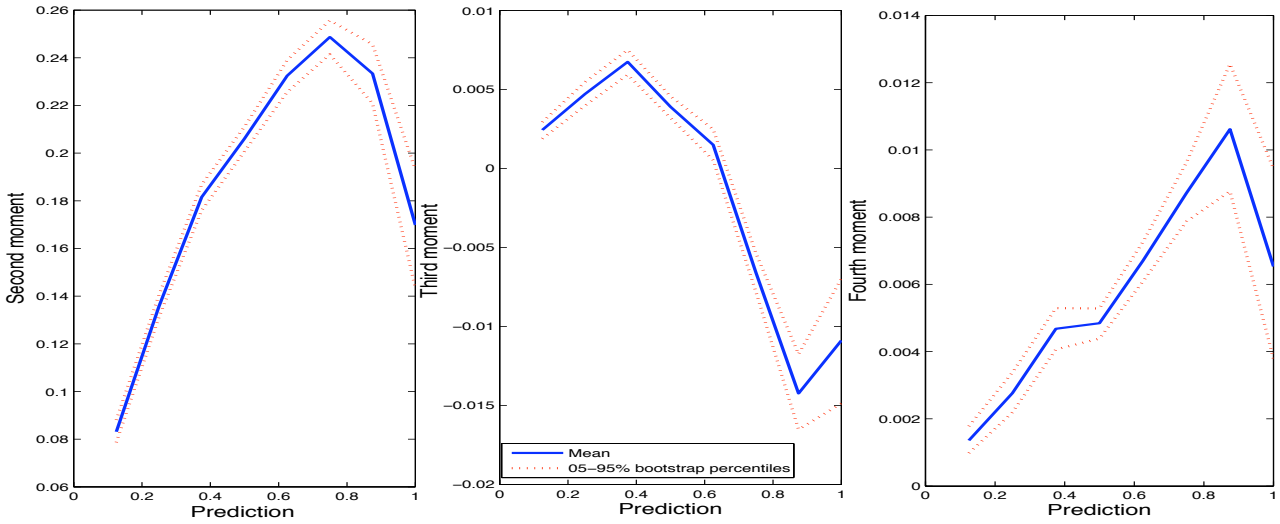


Figure 3: Mean of estimates of conditional moments for different horizons

2.2 Modeling the conditional central moments

The estimation of MEP density function make necessary the estimation of the jth conditional moments $m^{(j)}_{t+h|t}$. In this paper we propose the modelization of the conditional moments by means of adaptive regression models, with the independent variables being the powers of the point forecasts of wind power generation. The proposed model has the following form

$$e^{M}_{t+h|t} = \alpha_{0,t+h} + \alpha_{1,t+h}\widehat{p}_{t+h|t} + \alpha_{2,t+h}\widehat{p}^{2}_{t+h|t} + \alpha_{3,t+h}\widehat{p}^{3}_{t+h|t} + v_{t+h}, \tag{4}$$

where $M = 1, 2, ...$ In this way the different values of M are associated with different conditional moments, since the second conditional moment is associated with the square of the prediction error and so on.

The existing relation between the central moments of the wind power generation and its predicted values is checked by calculating the mean of the estimate of the central moments for different prediction horizons h. Figure 3 shows the bootstrap confidence intervals of the mean of the estimates. The central moments seem to have a similar pattern in function of the predicted values, it would apparently seem that the predicted values will be good regressor variable. The estimation of model (4) is done by means of the method of recursive least squares (see, for instance, Ljung and Söderstrom, 1983). The estimation is time adaptive, because a forgetting factor is used. This causes a progressive reduction in the importance of old data in the estimation. In this paper, a forgetting factor based on a recursive representation of Cook's distance is used (Sánchez, 2006).

3 Evaluation of predictive densities

Dawid (1984) proposed a method to evaluate if a sequence of predictive densities is suitable. The method is based on the probability integral transform (PIT) values which are defined by

$$u_{t+h} = \int_{-\infty}^{p_{t+h}} f_{t+h|t} dp_t. \tag{5}$$

If the sequence of the predictive densities $f_{t+h|t}$ coincides with the true underlying density, then the PIT values u_{t+h} will be a process independent and identically distributed according to uniform distribution $\mathrm{U}[0,1]$ (Rosenblatt, 1952). Figures 4 and 5 display the histogram of the PIT values obtained for the different estimated distributions for prediction horizons $h = 1, 6$. In this section the results showed are from Rokas wind farm. The results obtained for MEP distribution are compared with other parametric distributions: truncated normal, censored normal and beta.

Hamill (2001) showed that PIT method is a necessary condition to evaluate a sequence of predictive densities but it is not a sufficient condition. For this reason, a measure to compare different estimated predictive densities is needed. The great majority of existing measures are based on the Brier score (Brier, 1950). Let $F_{t+h|t}$ the cumulative distribution function and p_t the observed value, then the Brier score is defined by

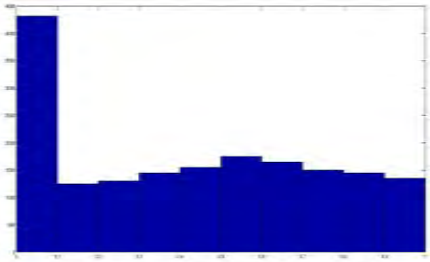
$$BS_h(x) = \frac{1}{N} \sum_{t=1}^{N-h} \left[F_{t+h|t} - H(x - p_{t+h}) \right]^2, \tag{6}$$

where $H(x - p_t)$ is the Heaviside function, which takes value 0 when $x < p_t$ and value 1 otherwise, and x is the domain of p_t. The Brier score gives a sequence of values, what makes complicated to evaluate which is the predictive density with lowest value. The continuous ranked probability score (CRPS) is a measure that assign the sequence of values given by BS and it is defined by (Matheson and Winkler, 1976)

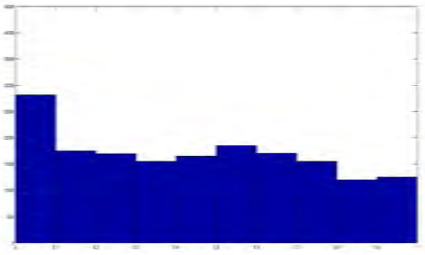
$$CRPS_h = \int_{-\infty}^{\infty} BS_{t+h|t}(x) \ \mathrm{d}x. \tag{7}$$

3.1 Choosing the best predictive density

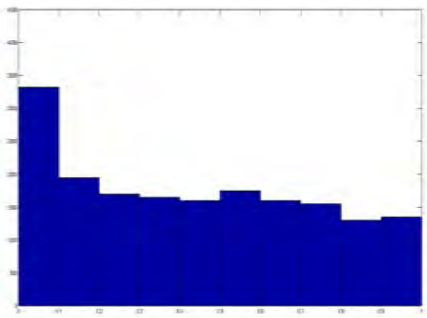
The nonlinear nature of the wind power data suggest that the best predictive distribution could change along the time. For this reason, we propose to use the predicted CRPS value to choose, among truncated normal, censored normal, beta and maximum entropy distributions, the best predictive



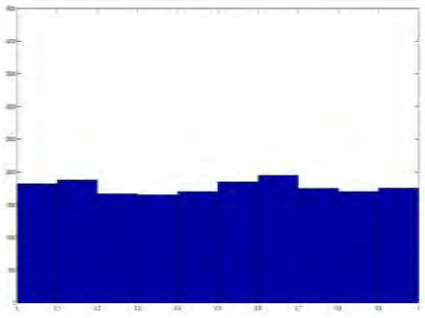
(a) Truncated normal



(b) Censored normal

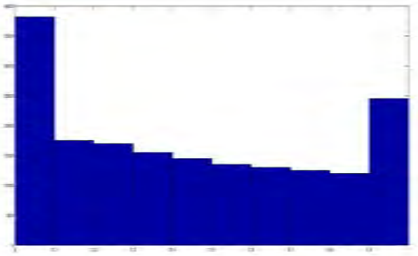


(c) Beta

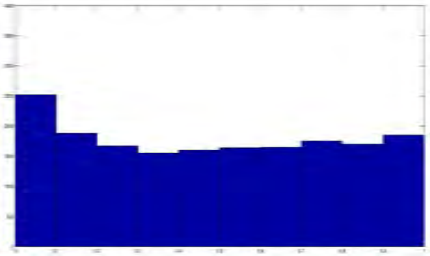


(d) MEP

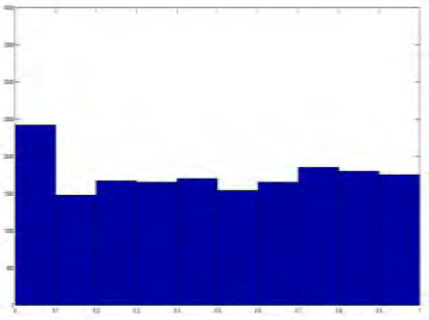
Figure 4: Histogram of the PIT values for horizon h=1.



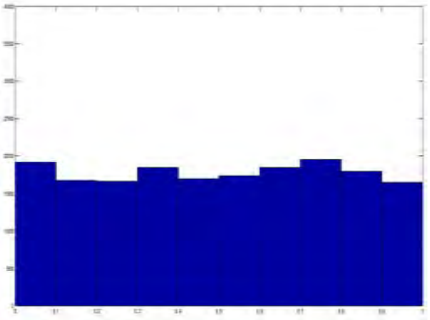
(a) Truncated normal



(b) Censored normal



(c) Beta



(d) MEP

Figure 5: Histogram of the PIT values for horizon h=6.

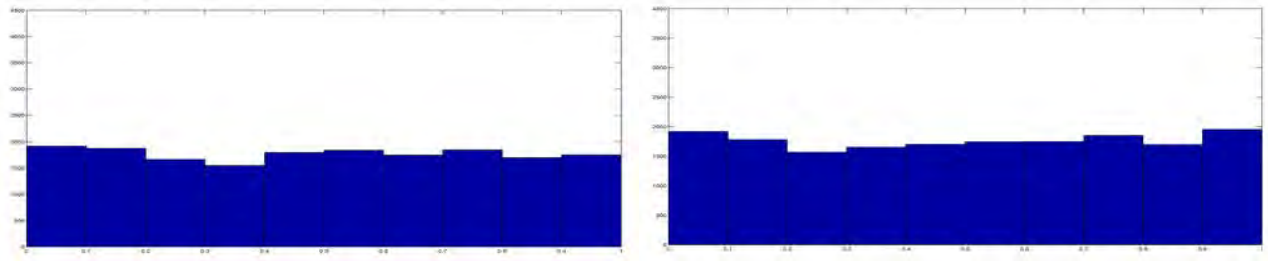
density in each t. Then, the minimum CRPS predictive density ($f_{t+h|t}^{min-CRPS}$) in $t+h$ is that one distribution that minimize the predicted CRPS, which is defined by

$$\widehat{CRPS}_{t+h|t}^{(D)} = \int_{-\infty}^{\infty} \widehat{BS}_{t+h|t}^{(D)}(x)\ \mathrm{d}x, \tag{8}$$

where

$$\widehat{BS}_{t+h|t}^{(D)}(x) = \left[F_{t+h|t}^{(D)} - H\left(x - \hat{p}_{t+h|t}\right)\right]^2, \tag{9}$$

where $F_{t+h|t}^{(D)}$ is the cumulative predictive distribution function and D can be the truncated normal, the censored normal, the beta or the MEP distribution. Figure 6 shows the histogram of the PIT values obtained for the minimum CRPS predictive density for different horizons. Table 1 shows the CRPS values for the different distributions. It can be seen that MEP and min-CRPS obtain the best results.



(a) h=1 (b) h=6

Figure 6: Histogram of the PIT values for minimum CRPS predictive density.

	h=1	h=6
Truncated Normal	0.1827	0.1883
Censored Normal	0.1402	0.1706
Beta	0.1206	0.1691
MEP	0.1113	0.1642
min-CRPS	0.1106	0.1639

Table 1: CRPS values by horizons $h = 1, 6$ obtained for the different used distributions.

References

[1] Akpinar S, Kavak Akpinar E. 2007. Wind Energy Analysis based on Maximum Entropy Principle (MEP)-type distribution function. *Energy Conversion and Management,* **48**: 1140-1149.

[2] Al-Awami AT, El-Sharkawi MA. 2009. Statistical Characterization of Wind Power Output for a given Wind Power Forecast. In *North American Power Symposum,* NAPS 2009.

[3] Bludszuweit H, Domínguez-Navarro JA, Llombart A. 2008. Statistical Analysis of Wind Power Forecast Error. *IEEE Transactions on Power Systems,* **23**: 983-991.

[4] Brier GW. 1950. Verification of forecasts expressed in terms of probability. *Monthly Weather Review* **78**: 1-3.

[5] Carta JA, Ramirez P, Velázquez S. 2009. 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.

[6] Dawid AP. 1984. The prequential approach. *Journal of the Royal Statistical Society, Series A* **147**: 278-292.

[7] Hamill TM. 2001. Interpretation of rank histograms for verifying ensemble forecasts. *Monthly Weather Reviews* **129**: 550-560.

[8] Jaynes ET. 1957. Information theory and statistical mechanics. *Physical Reviews* **106**: 620-630.

[9] Juban J, Fugon L, Kariniotakis, G. 2007. Probabilistic short-term wind power forecasting based on kernel density estimators. *Probabilistic wind power forecasting - European wind energy conference,* Milan, Italy.

[10] Lange M. 2005. On the Uncertainty of Wind Power Predictions: Analysis of the Forecast Accuracy and Statistical Distribution of Errors. *Journal of Solar Energy Engineering,* **127**: 177-184.

[11] Lau A, McSharry P. 2010. Approaches for Multi-Step Density Forecasts with Application to Aggregated Wind Power. *The Annals of Applied Statistics,* **4**: 1311-1341.

[12] Liu FJ, Chang TP. 2010. Validity analysis of maximum entropy distribution based on different moment constraint for wind energy assessment. *Energy,* doi:10.1016/j.energy.2010.11.033.

[13] Ljung L, Söderström T. 1983. *Theory and practice of recursive identification.* The MIT Press: Cambridge.

[14] Matheson JE, Winkler RL. 1976. Scoring rules for continuous probability distributions. *Management Science* **22**: 1087-1096.

[15] Nielsen HA, Madsen H, Nielsen TS, Badger J, Giebel G, Landberg L, Sattler K, Feddersen H. 2004. *Wind Power Ensemble Forecasting.* Technical report, Technical University of Denmark, DTU Informatics, Kgs. Lyngby, Denmark, 2004.

[16] Pinson P. (2010). *On probabilistic Forecasting of Wind Power Time-Series.* Technical report, Technical University of Denmark, DTU Informatics, Kgs. Lyngby, Denmark, 2010.

[17] Pinson P, Madsen H. 2009. Ensemble-based probabilistic forecasting at Horns Rev. *Wind Energy,* **12**: 137-155.

[18] Rosenblatt M. 1952. Remarks on a multivariate transformation. *The annals of mathematical statistic* **23**: 470-472.

[19] Sánchez I. 2006. Recursive estimation of dynamic models using Cook's distance, with application to wind energy forecast. *Technometrics* **48**: 61-73.

[20] Shannon CE. 1948. A mathematical theory of communication. *Bell System Technical Journal* **27**: 379-423, 623-656.

[21] Siddall JN, Diab Y. 1975. The use in probabilistic design of probability curves generated by maximizing the Shannon entropy function constrained by moments. *Journal of Industrial Engineering* **97**: 843-852.

[22] Taylor JW, McSharry PE, Buizza R. 2009. Wind Power Density Forecasting Using Wind Ensemble Predictions and Time Series Models. *IEEE Transactions on Energy Conversion,* **24**: 775-782.

WIND ENERGY
Wind Energ. 0000; **00**:1–14
DOI: 10.1002/we

RESEARCH ARTICLE

Forecasting Ramps of Wind Power Production with Numerical Weather Prediction Ensembles

Arthur Bossavy[1], Robin Girard[1], George Kariniotakis[1]

[1]MINES-ParisTech, Center for Energy & Processes,
Sophia Antipolis, France.

ABSTRACT

Today, there is a growing interest in developing short-term wind power forecasting tools able to provide reliable information about particular, so-called "extreme" situations. One of them is the large and sharp variation of the production a wind farm can experience within a few hours called *ramp event*. Developing forecast information specially dedicated to ramps is of primary interest both because of the difficulties usual models have to predict them, and the potential risk they represent in the management of a power system. First, we propose a methodology to characterize ramps of wind power production with a derivative filtering approach derived from the edge detection literature. Then, we investigate the skill of numerical weather prediction ensembles to make probabilistic forecasts of ramp occurrence. Through conditioning probability forecasts of ramp occurrence to the number of ensemble members forecasting a ramp in time intervals, we show how ensembles can be used to provide reliable forecasts of ramps with sharpness. Our study relies on 18 months of wind power measures from a 8 MW wind farm located in France and forecasts ensemble of 51 members from the Ensemble Prediction System (EPS) of the European Center for Medium-Range Weather Forecasts (ECMWF).

KEYWORDS

Wind power forecasting; extreme situations; ramp event; numerical weather prediction ensembles; renewable energies

Correspondence
MINES-ParisTech, Center for Energy & Processes,
Sophia Antipolis, France. E-mail: arthur.bossavy@mines-paristech.fr

Received . . .

1. INTRODUCTION

Wind energy is the fastest growing renewable energy source in the world. The total installed capacity has increased impressively during the last decade, passing from 10 GW in 1998 to 158 GW in 2009*. This is an encouraging development if we consider the issue of managing power systems with high penetration of wind power. Indeed, the variable nature of wind generation makes it difficult to reach the basic requirement of balancing the electricity demand by an equal production. Short-term forecasts of wind power production up to 2 or 3 days ahead can facilitate the management of power

*http://www.ewea.org

systems by operators. Wind power forecasts are useful for performing various power system management functions, like the dynamic quantification of reserves or the optimization of combined wind-hydro power plant scheduling [1, 2]. They also prove valuable when incorporated into bidding strategies for participating in electricity markets [3, 4].

Most existing wind power forecasting models are designed to provide point forecasts of expected future wind farm production. In recent years, research work has focused on associating uncertainty estimation with this type of point forecast. New forecasting approaches (often called probabilistic forecasting methods) have recently emerged that provide estimations of the entire distribution of future production. In such methods, forecasts may take the form of either quantile estimations [5, 6, 7] or density estimations [8, 9]. Two extensive reviews of the state of the art in wind power forecasting are given in [10, 11].

Nowadays, wind power forecasting systems are used operationally by end-users. It is increasingly apparent, however, that current forecasting technology cannot properly handle extreme situations related to wind generation. Extreme situations may take different forms, corresponding either to extreme weather phenomena or to critical periods for power system operation [12]. An example is the steep and high increase or drop in production that a wind farm can experience in the space of a few hours, commonly known as a *ramp event*. Operational systems' performance in forecasting large variations of wind power production is often reported to be significantly lower than usual [13, 14]. This means that we either need to improve forecasting tools' performance in these situations, or develop new dedicated forecasting tools. This paper investigates the development of a new tool to forecast ramps and the associated uncertainty.

In literature related to wind energy, the term "ramp" may refer to wind power variations taking place over different time scales. It sometimes refers to intra-hourly variations, e.g. 10 to 60 minute variations [15, 16, 17]. In this paper, we use it to refer to significant changes in wind generation over one or several hours. An accurate prediction of this type of variation for the next few hours up to several days ahead therefore relies on weather forecasts provided by Numerical Weather Prediction (NWP) models [13, 14, 18]. Making a reliable forecast of exactly when such events will occur is a significant challenge. Indeed, errors in predicting the underlying weather conditions made by NWP models often result in inaccurate forecasting of ramp timing [19]. An error in forecasting ramp timing is referred to as a *phase error*. There is a growing interest in developing forecasting approaches dedicated to ramps. A survey of these approaches is presented in [20].

The way standard forecasting approaches provide uncertainty estimations cannot be adapted when focusing on the temporal uncertainty of ramp events. For instance, prediction intervals are commonly provided on a per-horizon basis, and are dedicated to uncertainty estimations associated to power level forecasts (see [21] for instance). While such forecasts may be easily integrated into a decision-making process, they do not provide an explicit estimation of the temporal uncertainty associated with a particular ramp event. Pioneer work in [19] associates an estimation of the phase error distribution with point forecasts of wind power production. These estimations take the form of a probability density function with finite support. This kind of support is designed to contain all ramps. However, evaluation results indicate poor reliability since numerous ramps are observed outside the support. Another limitation of the methodology in [19] is the unconditional estimation of the phase error distribution.

In this paper, we propose to forecast ramps on the basis of a specific timing, plus the probability of observing a ramp within a set of time intervals (hereafter referred to as "prediction intervals") centered on that timing. Thus, our approach aims at providing a suitable uncertainty estimation associated with the forecast of a ramp timing. The proposed methodology relies on numerical weather prediction ensembles. Ensembles are alternative numerical weather predictions that are produced by perturbing the initial conditions, or a different parameterization of a numerical weather prediction model. The unperturbed prediction is referred to as control forecast and, in general, provides the most accurate wind production forecasts. Ensembles have already been used to derive probabilistic forecasts of wind generation [22, 23], and to assess future large-scale weather patterns [24]. Forecasts of wind power variations made by high-resolution ensembles have been evaluated to some extent in offshore conditions [25]. They have also been considered in a prototype ramp forecasting system [26]. In the latter, NWP ensembles are used to derive hourly distribution forecasts of wind generation. Then, statistical random sampling on two consecutive distribution forecasts makes it possible to estimate the distribution of production changes from one hour to the next. The main limitation of this methodology is that the correlation between

DOI: 10.1002/we
Prepared using weauth.cls

two consecutive wind generation distribution forecasts is not properly captured, resulting in unrealistic forecasts of hourly wind power changes. The methodology proposed here is completely different, since it is based on collecting ramp forecasts from the members of the wind power ensembles. These forecasts are then clustered into coherent groups, each of them resulting in a unique forecast event. The use of the ensembles from an event-based point of view allows us to go further in evaluating the merits of the ensembles themselves as weather forecast products [27].

The paper is organized as follows. In Section 2, we discuss the definition of a ramp and introduce a methodology to detect and characterize a ramp event from a wind power time series. The definition proposed relies on considerations from the signal processing field and uses a filtering/thresholding approach. In Section 3, we start by discussing how to identify a forecast ramp event among an ensemble of wind power forecast time series. Then, we introduce the proposed ramp forecasting methodology. Section 4 gives an evaluation of the proposed methodology. Conclusions and perspectives are given in Section 5.

2. DEFINITION AND CHARACTERIZATION OF A RAMP EVENT

In the related literature, a change in wind power production is considered as a ramp if it exceeds a given threshold in a period shorter than a given maximum duration. This threshold is sometimes expressed as a percentage of the nominal capacity of the wind farm. Thus, detecting a ramp relies on computing the difference between the minimum and maximum production during the considered period. This computing approach seems to be followed by [13, 14, 19], where it is explained in a descriptive way rather than using an explicit mathematical formulation. In general, this approach is expected to be highly noise sensitive given the high variability of wind power production. In this section, we propose a methodology for defining and detecting a ramp in a wind power time series, based on a linear filtering approach that aims to handle the above mentioned noise issue. As a result, we achieve a ramp characterization using a set of three parameters: the *support*, *timing* and *intensity* of the ramp.

2.1. Measuring time variations of wind power

The wind power production signal is characterized by high variability. For ramps, we need to propose a definition of such events that relies on a robust and computationally efficient way of measuring high and steep power variations, while appropriately handling the inherent noise in the process. Measuring variations, and detecting edges in a signal have been widely considered in the field of signal processing (for a literature overview we refer to [28]). The most common approach consists in estimating the first order derivative of the signal through filtering. The filtering considered associates the operations of smoothing and differentiating the signal. Following standard practice, we maintained a measure that would make interpretations easy and then chose to smooth the wind power signal using a moving average linear filter before computing the first order finite differences.

$$p_t^f = \frac{1}{n}\sum_{h=1}^{n} p_{t+h} - \frac{1}{n}\sum_{h=0}^{n-1} p_{t+h-n} \tag{1}$$

where p_t is a wind power time series and n is both the order of the moving average filter and the time step of finite differences. The filtered power p_t^f is merely the difference between the levels of average power from both sides of the instant t. Smoothing data allows us to regularize the differentiation operation, making it robust against noise [29]. The resulting variation measure can be rewritten by convolving the power signal p_t with the commonly known *difference of boxes* edge detector f_{DOB}

$$p_t^f = p_t * f_{DOB} \tag{2}$$

$$f_{DOB}(t) = \begin{cases} 0 & \text{if } t = 0 \text{ or } |t| > n \\ 1/n & \text{if } 0 < t \leq n \\ -1/n & \text{if } 0 > t \geq -n \end{cases} \tag{3}$$

Canny [30] introduced three criteria to evaluate edge detectors based on to their ability to detect and localize edges. Considering a model composed by an ideal step edge and white Gaussian noise, he theoretically showed the good performances of the considered edge detector f_{DOB} to detect step edges such as ramps.

In the literature on signal processing, the parameter n is generally introduced as a smoothing parameter controlling the trade off between noise reduction in the signal p_t and interference mitigation in detecting consecutive edges [31]. To avoid such interferences, n should be chosen to be no greater than the time period separating the two consecutive edges we are interested in. The choice of the value of n is described in more detail in the case-study presented in Section 4.1.

2.2. Ramp detection and characterization

Variations in the power signal p_t coincide with local extrema of the filtered power signal p_t^f (see Figure 1). Such extrema can be seen as a combined measure of both the magnitude and the steepness of the wind power variations. We consider that a variation is high and steep enough to be regarded as a ramp if the absolute value of p_t^f is higher than a given threshold $\tau > 0$

$$| p_t^f | \geq \tau \tag{4}$$

The complex geometry of a wind power variation may result in the absolute value of the filtered signal p_t^f fluctuating around the threshold τ. This phenomenon, which is responsible for breaking up edges, is known as "streaking" in literature on edge detection [30]. To palliate it, a common practice is to apply a so-called hysteresis thresholding. Rather than a single

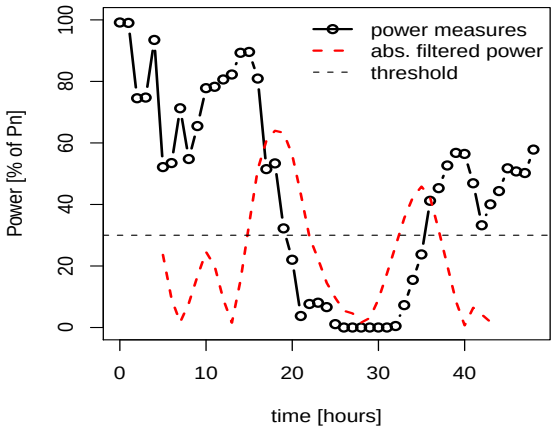


Figure 1. Example of an hourly wind power time series p_t covering 48 hours of production from a French wind farm. The measures have been normalized based on the wind farm nominal capacity P_n. The absolute value of the filtered production p_t^f (red dashed line) upwardly crosses the threshold τ (black dashed line) twice. First, a decreasing ramp is detected. Its timing is located at $t_r = 19$ hours, its intensity represents $64,0\%$ of the nominal capacity of the wind farm and its support is 6 hours long. It is followed by an increasing ramp at $t_r = 36$ hours, with an intensity 45.8% of P_n and a support of 4 hours.

Wind Energ. 0000; **00**:1–14 © 0000 John Wiley & Sons, Ltd.
DOI: 10.1002/we
Prepared using weauth.cls

value τ, two thresholds are then considered. Filtered outputs above the highest threshold are immediately considered as associated with an edge, as are those connected to them and lying between the two thresholds. In our study, we noticed that the streaking phenomenon occurs only rarely. For the sake of simplicity, we decided not to use this type of thresholding methodology.

A ramp is a time-limited event that can be characterized by a starting time t_s and an ending time t_e. They are defined here by the instants at which the absolute value of the filtered signal p_t^f crosses upwards and downwards the threshold τ. The period between these two instants $[\, t_s \, , \, t_e \,]$ then defines the *support* of the ramp. It is clear that the length of the support depends on the value of the threshold. In fact, ramp support should be considered as a requisite characteristic for developing our forecasting approach but is not essential to the definition of the event. Even when a ramp is not localized in time, it is useful to associate a particular timestamp to a ramp. We choose to associate to a ramp event the time t_r ($t_s \leq t_r \leq t_e$), called *timing* of the ramp, for which the absolute value of the filtered signal p_t^f reaches its maximal magnitude. If there are several maxima, the one associated to the nearest instant from the starting time t_s is chosen. This maximum defines the *intensity* of the ramp. To summarize, three quantities are defined to characterize a ramp event: support, timing and intensity.

3. FORECASTING RAMP EVENTS USING ENSEMBLE WIND POWER FORECASTS

The detection methodology of ramps that we introduced in the previous section relies on a filtering/thresholding approach applied to a measured wind power signal. In subsection 3.1, we discuss its use to detect ramps in a point forecast time series and then in an ensemble of forecast time series. In the latter case, an ensemble of ramp characteristics is produced, which may be related to different ramp events. In subsection 3.2, we propose solutions to cluster ramp forecasts obtained by different ensemble members into coherent groups. This enables us to discriminate between forecasts and associate them with separate events. The characteristics of the forecast events are then exploited in subsection 3.3 to produce probability forecasts of ramp occurrence in a set of prediction intervals.

3.1. Forecasting an ensemble of ramp characteristics

The methodology presented in the previous section was introduced for detecting ramps in a time series of measured wind power. Its application to a single or ensemble of wind power forecast time series is straightforward. In the case of ensembles, it is assumed that the methodology is applied to each member separately. However, it is not obvious whether the parameters, and namely the value of the threshold, used to detect ramps in a time series of measurements can be used as such to detect ramps in a time series of forecasts. Hereafter, we denote as $\hat{\tau}$ the threshold used to detect ramps in forecast time series.

Short-term wind power point forecasting models tend to overestimate production within a low wind speed range, and underestimate it at high wind speeds. This phenomenon, which is related to the wind-to-power conversion process, was demonstrated in [32] for five power prediction models based on "statistical" and "physical" approaches (with model output statistics correction), all using spot NWPs as input, and tuned to minimize the overall Root Mean Squared Error (RMSE). As a consequence, it is expected that the variations in power forecasts will on average be lower than those of the measurements. When such forecasting methods are applied to NWP ensembles, the resulting power ensembles may be poorly calibrated. In most applications, a form of recalibration is required [22, 23]. Here, in order to correctly identify ramps detected on both forecasts and observations as the same event, we can use a lower threshold value $\hat{\tau} \leq \tau$ to detect forecast ramps. One possible approach is to adjust this second threshold, so that the same amount of ramps are detected on both forecasts and observations. An alternative approach is to select the value of this threshold with respect to some other desirable properties of the considered ramp forecasting methodology. In this case, we need to keep in mind that increasing $\hat{\tau}$ decreases the number of forecast events, resulting in both less captured ramps and a reduced number of wrongly forecast

events. In Section 4, we investigate the performance of the proposed approach to forecast ramps for different values of this threshold.

Once we have chosen a suitable value for the threshold parameter $\hat{\tau}$, the filtering/thresholding approach can be applied to each member of the ensemble of wind power forecast time series. This procedure results in the production of an ensemble of ramp forecast characteristics (timing, support, intensity). These characteristics then need to be clustered into coherent groups, each of which corresponds to a unique forecast event.

3.2. Clustering an ensemble of ramp characteristics

When the ramp detection procedure is applied to each member of the forecast power ensembles, it is expected that the ramps detected using each member will correspond to the same ramp event when all or part of their characteristics (intensity, support and timing) are close enough to each other. Thus, for each lead time, we propose to compute the number of ensemble members with an absolute value of filtered power higher than $\hat{\tau}$. For a given lead time, this corresponds to the number of ensemble members with overlapping ramp supports. An example of the resulting signal is given in Figure 2, for respectively the increasing and decreasing forecast ramps.

Based on this signal, we propose two alternative approaches for clustering the ramp characteristics. In the first approach, denoted hereafter as **A1**, we cluster forecast ramps with overlapping supports into a unique event. In the second approach, denoted as **A2**, any local maximum in the number of overlapping ramp supports defines a single event. In the example of Figure 2, the first approach gives two forecast events: an increasing ramp followed by a decreasing one. On the other hand, the second approach gives three forecast events: an increasing ramp and two decreasing ones. Although the second approach may result in forecasting more ramps than observed, it may allow us to capture more events. The two approaches are evaluated in Section 4.

Once forecast events have been identified using one of the proposed clustering approaches, we count the number of forecasting members for each of them, taking care not to count each member more than once (e.g. a member could forecast the event two or more times because of the streaking phenomenon). We also forecast a mean timing $\bar{t}_r$ of the event by averaging the timings forecast by the ensemble members. If we consider the mean intensity, we end up obtaining a full characterization of the forecast event from the ensemble of forecast time series.

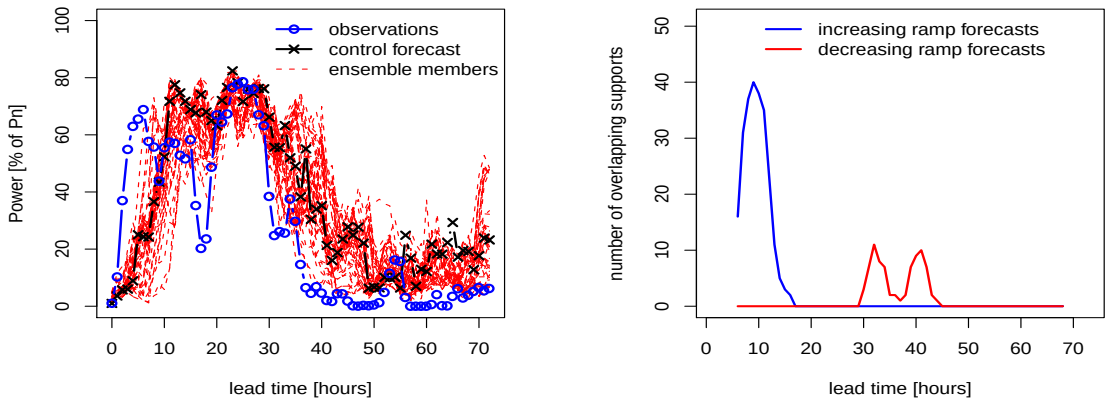


Figure 2. Example of an ensemble of wind power production forecasts composed by the control forecast and 50 members. The figure on the left depicts the hourly time series of measured and forecast wind power up to 72 hours ahead. The right-hand figure shows the number of ensemble members with overlapping ramp supports for increasing forecast ramps (blue curve) and decreasing forecast ramps (red curve). Ramps have been detected using a time scale parameter of $n = 5$ hours and a threshold parameter $\hat{\tau} = 30\%$ of the wind farm's nominal capacity P_n.

Wind Energ. 0000; **00**:1–14 © 0000 John Wiley & Sons, Ltd.
DOI: 10.1002/we
Prepared using weauth.cls

3.3. Making probabilistic forecasts of ramp occurrence using ensembles

In the proposed approach, we produce probabilistic forecasts of ramp occurrence using a set of prediction intervals $\{I_\delta,\, \delta = 1, \ldots, \delta_{max}\}$ centered on the average forecast timing $\bar{t}_r$, with associated probabilities p_δ. The intervals' radius δ are fixed beforehand and the probabilities of observing a ramp in these intervals are estimated. The maximum interval radius considered $\delta_{\max}$ is a parameter of the procedure which can be configured as a function of the end-user requirements. Unlike the proposal made in [19], characterizing the phase error distribution using a set of prediction intervals with associated probabilities allows us to account for the non-occurrence of a ramp.

To provide forecast information depending on forecast conditions, we propose to estimate p_δ conditioned to the number of ensemble members N_{mem} forecasting the ramp. For the sake of simplicity, and to produce parsimonious base-line models, we consider N_{mem} to be the only predictor. However, as a perspective of this work, one could imagine more complex models with additional inputs (e.g. average intensity forecast made by the ensemble members).

For a given interval's width 2δ, and since the center of the intervals is already forecast, it is possible to reformulate the problem of estimating p_δ in terms of a statistical regression problem. Consider the binary random variable Y_δ that equals 1 if a ramp occurs in I_δ and 0 otherwise. Thus, the probability $p_{\delta,m}$ of observing a ramp in I_δ given that m ensemble members are forecasting the ramp is given by the following regression formula

$$\begin{aligned} p_{\delta,m} &= \mathbb{P}(\, Y_\delta = 1 \,|\, N_{mem} = m \,) \\ &= \mathbb{E}[\, Y_\delta \,|\, N_{mem} = m \,] \end{aligned} \tag{5}$$

The estimation procedure Estimating $p_{\delta,m}$ as given by Equation (5) is a particular instance of a regression problem. As a consequence, there are numerous alternative estimation procedures that could be used here. In our case however, there are some particular properties that should be taken into account:

- $p_{\delta,m}$ is a probability and there are regression models dedicated to this kind of estimation problem.
- The distribution of the random variable N_{mem} used as an explanatory variable is strongly right-skewed. In other words, a ramp is more likely to be forecast by a small rather than a large number of ensemble members.

Based on these considerations, we choose to apply two alternative statistical estimation procedures, which are evaluated in Section 4.3. First, a Nadarya-Watson estimator is used with a varying bandwidth to introduce flexibility into our estimations and to adapt to the particular sampling of the explanatory variable N_{mem}. Then, we consider the alternative of a logistic regression model specifically dedicated to estimating probabilities.

Nadarya-Watson estimator A Nadarya-Watson estimator makes estimations according to the following generic formula

$$p_{\delta,m} = \frac{\sum_{i=1}^{n} \mathcal{K}_\lambda(m_i, m) Y_\delta^i}{\sum_{i=1}^{n} \mathcal{K}_\lambda(m_i, m)} \tag{6}$$

where n is the size of the data set dedicated to probability estimations and $\mathcal{K}_\lambda$ is a kernel function. Regression techniques using kernels fit a local model around each target point m. The value $\mathcal{K}_\lambda(m_i, m)$ gives the weighting of points m_i in the neighborhood of the target point m. The kernel's bandwidth λ gives the size of such a neighborhood and controls the estimation bias-variance trade off. Increasing λ results in averaging more observations and lowers the variance (but increases the bias) in each target point estimation. These techniques are very popular as they do not require any parametric assumption. In our study, we consider a tricube kernel and select λ by making it vary with m according to a nearest-neighbors procedure

$$\lambda(m) = |m - m_{[k]}| + 1 \tag{7}$$

where $m_{[k]}$ is the kth closest point from the target m. The parameter k is selected from a 10-fold cross-validation procedure. A local choice of λ allows us to incorporate more information about the probability law generating data points m_i and often gives better results when estimating curves with a complex shape [33]. For an overview on kernel smoothers and bandwidth selection procedures we refer to [34, 35].

Logistic regression The logistic regression model allows us to infer ramp occurrence probabilities from a linear function of the number of ensemble members m forecasting a ramp

$$\log \frac{p_{\delta,m}}{1-p_{\delta,m}} = \alpha_\delta m + \beta_\delta \tag{8}$$

where α_δ and β_δ are the model parameters to estimate. To ensure that the estimated probabilities remain within $[0,1]$, the modeling requires us to express such probabilities as a linear function of m through a logit transformation. The model is fit by maximum likelihood, assuming the conditional distribution of Y_δ given N_{mem} is binomial. For all values of δ, a statistical test (the Wald test) rejects the hypothesis that assumes a null value of coefficients α_δ and β_δ with a significance level of 5%. For an introduction to logistic and generalized linear models, we refer to [34, 36].

4. EVALUATION FRAMEWORK AND RESULTS

4.1. The Case-Study

In order to evaluate the proposed methodology, we considered a wind farm with a nominal capacity of 8MW, located on a complex terrain in the south of France (a few dozen kilometers from the Mediterranean Sea). Power measurements delivered by the SCADA system of the wind farm were considered. They covered a period of 18 months from July 2004 to December 2005 with a 10-min temporal resolution. For the purpose of the study, the data was averaged to hourly values.

In our study, ramp events correspond to variations detected with a filter (half) width of $n = 5$ hours. Such a value is reasonable if variations under study are related to somewhat persistent changes in weather conditions. Appropriate choices could be made by end-users focusing on shutdown events, for instance[†]. In this case, a lower value of n would be required. Further work could include a multi-scale analysis of the wind power signal using multi-scale edge detection techniques [28]. For a variation to be considered as a ramp, its minimum intensity τ has to be set according to which proportion of the most "extreme" (e.g. highest and steepest) variations we are interested in. We consider a value of $\tau = 30\%$ of the nominal capacity P_n of the considered wind farm. With such a value, we detected 168 ramps during the period between 1st April 2005 and 31st December 2005, that is an average 4.2 ramps a week. Note that our methodology is independent of the value of τ and a more specific choice could be made by end-users. As mentioned in Section 3.1, we will now investigate the performance of our approach to forecasting ramps considering different threshold values $\hat{\tau}$ to detect ramps on wind power forecast ensembles.

NWP ensembles of 51 members are provided by the EPS system of ECMWF with a spatial resolution of 1° in both longitude and latitude. This resolution corresponds to about $75-80$ km in the East-West direction and 110 km in the North-South direction. The NWP data is issued twice a day with a temporal resolution of 6 hours and for a horizon of 72 hours. Hourly predictions were obtained through interpolating the 6-hour values. To forecast wind power, we considered weather forecasts generated at the closest grid point to the wind farm located 10 meters above ground level. We used the *Random Forest* nonparametric estimation procedure to produce wind power forecasts using wind speed and direction NWPs as input. *Random Forest* is a machine-learning algorithm that aggregates predictions from a large number of regression trees, each tree being built from a bootstrap replica of the sample data and random selection of potential

[†]In situations with very high wind speeds, some or all of a wind farm's turbines can be temporary stopped to avoid damage due to strong mechanical constraints. Such situations can turn into high variations of decreasing and increasing wind power production within a short period of time

DOI: 10.1002/we
Prepared using weauth.cls

predictors. Recent work on meteorological downscaling and post-processing acknowledges the robustness of *Random Forest* against over-fitting and its ability to deal with nonlinearity and complex interactions between variables [37, 38]. For more details we refer to [34, 39]. To deal with ensembles, we applied the Random Forest procedure to each member of the NWP ensemble to obtain a corresponding power forecast ensemble.

Our wind power forecasting model was trained over a nine-month period (July 2004 - March 2005) and tested for another nine months (April to December 2005). In the testing set, the model provides a root mean squared error and a mean absolute error not greater than respectively 15% and 11% of the nominal capacity of the wind farm for the next 24h. This is a satisfactory performance given the state of the art [40], especially if one considers the complexity of the terrain in which the wind farm is embedded and the rather coarse resolution of the NWP data. The data sets used for training and testing the ramp forecasting models were four and a half months long each, covering the 9-month period between April and December 2005. We derive and hereafter evaluate ramp forecasts taking prediction intervals with a maximal radius value of $\delta_{\max} = 8$ hours. With such a value, we expect to describe most of the phase error distribution from the considered set of prediction intervals.

4.2. Evaluating the capture of ramp events from ensemble-based forecasts

In this section we focus on evaluating the rate of observed ramps that are effectively forecast, denoted as *ramp capture* ratio [19].

Figure 3 gives this ratio for ensemble-based forecasts and forecasts derived from the control member. As noticed in Section 3.1, predicting ramps from forecast variations of lower intensity (e.g. by decreasing $\hat{\tau}$), results in capturing more events. This property is valid when predicting ramps from a single wind power scenario, but may not hold when considering ensembles. Actually, it depends on the approach used to cluster forecasts issued by different ensemble members. When using the approach **A1**, low values of $\hat{\tau}$ result in ramp supports forecast by the ensemble to overlap, and ultimately in separated forecast events to be merged. This is the reason why in Figure 3, for $\hat{\tau} \leq 20\%$ of P_n and with clustering approach **A1**, the ramp capture decreases with $\hat{\tau}$.

As seen in Figure 3, ensembles generally outperform the control member in terms of ramp capture. However, such improvements may coincide with a higher number of false alarms, e.g. events that are forecast but not observed. In fact, the use of ensembles may not result in a better ramp capture/forecast accuracy trade off. Alternative approaches to cluster

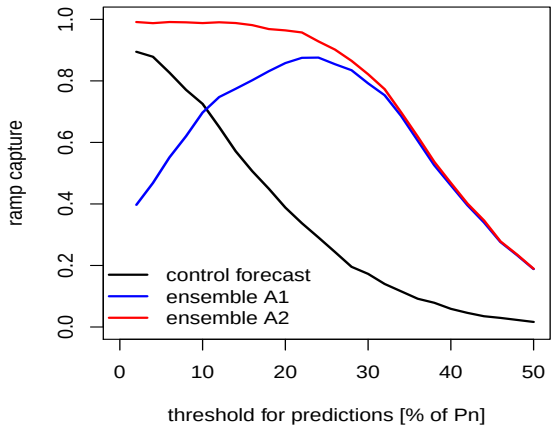


Figure 3. Ramp capture results when forecasting ramps using the control forecast (black curve) or the ensembles (blue curve is for clustering approach **A1**, red curve for approach **A2**). Results are for values of $\hat{\tau}$ ranging from 2% to 50% of the wind farm nominal capacity P_n, and a maximum delay between forecast and observed ramp events of $\delta_{\max} = 8$ hours.

ramp forecasts issued by different ensemble members could be envisaged. We have considered other clustering approaches, such as hierarchical agglomerative clustering with an euclidean distance [34]. However, such clustering approaches turn out to be far more computationally demanding when the number of ramp forecasts to cluster is increased (e.g. when decreasing $\hat{\tau}$).

4.3. Evaluating ensemble-based probability forecasts of ramp occurrence

Here we evaluate the probability forecasts for ramp occurrence. We first propose using the Brier score (BS) [41], which is a dedicated measure of accuracy. Consider $p_i = p^i_{\delta,m}$ as a probability forecast of ramp occurrence in a prediction interval I_δ. Let $y_i = y^i_\delta$ be equal to one if a ramp is observed in I_δ and zero otherwise, then

$$\mathrm{BS} = \frac{1}{n}\sum_{i=1}^{n}(p_i - y_i)^2 \tag{9}$$

where n is the test set sample size. As a reference approach, we consider the climatology which, for a given interval's width 2δ, derives unconditional probability forecasts from the empirical event frequency $\overline{y} = 1/n \sum_{i=1}^{n} y_i$. The relative improvement in terms of the decrease in forecasting errors, referred to as Brier skill score with respect to climatology (BSS_{clim}), is given in Table I. Conditioning clearly improves forecast accuracy. Such improvement is paramount for low values of $\hat{\tau}$, e.g. for high ramp capture. When focusing on the approach **A1**, such improvement also increases with the size of prediction intervals I_δ, which then results in better performances compared to the approach **A2** for large values of δ (e.g. $\delta = 8$ hours in Table I). However, the sampling uncertainty in the estimation of BSS_{clim} makes any further discrimination between approaches **A1** and **A2** difficult. For the same reason, it is not easy to conclude in favor of a particular estimation procedure between the logistic regression model and the Nadarya-Watson estimator.

Table I. Brier skill score with respect to climatology BSS_{clim} (in percentages) of ensemble-based probability forecasts of ramp occurrence in prediction intervals I_δ (δ is in hours). Results are given for both the clustering approaches **A1** and **A2**, and for the logistic regression (log. reg.) and the Nadarya-Watson (N.W.) estimation procedures. To account for sampling uncertainty, the standard deviation of BSS_{clim} (given in parenthesis) has been estimated following the theoretical work of Bradley et al. [42].

	log. reg.			N.W.		
$\hat{\tau}$ ($\%P_n$)	$\delta = 2$	$\delta = 5$	$\delta = 8$	$\delta = 2$	$\delta = 5$	$\delta = 8$
			Clustering approach **A1**			
10	5.5 (1.2)	11.1 (1.7)	16.6 (2.2)	5.4 (1.1)	10.8 (1.5)	16.6 (2.1)
20	3.8 (1.2)	9.9 (1.7)	14.3 (2.1)	4.5 (1.0)	9.8 (1.4)	13.8 (1.8)
30	3.7 (1.2)	8.0 (1.7)	9.5 (1.9)	4.2 (1.1)	8.2 (1.5)	9.4 (1.7)
40	2.9 (1.3)	1.5 (2.6)	3.3 (2.4)	2.8 (1.1)	3.7 (2.5)	4.8 (2.3)
			Clustering approach **A2**			
10	6.0 (1.0)	10.2 (1.3)	9.2 (1.3)	6.2 (0.9)	10.9 (1.3)	9.9 (1.3)
20	5.8 (1.1)	9.9 (1.4)	8.9 (1.4)	6.1 (1.0)	9.7 (1.3)	8.6 (1.4)
30	3.9 (1.2)	6.9 (1.3)	6.3 (1.4)	3.9 (1.2)	7.0 (1.4)	5.9 (1.5)
40	3.2 (1.21)	1.0 (1.8)	1.2 (1.7)	3.2 (1.3)	2.8 (2.0)	0.9 (2.1)

In addition to using the Brier score, we evaluated our forecasts according to a reliability/sharpness paradigm [43, 44]. Results for a particular threshold value $\hat{\tau} = 30\%$ of the wind farm nominal capacity, using the clustering approach **A2** and the logistic regression model, are shown in Figure 4. Forecasts show reliability (top left Figure), which does not deviate from perfect reliability (first diagonal) any more than the sampling uncertainty does. The latter was estimated with 90% confidence intervals (vertical segments) derived from the resampling technique proposed by Broecker and Smith [45]. An alternative was to derive them analytically considering the Poisson-Binomial distribution of $n\overline{y}$ [46]. The results from the two approaches were found to be very close.

Wind Energ. 0000; **00**:1–14 © 0000 John Wiley & Sons, Ltd.
DOI: 10.1002/we
Prepared using weauth.cls

The top right and bottom plots of Figure 4 show the sharpness property of our forecasts. We can see that the probability of a ramp occuring in prediction intervals increases with the number of members forecasting it. At the same time, ramps are seldom forecast with high probability. We quantified reliability and sharpness using measures derived from the decomposition of the Brier score proposed by Murphy [47]. Results showed that the decreasing skill of our forecasts (see Table I) while $\hat{\tau}$ increases was due to the simultaneous decrease in reliability and sharpness. In any case, high values of $\hat{\tau}$ (e.g. $\hat{\tau} > 40\%$ of P_n) make it possible to mitigate false alarms, which may be desirable when making deterministic forecasts of ramps. Following our probabilistic approach, the choice should be for lower threshold values leading to (sometimes large numbers of) forecasts with good reliability and sharpness.

To illustrate the forecast product of the proposed methodology, the 3 days ahead forecast scenario given in Figure 2 is represented in Figure 5, with prediction intervals and associated probability forecasts of ramp occurrence. Forecasts were derived from the clustering approach **A2** and the logistic regression model. As we described in Section 3.2, the situation considered should present two decreasing ramp forecasts. For more visibility, however, we removed one of them.

5. CONCLUSIONS AND PERSPECTIVES

There is growing demand for developing wind power forecasting models focusing on certain extreme situations. One well-known situation is the so-called ramp event, which involves sharp and high variations in production experienced by wind farms in the space of a few hours. These variations are critical for managing power systems with high wind penetration, as they can imply complex changes in power supply. For a power system operator, an estimation of the uncertainty of a forecast ramp should provide valuable information for making management decisions.

The first challenge in ramp forecasting is to correctly define a ramp. We proposed a definition based on filtering, guided by the significant signal processing literature on edge detection. Such a definition addresses several issues inherent to defining a ramp: at what level of power and within what duration does a variation represent a ”ramp” for the end-user? We discussed these questions and proposed a detection methodology using two parameters that make it possible to set at what time scale and magnitude a variation is considered to be a ramp. Such parameters can be set depending on either the modeler's or the end-user's opinion.

A single time series of wind power forecasts up to several days may not include enough information to make secure management decisions related to the potential occurrence of a ramp. In this paper, we studied the extent to which numerical weather prediction ensembles could provide information when forecasting ramps with associated uncertainty. We showed that ensembles used with our methodology can better capture ramp events than when considering a unique wind power scenario.

Then, we set out to estimate the uncertainty associated with ramp forecasts using a set of prediction intervals with associated probabilities of ramp occurrence. We introduced sharpness into our forecasts by conditioning them to the number of ensemble members forecasting a ramp. The resulting forecasts turned out to be reliable with greater accuracy regarding climatology. Our conclusions relate to data from a multi-megawatt wind farm located on a complex terrain in the south of France, and numerical weather prediction ensembles of 51 members from the EPS system of ECMWF. In further work, an evaluation of the methodology on more sites should help to validate our conclusions under different weather regimes.

An interesting perspective would be to compare the forecasting ramp performance of different wind power forecast ensembles. Indeed, the methodology we developed in this paper could be used to forecast ramps from either the commonly known *poor man's* ensembles, or from ensembles derived from multivariate statistical sampling, such as those proposed in [48]. The latter could provide a good alternative to numerical weather prediction ensembles for forecasting ramps, as their generation can be far less computationally expensive. Particular attention should also be paid to future potential applications of the proposed approach in forecasting wind generation ramps from a portfolio of wind farms. Finally, we

observed that different ramp typologies occur, and the limits of a single ramp definition will need to be explored in further work.

ACKNOWLEDGEMENTS

The work presented in this paper is part of the European R&D Project *SafeWind* (Grant No 213740) partly funded by the European Commission under the 7th Framework Programme. One of the aims of this project is to facilitate large-scale wind integration through developing advanced wind modeling and forecasting methods, with an emphasis on extreme situations. This work is also supported by *Agence de l'Environnement et la Maitrise de l'Energie* (ADEME). We thank EDF for providing some of the data used in this work, and Erik Holmgren for his support. Finally, we would like to thank the anonymous referees who reviewed this paper for their comments and suggestions.

REFERENCES

1. R. Doherty and M. O'Malley. A new approach to quantify reserve demand in systems with significant installed wind capacity. *Power Systems, IEEE Transactions on*, 20(2):587 – 595, 2005.
2. E.D. Castronuovo and J.A.P. Lopes. On the optimization of the daily operation of a wind-hydro power plant. *Power Systems, IEEE Transactions on*, 19(3):1599 – 1606, 2004.
3. J.M. Angarita and J.G. Usaola. Combining hydro-generation and wind energy: Biddings and operation on electricity spot markets. *Electric Power Systems Research*, 77(5-6):393 – 400, 2007.
4. P. Pinson, C. Chevallier, and G. Kariniotakis. Trading wind generation from short-term probabilistic forecasts of wind power. *Power Systems, IEEE Transactions on*, 22(3):1148 –1156, 2007.
5. B.J. Bremnes. A comparison of a few statistical models for making quantile wind power forecasts. *Wind Energy*, 9(1):3–11, 2006.
6. J.K. Moller, H.Aa. Nielsen, and H. Madsen. Time-adaptive quantile regression. *Computational Statistics & Data Analysis*, 52(3):1292 – 1303, 2008.
7. H.Aa. Nielsen, H. Madsen, and T.S. Nielsen. Using quantile regression to extend an existing wind power forecasting system with probabilistic forecasts. *Wind Energy*, 9(1-2):95–108, 2006.
8. J. Juban, N. Siebert, and G. Kariniotakis. Probabilistic short-term wind power forecasting for the optimal management of wind generation. In *Proceedings of the IEEE Power Tech Conference, Lausanne, Swizerland*, 2007.
9. J.W. Taylor, P.E. McSharry, and R. Buizza. Wind power density forecasting using ensemble predictions and time series models. *Energy Conversion, IEEE Transactions on*, 24(3):775 –782, 2009.
10. A. Costa, A. Crespo, J. Navarro, G. Lizcano, H. Madsen, and E. Feitosa. A review on the young history of the wind power short-term prediction. *Renewable and Sustainable Energy Reviews*, 12(6):1725 – 1744, 2008.
11. G. Giebel, G. Kariniotakis, and R. Brownsword. The state of the art on short-term wind power prediction - A literature overview. Technical report, ANEMOS EU project, deliverable report D1.1. [Available online: http://www.anemos-project. eu], 2003.
12. P. Pinson. Catalogue of complex to extreme situations. Technical report, EU Project SafeWind, Deliverable Dc1.2. [Available online: http://www.safewind. eu], 2009.
13. N.J. Cutler, M. Kay, K. Jacka, and T.S. Nielsen. Detecting, categorizing and forecasting large ramps in wind farm power output using meteorological observations and WPPT. *Wind Energy*, 10(5):453–470, 2007.
14. N.J. Cutler. *Characterizing the uncertainty in potential large rapid changes in wind power generation*. PhD thesis, Electrical Engineering & Telecommunications, Faculty of Engineering, UNSW. [Available online: http://handle.unsw.edu.au/1959.4/43570], 2009.

15. H. Zheng and A. Kusiak. Prediction of wind farm power ramp rates: A data-mining approach. *Journal of solar energy engineering*, 131:031011.1–031011.8, 2009.
16. H. Zareipour. Wind power ramp events classification and forecasting: A data mining approach. In *Proceedings of the 2011 IEEE Power and Energy Society (PES) Annual General Meeting, Detroit, USA*, 2011.
17. C. Kamath. Associating weather conditions with ramp events in wind power generation. In *IEEE PES Power Systems Conference & Exposition, Phoenix, Arizona*, 2011.
18. N. J. Cutler, H. R. Outhred, I. F. MacGill, M. J. Kay, and J. D. Kepert. Characterizing future large, rapid changes in aggregated wind power using Numerical Weather Prediction spatial fields. *Wind Energy*, 12(6):542–555, 2009.
19. B. Greaves, J. Collins, J. Parkes, and A. Tindal. Temporal forecast uncertainty for ramp events. *Wind Engineering*, 33(11):309–319, 2009.
20. C. Ferreira, J. Gama, L. Matias, A. Botterud, and J. Wang. A survey on wind power ramp forecasting. A report from the Argonne U.S. Department of Energy Laboratory. [Available online at http://www.dis.anl.gov/], 2010.
21. P. Pinson and G. Kariniotakis. Conditional prediction intervals of wind power generation. *Power Systems, IEEE Transactions on*, 25(4):1845 –1856, 2010.
22. H.Aa. Nielsen, T.S. Nielsen, H. Madsen, J. Badger, G. Giebel, L. Landberg, K. Sattler, L. Voulund, and J. Tøfting. From wind ensembles to probabilistic information about future wind power production - Results from an actual application. In *Proceedings of the IEEE PMAPS 2006 Conference, Probabilistic Methods Applied to Power Systems, Stockholm, Sweden*, 2006.
23. P. Pinson and H. Madsen. Ensemble-based probabilistic forecasting at Horns Rev. *Wind Energy*, 12(2):137–155, 2009.
24. P.A. Chessa and F. Lalaurette. Verification of the ECMWF ensemble prediction system forecasts: A study of large-scale patterns. *Weather and Forecasting*, 16(5):611–619, 2001.
25. WEPROG ApS. High resolution ensemble for Horns Rev - HRensembleHR -. Technical report, Project funded by the Danish PSO F&U Program, Final Report. [Available online: http://www.hrensemble.net], 2010.
26. E. Grimit and C. Potter. A prototype day-ahead forecast system for rapid wind ramp events. In *Proceedings of Windpower 2008 Conference and Exhibition, Houston, Texas*, 2008.
27. R. Girard and P. Pinson. Evaluation of time trajectories - Application to wind power forecasting. Submitted to Applied Energy, 2011.
28. D. Ziou and S. Tabbone. Edge detection techniques - An overview. *International Journal of Pattern Recognition and Image Analysis*, 8:537–559, 1998.
29. V. Torre and T.A. Poggio. On edge detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, PAMI-8(2):147 –163, 1986.
30. J. Canny. A computational approach to edge detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, PAMI-8(6):679 –698, 1986.
31. D. Demigny. On optimal linear filtering for edge detection. *Image Processing, IEEE Transactions on*, 11(7):728 – 737, 2002.
32. P. Pinson. *Estimation of the uncertainty in wind power forecasting*. PhD thesis, MINES ParisTech. [Available online: http://pastel.paristech.org], 2006.
33. J. Fan and I. Gijbels. Variable bandwidth and local linear regression smoothers. *Annals of Statistics*, 20:2008–2036, 1992.
34. T. Hastie, R. Tibshirani, and J. Friedman. *The elements of statistical learning, second edition: Data mining, inference, and prediction*. Springer Series in Statistics. Springer, 2nd ed. 2009. corr. 3rd printing edition, 2009.
35. W. Schucany. Kernel smoothers: An overview of curve estimators for the first graduate course in nonparametric statistics. *Statistical Sciences*, 19:663–675, 2004.
36. H. Madsen and P. Thyregod. *An introduction to general and generalized linear models*. Chapman & Hall, 2010.

37. E. Eccel, L. Ghielmi, P. Granitto, R. Barbiero, F. Grazzini, and D. Cesari. Prediction of minimum temperatures in an alpine region by linear and non-linear post-processing of meteorological models. *Nonlinear Processes in Geophysics*, 14(3):211–222, 2007.
38. R. Davy, M. Woods, C. Russell, and P. Coppin. Statistical downscaling of wind variability from meteorological fields. *Boundary-Layer Meteorology*, 135:161–175, 2010.
39. L. Breiman. Random Forests. *Machine Learning*, 45:5–32, 2001.
40. G. Kariniotakis, I. Marti, D. Casas, P. Pinson, T.S. Nielsen, H. Madsen, G. Giebel, J. Usaola, I. Sanchez, A.M. Palomares, R. Brownsword, J. Tambke, U. Focken, M. Lange, P. Loucka, G. Kallos, C. Lac, G. Sideratos, and G. Descombes. What performances can be expected by short-term wind power prediction models depending on site characteristics? In *Proceedings of the 2004 European Wind Energy Conference EWEC'04, London, UK*, 2004.
41. G. Brier. Verification of forecast expressed in terms of probability. *Monthly weather review*, 78:1–3, 1950.
42. A. Bradley, S.S. Schwartz, and T. Hashino. Sampling uncertainty and confidence intervals for the Brier score and Brier skill score. *Weather and Forecasting*, 23(5):992–1006, 2008.
43. T. Gneiting, F. Balabdaoui, and A. E. Raftery. Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(2):243–268, 2007.
44. P. Pinson, H. Aa. Nielsen, J. K. Moller, H. Madsen, and G. Kariniotakis. Non-parametric probabilistic forecasts of wind power: required properties and evaluation. *Wind Energy*, 10(6):497–516, 2007.
45. J. Broecker and L.A. Smith. Increasing the reliability of reliability diagrams. *Weather and Forecasting*, 22(3):651–661, 2007.
46. Y. Hong. On computing the distribution function for the sum of independent and non-identical random indicators. Technical report, Department of Statitics, Virginia Tech, Blacksburg, VA, 2011.
47. A.H. Murphy. A new vector partition of the probability score. *Journal of Applied Meteorology*, 12:595–600, 1973.
48. P. Pinson, H. Madsen, H.Aa. Nielsen, G. Papaefthymiou, and B. Klöckl. From probabilistic forecasts to statistical scenarios of short-term wind power production. *Wind Energy*, 12(1):51–62, 2009.

DOI: 10.1002/we
Prepared using weauth.cls

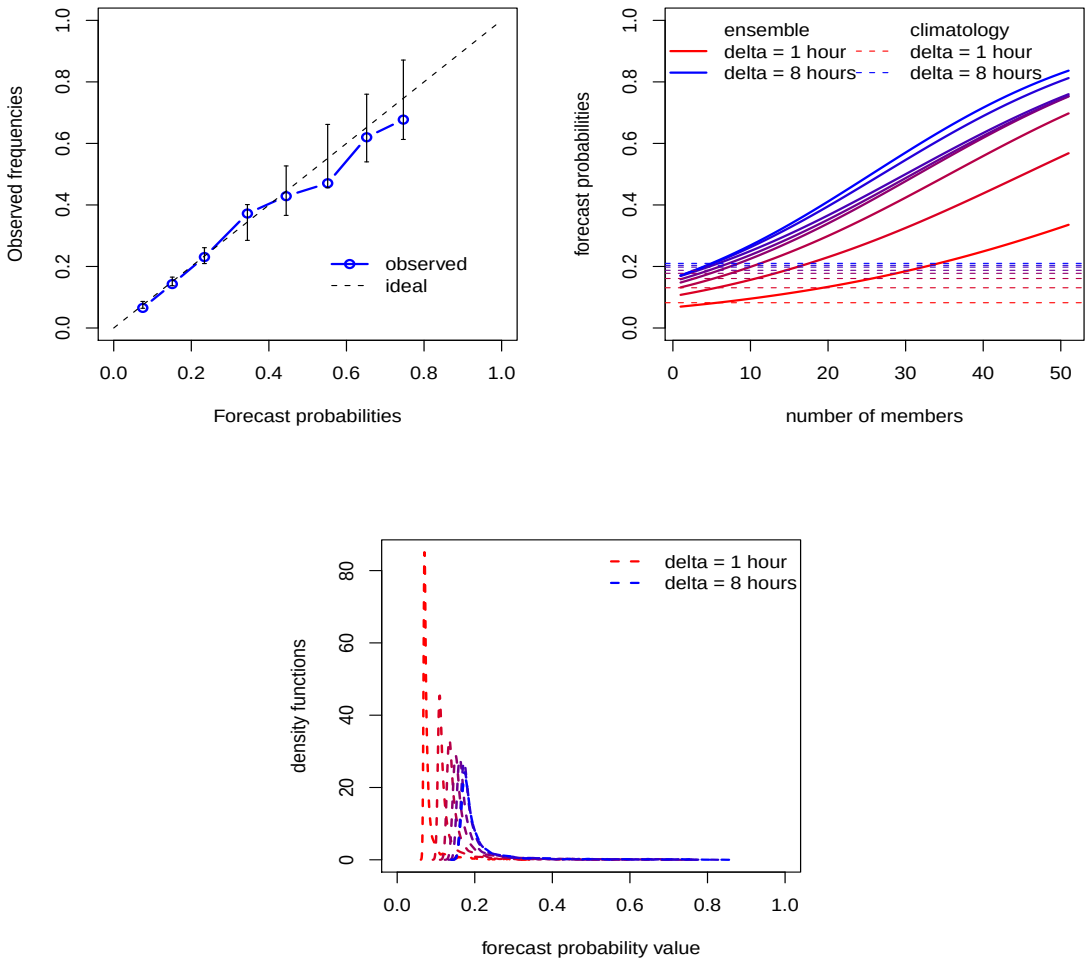


Figure 4. Reliability (top left Figure) and sharpness (top right and bottom Figures) of ensemble-based probability forecasts of ramp occurrence in prediction intervals I_δ. Results are for forecasts derived from the clustering approach **A2**, with $\hat{\tau} = 30\%$ of P_n and the logistic regression model. To take the sampling uncertainty into account when estimating the event frequencies (top left Figure), 90% centered confidence intervals were derived using the resampling technique proposed in [45].

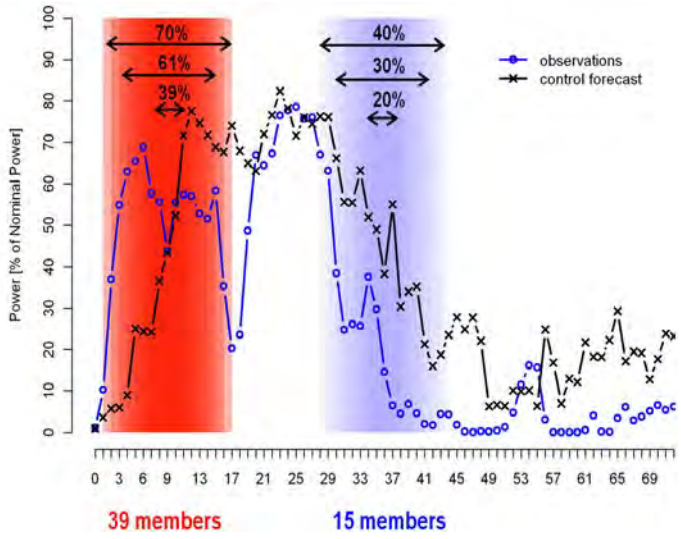


Figure 5. Prediction intervals $I_\delta, \delta = 1, \ldots, 8$ hours for an increasing ramp forecast by 39 members, followed by a decreasing one forecast by 15 members. Probability forecasts of ramp occurrence (given in percentages at the top of the figure) were derived using the clustering approach **A2**, $\hat{\tau} = 30\%$ of P_n and the logistic regression model. The change in forecast probability depending on the number of members forecasting a ramp provides a good illustration of the sharpness property of the proposed methodology.

Wind Energ. 0000; **00**:1–14 © 0000 John Wiley & Sons, Ltd.
DOI: 10.1002/we
Prepared using weauth.cls

Using Conditional Kernel Density Estimation for Wind Power Density Forecasting

Jooyoung Jeon*
Smith School of Enterprise and the Environment, University of Oxford

James W. Taylor
Saïd Business School, University of Oxford

Journal of the American Statistical Association, forthcoming

* Address for Correspondence:

Jooyoung Jeon
Smith School of Enterprise and the Environment
University of Oxford
Hayes House, 75 George Street
Oxford OX1 2BQ, UK

Tel: +44 (0)1865 288927
Fax: +44 (0)1865 288805
Email: joo.jeon@smithschool.ox.ac.uk

Abstract

Of the various renewable energy resources, wind power is widely recognized as one of the most promising. The management of wind farms and electricity systems can benefit greatly from the availability of estimates of the probability distribution of wind power generation. However, most research has focused on point forecasting of wind power. In this paper, we develop an approach to producing density forecasts for the wind power generated at individual wind farms. Our interest is in intraday data and prediction from 1 to 72 hours ahead. We model wind power in terms of wind speed and wind direction. In this framework, there are two key uncertainties. First, there is the inherent uncertainty in wind speed and direction, and we model this using a bivariate VARMA-GARCH model, with a Student *t* distribution, in the Cartesian space of wind speed and direction. Second, there is the stochastic nature of the relationship of wind power to wind speed (described by the power curve), and to wind direction. We model this using conditional kernel density (CKD) estimation, which enables a nonparametric modeling of the conditional density of wind power. Using Monte Carlo simulation of the VARMA-GARCH model and CKD estimation, density forecasts of wind speed and direction are converted to wind power density forecasts. Our work is novel in several respects: previous wind power studies have not modeled a stochastic power curve; to accommodate time evolution in the power curve, we incorporate a time decay factor within the CKD method; and the CKD method is conditional on a density, rather than a single value. The new approach is evaluated using datasets from four Greek wind farms.

Key words: Wind energy; Predictive distribution; Kernel estimation; Wind speed; Wind direction; Bivariate GARCH.

1. INTRODUCTION

Wind power is the fastest growing form of renewable energy. Efficient management of wind farms and electricity systems requires forecasts of the power generated. This is a challenging forecasting task due to the erratic nature of wind and the nonlinear relationship between the variables involved. In an effort to enhance the information provided by the forecasters, probabilistic forecasting has been a recent area of development in wind energy forecasting (Bremnes 2004; Pinson and Kariniotakis 2010). The extension from deterministic forecasting to probabilistic allows more efficient management of the transmission system through improved load dispatch and better scheduling of spinning reserve. For a wind farm operator, a probabilistic forecast enables improved decision making regarding the amount of energy to commit to the electricity grid for a future period, which is important given the financial penalties that are incurred for the shortfall. Ultimately, a probabilistic assessment of the uncertainty in wind energy generation will bring societal benefits by enabling more efficient energy management resulting in a reduced reliance on power generation from fossil fuels.

Although there has been substantial attention paid by researchers to methods for wind power point forecasting, methods for wind power density forecasting are far less developed. One approach to density forecasting would be to fit a univariate statistical model to the wind power time series, but this is not straightforward due to the inherent nonlinear evolution of wind power. The alternative is to convert wind speed density forecasts to wind power density forecasts using the power curve, which relates wind power to wind speed. This approach is the focus of this paper. It relies on the availability of wind speed density forecasts, and these can be derived from a time series model or from ensemble predictions produced by an atmospheric model (see, for example,

Gneiting et al. 2006; Taylor et al. 2009; Sloughter et al. 2010; Gneiting 2011a). As for the power curve, historical data indicates that there is substantial variability in the relationship between wind speed and wind power (Sánchez 2006). It would seem wise to take this additional uncertainty into consideration when converting wind speed density forecasts to wind power density predictions.

This paper is the first to produce wind power density forecasts by explicitly modeling both the wind speed uncertainty and the stochastic power curve. With our focus on hourly data and relatively short forecast lead times, we use a time series model for wind speed density forecasting, rather than ensemble predictions from an atmospheric model. Advantages of using time series models are: (i) they are no less accurate than atmospheric modeling for short-term predictions (Focken et al. 2002); (ii) they can produce forecasts from any time origin and for any lead time, which contrasts with atmospheric model ensemble predictions (see, for example, Taylor et al., 2009); (iii) acquiring forecasts from an atmospheric model can be costly; and (iv) predictions from such models are often not available for the wind farm location of interest. Following the work of Cripps and Dunsmuir (2003) and Hering and Genton (2010), we use a bivariate VARMA-GARCH time series model, which enhances wind speed prediction through the joint modeling of wind speed and direction. We convert the resulting wind speed density predictions into wind power density forecasts using Monte Carlo simulation and conditional kernel density (CKD) estimation (see Rosenblatt 1969; Hyndman et al. 1996), which enables a nonparametric modeling of the conditional density of wind power. The method involves kernel weighting over the conditioning variable, wind speed, and kernel density estimation for the target variable, wind power. Our use of CKD estimation has two novel features: (i) the CKD estimation is conditional on a density, rather than a single

value, and (ii) to model time evolution in the relationship between wind power and the conditioning variables, we incorporate a time decay factor within CKD estimation.

To illustrate implementation of our approach, and to compare it to alternatives, we use data from four Greek wind farms, and evaluate forecast accuracy from 1 to 72 hours ahead. Due to the intermittency and non-dispatchable nature of wind energy, accurate short-term wind power forecasts from 15 minutes up to several days ahead are vital to transmission system operators, who maintain the balance between load and generation. For wind farm operators, the accuracy of short-term wind power forecasts is crucial for minimizing the penalties for failing to meet the commitment. Very short-term forecast horizons from a few seconds to an hour are related to turbine active control in wind farms. Long-term horizons up to 7 days are useful for maintenance planning of wind turbines.

Section 2 presents the Greek wind power dataset. Section 3 describes the VARMA-GARCH model that we use to produce wind speed and direction density forecasts. Section 4 explores the stochastic nature of the relationship between these two variables and wind power. Section 5 introduces CKD estimation, and Section 6 describes our use of CKD estimation for wind power density forecasting. Section 7 presents empirical results, and Section 8 provides a brief summary and conclusion.

2. THE WIND DATA

For the empirical work in this paper, we used hourly data from four wind farms in Crete: Aeolos, Rokas, Enteka and Iweco. Crete is the largest island in the Aegean Sea. It has high wind energy potential and an autonomous electricity grid. The datasets from Aeolos, Enteka and Rokas, which are in the east of Crete, consist of data from January 1,

2006 to December 31, 2006, which amounts to 8,760 hourly observations. The data from Iweco wind farm, which is located in the center of Crete, is over a shorter period from January 1, 2006 to October 30, 2006, which amounts to 7,272 observations. The wind speed, direction and power data were provided rounded to the nearest 0.1 m/s, 0.1° and 0.1 MW, respectively. For each wind farm, the wind speed and direction were recorded at a meteorological tower near to the wind farm at the hub height of the wind farm's turbines. The wind power data corresponds to total power generated from the whole wind farm. The capacities of the Aeolos, Enteka, Iweco and Rokas wind farms, at the end of 2006, were 11.6 MW, 2.8 MW, and 4.3 MW and 16.3 MW, respectively. In our empirical forecasting comparison of Section 7, the last 25% of each wind power series is used as the post-sample period. Figure 1 shows the time series plots for wind speed, direction and power for the Aeolos wind farm. All three series exhibit substantial variability. The wind power series is bounded above by the capacity of the wind farm. Interestingly, the wind power plot seems to indicate shifts in the capacity during January and February, and in the first half of July. The rise in the capacity around the end of February was due to new turbine installations, but we do not have explanations for the other apparent capacity changes. In Sections 3 and 4, we provide further descriptive plots to gain insight into the three wind variables and the relationships between them.

3. BIVARIATE VARMA-GARCH MODEL FOR WIND SPEED AND DIRECTION

In fitting a statistical model to wind speed time series, it has been suggested that wind direction can enhance the model, and hence improve wind speed forecast accuracy. Observing that wind direction is related to the fluctuations of sea breeze in Sydney Harbor, Cripps and Dunsmuir (2003) transform the minute-by-minute time series of wind

speed and direction to Cartesian coordinates, and model the pair of wind velocity variables using a bivariate vector autoregressive moving average (VARMA) model with error variances described by a bivariate generalized autoregressive conditional heteroskedastic (GARCH) process and a Student t distribution for the error terms. Gneiting et al. (2006) identify two distinct forecast regimes, split by wind direction, in hourly time series of wind speed and direction in the U.S. Pacific Northwest. They propose spatial and temporal regime switching models using a vector autoregressive idea and a truncated Gaussian distribution for two-hour ahead prediction. For the same data, Hering and Genton (2010) consider a univariate wind speed model incorporating wind direction in trigonometric forms using the truncated Gaussian distribution, and also a bivariate model for wind speed and direction, in Cartesian coordinate form, using a skewed t distribution.

A major benefit of modeling wind speed and wind direction, after transformation to Cartesian coordinates, is that non-negativity of wind speed prediction is automatically ensured. The two axes of the coordinates represent wind velocities corresponding to the east-west and north-south wind components. In this paper, we use a bivariate model for $\boldsymbol{Z}_t = (U_t, V_t)$, where $U_t = X_t \sin(\theta_t + \delta)$, $V_t = X_t \cos(\theta_t + \delta)$, X_t is the wind speed at time t, θ_t is the wind direction at time t, and δ is a constant parameter that is estimated along with the model parameters using maximum likelihood. Our inclusion of the parameter δ is new. It is intended to rotate the coordinate axes, so that one of them is aligned with a commonly occurring wind direction. This reduces the covariance between U_t and V_t, which should simplify the bivariate modeling. For the Aeolos wind farm, we plot the angular histogram of wind direction on the left hand side of Figure 2, and wind speed and direction on the right hand side of the figure. The plots indicate that, for this particular wind farm, wind

blows more frequently at approximately 20° (north-easterly) and 250° (south-westerly). In the model that we use for Aeolos in our empirical analysis of Section 7, the optimized value of the parameter δ was estimated as 30.1°. This implies that the axes are rotated anticlockwise by 30.1°, which results in the U_t axis approximately aligning with the commonly occurring Aeolos south-westerly wind direction.

We use a similar model to that presented by Cripps and Dunsmuir (2003). Using the VEC form of the bivariate GARCH model of Bollerslev et al. (1988), we estimated the entire covariance matrix of VEC, rather than the diagonal covariance matrix assumed by Cripps and Dunsmuir (2003). The model is given in expressions (1)-(4). Expression (1) is the VARMA part of the model and expression (3) is the bivariate GARCH part.

$$\boldsymbol{Z}_t = \boldsymbol{s}(\boldsymbol{\mu}, t) + \sum_{i=1}^{m} \boldsymbol{A}_i \boldsymbol{\varepsilon}_{t-i} + \sum_{j=1}^{r} \boldsymbol{B}_j \boldsymbol{Z}_{t-j} + \boldsymbol{\varepsilon}_t, \tag{1}$$

$$\boldsymbol{\varepsilon}_t = \boldsymbol{H}_t^{1/2} \boldsymbol{\eta}_t, \tag{2}$$

$$\operatorname{vech}(\boldsymbol{H}_t) = \boldsymbol{s}(\boldsymbol{\omega}, t) + \sum_{i=1}^{q} \boldsymbol{D}_i \operatorname{vech}(\boldsymbol{\varepsilon}_{t-i} \boldsymbol{\varepsilon}'_{t-i}) + \sum_{j=1}^{p} \boldsymbol{G}_j \operatorname{vech}(\boldsymbol{H}_{t-j}), \tag{3}$$

$$\boldsymbol{s}(\boldsymbol{\gamma}, t) = \boldsymbol{\gamma}_0 + \sum_{i=1}^{N_\gamma} \left[\boldsymbol{\gamma}_{i,1} \sin\left(2i\pi \tfrac{h(t)}{24}\right) + \boldsymbol{\gamma}_{i,2} \cos\left(2i\pi \tfrac{h(t)}{24}\right) \right], \tag{4}$$

where $\boldsymbol{Z}_t$ is the (2×1) vector of wind velocities; $\boldsymbol{\varepsilon}_t$ is a vector of error terms; $\boldsymbol{H}_t$ is the conditional covariance matrix of $\boldsymbol{\varepsilon}_t$; $\boldsymbol{\eta}_t$ is an i.i.d. vector of error terms; vech(·) denotes the column stacking operator of the lower triangular part of the symmetric matrix that is its argument; $\boldsymbol{A}_i$ and $\boldsymbol{B}_j$ are (2×2) matrices of parameters; $\boldsymbol{D}_i$ and $\boldsymbol{G}_j$ are (3×3) matrices of

parameters; *m*, *r*, *q* and *p* are non-negative integer valued constants that indicate the order of the VARMA and GARCH models; $\boldsymbol{s}(\boldsymbol{\mu}, t)$ is a (2×1) vector of deterministic intraday seasonal terms with (2×1) constant parameter vectors $\boldsymbol{\mu}_i$; $\boldsymbol{s}(\boldsymbol{\omega}, t)$ is a (3×1) vector of deterministic intraday seasonal terms with (3×1) constant parameter vectors $\boldsymbol{\omega}_i$; N_γ is a non-negative integer indicating the number of terms in the summation of $\boldsymbol{s}(\boldsymbol{\gamma}, t)$; and $h(t)$ is a repeating step function that numbers the hours from 1 to 24 within each day.

We used the Schwarz's Bayesian Criterion to select the order of the VARMA and GARCH components of the model, and the N_r. We considered VARMA and GARCH lags up to order 24, and $N_r \leq 8$. For our data, we implemented the VARMA-GARCH model using multivariate Gaussian, Student *t* and skewed *t* (Azzalini and Genton 2008) distributions for $\boldsymbol{\varepsilon}_t$. Table 1 summarizes the three models fitted to the Aeolos wind farm data. The axes-rotation δ parameter is significant for each of the three models. The table shows that the values of the degrees of freedom were low for the models with Student *t* and skewed *t* distributions, suggesting that the distribution is not close to Gaussian. In the table, the orders of the models are reasonably low, especially for the two non-Gaussian distributions, and it is interesting to see that diurnal seasonality only features in the Gaussian model. We investigated stationarity by considering, for the ARMA part of the model, the eigenvalues of the sum of the $\boldsymbol{A}_i$ and $\boldsymbol{B}_j$ matrices, and for the GARCH part, the eigenvalues of the sum of the $\boldsymbol{D}_i$ and $\boldsymbol{G}_j$ matrices (see Fountis and Dickey 1989; Bauwens et al. 2006). These eigenvalues are shown in the bottom rows of Table 1. Stationarity is indicated by the eigenvalues being less than one in modulus. The values for the GARCH part suggest that the model has persistence in its modeling of volatility. When assessed

across all four wind farms, the Student *t* distribution led to the most accurate wind speed density forecasts, and so we use this in the empirical analysis in the rest of this paper.

Recent research has considered how bootstrap procedures can be used to include parameter uncertainty in density forecasts generated from GARCH models (Reeves 2005; Pascual et al. 2006). However, such procedures are highly computational as they require repeated maximum likelihood estimation. Chen et al. (2011) suggest approximating a GARCH model as a linear process written in terms of squared residuals, which can then be estimated using least squares. This idea is not readily applicable to the multivariate GARCH model in this paper. In addition, bootstrapping such a model would seem to be computationally challenging. In view of this, in producing density forecasts from our VARMA-GARCH model, we did not account for the parameter estimation uncertainty. However, this is an aspect of our analysis that could be developed in future work.

4. THE POWER CURVE

Our proposed methodology for wind power density forecasting requires an understanding of the relationship between wind power and wind speed, which is described by the power curve. Although wind power theoretically depends on wind speed, air density and the area swept by the turbine blades, turbine manufacturers typically provide information regarding the power curve for an individual wind turbine assuming fixed air density. A power curve has a 'cut-in speed', at which the turbine blades begin to rotate; a 'rated speed', which is the lowest speed at which the maximum power output of the turbine is generated; and a 'cut-out speed', beyond which the turbine is shut down to prevent damage. An idealized deterministic curve, of this form, is used in the work of Taylor et al. (2009), Hering and Genton (2010) and Gneiting

(2011a). However, as these authors acknowledge, in practice, the power curve is not deterministic.

Sánchez (2006) explains that, in reality, the form of the power curve depends on meteorological variables such as wind direction, temperature, local air density and precipitation. He notes that the behavior of power curves when the wind speed increases can be different from the behavior when the speed decreases. Furthermore, the task is to predict wind power for an entire wind farm, and for this, the power curve of the whole wind farm is needed. The choice of a deterministic power curve is then complicated by the fact that the wind turbines in a wind farm can have different cut-in and rated speeds, and that there may be changes in the power capacity of the wind farm due to the addition of new turbines and turbine maintenance. Also, apparent changes in the capacity can be due to down-ramping to reduce the amount of wind energy entering the utility system. Given the complexity of the power curve, in practice, a deterministic power curve is often derived from historical wind speed and power data recorded at the wind farm level.

In Figure 3, we plot the historical wind power and wind speed data for two of the Greek wind farms. The typical features of a power curve are evident, along with substantial variability. In this paper, our approach to wind power density forecasting aims to capture the stochastic nature of the power curve. In Figure 4, we again plot wind power against wind speed, but, in this plot, we use different symbols to show the data points for selected months. The figure seems to show the wind farms capacities changing over time. As noted in Section 2, the capacity of the Aeolos wind farm increased around the end of February 2006 due to the installation of new turbines. However, we do not have explanations for the other apparent capacity changes. Unfortunately, information regarding idiosyncrasies in a wind power time series may well not be available to the

forecaster. Therefore, our approach to wind power density forecasting should try to accommodate the potentially time-varying nature of the power curve.

Although wind turbines are able to turn to face the wind, it has been suggested that the relationship between wind power and wind speed is, to some extent, dependent on the wind direction. Potter et al. (2007) find that the uncertainty in the relationship depends on the wind direction. Nielsen et al. (2006) include a wind direction variable into the relationship to explain turbine wake effects and direction dependent bias of the meteorological forecasts. Sánchez (2006) recognizes that wind direction influences the performance of a wind farm, and so uses it in a wind power prediction model. In Figure 5, we plot wind power against wind speed using different symbols to show the data points for selected wind directions. The plots suggest that the variability in the relationship can depend on wind direction. For Aeolos, south-westerly wind seems to produce a higher degree of variability in the relationship, and for Rokas, south-westerly wind shows higher variability than north-westerly when the wind speed is below about 13 m/s.

As mentioned in Section 3 for the Aeolos wind farm, wind blows more frequently around 20° and 250°. The plot on the right hand side of Figure 2 indicates that, in these directions, the level and variability of the wind speed is lower than from other directions. This might be due to the inherent characteristics of the wind, or to the terrain surrounding the wind farm. We investigate the influence of wind direction on the relationship of wind speed and power by plotting power against both speed and direction on the left hand side of Figure 6. A smooth surface, fitted by the Nadaraya-Watson estimator (see Nadaraya 1964; Watson 1964), is shown on the right hand side of the figure. The surface indicates that, for speed above about 20 m/s, the wind blowing around 30°, 80° and 240° tends to generate power more persistently. In summary, Figures 5 and 6 suggest that wind power,

its relationship to wind speed, and wind speed itself are dependent, to some extent, on wind direction. This supports our joint modeling of wind speed and direction in Section 3, and motivates the inclusion of wind direction in our modeling of wind power.

The conversion of a wind speed density forecast into a wind power density forecast is a conditional density estimation problem. If the conditional wind power density were to follow a Gaussian distribution with a constant variance and a mean that is a linear function of explanatory variables, it would be a standard multiple regression problem. However, the relationship between wind power and wind speed is nonlinear, and the conditional density of wind power is often skewed and can be bimodal, and, importantly, its shape is conditional on the value of the wind speed. This dependence on wind speed is illustrated by Figure 7, which shows smoothed histograms of wind power for different values of wind speed. In this paper, we use conditional kernel density estimation to overcome the issues of nonlinearity and a non-Gaussian conditional density.

5. CKD ESTIMATION

Conditional kernel density (CKD) estimation enables density estimation of a variable conditional on the value of one or more explanatory variables. The method is nonparametric in two senses; it involves no parametric assumption for the density of the target variable, and it makes no parametric assumption regarding the form of the relationship between target and explanatory variables. This is enabled through the use of double kernel estimation. Since the method involves no distributional assumption, the conditional kernel density estimation is particularly advantageous in estimating the density when the conditional distribution is multimodal or skewed, as is often the case in non-linear or non-Gaussian situations. In view of the nonlinear and stochastic

dependency of wind power on wind speed, CKD would seem to be very suitable for estimating wind power densities conditional on values for wind speed.

We define Y_t as dependent variable and X_t as explanatory variable. Let $f(y \mid x)$ be the conditional density function of Y_t given $X_t = x$. The Rosenblatt CKD estimator (Rosenblatt 1969) of $f(y \mid x)$ is written as

$$\hat{f}(y \mid x) = \frac{\sum_{t=1}^{n} K_{h_x}(X_t - x)\, K_{h_y}(Y_t - y)}{\sum_{t=1}^{n} K_{h_x}(X_t - x)},$$

(5)

where n is the sample size, and $K_h(\cdot)=K(\cdot/h)/h$ is a kernel function with bandwidth h. This formulation contains two bandwidths, h_x and h_y. They are scale parameters that control the amount of smoothing. We discuss our approach to bandwidth optimization in Section 6. An estimate of the full density function can be built up by repeating the CKD estimation for a range of y values. The CKD estimator involves double kernel estimation, with kernel density estimation in the y direction and kernel smoothing in the x direction. For a given x, the density function of Y_t at the value y is constructed by applying kernel density estimation to the sample of values of Y_t, with each Y_t value weighted in accordance with the proximity of the corresponding X_t relative to the value x.

Hyndman et al. (1996) note that the mean of the Rosenblatt estimator is a biased estimator of the conditional mean. To address this, they propose a two-step CKD estimator. In the first step, the conditional mean is estimated using some form of unbiased kernel smoothing. Subtracting the resulting mean estimates from the observed values for Y_t delivers a set of residuals e_t. In the second step, the Rosenblatt CKD estimator is

applied to these residuals. The Hyndman et al. two-step CKD estimator is given in expressions (6) and (7), where $\hat{m}(X_t)$ is the unbiased conditional mean estimator.

$$\hat{f}(y \mid x) = \frac{\sum_{t=1}^{n} K_{h_x}(X_t - x) K_{h_e}(\hat{e}_t - e)}{\sum_{t=1}^{n} K_{h_x}(X_t - x)}, \tag{6}$$

where $\hat{e}_t = Y_t - \hat{m}(X_t)$ and $e = y - \hat{m}(x)$ (7)

This formulation contains the bandwidths, h_x and h_e, and possibly additional bandwidths associated with the estimation of the conditional mean. As with the Rosenblatt CKD estimator, an estimate of the full density function can be produced by repeating the Hyndman et al. CKD estimation for different values of y.

There are not many examples of the use of CKD estimation in a time series context. Hyndman et al. use their density estimator for daily temperature conditional on lagged temperature and a seasonal variable. Bashtannyk and Hyndman (2001) use the estimator for the density of the eruption duration of the Old Faithful geyser conditional on waiting time. Fan and Yim (2004) extend the Rosenblatt estimator using a local linear regression, and forecast the density of the yield change of treasury bills conditional on the current yield. Juban et al. (2007) use the Rosenblatt CKD estimator to predict wind power densities conditional on the most recent wind power observation and point forecasts of wind speed and direction produced by an atmospheric model. Our use of CKD differs from this in that we produce wind power density forecasts conditional on density forecasts of wind speed and direction.

6. A NEW APPROACH TO CKD ESTIMATION FOR WIND POWER MODELING

6.1. CKD Estimation Conditional on Wind Speed

For wind power density forecasting, we implemented the two-step CKD estimator of Hyndman et al. (1996), in expressions (6)-(7), with Y_t and X_t specified as wind power and speed, respectively. We considered the Loess mean estimator of Cleveland (1979) and the Nadaraya-Watson mean estimator. However, the two-step CKD estimator led to wind power density forecasts that were only slightly more accurate than those produced by the Rosenblatt CKD estimator of expression (5). In view of this, we felt we could not justify the use of the more complex two-step estimator, and so, in this section, we report the results for the simpler Rosenblatt CKD estimator.

In our work, the full wind power density function was constructed by repeating the CKD estimation for values of y from zero to the wind farm's capacity with increments equal to 1% of the capacity. Linear interpolation between the resulting values of the density function delivered an estimate of the complete wind power density. Using a finer increment than 1% of the capacity led to increased computational burden with very little improvement in density forecast accuracy. We used a Gaussian kernel. We also tested a truncated Gaussian kernel, since the wind power is a bounded variable, but we found no practical benefit, which is consistent with the comments of Hyndman and Yao (2002).

The CKD estimator enables the density function of wind power to be estimated for a given value of the explanatory variable, wind speed. However, for any future period, the value of wind speed is of course unknown. Indeed, the purpose of the VARMA-GARCH model of Section 3 is to produce wind speed density forecasts. We are not aware of any previous studies that have considered the application of the CKD estimator when the explanatory variable is stochastic. In this paper, we introduce a CKD-based approach

to estimating the wind power density *conditional on a density* for the explanatory variable, wind speed. The approach involves the following three stages:

1. The CKD estimator is used to produce an estimate of the full wind power density conditional on each value for wind speed from zero to 30 m/s with an increment of 0.1 m/s. The result is 301 different wind speed values and their corresponding conditional wind power density estimates, which are stored for use in Stage 2.

2. Monte Carlo simulation of the bivariate VARMA-GARCH model of Section 3 is performed to generate 1,000 realizations of wind speed for a selected lead time. These 1,000 values can be considered as sampled values from the model's wind speed density forecast. We rounded to the nearest 0.1 m/s each of the 1,000 wind speed simulated values, and then obtained the corresponding conditional wind power density estimates, which had been stored in Stage 1.

3. The 1,000 wind power density estimates from Stage 2 are averaged to give a single wind power density forecast.

In the approach, 30 m/s was chosen, as it was perceived as being the maximum possible wind speed. We considered finer increments than 0.1 m/s, but this led to increased computational cost, without noticeably improving density forecast accuracy.

With regard to bandwidth selection for CKD estimation, the literature can be divided into two categories, namely a rule-based approach (Hall et al. 1999; Bashtannyk and Hyndman 2001; Hyndman and Yao 2002) and a data-driven approach (Fan and Yim 2004; Hall et al. 2004). Hall et al. (2004) comment that there is no general rule for the optimal bandwidth parameter, and that cross-validation delivers more appropriate bandwidths, particularly when multiple explanatory variables are involved, such as in Section 6.2, where we use wind speed and direction. Fan and Yim (2004) find that cross-

validation outperforms rule-based approaches. Holmes et al. (2007) write that rule-based approaches tend to perform poorly in finite samples when the reference distribution is not suitable. In this paper, we use cross-validation. To reduce the computational time, the parallel computing toolbox of Matlab is used. The source code is available upon request.

In the cross-validation, we selected the kernel bandwidths that led to the most accurate wind power density estimates, conditional on observed wind speed, where accuracy was measured by the mean continuous ranked probability score (CRPS) calculated over the cross-validation evaluation period for one hour ahead prediction. The CRPS captures the two important characteristics of a density forecast: its location relative to the observed value and its sharpness around that value (see Gneiting et al. 2007).

For the kernel density estimation methods, we used a rolling window of six months. This amounted to 50% of the length of the entire dataset. The penultimate 25% of the dataset was used as the cross-validation evaluation period, and the last 25% of the dataset was used for post-sample evaluation of the wind power density forecasts. The bivariate VARMA-GARCH model was estimated using the first 75% of the data. Elsewhere in the paper, we refer to this as the 'in-sample' period of data.

We used a fixed length rolling window in the kernel density estimation methods because the optimized bandwidths tended to vary with the length of the rolling window, with the bandwidths tending to be larger for shorter windows. The six-month rolling window was not updated every forecast origin, but updated every 24 hours in order to reduce computational running time. We experimented with rolling window lengths of three months and one month, and found that the resultant wind power density forecast accuracy was very similar to that for the six-month window. However, this was not the case when conditioning CKD estimation on wind speed and wind direction, which is

described in the next section. For this, we found that density forecast accuracy for prediction up to one day ahead reduced when the shorter window lengths were used.

Hyndman et al. (1996) suggest the use of different bandwidths for different values of the explanatory variable x. We experimented with a different value of h_x for different ranges of wind speed values. However, this did not lead to a clear benefit in accuracy.

6.2. CKD Estimation Conditional on Wind Speed and Wind Direction

In Section 2, we discussed how the relationship between wind power and wind speed can vary with wind direction. Therefore, it seems sensible to consider the use of a wind power CKD estimator with conditioning on both wind speed and direction. As we discussed briefly in Section 5, Hyndman et al. (1996) and Juban et al. (2007) provide applications of CKD estimation with more than one explanatory variable. In our application, we use the transformation to Cartesian coordinates discussed in Section 3, and condition wind power density estimation on the wind velocities U_t and V_t. In Figure 8, using the in-sample data, we plot the historical relationship between wind power and the two velocities. In expression (8), we present the Rosenblatt CKD estimator for wind power conditional on U_t and V_t.

$$\widetilde{f}(y \mid u,v) = \frac{\sum_{t=1}^{n} K_{h_{uv}}(U_t - u)\, K_{h_{uv}}(V_t - v)\, K_{h_y}(Y_t - y)}{\sum_{t=1}^{n} K_{h_{uv}}(U_t - u)\, K_{h_{uv}}(V_t - v)}.$$

(8)

This estimator involves two bandwidths, h_{uv} and h_y. We obtained slightly improved wind power density forecast accuracy when using different bandwidths for the kernels in the u and v directions. However, as the improvement was not substantial, for the sake of simplicity, in this paper, we treat these bandwidths as being identical.

We implemented the three-stage CKD-based approach described in Section 6.1. In that section, we considered wind speed values from zero to 30 m/s with an increment of 0.1 m/s. In this section, we replace this with *u* and *v* values from -30 m/s to 30 m/s with an increment of 0.5 m/s. We used the larger increment to compensate for the additional computational cost due to the increased dimension size. (Using an increment of 0.1 m/s did not lead to noticeable improvement in wind power density forecast accuracy.)

6.3. CKD Estimation with Time Decay

The relationship between wind power and the explanatory variables, wind speed and direction, can evolve over time, as discussed in Section 4 and suggested by Figure 4. One way of addressing the time-variation is to use only recent information. This could be enabled by basing estimation on a rolling window. An alternative, which uses all historical data, is to employ a time decay factor. After observing that the power curve changes over time, Sánchez (2006) employs a time decay factor within a recursive least square technique for estimating the time-varying parameters of a model to be used for wind power point forecasting. In expression (9), we present the Rosenblatt CKD estimator of expression (5) with an exponential time decay parameter λ ($0<\lambda\leq 1$).

$$\widetilde{f}(y \mid x) = \frac{\sum_{t=1}^{n} \lambda^{n-i} K_{h_x}(X_t - x) K_{h_y}(Y_t - y)}{\sum_{t=1}^{n} \lambda^{n-i} K_{h_x}(X_t - x)}.$$

(9)

A lower value of λ implies faster exponential decay, and hence more weight is given to the recent observations. This allows for a power curve that evolves due to factors that cannot be modeled explicitly in the CKD estimation framework. We optimized λ,

along with the bandwidths, using cross-validation. We also considered the use of the time decay factor within the multi-dimensional CKD estimator of expression (8).

7. EMPIRICAL COMPARISON OF POST-SAMPLE FORECAST ACCURACY

We compared the accuracy of wind power density forecasts from the CKD-based approach with the accuracy of simpler and more traditional methods. As stated previously, we used the last 25% of each of the four wind power series for post-sample forecast evaluation. We considered forecasts from 1 to 72 hours ahead. For each wind power series, we rolled the forecast origin forward (one hour at a time) through the post-sample evaluation period to produce a collection of forecasts from each method for each horizon.

7.1. Methods

We implemented three different categories of density forecasting methods:

1. Simple kernel density estimation – As a relatively simple benchmark method, we applied kernel density estimation to a moving window of recent historical wind power observations. This relatively simple estimator is written as

$$\hat{f}(y \mid x) = \sum_{t=n-l+1}^{n} K_{h_y}(Y_t - y).$$

We considered three versions of the method using the following different lengths, *l*, for the moving window: (i) 24 hours, (ii) 10 days and (iii) 6 months.

2. Conditioning on wind speed – We implemented the following three methods that produced wind power density forecasts conditional on just the wind speed variable:

(i) Deterministic with wind speed – The power curve is assumed to be deterministic, and is estimated using the Nadaraya-Watson estimator for the conditional

mean. Based on this deterministic power curve, wind speed density forecasts are converted, using Monte Carlo simulation, into wind power density forecasts. This estimator captures the uncertainty in wind power due to wind speed uncertainty, but not the uncertainty due to the power curve. This is a relatively sophisticated benchmark against which to compare the CKD-based methods.

(ii) CKD with wind speed – This is the three-stage method of Section 6.1 based on the CKD estimator of expression (5).

(iii) CKDλ with wind speed – This is the three-stage method of Section 6.1 based on the CKD estimator with exponential decay of expression (9).

3. Conditioning on wind velocities – We implemented the three methods just described with conditioning on wind speed replaced by conditioning on the two wind velocities, U_t and V_t. For these three methods, Table 2 presents the bandwidths and exponential weights optimized using cross-validation. Four of the bandwidths for CKDλ are larger than the corresponding bandwidths for CKD, and the opposite is true for two of the bandwidths. We would suggest larger bandwidths for CKDλ is intuitive because exponential decay leads to less historical information being captured, and so there is a need for a greater degree of kernel smoothing, and this is manifested in larger values for the bandwidths. This effect was observed by Taylor (2008) for an exponentially weighted kernel quantile estimator.

To gain insight into the density forecasts produced by the various methods, in Figure 9, we plot the forecasts of the cumulative distribution function (cdf) produced for Aeolos and Rokas with forecast origin set as the final period in the in-sample set of data. We would expect the cdf forecasts from the CKD-based methods to be wider than those from a method based on a deterministic power curve, because the CKD-based methods

involve the modeling of an additional uncertainty, namely the stochastic power curve. This is far more apparent in the plot for the Rokas wind farm than the one for Aeolos. This is because the corresponding wind speed density forecast for Aeolos was relatively wide, with the effect that this uncertainty substantially dominated the uncertainty due to the power curve. For Rokas, the cdf forecast from the CKDλ method is noticeably different to that from the other CKD-based method. This is not the case for Aeolos, and this is because the value of the CKDλ method exponential time decay parameter is 0.999, which is relatively high, implying slow decay, while the value for Rokas is 0.995.

7.2. Point Forecasting

Although our main focus is density forecasting, we also evaluate point forecast accuracy, as this provides insight into the accuracy of the central locations of the density forecasts. We evaluated point forecasts using the mean absolute error (MAE) and the root mean squared error (RMSE). Gneiting (2011a,b) notes that the median of a density forecast is the optimal point forecast if the loss function is symmetric piecewise linear, and the mean is the optimal point forecast for a quadratic loss function. In view of this, we used the MAE for point forecasts produced as the medians of the density forecasts, and the RMSE for point forecasts produced as the means of the density forecasts.

For each of the four Greek datasets, and for each method, we calculated the MAE and RMSE of point forecasts for each forecast horizon from 1 to 72 hours ahead. Table 3 presents the MAE, averaged over the four wind farms, for various lead times and groups of lead times. We report accuracy for the earlier lead times in more detail because it seems likely that statistical methods will have more to offer over atmospheric models for short lead times. The final column of the table summarizes accuracy across all 72 lead

times. In the table, the bold font indicates the best performing method at each lead time. For the RMSE, the rankings of the methods were very similar to those for the MAE, and so for conciseness we do not report the RMSE results.

Table 3 shows that the three simple kernel density estimation methods performed relatively poorly in terms of point forecast accuracy. All three versions of the method were comfortably outperformed by the other methods at all lead times. Turning to the three methods that produce forecasts conditional on wind speed, Table 3 shows that the three produced fairly similar results, with the CKD-based method with exponential time decay being a little more accurate than the other two methods. It is interesting to note that each of these three methods, conditional on just wind speed, is outperformed at all lead times by the corresponding method that produces forecasts conditional on the two wind velocity variables.

7.3. Density Forecasting

To evaluate density forecast accuracy at each lead time, we calculated the CRPS averaged across the four wind farms. In Table 4, we summarize these results using the same format as in Table 3 for point forecasting. As with the point forecasting, the three simple kernel density estimation methods performed relatively poorly in terms of density forecasting. Also consistent with the point forecasting results is the superiority of each method conditional on wind velocities when compared with the corresponding method conditional on just wind speed. It would seem that wind power density forecast accuracy does benefit by modeling wind power in terms of both wind speed and direction.

If we focus on the three methods that are conditional on wind velocities, we see from Table 4 that the two CKD-based methods outperform the method that assumed a deterministic power curve. The same comment can be made regarding the three methods

conditional on just wind speed. Comparing the methods based on CKD, we can see that accuracy improved with the inclusion of the exponential time decay parameter.

Figure 10 provides further post-sample evaluation of the density forecasting methods by showing histogram plots of the probability integral transform (PIT) (see Diebold et al. 1998; Gneiting et al. 2007). We show the PIT values for lead times of 1, 4, 12, 24 and 72 hours, and for four methods: the simple kernel density estimation method using a 10-day moving window, and the three methods conditional on the wind velocities. The histograms correspond to the PIT values for all four Greek wind farms. The optimal shape of a PIT histogram is a uniform distribution. The simple kernel estimation method and the method based on the deterministic power curve show a high peak in both tails, demonstrating that their density forecasts are too narrow, underestimating the tail risk. The histogram plots for the two CKD-based methods are closer to uniform distributions.

7.4. Quantile Forecasting

Interest often lies in the accurate estimation of tail quantiles or a certain prediction interval. For example, the estimation of tail quantiles provides useful information to support trading based on future production (Pinson et al. 2007). Furthermore, Gneiting (2011a) and Pinson et al. (2007) show that a quantile forecast, other than the mean or median, can be the optimal point forecast in situations where there is an asymmetric cost function.

To evaluate the quantile forecasts, we use the hit percentage. This measure assesses the unconditional coverage of a θ conditional quantile estimator. It is the percentage of observations falling below the estimator. Ideally, the percentage should be θ. The measure can be viewed as complimenting the CRPS and PIT in evaluating the

density forecasts. For all the density forecasting methods, we obtained the hit percentage for the 5% quantile forecasts from the density forecasts of each method. We then calculated the absolute value of the difference between the hit percentage and the ideal value of 5%. Averaging this value across the four Greek wind farm datasets delivered the values reported in Table 5. As with point forecast evaluation, smaller values of this mean absolute error measure are better. Table 6 reports the analogous measure for the evaluation of forecasts of the 95% quantile.

Looking first at the results for the simple kernel density estimation methods, we see that the methods performed relatively poorly for the 5% quantile, but for the 95% quantile, the version of the method based on a moving window of 10 days was more competitive. It is interesting to see that all the CKD-based methods clearly outperformed the deterministic power curve methods at each lead time for both the 5% and 95% quantiles.

8. CONCLUSION

In this paper, we have introduced an approach to wind power density forecasting that captures the uncertainty due to wind speed, as well as the uncertainty due to the stochastic nature of the power curve. The approach involves Monte Carlo simulation of a statistical model and CKD estimation. We considered extensions of the approach that allow for conditioning on both wind speed and direction, and allow the inclusion of exponential time decay. Post-sample density forecasting results show that the new approach was able to outperform a simpler version based on a deterministic power curve, as well as simple benchmark methods.

In terms of future work, it would be interesting to consider additional explanatory variables, such as temperature and air pressure, and perhaps also weather information at upwind locations. In this paper, we have used time series models to produce density forecasts for the meteorological explanatory variables, but an alternative would be to base these forecasts on ensemble predictions from an atmospheric model. One might anticipate that this would be particularly advantageous for the longer lead times. If it is only certain quantiles of the wind power density that are needed (see Pinson et al. 2007), then it may be beneficial to optimize the CKD bandwidths separately for each quantile of interest, rather than for the whole density using CRPS. It would also be interesting to consider the use of the CKD-based approach in this paper for predicting the density of the total wind power produced from many wind farms. This could be used by an electricity system operator to make decisions regarding operating reserve. An additional application of the CKD-based approach would be to generate a density forecast for the electricity load conditional on density forecasts for various meteorological variables.

ACKNOWLEDGEMENTS

The authors would like to thank Tilmann Gneiting, Amanda S. Hering, Max Little and Nigel Meade for their helpful comments and suggestions, as well as George Sideratos of the National Technical University of Athens and the SafeWind Project of the European Union for providing the Greek wind farm data. We are also grateful for the helpful comments of two referees.

REFERENCES

Azzalini, A., and Genton, M. G. (2008), "Robust Likelihood Methods Based on the Skew-t and Related Distributions," *International Statistical Review*, 76, 106-129.

Bashtannyk, D. M., and Hyndman, R. J. (2001), "Bandwidth Selection for Kernel Conditional Density Estimation," *Computational Statistics and Data Analysis*, 36, 279-298.

Bauwens, L., Laurent, S. and Rombouts, J. V. K. (2006), "Multivariate GARCH Models: A Survey," *Journal of Applied Econometrics*, 21, 79-109.

Bollerslev, T., Engle, R. F., and Wooldridge, J. M. (1988), "A Capital Asset Pricing Model with Time-Varying Covariances," *Journal of Political Economy*, 96, 116.

Bremnes, J.B. (2004), "Probabilistic Wind Power Forecasts using Local Quantile Regression," *Wind Energy*, 7, 47-54.

Chen, B., Gel, Y. R., Balakrishna, N. and Abraham, B. (2011), "Computationally Efficient Bootstrap Prediction Intervals for Returns and Volatilites in ARCH and GARCH Processes," *Journal of Forecasting*, 30, 51-71.

Cleveland, W. S. (1979), "Robust Locally Weighted Regression and Smoothing Scatterplots," *Journal of the American Statistical Association*, 74, 829-836.

Cripps, E., and Dunsmuir, W. T. M. (2003), "Modelling the Variability of Sydney Harbour Wind Measurements," *Journal of Applied Meteorology*, 42, 1131-1138.

Diebold, F. X., Gunther, T. A., and Tay, A. S. (1998), "Evaluating Density Forecasts with Applications to Financial Risk Management," *International Economic Review*, 39, 863–883.

Fan, J., and Yim, T. H. (2004), "A Crossvalidation Method for Estimating Conditional Densities," *Biometrika*, 91, 819-834.

Focken, U., Lange, M., Mönnich, K., Waldl, H. P., Beyer, H. G., and Luig, A. (2002), "Short-Term Prediction of the Aggregated Power Output of Wind Farms - A Statistical Analysis of the Reduction of the Prediction Error by Spatial Smoothing Effects," *Journal of Wind Engineering and Industrial Aerodynamics*, 90, 231-246.

Fountis, N. G., and Dickey, D. A. (1989), "Testing for a Unit Root Nonstationarity in Multivariate Autoregressive Time Series," *The Annals of Statistics*, 17, 419-428.

Gneiting, T., Balabdaoui, F., and Raftery, A. E. (2007), "Probabilistic Forecasts, Calibration and Sharpness," *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 69, 243-268.

Gneiting, T., Larson, K., Westrick, K., Genton, M. G., and Aldrich, E. (2006), "Calibrated Probabilistic Forecasting at the Stateline Wind Energy Center: The Regime-Switching Space-Time Method," *Journal of the American Statistical Association*, 101, 968-979.

Gneiting, T. (2011a), "Quantiles as Optimal Point Forecasts," *International Journal of Forecasting*, 27, 197-207.

Gneiting, T. (2011b), "Making and Evaluating Point Forecasts," *Journal of the American Statistical Association*, 106, 746-762.

Hall, P., Racine, J., and Li, Q. (2004), "Cross-Validation and the Estimation of Conditional Probability Densities," *Journal of the American Statistical Association*, 99, 1015-1026.

Hall, P., Wolff, R. C. L., and Yao, Q. (1999), "Methods for Estimating a Conditional Distribution Function," *Journal of the American Statistical Association*, 94, 154-163.

Hering, A. S., and Genton, M. G. (2010), "Powering Up with Space-Time Wind Forecasting," *Journal of the American Statistical Association*, 105, 92-104.

Hyndman, R. J., Bashtannyk, D. M., and Grunwald, G. K. (1996), "Estimating and Visualizing Conditional Densities," *Journal of Computational and Graphical Statistics*, 5, 315-336.

Hyndman, R. J., and Yao, Q. (2002), "Nonparametric Estimation and Symmetry Tests for Conditional Density Functions," *Journal of Nonparametric Statistics*, 14, 259-278.

Juban, J., Fugon, L., and Kariniotakis, G. (2007), "Probabilistic Short-term Wind Power Forecasting Based on Kernel Density Estimators," *European Wind Energy Conference: Milan, Italy.*

Nadaraya, E. A. (1964), "Remarks on Nonparametric Estimates for Density Functions and Regression Curves," *Theory of Probability and its Applications*, 15, 134-137.

Nielsen, H. A., Nielsen, T. S., Madsen, H., Giebel, G., Badger, J., Landbergt, L., Sattler, K., Voulund, L., and Tofting, J. (2006), "From Wind Ensembles to Probabilistic Information about Future Wind Power Production - Results from an Actual Application," in *Proceedings of the 9th International Conference on Probabilistic Methods Applied to Power Systems.*

Pascual L., Romo J., Ruiz E. (2006), "Bootstrap Prediction for Returns and Volatilities in GARCH models," *Computational Statistics and Data Analysis*, 50, 2293–2312.

Pinson, P., Chevallier, C., and Kariniotakis, G. N. (2007), "Trading Wind Generation from Short-Term Probabilistic Forecasts of Wind Power," *IEEE Transactions on Power Systems*, 22, 1148-1156.

Pinson, P., and Kariniotakis, G. N. (2010), "Conditional prediction intervals of wind power generation," *IEEE Transactions on Power Systems*, 25, 1845-1856.

Potter, C. W., Gil, H. A., and McCaa, J. (2007), "Wind Power Data for Grid Integration Studies," in *Proceedings of the IEEE/PES General Meeting, Tampa Bay, US, Paper Number: 07GM0808.*

Reeves, J.J. (2005), "Bootstrap Prediction Intervals for ARCH Models," *International Journal of Forecasting*, 21, 237–248.

Rosenblatt, M. (1969), "Conditional Probability Density and Regression Estimators," in Multivariate Analysis II, ed. P. R. Krishnaiah, New York: Academic Press, pp. 25-31.

Sánchez, I. (2006), "Short-Term Prediction of Wind Energy Production," *International Journal of Forecasting*, 22, 43-56.

Sloughter, J. M., Gneiting, T., and Raftery, A. E. (2010), "Probabilistic Wind Speed Forecasting using Ensembles and Bayesian Model Averaging," *Journal of the American Statistical Association*, 105, 25-35.

Taylor, J. W. (2008), "Using Exponentially Weighted Quantile Regression to Estimate Value at Risk and Expected Shortfall," *Journal of Financial Econometrics*, 6, 382-406.

Taylor, J. W., McSharry, P. E., and Buizza, R. (2009), "Wind Power Density Forecasting using Ensemble Predictions and Time Series Models," *IEEE Transactions on Energy Conversion*, 24, 775-782.

Watson, G. S. (1964), "Smooth Regression Analysis," *Shankya Series A*, 26, 359-372.

Table 1. Summary of the VARMA-GARCH model of expressions (1)-(4) fitted to the in-sample data for the Aeolos wind farm. Standard errors are given in parentheses.

	Gaussian	**Student *t***	**skewed *t***
Axes-rotation parameter δ	15.2° (0.4°)	30.1° (2.7°)	32.3° (2.1°)
Degrees of freedom		4.05 (0.10)	3.98 (0.09)
Skewness parameter for U_t			-0.009 (0.028)
Skewness parameter for V_t			-0.084 (0.029)
AR order (r)	6	1	2
MA order (m)	1	2	3
ARMA diurnal (N_μ)	1	0	0
ARCH order (q)	2	1	1
GARCH order (p)	1	1	1
GARCH diurnal (N_ω)	0	0	0
ARMA eigenvalues	0.89 0.44	0.93 0.72	0.75 0.75
GARCH eigenvalues	1.00 0.82 0.82	1.00 0.78 0.71	1.00 0.83 0.65

Table 2. Bandwidths and decay parameters optimized using cross-validation for the three density forecasting methods that are conditional on the two wind velocities, U_t and V_t.

Method	Wind farm	Bandwidth h_{uv} (m/s)	Bandwidth h_e (MW)	λ (half-life)
Deterministic	Aeolos	0.74		
	Rokas	0.51		
	Enteka	0.63		
	Iweco	0.55		
CKD	Aeolos	0.93	0.21	
	Rokas	0.68	0.25	
	Enteka	0.59	0.04	
	Iweco	0.71	0.04	
CKDλ	Aeolos	1.03	0.14	0.999 (28.9 days)
	Rokas	0.61	0.64	0.995 (5.8 days)
	Enteka	0.70	0.05	0.997 (9.6 days)
	Iweco	0.71	0.04	0.999 (28.9 days)

Table 3. Evaluation of post-sample wind power point forecast accuracy in MW using MAE averaged over the four Greek datasets. Smaller values are better. Point forecasts are medians of density forecasts.

Lead Time (hours):	1	2	3-4	5-6	7-8	9-12	13-24	25-48	49-60	61-72	1-72
Simple kernel density estimation											
24 hours	2.44	2.45	2.48	2.51	2.54	2.56	2.60	2.73	2.83	2.88	2.71
10 days	2.66	2.66	2.66	2.66	2.66	2.67	2.67	2.66	2.65	2.63	2.66
6 months	3.02	3.02	3.02	3.02	3.02	3.02	3.03	3.04	3.04	3.04	3.04
Conditioning on wind speed											
Deterministic	0.91	1.07	1.23	1.37	1.50	1.64	1.94	2.35	2.57	2.59	2.20
CKD	0.93	1.08	1.23	1.38	1.50	1.63	1.92	2.32	2.52	2.55	2.17
CKDλ	0.91	1.05	1.20	1.35	1.46	1.59	1.86	2.25	2.44	2.46	2.11
Conditioning on wind velocities											
Deterministic	0.90	1.04	1.19	1.34	1.45	1.61	1.89	2.31	2.50	2.51	2.17
CKD	0.91	1.05	1.19	1.34	1.44	1.59	1.87	2.27	2.45	2.46	2.11
CKDλ	**0.89**	**1.03**	**1.17**	**1.31**	**1.42**	**1.56**	**1.84**	**2.23**	**2.40**	**2.41**	**2.07**

NOTE: The best performing model at each lead time is in bold.

Table 4. Evaluation of post-sample wind power density forecast accuracy in MW using CRPS averaged over the four Greek datasets. Smaller values are better.

Lead Time (hours):	1	2	3-4	5-6	7-8	9-12	13-24	25-48	49-60	61-72	1-72
Simple kernel density estimation											
24 hours	1.74	1.75	1.77	1.80	1.82	1.83	1.86	1.95	2.03	2.07	1.95
10 days	1.79	1.80	1.80	1.80	1.80	1.80	1.81	1.80	1.79	1.78	1.80
6 months	2.04	2.04	2.04	2.04	2.04	2.04	2.05	2.05	2.05	2.05	2.05
Conditioning on wind speed											
Deterministic	0.72	0.82	0.91	1.00	1.07	1.17	1.34	1.56	1.64	1.65	1.47
CKD	0.66	0.76	0.86	0.95	1.0	1.13	1.30	1.54	1.61	1.62	1.44

CKDλ	0.64	0.74	0.83	0.93	1.03 1.01	1.10	1.27	1.50	1.58	1.59	1.40
Conditioning on wind velocities											
Deterministic	0.70	0.79	0.88	0.97	1.03	1.12	1.30	1.52	1.60	1.61	1.43
CKD	0.65	0.74	0.83	0.93	1.00	1.09	1.27	1.49	1.57	1.57	1.39
CKDλ	**0.63**	**0.72**	**0.81**	**0.91**	**0.98**	**1.07**	**1.25**	**1.47**	**1.55**	**1.56**	**1.37**

NOTE: The best performing model at each lead time is in bold.

Table 5. Evaluation of post-sample forecast accuracy for the 5% wind power quantile using absolute hit percentage error averaged over the four Greek datasets. Smaller values are better.

Lead Time (hours):	1	2	3-4	5-6	7-8	9-12	13-24	25-48	49-60	61-72	1-72
Simple kernel density estimation											
24 hours	24.7	24.7	25.0	25.0	25.5	25.5	25.7	27.1	27.8	28.1	26.8
10 days	23.0	23.0	23.1	23.1	23.3	23.2	23.6	23.7	23.7	23.5	23.6
6 months	27.2	27.3	27.1	27.2	27.2	27.4	27.6	27.9	27.8	27.9	27.7
Conditioning on wind speed											
Deterministic	26.6	23.9	22.1	20.7	19.7	18.5	17.7	16.6	15.8	15.5	17.1
CKD	2.4	1.9	**2.1**	**2.0**	**2.2**	2.6	**2.7**	**2.8**	**3.3**	**3.6**	**2.9**
CKDλ	2.7	2.7	2.9	2.6	2.4	**2.3**	2.8	4.2	5.0	5.5	4.0
Conditioning on wind velocities											
Deterministic	26.6	23.7	22.0	20.9	19.9	18.8	17.8	17.1	16.6	16.2	17.7
CKD	**1.8**	**1.8**	2.2	2.4	2.6	2.8	3.0	3.2	3.6	4.1	3.2
CKDλ	2.8	3.1	3.1	2.9	2.8	2.5	**2.7**	4.0	4.8	5.3	3.9

NOTE: The best performing model at each lead time is in bold.

Table 6. Evaluation of post-sample forecast accuracy for the 95% wind power quantile using absolute hit percentage error averaged over the four Greek datasets. Smaller values are better.

Lead Time (hours):	1	2	3-4	5-6	7-8	9-12	13-24	25-48	49-60	61-72	1-72

Simple kernel density estimation											
24 hours	3.8	4.0	4.2	4.5	4.6	4.7	4.8	5.7	6.7	7.1	5.8
10 days	1.6	1.6	1.8	1.7	1.8	1.7	1.8	1.8	**2.0**	**2.2**	1.9
6 months	4.4	4.5	4.5	4.5	4.5	4.5	4.4	4.4	4.4	4.5	4.4
Conditioning on wind speed											
Deterministic	9.0	7.9	7.3	6.8	6.5	6.5	6.1	5.7	5.7	5.9	6.0
CKD	1.8	1.5	1.1	0.8	0.8	0.4	0.9	1.8	2.6	2.9	1.8
CKDλ	**1.3**	**0.9**	**0.8**	**0.7**	**0.4**	**0.3**	1.0	2.0	2.8	3.1	1.9
Conditioning on wind velocities											
Deterministic	7.3	6.4	5.6	5.0	4.8	4.8	4.4	4.6	4.8	5.1	4.7
CKD	1.9	1.7	1.3	1.1	0.9	0.7	**0.7**	**1.7**	2.3	2.6	1.7
CKDλ	1.6	1.2	0.9	0.9	0.6	0.5	1.1	1.6	2.1	2.3	**1.6**

NOTE: The best performing model at each lead time is in bold.

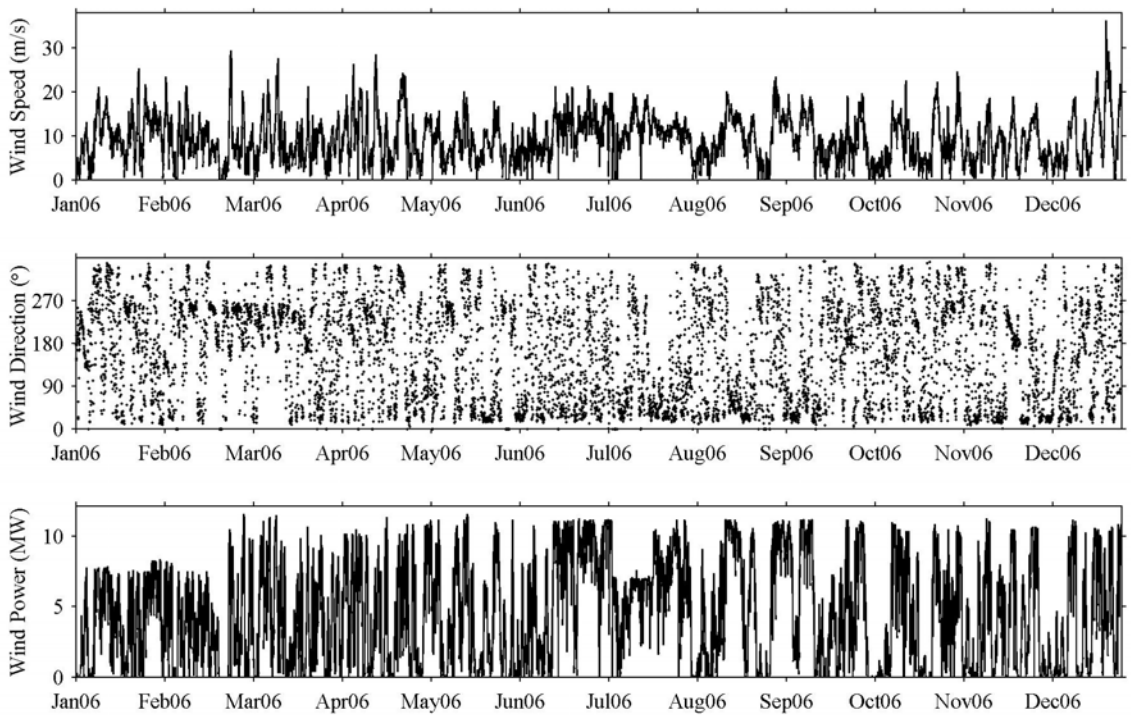


Figure 1. Wind speed, direction and power time series for the Aeolos wind farm.

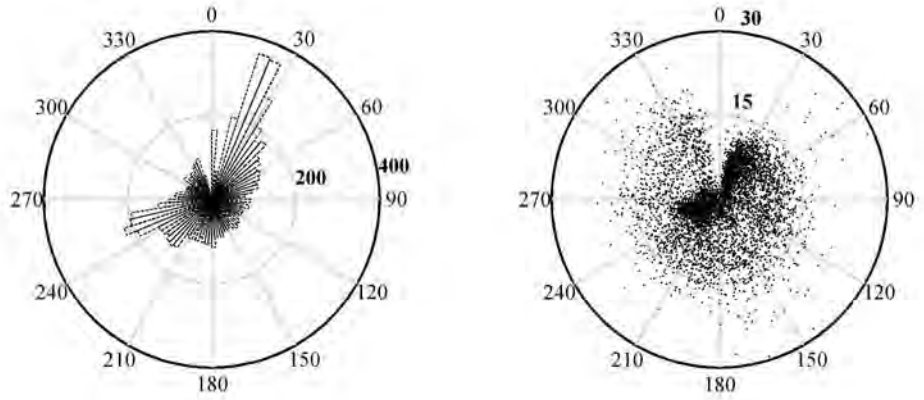


Figure 2. Angular histogram of wind direction (left). A Cartesian plot of wind speed and direction (right), where the distance of each observation from the origin is the strength of the wind speed. The plots use in-sample data for the Aeolos wind farm.

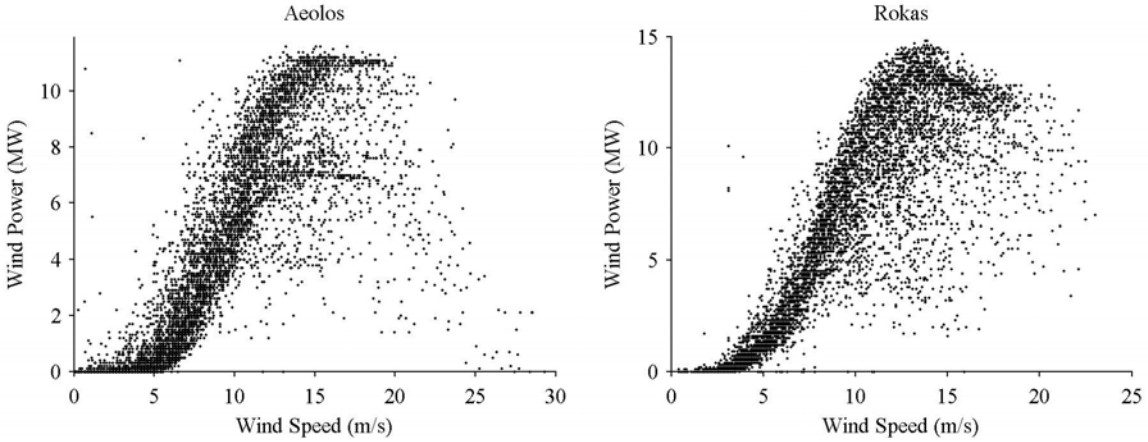


Figure 3. Plot of wind power against wind speed using the in-sample data for the Aeolos and the Rokas wind farms.

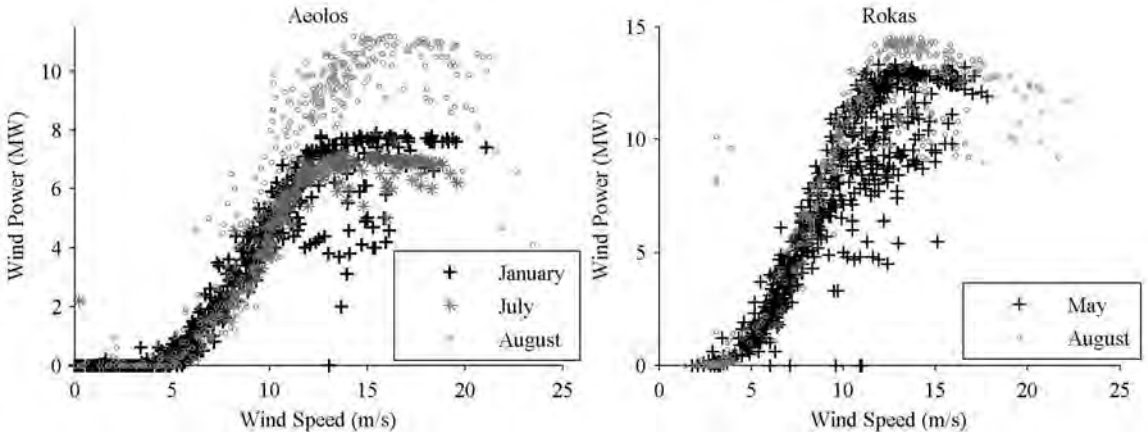


Figure 4. Plot of wind power against wind speed for selected months of the in-sample data for the Aeolos and the Rokas wind farms.

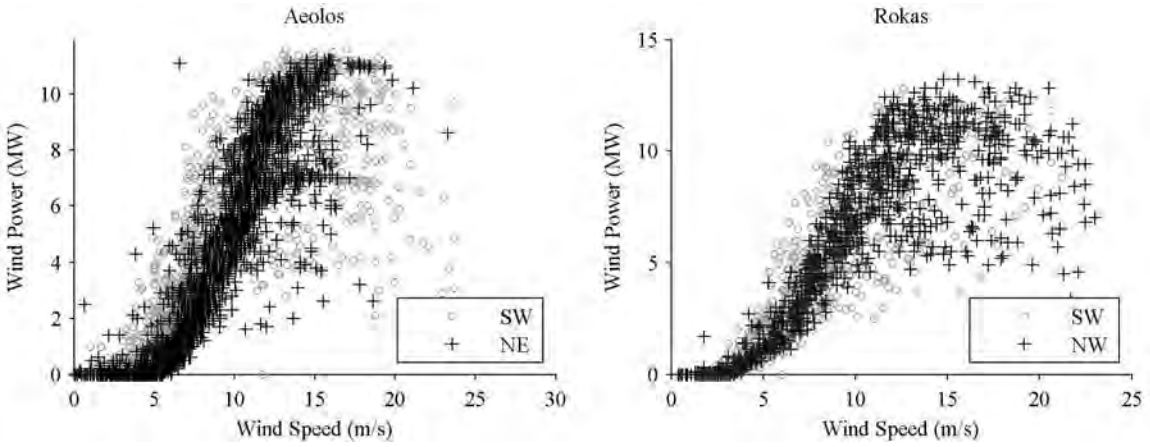


Figure 5. Plot of wind power against wind speed for two selected wind directions using the in-sample data for the Aeolos and the Rokas wind farms.

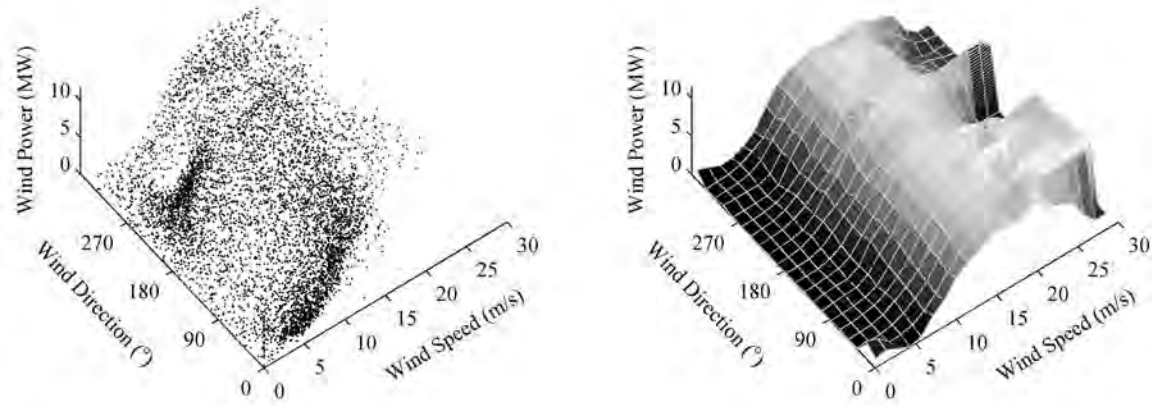


Figure 6. Plot of wind power against wind speed and wind direction (left) and, to help interpretation of this plot, a smooth surface fitted using a Nadaraya-Watson estimator (right). The plots use in-sample data for the Aeolos wind farm.

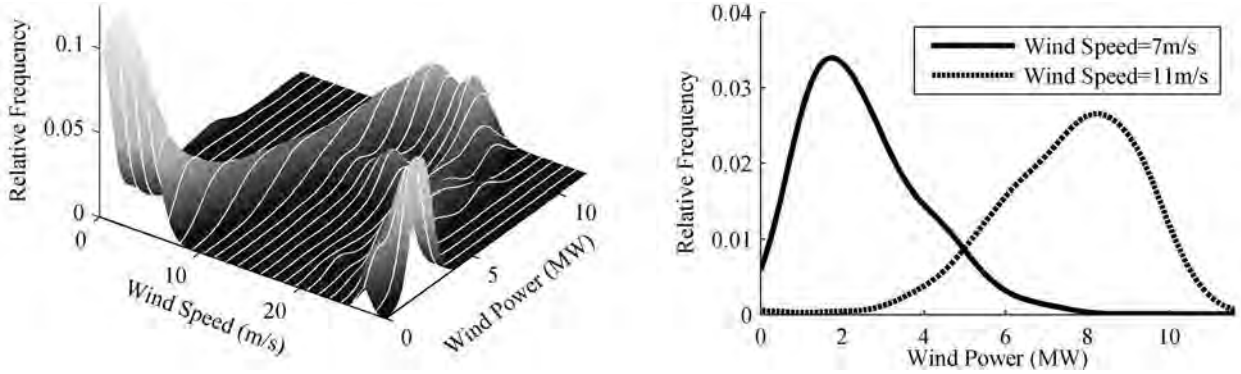


Figure 7. Smoothed histograms of wind power for different values of wind speed. Smoothing performed using a Nadaraya-Watson estimator. The plots use in-sample data for the Aeolos wind farm.

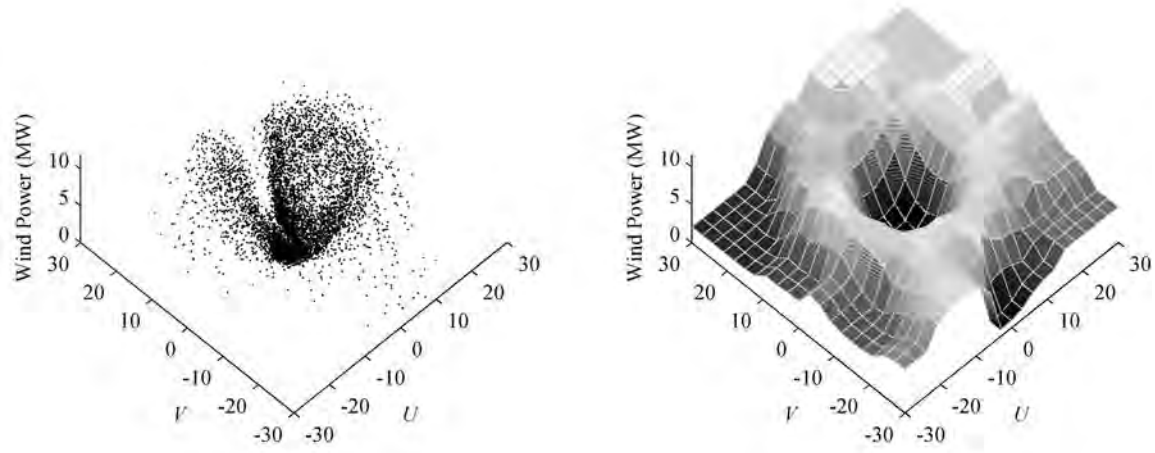


Figure 8. Plot of wind power against wind velocity variables (left) and, to help interpretation of the left plot, a smooth surface fitted using a Nadaraya-Watson estimator (right). The plots use in-sample data for the Aeolos wind farm.

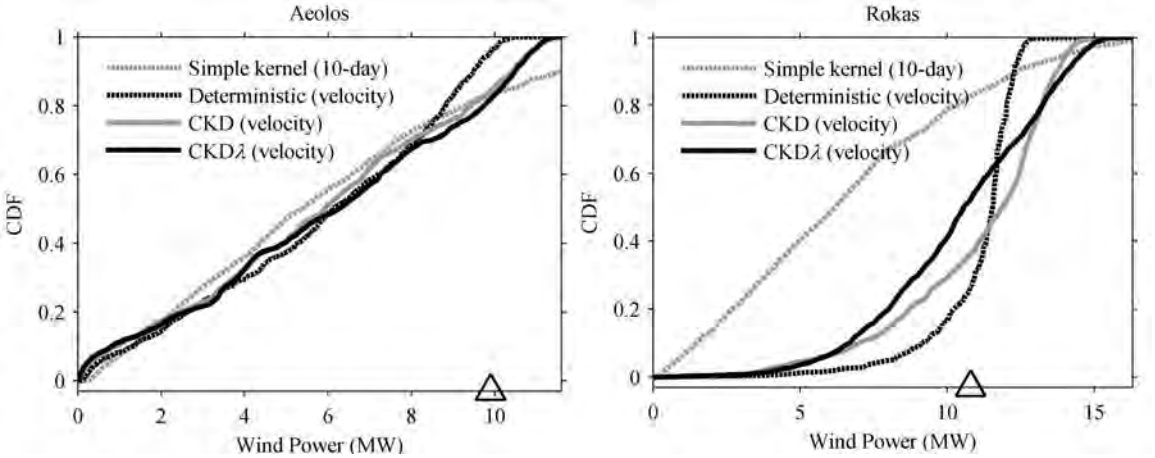


Figure **9. For the Aeolos** and the **Rokas** wind farms, **four hour-ahead forecasts of the wind power cdf. The forecast origin is at 7pm on October 01, 2006, which was the final period of the in-sample set of data. The wind power observation is indicated by a triangular symbol on the *x*-axis.**

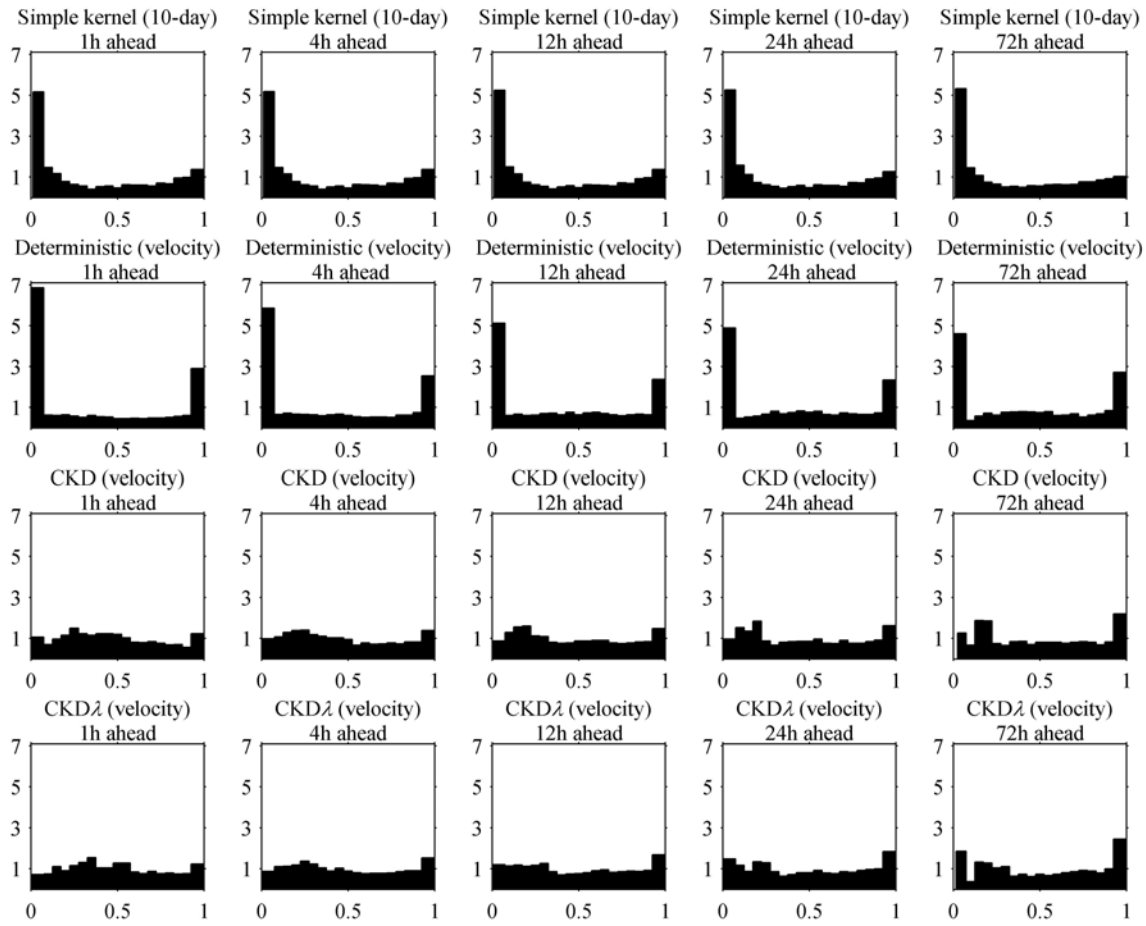


Figure 10. Summary PIT histograms for the post-sample density forecasts for the Greek datasets.

Very short-term probabilistic forecasting of wind power with generalised logit-Normal distributions

Pierre Pinson

DTU Informatics, Technical University of Denmark, Kgs. Lyngby, Denmark

Abstract

Very short-term probabilistic forecasts, which are essential for an optimal management of wind generation, ought to account for the nonlinear and double-bounded nature of that stochastic process. They take here the form of discrete-continuous mixtures of generalised logit-Normal distributions and probability masses at the bounds. Both autoregressive and conditional parametric autoregressive models are considered for the dynamics of their location and scale parameters. Estimation is performed in a recursive least-squares framework with exponential forgetting. The superiority of this proposal over classical assumptions about the shape of predictive densities e.g. Normal and Beta, is demonstrated based on 10-minute ahead point and probabilistic forecasting at the Horns Rev wind farm in Denmark.

Keywords: bounded time-series, logit transformation, mixture model, probabilistic forecasting, wind power

* Corresponding author:
P. Pinson, DTU Informatics, Technical University of Denmark,
Richard Petersens Plads (bg. 305 - 212), DK-2800 Kgs. Lyngby, Denmark.
Tel: +45 4525 3428, fax: +45 4588 2673, email: pp@imm.dtu.dk, webpage: pierrepinson.com

1 Introduction

In different areas of forecasting, both theoretical and practical developments are pointing towards various forms of probabilistic forecasting. This has been the case in economics and finance[46,48], earthquake prediction[20], meteorology[25], as well as for weather-related processes like renewable energy production[4,40] or floods[9]. When considering continuous variables such as wind power, they optimally take the form of predictive densities—equivalently referred to as predictive distributions or density forecasts, giving the full probability density function of the power random variable for a set of lead times. The optimal management and trading of wind power calls for probabilistic forecasts[35]. This follows from a more general result: for a large class of decision-making problems, optimal decisions directly relate to quantiles of conditional predictive distributions[15].

Lead times in the range of hours to days correspond to the needs of decision-makers in view of the structure and operation of European electricity markets[10,14]. They have thus attracted the most attention so far. Density forecasts for these horizons are commonly based on statistical models for the "dressing" of point predictions, conditional to weather conditions[7,31,36]. Alternatively, they result from the dynamic post-processing of ensemble forecasts of meteorological variables[33,37,47]. When wind power penetration reaches a certain level, and with large-scale wind farms injecting power at a single point of the electricity network, it becomes crucial to also have forecasts for horizons of a few minutes ahead. Characterising and modelling the power fluctuations at these time scales is recognised as a current challenge. Firstly, this is required by the Transmission System Operator (TSO) in order to optimally operate reserves for the continuous balance of the power system. In that case, the relevant horizons are defined as 1, 10 and 30 minutes[44]. More specifically for the case of Denmark, the TSO Energinet.dk has defined the 10-minute lead time as the most important one since power fluctuations at this time scale are those that most seriously affect the balance in the power system[3]. This aspect directly motivates our choice to focus on 10-minute ahead forecasts only. Secondly, very short-term predictions are needed as input to the (offshore) wind farm controllers themselves with typical time steps between 5 and 15 minutes[21,45]. Finally, the operation of wind-storage systems at temporal scales consistent with the temporal resolution at which regulation is carried out in practise in some electricity markets e.g. 15 minutes in the Netherlands, or 5 minutes in New Zealand and Australia, calls for very short-term forecasts. Even though emphasis is placed on the 10-minute lead time only, extension to further horizons can be the topic of future work. This may be based on the same predictive distributions but with different input data and dynamics models.

For such short lead times, statistical models based on historical measurements only, though taking advantage of meteorological and physical expertise on the problem at hand, should be preferred. Running Numerical Weather Prediction (NWP) models for these temporal resolutions with frequent updates would be clearly too expensive if not impossible today. Also statistical approaches are known to outperform NWP-based ones for the forecasting of the power output of a wind farm at lead times less than a few hours[14]. Examples of recent works on improving such statistical approaches, focusing on the wind (speed and potentially direction) variable, include the regime-switching space-time method introduced by Gneiting *et al.* [17], and the multivariate wind vector models of Hering and Genton [18]. Some other works have been focusing on capturing changes in the wind field dynamics using e.g. Markov-switching approaches[1]. In parallel, Pinson and Madsen [38] discuss the benefits of Markov-switching autoregressive models for point and probabilistic forecasting of wind power generation at forecast horizons of 10-minutes ahead.

We place ourselves in a parametric probabilistic forecasting framework, for which very few proposals exist in the literature, especially at these lead times. Our core objective is to discuss how wind power characteristics motivate the choice for specific types of predictive distributions. A first crucial feature of the wind power

variable is that it is double-bounded between a minimum production of zero and a maximum one being equal to the nominal capacity (denoted P_n) of the wind turbine, wind farm, or wind energy portfolio considered. It is in parallel a nonlinear function of wind speed in the form of a sigmoid. Wind power measurements and forecasts are normalised by P_n in the following. They then take values in the unit interval $[0, 1]$. The time variations of P_n, if any, are commonly recorded for recent wind farms. They can be used for an adaptive scaling of measurements and forecasts. A consequence of the above is that predictive densities cannot be Gaussian. This is valid whatever the temporal resolution of wind power time-series or the forecast horizon considered. Their higher-order moments are also directly related to their mean, and potentially to some external signal. These two points have already been discussed in a number of publications[5,22,34]. To our knowledge however, no work exists in the literature where these aspects have been appropriately taken into consideration in a probabilistic forecasting framework. For illustrative purposes, Figure 1 depicts an episode with 2 days of successive 10-minute average power output at the Horns Rev offshore wind farm which will be the focus of the application and results part of the paper.

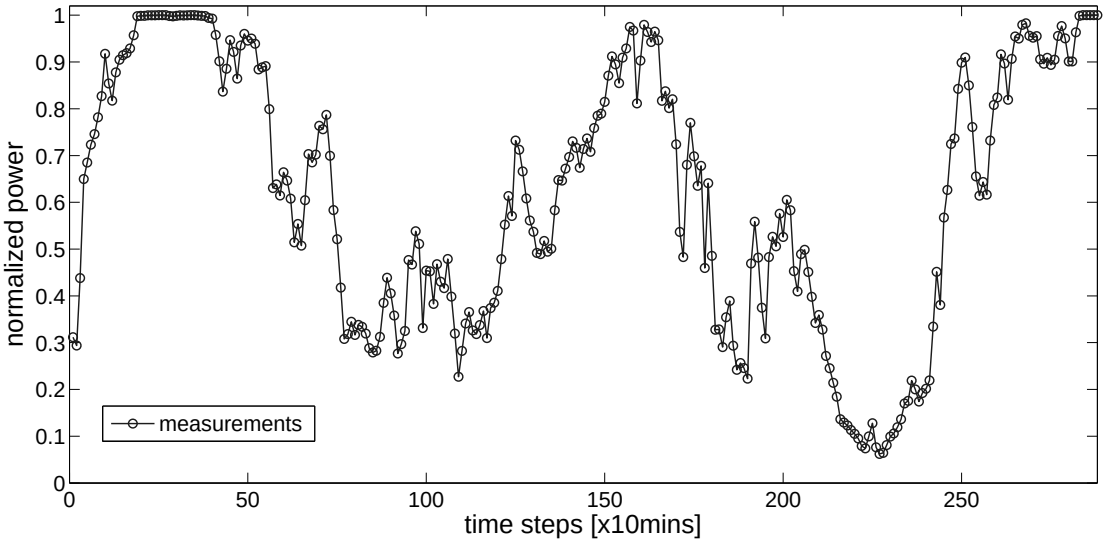


FIGURE 1: *Episode of 2 days (288 time-steps, with a temporal resolution of 10 minutes) with wind power measurements at the Horns Rev wind farm in Denmark.*

We propose here a methodology for the probabilistic forecasting of wind power time-series that account for these characteristics. Predictive densities take the form of discrete-continuous mixtures consisting of a generalised logit-Normal (GL-Normal) distribution with potential concentration of probability mass at the bounds of the unit interval $[0, 1]$. Given a shape parameter, these predictive densities are fully characterised by a location and a scale parameter only. By employing the related generalised logit transform for variance stabilisation, the obtained variable can be modelled with censored Gaussian distributions. After preliminary definitions in Section 2, the forecasting methodology is described in Section 3. Both model building and estimation aspects are subsequently dealt with in Section 4, with focus on simple autoregressive (AR) models or more advanced conditional parametric autoregressive (CP-AR) models. These models are considered for the sake of example only and could be further improved in the future. In the case of further lead times, say up to a few days ahead, it should be envisaged to build models additionally accounting for the weather dynamics as given by meteorological forecasts. The methodology is applied in Section 5 for the probabilistic forecasting of wind power generation at the Horns Rev wind farm in Denmark, at lead times of 10 minutes. It is evaluated against some simple benchmarks, as well as against other assumptions that could be made regarding the shape of predictive

densities e.g. Normal or Beta. The paper ends with concluding remarks in Section 6.

2 Preliminary definitions

Transformations are commonly employed for variance stabilisation, and to work in a Gaussian framework. Besides the general class of power or Box-Cox transformations [6], Johnson [19] originally introduced the so-called logistic (or logit) transform, subsequently used for e.g. bounded economic variables [52]. It directly relates to using logit-Normal distributions for modelling the original effect [2]. A generalisation of both the logit transform and of logit-Normal distributions is given below, being appropriate for the case of wind power modelling and forecasting applications.

Wind power generation is a double-bounded process, thus making the logistic transformation appealing. Owing to the different inflection in power curves for low and high power values, a suitable transformation should allow for asymmetry. The generalised logit (GL) transformation inspired by Mead [30] hence comprises an ideal candidate. For an original time-series $\{x_t\}$, the generalised logit transform y_t is given by

$$y_t = \gamma(x_t;\nu) = \ln\left(\frac{x_t^\nu}{1-x_t^\nu}\right), \quad \nu > 0, \quad x_t \in (0,1) \tag{1}$$

while the inverse transformation, referred to as inverse generalised logit (IGL), is defined as

$$x_t = \gamma^{-1}(y_t;\nu) = \left\{1 + \frac{1}{\exp(y_t)}\right\}^{-1/\nu}, \quad \nu > 0, \quad y_t \in \mathbb{R} \tag{2}$$

The GL transformation generalises the logistic one with an additional shape parameter ν only. Its purpose is to influence the evolution of variance and skewness of these distributions as a function of their mean. For $\nu = 1$, one retrieves the more classical logit transformation, recently used by Lau and McSharry [23] for wind power prediction. The aim of employing such transformation is to work with a time-series $\{y_t\}$ for which the assumption that conditional densities may be Gaussian is more appropriate. In parallel, in view of the known relationship between conditional expectation and variance of wind power densities [5,34], it is expected that the conditional variance of the predictive densities for the transformed variable will be independent of their mean.

We introduce the generalised logit-Normal distribution (GL-Normal) based on the aforementioned GL transform. It is denoted by $L_\nu(\mu,\sigma^2)$, where μ and σ are the location and scale parameters, respectively. For a Gaussian variable $Y \sim \mathcal{N}(\mu,\sigma^2)$, the transformed variable $X = \gamma(Y;\nu)$ is such that $X \sim L_\nu(\mu,\sigma^2)$. By generalising the probability density function of a univariate logit-Normal variable given by Aitchison and Shen [2] and Frederic and Lad [12], we obtain that for the GL-Normal variable,

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\left(\frac{\nu}{x(1-x^\nu)}\right)\exp\left\{-\frac{1}{2}\left[\frac{\gamma(x;\nu)-\mu}{\sigma}\right]^2\right\}, \quad x \in (0,1) \tag{3}$$

We do not discuss here the moments of a variable that would be distributed GL-Normal. One could carry out theoretical developments in the spirit of Aitchison and Shen [2] to derive some properties and perform comparisons with the class of logit-Normal and Dirichlet distributions. It could also be envisaged to perform similar numerical developments as in Frederic and Lad [12] in order to visualize the moments of GL-Normal distributions as a function of their location and scale parameters.

3 Forecasting methodology based on generalised logit-Normal distributions

Several authors e.g. Lange [22] and Pinson [34], have observed that the standard deviation of the distributions of forecast errors was directly linked to its conditional expectation, whatever the time of the year, horizon and forecast model. Appropriate predictive densities of wind power generation should account for and reproduce this observed feature. Recently, some models for this relationship have been proposed[5,37,49]. More precisely, Bludszuweit *et al.* [5] looked at second-order polynomial mean-variance models for obtaining the parameters of Beta distributions when analysing the uncertainty of wind power forecasts, a posteriori. The models they obtained had the general form of asymmetric logistic models. Based on a similar observation, Usaola [49] modeled moments and cumulants of forecast error distributions as a function of the level of predicted power, as input to power flow models. Finally Pinson and Madsen [37] introduced a logistic mean-variance model for the Gaussian kernels used in the dressing of ensemble forecasts of wind power generation. Applying an appropriate GL-transform to wind power data allows dampening the effect: the variance of the predictive densities for the transformed variable can then be assumed to be independent of its mean.

Write $\{X_t\}$ and $\{Y_t\}$ the stochastic processes whose realisations define the time-series $\{x_t\}$ and $\{y_t\}$, following the preliminary definitions of Section 2. Wind power generation is null for wind speed values below cut-in speed, while being equal to nominal capacity when wind speed is above the rated speed. The values for cut-in and rated wind speeds may well vary depending upon the turbines and wind power portfolio considered. We consequently propose to write the predictive density for the wind power generation X_{t+k} at time $t+k$ as

$$X_{t+k} \sim \omega^0_{t+k}\delta_0 + (1-\omega^0_{t+k}-\omega^1_{t+k})L_\nu(\mu_{t+k},\sigma^2_{t+k}) + \omega^1_{t+k}\delta_1 \tag{4}$$

with δ_0 and δ_1 Dirac delta functions at 0 and 1, respectively, representing the potential concentration of probability mass at the bounds of the unit interval. To be consistent, one has

$$\omega^0_{t+k}, \omega^1_{t+k} \in [0,1], \quad \omega^0_{t+k} + \omega^1_{t+k} \in [0,1]$$

The main reason for the introduction of this concentration of probability mass at the bounds is that the original variable to be GL-transformed must take values in $(0,1)$ only (as expressed in (1)). For the case of wind power, these bounds are obviously reached on a regular basis. Consider for instance issuing 10-minute ahead point forecasts of wind power generation at Horns Rev (the test case site of Section 5). Point forecasts correspond to the 1-step ahead conditional expectation of wind power production based on an AR model. The distributions of observed wind power generation conditioned upon the prediction level are depicted in Figure 2, for various power classes close to the natural generation bounds for that wind farm. The non-negligible concentration of probability mass at these bounds is visible from this Figure. In a general manner, other forms of distributions e.g. exponential ones, could be used in (4) instead of probability masses at the bounds. This would reflect that even if predicting zero (or nominal) power, there still may be some uncertainty.

Based on (4), but now considering Y_{t+k} the GL-transform of X_{t+k}, the form of its predictive density is given by

$$Y_{t+k} \sim \omega^0_{t+k}\delta_{-\infty} + (1-\omega^0_{t+k}-\omega^1_{t+k})\mathcal{N}(\mu_{t+k},\sigma^2_{t+k}) + \omega^1_{t+k}\delta_{+\infty} \tag{5}$$

Similarly to the coarsening approach of Lesaffre *et al.* [24] for the modelling of outcome scores in $[0,1]$ using classical logit-Normal distributions, we define a threshold value ϵ, being in the order of measurement precision, say $\epsilon \leq 10^{-2}$. Wind power values below or equal to ϵ (resp. above or equal to $1-\epsilon$) are considered to be 0

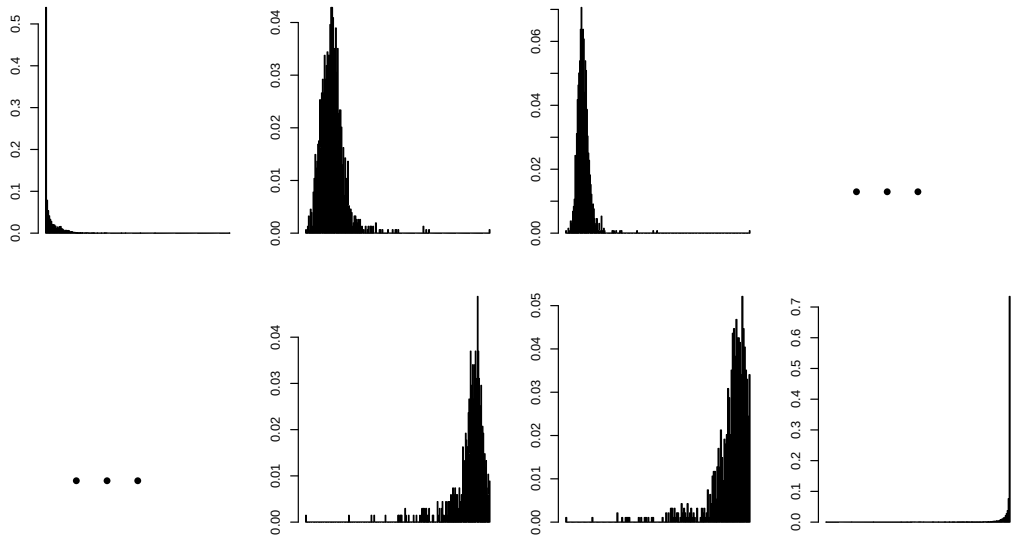


FIGURE 2: *Distributions of observed power conditional to the predicted expected power (within [0,2] % of P_n, [2,4], [4,6], [94,96], [96,98], [98,100]). The same scale is used for the x axis in all bar plots.*

(resp. 1). This translates to constraining the range of potential variations of the GL-transformed variable, i.e. $y_t \in \mathcal{D}_y = [\gamma(\epsilon;\nu), \gamma(1-\epsilon;\nu)]$. One can consequently rewrite (5) as

$$Y_{t+k} \;\sim\; \omega^0_{t+k}\delta_{\gamma(\epsilon;\nu)} + \mathcal{N}(\mu_{t+k}, \sigma^2_{t+k})\mathbf{1}_{\bar{\mathcal{D}}_y} + \omega^1_{t+k}\delta_{\gamma(1-\epsilon;\nu)} \tag{6}$$

where $\bar{\mathcal{D}}_y$ is the complement of $\mathcal{D}_y$ and $\mathbf{1}_{\bar{\mathcal{D}}_y}$ the indicator function for this open interval, i.e.

$$\mathbf{1}_{\bar{\mathcal{D}}_y} = \begin{cases} 1, & y \in \bar{\mathcal{D}}_y \\ 0, & \text{otherwise} \end{cases} \tag{7}$$

The weight $(1 - \omega^0_{t+k} - \omega^1_{t+k})$ does not appear anymore in (6). This is because the predictive density for Y_{t+k} takes the form of a censored Normal instead. This permits to straightforwardly derive the expressions for the weights ω^0_{t+k} and ω^1_{t+k},

$$\omega^0_{t+k} = \Phi\left(\frac{\gamma(\epsilon;\nu) - \mu_{t+k}}{\sigma_{t+k}}\right), \quad \omega^1_{t+k} = 1 - \Phi\left(\frac{\gamma(1-\epsilon;\nu) - \mu_{t+k}}{\sigma_{t+k}}\right) \tag{8}$$

with Φ the cumulative distribution function of a standard Normal variable. Predictive densities are then fully characterised by their location (μ_{t+k}) and scale (σ_{t+k}) parameters only.

4 Dynamic models for the location and scale parameters

We aim to show the advantages of modelling and forecasting wind power time-series with GL-Normal variables, not to discuss the superiority of certain dynamic models. Therefore the developments gathered below concentrate on fairly simple dynamic models only. They could however be extended to account for additional effects e.g. conditional heteroskesdaticity or regime-switching dynamics. It would clearly call for modification of the estimation methods described. GL-Normal densities could also be envisaged for further lead times (i.e. from a few hours to a few days ahead), for which appropriate statistical approaches should account for the

nonlinear transformation of the dynamic information provided by meteorological predictions.

To give a brief overview of our approach, AR and CP-AR models are proposed for the location parameter. In parallel, the scale parameter is viewed as constant in the former case, or as a nonlinear function of the covariate conditioning the AR dynamics in the latter one. Only AR types of dynamics are considered, as earlier results on similar offshore data suggested that such dynamics for the mean are appropriate when considering short lead times[38]. Extensive data analysis reported by Vincent *et al.* [50] and Vincent *et al.* [51] showed that seasonalities were not observed at the Horns Rev wind farm considered as a case study in the following. They are thus not accounted for in the various models below. Model coefficients are adaptively and recursively estimated based on an exponential forgetting of past observations. This proposal follows from various results reported in the literature[17,38,42,50] e.g. showing that parameters in dynamic models should be seen as slowly changing over time. This is because the dynamics of the processes involved may well vary at time scales of months, seasons and years, with features that cannot easily be accounted for in the models themselves in a parametric framework.

4.1 Autoregressive dynamics

In the case of AR dynamics, the expression relating the location parameter μ_{t+k} to previous observations writes

$$\mu_{t+k} = \boldsymbol{\theta}^\top \tilde{\mathbf{y}}_t \tag{9}$$

where $\boldsymbol{\theta} = [\theta_0 \ \theta_1 \ \dots \ \theta_l]^\top$ and $\tilde{\mathbf{y}}_t = [1 \ y_t \ \dots \ y_{t-l+1}]^\top$. This is while the scale parameter is defined as a constant, i.e. $\sigma^2_{t+k} = \beta$.

The maximum lag l can be determined after examining the autocorrelation and partial autocorrelation functions of the $\{y_t\}$ time-series, by minimisation of either Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), or finally through a cross-validation exercise. Note that in the case of censored Normal variables, the location parameter may take values outside of the range within which y_t is constrained.

The $\{y_t\}$ time-series corresponds to realisations of censored Normal variables. It may be argued that the vector of AR parameters $\boldsymbol{\theta}$ can be estimated with either Least-Squares (LS) or Maximum-Likelihood (ML) techniques. In the case of censored Normal variables, the LS-estimator for the location parameter is certainly more subject to estimation bias than the ML-estimator. Experiments on wind power generation data showed that the LS-estimator of the location parameter was acceptable, while having the advantage of simplicity of implementation if compared with the ML-estimator. This is the reason why we only consider and describe the LS-estimator below.

The parameters $\boldsymbol{\theta}$ are estimated within a Recursive Least Squares (RLS) framework[27,28]. The scale parameter σ_{t+k} can similarly be tracked accordingly. Recursive estimation offers the advantage that only the last available measurements are used at each time step for updating the model parameters. This contrasts with the more computationally-intensive idea of employing a sliding window, as done by Gneiting *et al.* [17] and Hering and Genton [18], for which estimation is performed on a bulk of data at each time step.

The estimate of the parameters $\boldsymbol{\theta}_t$ at time t is defined as

$$\hat{\boldsymbol{\theta}}_t = \arg\min_{\boldsymbol{\theta}} \sum_{i=1}^{t} \lambda^{t-i} \left(y_i - \boldsymbol{\theta}^\top \tilde{\mathbf{y}}_{i-k} \right)^2 \tag{10}$$

where λ is the forgetting factor, $\lambda \in (0,1)$, allowing for the exponential forgetting of past observations, for

which the corresponding effective number of observations is defined as $n_\lambda = (1-\lambda)^{-1}$. The value of the forgetting factor is typically slightly below 1.

In a recursive framework, the formulae for updating the LS-estimate of $\boldsymbol{\theta}_t$ based on the newly available information at time t summarises to

$$\hat{\mathbf{R}}_t = \lambda \hat{\mathbf{R}}_{t-1} + \tilde{\mathbf{y}}_{t-k}\tilde{\mathbf{y}}_{t-k}^\top \tag{11}$$

$$\hat{\boldsymbol{\theta}}_t = \hat{\boldsymbol{\theta}}_{t-1} + \hat{\mathbf{R}}_t^{-1}\tilde{\mathbf{y}}_{t-k}\varepsilon_t \tag{12}$$

where ε_t is defined as $\varepsilon_t = y_t - \boldsymbol{\theta}_{t-1}^\top \tilde{\mathbf{y}}_{t-k}$. Note that there is a direct correspondence of the RLS estimator in the above with what would be the Kalman filter estimator of $\hat{\theta}_t$ in a state-space framework. These aspects are extensively discussed by e.g. Madsen [28] and Li [26]. Consequently, it is interesting to note that the posterior distribution of the true parameter $\boldsymbol{\theta}_t$ given the available information up to time $t-1$ is Gaussian with mean $\hat{\boldsymbol{\theta}}_{t-1}$ and covariance $\hat{\mathbf{R}}_{t-1}^{-1}$ [27] pp. 55-56.

The update (12) of the parameter estimates does not ensure that they will remain within the stationary region. This does not appear to be a problem for the 1-step ahead forecasting application. It would be more problematic if aiming at simulating the process over a period of time or at issuing multi-step ahead forecasts. With time-varying parameters, it may also well be that the parameter estimates cross into the nonstationary region and then get back into the stationary one after few time steps.

Since the scale parameter is given by a constant only, its adaptive estimation at time t can be performed thanks to an exponential smoothing scheme, being consistent with the way we update the parameters of the dynamic model for μ_{t+k}. Owing to the sensitivity of the GL transform close to the bounds of the unit interval, we have noticed that it would be beneficial to robustify the exponential smoothing of the scale parameter. This can be carried out by down-weighting observations when the location parameter gets closer to the bounds. Our proposal for this weight is given by

$$w_t = 4\boldsymbol{\theta}_t^\top \tilde{\mathbf{y}}_{t-k}\left(1 - \boldsymbol{\theta}_t^\top \tilde{\mathbf{y}}_{t-k}\right), \quad w_t \in [0,1] \tag{13}$$

And the resulting exponential smoothing scheme writes

$$\hat{\beta}_t = \lambda_t^* \hat{\beta}_{t-1} + (1-\lambda_t^*)\left(y_t - \boldsymbol{\theta}_t^\top \tilde{\mathbf{y}}_{t-k}\right)^2 \tag{14}$$

with $\lambda_t^* = 1 - (1-\lambda)w_t$.

It appears appropriate to initialise $\boldsymbol{\theta}_t$ with a vector of zeros, the resulting model hence corresponding to white noise. β_t may be initialised with a small value or alternatively with an expert guess. In parallel, the inverse covariance matrix $\hat{\mathbf{R}}_t$ can be initialised as an identity matrix times a small constant, yielding a small load on its diagonal. Only the updating formula (11) is used for the first steps of the recursive estimation procedure in order to gather sufficient information on the process characteristics. When that point is reached (say, for $t > 50$), both (12) and (14) are also employed at each time step for updating the models for the location and scale parameters. Note that $\hat{\mathbf{R}}_t$ is necessarily invertible in view of the updating performed in (11), as long as the underlying model is not over-parametrized.

4.2 Conditional parametric autoregressive dynamics

If considering CP-AR dynamics instead, the constant parameters in (9) are replaced by coefficient functions of some covariate, these functions being of unknown form,

$$\mu_{t+k} = \boldsymbol{\theta}(\omega)^\top \tilde{\mathbf{y}}_t \tag{15}$$

where $\boldsymbol{\theta}(\omega) = [\theta_0(\omega)\ \theta_1(\omega)\ \ldots\ \theta_l(\omega)]^\top$, while $\tilde{\mathbf{y}}_t$ is defined as in (9). In parallel the model for the scale parameter becomes an unknown function of this same covariate, i.e. $\sigma^2_{t+k} = \beta(\omega)$. The maximum lag l in the dynamic model above can be chosen similarly to that for the more simple AR dynamics. Recent developments on conditional parametric and the more general varying-coefficient models are reported in Fan and Zhang [11].

The most obvious choice for ω is wind direction, since it directly induces shadowing effects within a wind park, thus affecting the dynamic response of that wind park to meteorological conditions. Wind direction might also influence the wind speed dynamics themselves, as is the case at Horns Rev for instance[50]. We denote by ω_t the wind direction measurement at time t, $\omega_t \in [0, 2\pi]$. In view of the short lead times considered in the test case application, conditioning the AR dynamics upon the wind direction measurement available at time t can be seen as reasonable. If focusing on further lead times, it would certainly be beneficial to use wind direction forecasts instead.

The coefficient functions $\boldsymbol{\theta}(\omega)$ and $\beta(\omega)$ are estimated in a nonparametric framework, i.e. without assuming a shape for these functions. The method for their adaptive and recursive estimation combines local kernel regression and RLS with exponential forgetting. The estimation procedure described below is in the spirit of Nielsen *et al.* [32] and Pinson *et al.* [39], to which the reader is referred to for more details. The only assumption about the coefficient functions is that they are sufficiently smooth to be locally approximated with constants. The estimation problem reduces to locally fitting linear models at a number m of fitting points $\omega_{(j)}$. Defining these fitting points is best done by using information on the distribution of ω.

Let us focus on a single fitting point $\omega_{(j)}$ only. The objective function to be minimised at each time t consists of a modified version of that used for adaptive estimation of the AR-dynamic model considered above. The estimate of the set of local coefficients at $\omega_{(j)}$ and at time t is defined as

$$\hat{\boldsymbol{\theta}}_{(j),t} = \arg\min_{\boldsymbol{\theta}_{(j)}} \sum_{i=1}^{t} \Lambda_{(j),t}(i) c_{(j),i} \left(y_i - \boldsymbol{\theta}_{(j)}^\top \tilde{\mathbf{y}}_{i-k} \right)^2 \tag{16}$$

where the weights $c_{(j),i}$ to be assigned to past observations are given by a Kernel function of the form

$$c_{(j),i} = T\left(\frac{|\omega_i - \omega_{(j)}|}{\hbar_{(j)}} \right) \tag{17}$$

In the above, $|.|$ denotes a polar distance, being consistent with ω being a circular variable. $\hbar_{(j)}$ is the bandwidth of the kernel assigned to $\omega_{(j)}$. T may for instance be chosen as a tricube function,

$$T(v) = \begin{cases} \left(1 - v^3\right)^3, & v \in [0,1] \\ 0 \qquad\quad , & v > 1 \end{cases} \tag{18}$$

In parallel in (16), $\Lambda_{(j),t}$ is the function that permits the exponential forgetting of past observations, i.e.

$$\Lambda_{(j),t}(i) = \begin{cases} \lambda^{\text{eff}}_{(j),t} \Lambda_{(j),t-1}(i), & 1 \leq i \leq t-1 \\ 1 \qquad\qquad\qquad\quad , & i = t \end{cases} \tag{19}$$

$\lambda^{\text{eff}}_{(j),t}$ is the effective forgetting factor for the fitting point $\omega_{(j)}$. It follows the definition given by[32], which links

$\lambda^{\text{eff}}_{(j),t}$ and $c_{(j),t}$ so that

$$\lambda^{\text{eff}}_{(j),t} = 1 - (1-\lambda)c_{(j),t} \tag{20}$$

where λ is the classical user-defined forgetting factor, $\lambda \in (0,1)$. This effective forgetting factor ensures that old observations are down-weighted only when new information is available.

A little algebra, which is skipped here since extensively covered by Nielsen *et al.* [32] and Pinson *et al.* [39], yields the set of recursive formulae for the adaptive estimation of the local coefficients $\hat{\boldsymbol{\theta}}_{(j),t}$,

$$\begin{aligned}
\varepsilon_{(j),t} &= y_t - \hat{\boldsymbol{\theta}}^{\top}_{(j),t-1}\tilde{\mathbf{y}}_{t-k} && (21)\\
\mathbf{R}_{(j),t} &= \lambda^{\text{eff}}_{(i),t}\mathbf{R}_{(j),t-1} + c_{(j),t}\tilde{\mathbf{y}}_{t-k}\tilde{\mathbf{y}}^{\top}_{t-k} && (22)\\
\hat{\boldsymbol{\theta}}_{(j),t} &= \hat{\boldsymbol{\theta}}_{(j),t-1} + c_{(j),t}\mathbf{R}^{-1}_{(j),t}\tilde{\mathbf{y}}_{t-k}\varepsilon_{(j),t} && (23)
\end{aligned}$$

One sees that when the weight $c_{(j),t}$ equals 0 (thus meaning that the local estimates should not be affected by new information), then one has $\hat{\boldsymbol{\theta}}_{(j),t} = \hat{\boldsymbol{\theta}}_{(j),t-1}$ and $\mathbf{R}_{(j),t} = \mathbf{R}_{(j),t-1}$. This confirms the role of the effective forgetting factor, i.e. to down-weight old observations, but only when new information is available.

In parallel, one has to estimate $\hat{\beta}_{(j),t}$, the local constant that defines the scale parameter at time t and at $\omega_{(j)}$. As an extension to (14), it is given by

$$\hat{\beta}_{(j),t} = \lambda^{*}_{(j),t}\hat{\beta}_{(j),t-1} + (1-\lambda^{*}_{(j),t})\left(y_t - \boldsymbol{\theta}^{\top}_{(j),t}\tilde{\mathbf{y}}_{t-k}\right)^2 \tag{24}$$

with $\lambda^{*}_{(j),t} = 1-(1-\lambda)w_t c_{(j),t}$, w_t still being defined by (13).

When starting the recursive process, the local inverse covariance matrices $\mathbf{R}_{(j),t}$, as well as the local parameter estimates $\hat{\boldsymbol{\theta}}_{(j),t}$ and $\hat{\beta}_{(j),t}$, can be initialised for all fitting points as in Section 4.1 for the case of AR dynamics. The value of the parameter estimates $\hat{\boldsymbol{\theta}}_t(\omega)$ and $\hat{\beta}_t(\omega)$ for any ω can then be obtained by interpolation through the estimated local parameters at the various fitting points $\omega_{(j)}$ ($j = 1, \ldots, m$).

5 Application and results

Focus is given here to the application of the methodology and models described above to the test case of an offshore wind farm, for 10-minute ahead forecasting of wind power production. Such lead time corresponds to one-step ahead forecasts in view of the temporal resolution of the time series considered. Note though that the methodology has been introduced for general direct k-step ahead forecasting. Indeed, one could work along the lines of Gneiting *et al.* [17] for instance, and build models for direct 2-step ahead forecasting. Statistical models based only on historical measurements are in order for forecast horizons up to 6-8 hours ahead. However for horizons further than one hour ahead, dynamic information from meteorological forecasts permits to significantly improve forecast accuracy [10,14].

The test case and available data are introduced in the first stage, followed by a detailed description of the model configuration and estimation setup employed. The paper also discusses the way forecasts are evaluated, as well as the benchmarks considered. Application results and related comments are gathered at the end.

5.1 Test case and available data

We consider the case of the offshore wind farm located at Horns Rev, off the west coast of Jutland in Denmark. This wind farm has a nominal capacity P_n of 160 MW. The original wind and power measurement data includes raw one-second measurements for each wind turbine. Focus is given to the total power output, normalised by P_n, and to the average wind direction for the wind farm. An averaging procedure has been developed in order to obtain time-series of 10-minute wind direction and power averages. This averaging rate makes the very fast fluctuations related to the turbulent nature of the wind disappear and reveal slower fluctuations at the minute scale. Because there may be some erroneous or suspicious data in the raw measurements, the averaging procedure has a threshold parameter τ_v, which corresponds to the minimum percentage of data needed to be considered as valid within a given time interval, so that the related power (or wind direction) average is also considered as valid. Choosing a sufficiently high value for τ_v permits to avoid sampling effects in the averaging procedure that would impact the representativeness of the average values obtained. The threshold chosen is $\tau_v = 75\%$. The available raw data is from 16$^{\text{th}}$ February 2005 to 25$^{\text{th}}$ January 2006, consisting of 49536 data points. As a result of the averaging procedure, the wind direction and power generation dataset consists of 41651 valid data points.

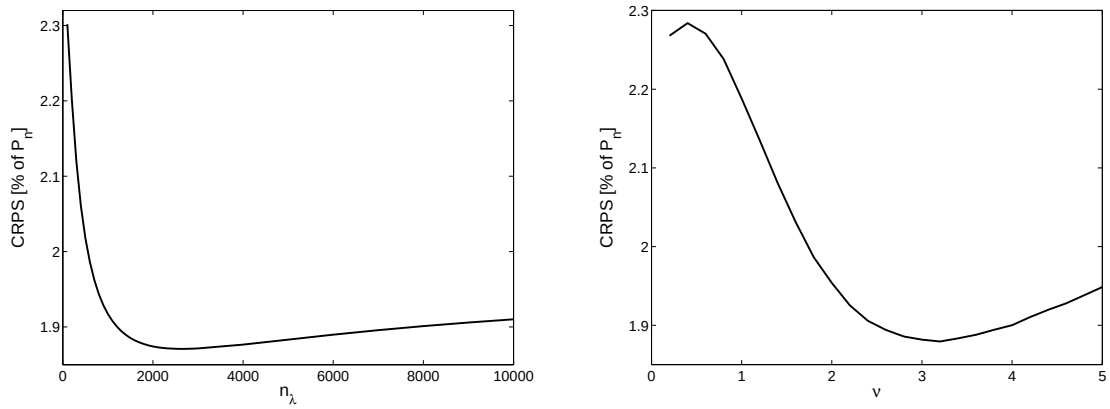
5.2 Model configuration, estimation setup and benchmarking

From the available data, two periods are defined, the first one being used for identification (and initial training) of the statistical models, and the second one for evaluating what the performance of these models may be in operational conditions. The first 15120 data points (with 11968 valid ones) are employed as the learning set, exactly covering the months from February to May 2005. The remainder of the dataset, covering a period from June 2005 to January 2006, is used for out-of-sample evaluation of 1-step ahead forecast performance. This evaluation set contains 34536 data points, including 29683 valid ones.

Over the learning period, a part of the data is used for one-fold cross validation (the last 9000 points) in order to select an optimal autoregressive order for the dynamic models, the values of the forgetting factors, the shape parameter of the GL-transformation, as well as the fitting points and bandwidths for the case of the conditional parametric models. Actually, instead of considering the forgetting factor itself, it is preferred to use the corresponding effective number of observations n_λ. It allows one to better appraise the size of the equivalent 'sliding window' in the adaptive estimation of the dynamic model parameters, such as that considered by Gneiting *et al.* [17] and Hering and Genton [18]. In parallel, the censoring parameter ϵ is arbitrarily set to $\epsilon = 0.001$, corresponding to the resolution of the power measurements. Selection of optimal values for the model structure and parameters n_λ and ν is done in a trial-and-error manner, by evaluating the results obtained from a set of different setups. For more information on cross validation, we refer to Stone [43]. The criterion to be minimised over the cross-validation set is the Continuous Rank Probability Score (CRPS) of 1-step ahead density forecasts. The CRPS is a common proper skill score used for the evaluation of density forecasts[16]. For the CP-AR dynamics, the bandwidth values are selected similarly.

Focusing on the order of the various models first, results from Pinson and Madsen [38] suggest that a maximum lag $l = 3$ is sufficient for capturing the dynamics in offshore wind power fluctuations with a 10-minute resolution, at the Horns Rev wind farm. We verified that for all models considered, a significant decrease in 1-step ahead CRPS values can be observed when augmenting l up to $l = 3$, and that no further benefit is obtained by further increasing the value of l. Applying a parsimonious model principle, we then decide to concentrate on models of order $l = 3$ only. In all cases, the AR models are initialised as simple random walk models, with very low variance of predictive densities.

In the case of the GL-Normal densities and AR dynamics, one may consider the selection of n_λ and ν to be performed simultaneously or iteratively. It has been observed for the data of the present case-study that both ways yielded similar results. We thus present them below in Figure 3 after n_λ and ν were optimised one after the other, iteratively. More precisely, Figure 3(a) depicts the variations in 1-step-ahead CRPS for ν fixed to its optimal value, and with n_λ varying between 100 and 10.000, with steps of 100 observations. Similarly, Figure 3(b) shows the evaluation of the 1-step-ahead CRPS for n_λ fixed to its optimal value, and with ν varying between 0.2 and 5, with increments of 0.2. The overall minimum 1-step-ahead CRPS value on the cross-validation set is obtained for $n_\lambda = 2500$ (corresponding to approximately 17.5 days) and $\nu = 3.2$. This optimal value for n_λ is smaller than the 45 days deemed optimal for the sliding estimation windows of Gneiting *et al.* [17] and Hering and Genton [18] when considering lead times of 2 hours. The speed of change in the characteristics of wind and power dynamics certainly depends upon the temporal resolution of the time-series. It may also be related to the local wind climate: the offshore wind farm at Horns Rev is known for particular characteristics of its wind and power dynamics[38,51].



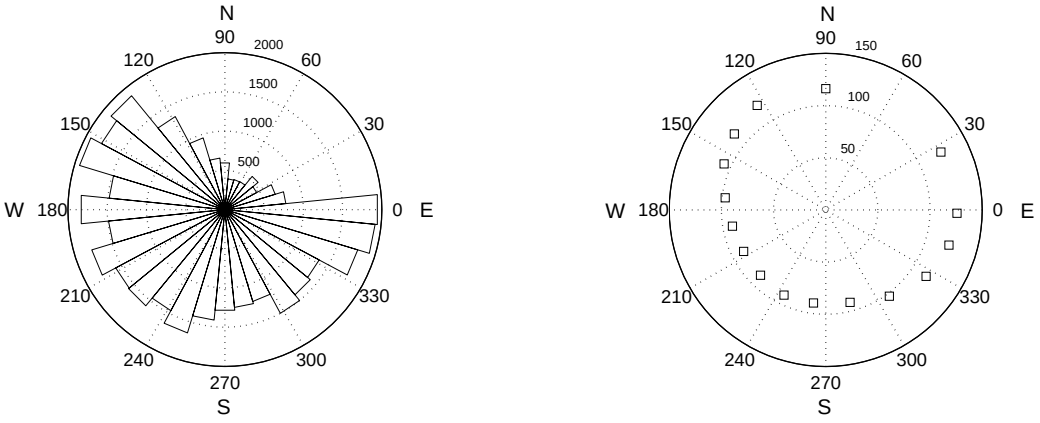
(a) As a function of n_λ (for $\nu = 3.2$) (b) As a function of ν (for $n_\lambda = 2500$)

FIGURE 3: *Evolution of the 1-step-ahead CRPS of GL-Normal density forecasts of wind power generation based on simple AR-dynamics. These results are obtained on the cross-validation set, as a function of the parameters n_λ and ν.*

Figure 3(b) demonstrates the interest of introducing the generalised form of the logit transformation and hence to generalise the use of the more classical logistic transform by Lau and McSharry [23]: a significant decrease in CRPS is seen for $\nu > 1$, i.e. when considering negatively-skewed mean-variance relationships for the GL-Normal predictive densities. Our proposal probabilistic forecasting methodology seems to be fairly robust with respect to the choice of n_λ and ν values, as the 1-step-ahead CRPS of predictive densities does not vary much for $n_\lambda \in [2000, 4000]$ and for $\nu \in [2.5, 3.5]$.

Turning our attention to the GL-Normal predictive densities with CP-AR dynamics, an exercise similar to that performed above led to the same values for n_λ and ν, i.e. $n_\lambda = 2500$ and $\nu = 3.2$. In addition, the location of the fitting points and related bandwidth values for the local fitting of coefficient functions have to be selected. It has been arbitrarily decided to use 16 fitting points, as such a number allows a fairly good coverage of potential wind directions values, while limiting the computational power needed. The spreading of these fitting points has been performed in view of the distribution of wind direction values over the learning period, i.e. such that there is the same frequency of data between any two successive fitting points. Regarding the Kernel bandwidth in (17), both fixed and nearest-neighbour bandwidths have been envisaged. The latter concept was preferred, in order to be consistent with the idea introduced above, so it would be more consistent to

respect the distribution of wind direction data when defining the fitting points. Following Nielsen *et al.* [32], the nearest-neighbour bandwidth is defined by the share of the (wind direction) data that should be covered by the Kernel attached to any single fitting point. This share is chosen here to be 60% in order to obtain sufficiently smooth variations of the coefficient functions in the CP-AR model. Smaller values have been tested, though not improving the forecast performance of the model over the cross-validation set. The resulting setup for the fitting points and related bandwidth values is depicted in Figure 4, along with the corresponding wind rose at Horns Rev. In the polar representation of Figure 4(b), the polar coordinates give the location of the fitting points, while the bandwidths values are given by the ordinate of the various points. It is visually straightforward to recognise that less fitting points and higher bandwidth values are present in areas with less data available.



(a) wind rose

(b) fitting points and related bandwidth values

FIGURE 4: *Wind rose, fitting points and related bandwidths values based on the distribution of wind direction values over the learning set. All values are in degrees, following the trigonometric convention.*

As a basis for the evaluation of the forecast performance of the methodology introduced in the paper, different types of benchmark methods are considered, some of them being reviewed by Giebel *et al.* [14], Costa *et al.* [10] and Pinson and Madsen [38]. First of all for point forecasting the persistence forecast method, for which the forecast at time $t + k$ is given by the measurement available at time t, is known to be difficult to outperform for short lead times. Another basic benchmark is an extension of persistence based on the average of the last few measurements. To be consistent with the autoregressive orders of the models employed here, this moving average benchmark consists of the average of the last 3 measurements. In parallel the most common benchmark for probabilistic forecasts of meteorological or weather-related processes is climatology, which is based on always issuing the same unconditional predictive density built from all historical observations available. Since climatology is fairly easy to outperform for short lead times, we also consider the probabilistic generalization of persistence based on a random walk model as a benchmark. At a given time this probabilistic forecast takes the form of a Gaussian predictive density with its mean given by the last available power measurement and its variance resulting from exponential smoothing of past squared residuals. For consistency, this exponential smoothing uses the same forgetting factor as for the more advanced approaches considered, i.e. with $n_\lambda = 2500$.

The main objective of the present paper is to demonstrate the interest of considering GL-Normal distributions for wind power prediction. As a basis for comparison, we therefore consider similar forecasting methodologies based on more classical assumptions about the shape of predictive densities, but with the same dynamical models for their location and shape parameters. For point forecasting the common assumption is that of a

TABLE 1: *Monthly results for the evaluation of point forecasts extracted from density forecasts: focus on the NMAE for the median (in % of the nominal capacity P_n). Best score values are highlighted using bold fonts.*

Month	Jun.	Jul.	Aug.	Sept.	Oct.	Nov.	Dec.	Jan.	All
Moving average	3.44	3.28	3.47	3.62	3.47	3.57	3.60	3.62	3.50
Persistence	2.40	2.43	2.50	2.50	2.40	2.59	2.61	2.47	2.48
Normal AR	2.34	2.42	2.48	2.43	2.33	2.64	2.63	2.45	2.45
GL-Normal AR	2.32	2.35	2.43	2.38	2.28	**2.47**	2.54	2.40	2.38
GL-Normal CP-AR	**2.31**	**2.34**	**2.42**	**2.37**	**2.26**	**2.47**	**2.53**	**2.36**	**2.36**

Normal distribution, while in a probabilistic forecasting framework, the most common ones are (censored) Normal or Beta distributions[5,38]. Additional point and probabilistic forecasts based on these assumptions will be seen as a benchmark. Results will be given for AR models only, and labelled as 'Normal AR' and 'Beta AR'. The optimal value found for the effective number of observations n_λ is $n_\lambda = 2000$ (corresponding to approximately 14 days) for the case of both assumptions on predictive densities.

5.3 Results and comments

A fully probabilistic approach to forecast evaluation is employed. This means that focus is given to evaluating the quality of predictive densities, including their reliability and overall skill, as well as to assessing the quality of some point forecasts that can be extracted from such predictive densities. Following the point of Gneiting [15] among others, one should extract the optimal point forecasts from predictive densities based on the target evaluation score. Indeed, the median of predictive densities should be selected if the target evaluation score is of the Mean Average Error type (abbreviated MAE, or NMAE for its normalised version). Similarly, if the target score is of quadratic nature like the Root Mean Square Error (RMSE, or NRMSE for its normalized version), the optimal point forecast to be extracted is the expectation of predictive densities.

In view of these aspects, we focus first on the quality of point forecasts related to the median and expectation of predictive densities, since the corresponding NMAE and NRMSE scores appear to be the most employed in the wind power forecasting literature[10,14,29]. Owing to the length of the evaluation set (approximately 8 months), it has been chosen to study these scores overall and on a monthly basis. The evaluation results for point forecasts based on the persistence and moving-average benchmarks, as well as for the Normal-AR, and the proposed GL-Normal AR and CP-AR predictive densities, are collated in Tables 1 and 2. Whatever the score considered, the persistence benchmark indeed seems competitive: the more advanced approaches only propose overall improvements up to 5%. This is due to the inertia in local atmospheric processes at such temporal scales. Persistence has a clear advantage over the other moving-average benchmark, for which all score values are much higher. Even though monthly fluctuations can be observed in the NMAE and NRMSE scores, it appears that the Normal AR benchmark almost consistently outperforms the persistence and moving-average benchmarks, while being outperformed by the point forecasts extracted from GL-Normal predictive densities. A slight gain in point forecast accuracy is also seen if considering that the moments of GL-Normal predictive densities may vary as a function of wind direction, as is the case with the CP-AR dynamic model. The recent work of Gallego *et al.* [13], where it was shown that for short lead times wind direction influences power dynamics in a smooth manner and that this should be accounted for in statistical approaches to forecasting, confirms our results.

In the second stage, we look at the quality of predictive densities, concentrating first on their overall skill

TABLE 2: *Monthly results for the evaluation of point forecasts extracted from density forecasts: focus on the NRMSE for the expectation (in % of the nominal capacity P_n). Best score values are highlighted using bold fonts.*

Month	Jun.	Jul.	Aug.	Sept.	Oct.	Nov.	Dec.	Jan.	All
Moving average	5.70	6.06	5.93	6.35	5.79	8.06	6.35	5.88	6.31
Persistence	3.98	4.68	4.47	4.66	4.17	6.23	4.76	4.28	4.70
Normal AR	3.80	4.56	4.36	4.46	3.97	6.00	4.62	4.20	4.54
GL-Normal AR	3.81	4.57	4.32	4.43	3.96	**5.87**	**4.55**	4.12	4.50
GL-Normal CP-AR	**3.79**	**4.53**	**4.30**	**4.42**	**3.93**	**5.87**	4.56	**4.08**	**4.47**

TABLE 3: *Monthly results for the evaluation of density forecasts with a CRPS criterion (in % of the nominal capacity P_n). Best score values are highlighted using bold fonts.*

Month	Jun.	Jul.	Aug.	Sept.	Oct.	Nov.	Dec.	Jan.	All
Climatology	18.67	19.75	20.07	18.61	19.79	24.65	21.81	21.63	20.47
Persistence	2.30	2.06	2.03	2.06	1.96	2.35	2.29	2.02	2.12
Normal AR	1.85	2.01	1.99	1.98	1.87	2.31	2.23	1.97	2.01
Beta AR	1.81	1.91	1.96	1.93	1.78	2.08	2.08	1.90	1.92
GL-Normal AR	1.76	1.80	1.86	1.86	1.81	**1.93**	1.92	2.03	1.86
GL-Normal CP-AR	**1.73**	**1.78**	**1.82**	**1.81**	**1.71**	1.94	**1.88**	**1.79**	**1.80**

by employing the CRPS criterion, being a proper skill score. Extensive discussion about the evaluation of probabilistic forecasts through the use of proper scoring rules can be found in Gneiting *et al.* [16]. Similar to the above, the CRPS criterion is calculated on a monthly basis as well as globally. Evaluation results are gathered in Table 3, for the two basic benchmarks (climatology and the probabilistic extension of persistence), Normal-AR and Beta-AR predictive densities, and finally for the proposed GL-Normal AR and CP-AR probabilistic forecasts.

The improvements in terms of the CRPS criterion are significant when going from the Normal predictive densities to GL-Normal predictive densities, in the order of 7.5%, and consistent over the evaluation period. The fact that Beta predictive distributions have a varying shape consistent with the double-bounded nature of wind power is reflected by the CRPS values for the Beta-AR predictive densities. The additional feature of GL-Normal distributions consisting of their natural mean-variance relationship still gives them an advantage though. Like for the case of point forecasts in the above, an additional improvement is observed if considering that the parameters of GL-Normal densities should be made a function of wind direction. The overall decrease in the CRPS criterion when going from Normal predictive densities with AR dynamics to GL-Normal predictive densities with CP-AR dynamics is of 10.5%. The climatology benchmark is dramatically outperformed in all cases, while the probabilistic extension of persistence still appears as a competitive benchmark. Overall, non-negligible benefits may be achieved by proposing adequate models for the wind power dynamics combined with an improved assumption about the shape of predictive densities.

We somehow expect that such improvement comes from the fact that GL-Normal predictive densities are more probabilistically reliable (or in other words, calibrated) than the Normal ones. Calibration of predictive densities can be assessed through the use of e.g. Probability Integral Transform (PIT) histograms, as carried out by Gneiting *et al.* [17], or with reliability diagrams in the form of quantile-quantile plots, as in Figure 5. Such reliability diagrams depict the observed proportions of a set of quantile composing predictive densities of wind power generation against the nominal ones[40]. Here the set of quantiles is defined with a step of 5% in their

nominal proportions. This assessment of the reliability of calibration of the quantiles composing predictive densities directly builds on their very definition: a quantile forecast with nominal proportion α should cover the actual observation $\alpha\%$ of the time. Climatology is not represented since it has perfect marginal calibration.

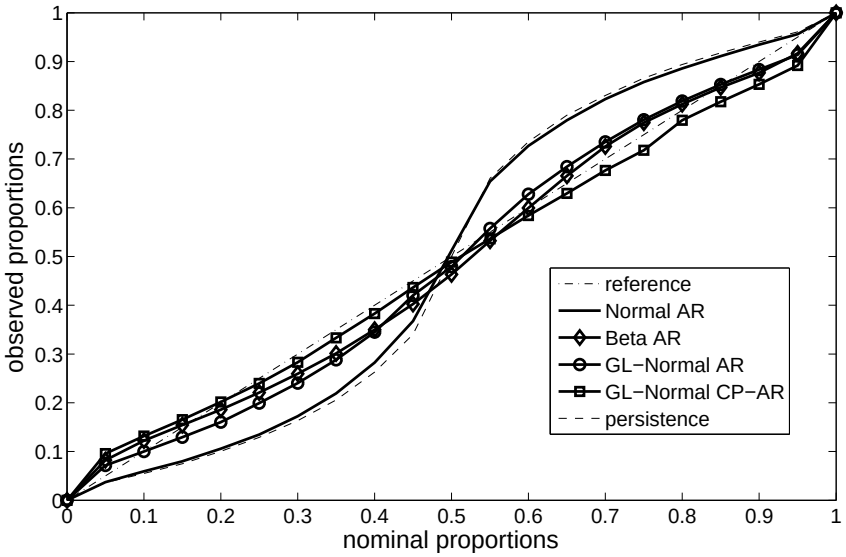


FIGURE 5: *Evaluation of the reliability of predictive densities with reliability diagrams in the form of quantile-quantile plots.*

Normal-AR predictive densities introduce a systematic and significant probabilistic bias in the probabilistic forecasts they provide. Predictive densities seem to be too wide on average, since observed proportions are larger than the nominal ones for nominal proportions above 0.5, and inversely for nominal proportions below 0.5. A simple explanation is that the variance of predictive densities is not conditional to the level of wind generation. Since wind generation is quite often at a medium level, the exponentially smoothed error variance used for the variance of predictive densities is fairly high. The variance is consequently over-predicted when wind generation is at a low and high level. The calibration results for the persistence benchmark are very close to that of Normal-AR predictive densities owing to the similarity in their definition. In contrast when employing GL-Normal distributions for probabilistic forecasting, predictive densities do not tend to be systematically over- or under-dispersive due to their natural mean-variance relationship. This appears to be valid for both types of dynamics considered for the mean of predictive densities. Similarly to GL-Normal predictive densities, the Beta-AR densities have an acceptable level of reliability due to their natural shape variations as a function of their location parameter.

It was argued that the doubled-bounded nature of wind power generation causes the fact that the shape of predictive densities ought to vary depending upon the power level. This motivated the proposal of GL-Normal predictive densities. In parallel following Gneiting *et al.* [16] and other authors, the correspondence of predicted probabilities and observed frequencies should be verified for the evaluation set as a whole but potentially also for any subset. The calibration of predictive densities is hence further investigated as a function of the level of power, in a spirit similar to that of Pinson *et al.* [40]. Mutually exclusive subsets of the evaluation set are defined based on measurements, so that they are the same for all probabilistic forecasting approaches, following the proposal for calibration assessment based on forecast strata introduced and extensively discussed by Bröcker [8]. The criterion employed for quantifying the lack of calibration of predictive densities is the Mean Absolute Deviation from Calibration (MADC), defined as the integrated absolute difference between the reference case

TABLE 4: *Results for the evaluation of the calibration of predictive densities conditional to the level of measured power and overall, with the MADC criterion (in %).*

Level	[0,0.2]	[0.2,0.4]	[0.4,0.6]	[0.6,0.8]	[0.8,1]	Overall
Climatology	38.86	26.64	20.03	23.33	34.91	0
Persistence	7.85	6.92	5.99	5.78	5.99	5.82
Normal AR	19.38	10.71	2.99	2.70	11.61	7.91
Beta AR	21.71	10.56	3.28	2.26	16.75	2.41
GL-Normal AR	6.50	3.43	6.34	5.80	6.85	2.85
GL-Normal CP-AR	9.19	1.78	3.45	3.17	11.15	2.22

of perfectly calibrated forecast and the observed proportions of the quantiles of predictive densities. It then represents the area between the reliability curves and the diagonal as in the case of the reliability diagram of Figure 5. Lower MADC values indicate more reliable predictive densities. It is calculated in the case of the whole evaluation set and over subsets defined by various ranges of measured wind power, e.g. when normalized wind power generation is between 0 and 0.2. The corresponding results are summarised in Table 4, with the "overall" column giving the MADC values for the reliability assessment presented in Figure 5.

Climatology has perfect reliability when evaluated over the whole evaluation set since predictive densities are unconditional and directly derived from the collected measurements. This unconditional nature is clearly affecting its reliability when assessed over the various subsets defined by the level of observed wind power owing to the level-varying characteristics of the stochastic process. A similar result can be observed for the case of Normal-AR and Beta-AR predictive densities for low and high power values. For medium power values, the low MADC values indicate that both Normal and Beta assumptions, as well as the related variance estimation, are fairly acceptable. When looking across the various levels of measured power and at overall results, both persistence and GL-Normal predictive densities seem to be better calibrated. One remembers, however, that calibration is only one aspect of probabilistic forecasting skill [16]. Appropriate ranking of competing approaches can only be carried out based on the skill score values of Table 3. Based on all the results presented, GL-Normal predictive densities appear to be best, with a fair advantage to the CP-AR dynamics for their location and scale parameters. And even if persistence seems to have an acceptable calibration, this does not translate to these benchmark probabilistic forecasts having a competitive skill level (as seen from Table 3).

For illustrative purposes, Figure 6 depicts an episode with 375 successive time-steps of quantile forecasts (with nominal proportions 0.05 and 0.95) extracted from GL-Normal predictive densities with CP-AR dynamics, along with the corresponding measurements. One can easily observe that the predictive densities are naturally bounded between 0 and 1, and that the variance of GL-Normal densities is larger when the conditional expectation of the process is at a medium level. It is harder though to see that both dynamics and variance of the predictive densities are a function of the wind direction onsite. Let us have a look at some of the parameters of the CP-AR dynamic model, more specifically the coefficient for the first autoregressive lag, as well as the variance of the GL-Normal predictive densities, at the end of the evaluation set. These various quantities are represented in polar coordinates in Figure 7. Variations as a function of the wind direction ω can be observed. The variance of predictive densities appears to be significantly higher for Westerly and North-Westerly winds, and lower for the other sectors. This is in line with the extensive wind data analysis performed by Vincent *et al.* [50] and Vincent *et al.* [51] based on meteorological measurements obtained at Horns Rev. It was observed that wind fluctuations (and by extension wind power fluctuations) tend to have a higher variance for winds coming from the North Sea, thus from Westerly and North-Westerly directions, than for winds coming from land.

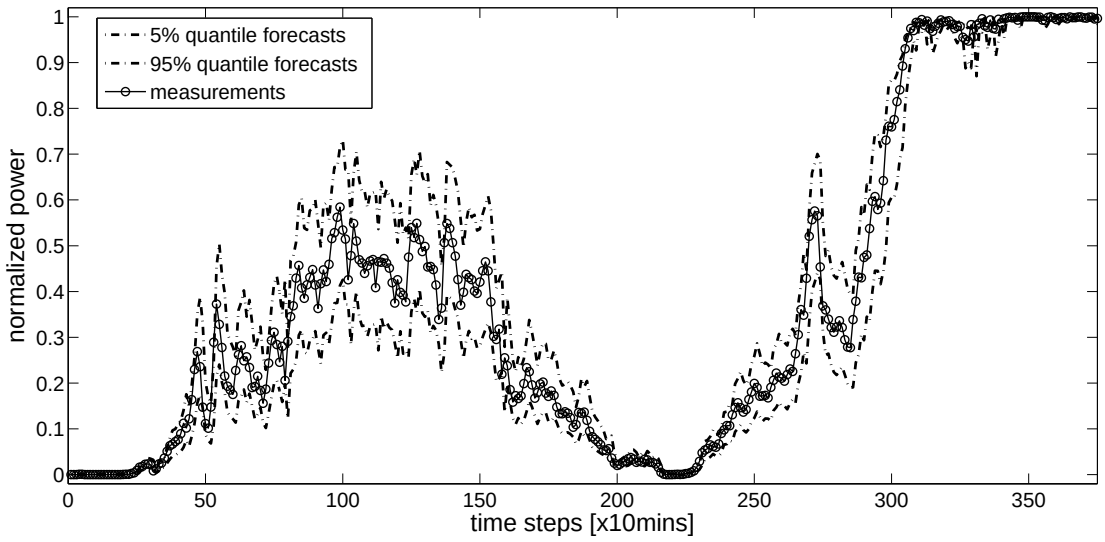
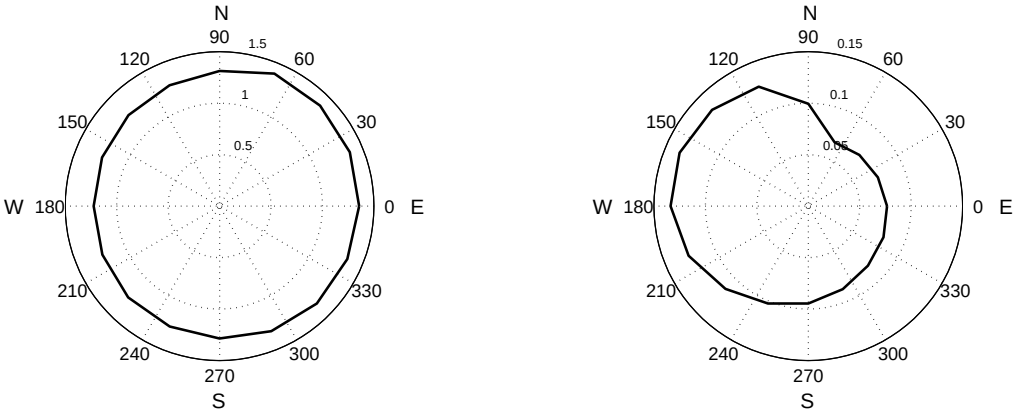


FIGURE 6: *Episode of 375 time-steps (with a time resolution of 10 minutes), with quantile forecasts of nominal proportions 0.05 and 0.95, along with related normalised wind power measurements. These quantile forecasts are extracted from the predictive densities with CP-AR dynamics.*



(a) $\theta_1(\omega)$ (b) $\sigma^2(\omega)$

FIGURE 7: *Some of the CP-AR model parameters as a function of the wind direction ω at the end of the evaluation set.*

6 Concluding remarks

A generalization of the logit-Normal distribution was introduced in order to properly account for the nonlinear and double-bounded nature of wind power generation in short-term probabilistic forecasting. It was observed that its characteristics make it more appropriate than the more commonly assumed Normal and Beta distributions. We believe that GL-Normal distributions could be viewed as a basic feature of parametric probabilistic forecasting methodologies for lead times between a few minutes and a few days ahead. Being fully characterised by a location and a scale parameter only (assuming that the shape parameter is a meta-parameter estimated once and for all from available data), fairly simple models can be employed for the description of its

dynamics.

In our proposal forecasting methodology, predictive densities of wind power take the form of discrete-continuous mixtures consisting of these GL-Normal distributions with potential concentration of probability mass at the bounds of the unit interval $[0, 1]$. These mixtures could be rethought in the future based on further data analysis and if looking at longer lead times. Since forecast uncertainty may not be null when predicting zero or nominal power, probability masses should be replaced by exponential distributions, the parameters of which would then also have to be estimated. For these more general mixtures, it will certainly be beneficial to develop estimation methods in a dynamic (Bayesian) model averaging framework, as in Raftery *et al.* [41].

The models employed for the location and scale parameters of the predictive densities account for the autoregressive nature of the wind power generation process, as well as the non-linear influence of wind direction. For further lead times and multi-step ahead forecasting some of the numerous models described in the literature could be employed for the dynamics of GL-Normal distribution parameters. They could for instance account for additional effects and covariates, like meteorological measurements which are known to be important for short lead times [13], or meteorological forecasts for lead times further than 6 hours ahead. More particularly for very short-term forecasts, the proposal of new dynamic models should be supported by some of the recent and extensive data analysis studies performed based on wind, wind power and other meteorological data. For the case of this offshore site, the results from Pinson and Madsen [38] suggest that regime-switching may be present in the time-series of wind power generation. This is confirmed by the results of Vincent *et al.* [51] explaining that these switches relate to local thermal effects, atmospheric stability and to the presence of convective cells in the vicinity or over this site. Regime-switching approaches are also relevant for onshore sites especially in complex terrain environments, as investigated by Gneiting *et al.* [17] when focusing on wind channelling effects.

GL-Normal distributions should also be considered for the modelling and forecasting of other nonlinear and bounded processes where it is expected or known that the effects of the bounds are not symmetrical, as for bounded economic variables [52]. For the wind power application, the shape parameter ν of GL-Normal distributions will surely vary from site to site. In a more general manner, this parameter will depend upon the process considered and should be estimated from data. The cross-validation approach employed here yielded acceptable results. It may however be envisaged to develop a more rigorous framework based on e.g. maximum likelihood estimation for selecting ν.

Acknowledgements

The work presented has been partly supported by the European Commission under the SafeWind project (ENK7-CT2008-213740), by the Danish Public Service Obligation (PSO) programme through the project 'Radar@Sea' (PSO-2009-1-0226), and finally by the Danish Research Council for Technology and Production Sciences (FTP-274-08-0573), which are hereby acknowledged. We are thankful to Vattenfall Denmark for originally providing the wind power measurements for the Horns Rev wind farm, Pierre-Julien Trombe for the data processing and quality checking, as well as Tilmann Gneiting, Niels K. Poulsen and Henrik Madsen for inspiring discussions on probabilistic forecasting and bounded time-series. We finally acknowledge two anonymous reviewers, an associate editor, as well as Richard Chandler, for useful comments and suggestions.

References

1. Ailliot, P., Monbet, V. and Prevosto M. (2006) An autoregressive model with time-varying coefficients for wind fields. *Environmetrics*, **17**, 107–117.
2. Aitchison, J. and Shen, S. (1980) Logistic-normal distributions: some properties and uses. *Biometrika*, **67**, 261–272.
3. Akhmatov, V. (2007) Influence of wind direction on intense power fluctuations in large offshore wind farms in the North Sea. *Wind. Eng.*, **31**, 59–64.
4. Bacher, P., Madsen, H. and Nielsen, H.Aa. (2009) Online short-term solar power forecasting. *Sol. Energy*, **83**, 1772–1783.
5. Bludszuweit, H., Domínguez-Navarro, J.A. and Llombart A. (2008) Statistical analysis of wind power forecast errors. *IEEE T. Power Syst.*, **23**, 983–991.
6. Box, G.E.P. and Cox, D.R. (1964) An analysis of transformations. *J. Royal Stat. Soc. B*, **26**, 211–243.
7. Bremnes, J.B. (2006) A comparison of a few statistical models for making quantile wind power forecasts. *Wind Energy*, **9**, 3–11.
8. Bröcker, J. (2008) On reliability analysis of multi-categorical forecasts. *Nonlin. Proc. Geophys.*, **15**, 661–673.
9. Cloke, H.L. and Pappenberger, F. (2009) Ensemble flood forecasting: a review. *J. Hydrol.*, **375**, 613–626.
10. Costa, A., Crespo, A., Navarro, J., Lizcano, G., Madsen, H. and Feitosa E (2008) A review on the young history of the wind power short-term prediction. *Renew. Sust. Energ. Rev.*, **12**, 1725–1744.
11. Fan, J.Q. and Zhang, W. (2008) Statistical methods with varying-coefficient models. *Stat. Interf.*, **1**, 179–195.
12. Frederic, P. and Lad, F. (2008) Two moments of the logit-normal distribution. *Comm. Stat. B-Simul.*, **37**, 1263–1269.
13. Gallego, C., Pinson, P., Madsen, H., Costa, A. and Cuerva, A. (2011). Influence of local wind speed and direction on wind power dynamics Application to offshore very short-term forecasting. *Appl. Energ.*, **88**, 4087–4096.
14. Giebel, G., Brownsword, R., Kariniotakis, G., Denhard, M. and Draxl C. (2011) The state of the art in short-term prediction of wind power - A literature overview, 2nd edition. Technical report, EU project Anemos.plus. [Available online: http://orbit.dtu.dk]
15. Gneiting, T. (2011) Quantiles as optimal point predictors. *Int. J. Forecasting*, **27**, 197–207.
16. Gneiting, T., Balabdaoui, F. and Raftery A.E. (2007) Probabilistic forecasts, calibration and sharpness. *J. Royal Stat. Soc. B*, **69**, 243–268.
17. Gneiting, T., Larson, K., Westrick, K., Genton, M.G. and Aldrich, E. (2006) Calibrated probabilistic forecasting at the stateline wind energy center - The regime-switching space-time method. *J. Am. Stat. Assoc.*, **101**, 968–979.
18. Hering, A.S. and Genton, M.G. (2009) Powering up with space-time wind forecasting. *J. Am. Stat. Assoc.*, **105**, 96–104.
19. Johnson, N. (1949) Systems of frequency curves generated by methods of translation. *Biometrika*, **36**, 146–176.

20. Kagan, Y.Y. and Jackson, D.D. (2000) Probabilistic forecasting of earthquakes. *Geophys. J. Int.*, **143**, 438–453.
21. Kristoffersen, J.R. and Christiansen, P. (2003) Horns Rev offshore wind farm: its main controller and remote control system. *Wind. Eng.*, **27**, 351–359.
22. Lange, M. (2005) On the uncertainty of wind power predictions - Analysis of the forecast accuracy and statistical distribution of errors. *J. Solar Energ.-T. ASME*, **127**, 177–184.
23. Lau, A. and McSharry, P. (2010). Approaches for multi-step density forecasts with application to aggregated wind power. *Ann. Appl. Stat.*, **4**, 1311–1341.
24. Lesaffre, E., Rizopoulos, D. and Tsonoka, R. (2007) The logistic-transform for bounded outcome scores. *Biostatistics*, **8**, 72–95.
25. Leutbecher, M. and Palmer, T.N. (2008) Ensemble forecasting. *J. Comput. Phys.*, **227**, 3515–3539.
26. Li, T.H. (2008) On exponentially weighted recursive least squares for estimating time-varying parameters. *J. Stat. Theo. Pract.*, **2**, 339–354.
27. Ljung, L. and Söderström, T. (1983) *Theory and Practice of Recursive Estimation*. MIT Press: Boston.
28. Madsen, H. (2007) *Time Series Analysis*. Chapman & Hall/CRC: London.
29. Madsen, H., Pinson, P., Nielsen, T.S., Nielsen, H.Aa. and Kariniotakis, G. (2005) Standardizing the performance evaluation of short-term wind power prediction models. *Wind Eng.*, **29**, 475–489.
30. Mead, R. (1965) A generalized logit-normal distribution. *Biometrics*, **21**, 721–731.
31. Møller, J.K., Nielsen, H.Aa. and Madsen, H. (2008) Time-adaptive quantile regression. *Comput. Stat. Data An.*, **52**, 1292–1303.
32. Nielsen, H.Aa., Nielsen, T.S., Joensen, A.K. and Madsen, H. and Holst, J. (2000) Tracking time-varying coefficient functions. *Int. J. of Adapt. Control*, **14**, 813–828.
33. Nielsen, H.Aa., Nielsen, T.S., Madsen, H., Badger, J., Giebel, G., Landberg, L., Sattler, K., Voulund, L. and Tøfting, J. (2006) From wind ensembles to probabilistic information about future wind power production - Results from an actual application. in *Proc. IEEE PMAPS Conference, Probabilistic Methods Applied to Power Systems*, Stockholm, Sweden.
34. Pinson, P. (2006) *Estimation of the uncertainty in wind power forecasting*. Ph.D. thesis, Ecole des Mines de Paris, Paris, France. [Available online: http://pastel.paristech.org]
35. Pinson, P., Chevallier, C. and Kariniotakis, G. (2007) Trading wind generation from short-term probabilistic forecasts of wind power. *IEEE T. Power Syst.*, **22**, 1148–1156.
36. Pinson, P. and Kariniotakis, G. (2010) Conditional prediction intervals of wind power. *IEEE T. Power Syst.*, **25**, 1845–1856.
37. Pinson, P. and Madsen, H. (2009) Ensemble-based probabilistic forecasting at Horns Rev. *Wind Energy*, **12**, 137–55.
38. Pinson, P. and Madsen, H. (2011) Adaptive modeling and forecasting of wind power fluctuations with Markov-switching autoregressive models. *J. Forecasting*, available online.
39. Pinson, P., Nielsen, H.Aa., Madsen, H. and Nielsen, T.S. (2008) Local linear regression with adaptive orthogonal fitting for the wind power application. *Stat. Comput.*, **18**, 59–71.
40. Pinson, P., Nielsen, H.Aa., Møller, J.K., Madsen, H. and Kariniotakis, G. (2007) Nonparametric

probabilistic forecasts of wind power: required properties and evaluation. *Wind Energy*, **10**, 497–516.

41. Raftery, A.E., Karny, M. and Ettler, P. (2010). Online prediction under model uncertainty via dynamic model averaging: Application to a cold rolling mill. *Technometrics*, **52**, 52–66.
42. Sanchez, I. (2006) Short-term prediction of wind energy production. *Int. J. Forecasting*, **22**, 43–56.
43. Stone, M. (1974) Cross-validation and assessment of statistical predictions (with discussion). *J. Royal Stat. Soc. B*, **36**, 111–147.
44. Sørensen, P.E., Cutululis, N.A., Vigueras-Rodriguez, A., Jensen, L.E., Hjerrild, J., Donovan, M.H. and Madsen, H. (2007) Power fluctuations from large wind farms. *IEEE T. Power Syst.*, **22**, 958–965.
45. Sørensen, P. and co-authors (2005) Operation and control of large wind turbines and wind farms - Final report. Technical report, Risø National Laboratories for Sustainable Energy, Technical University of Denmark, Risø-R-1532(EN).
46. Tay, A.S. and Wallis, K.F. (2000) Density forecasting: a survey. *J. Forecasting*, **19**, 235–254.
47. Taylor, J.W., McSharry, P.E. and Buizza, R. (2009) Wind power density forecasting using wind ensemble predictions and time series models. *IEEE T. Energy Conver.*, **24**, 775–782.
48. Timmermann A. (2000) Density forecasting in economics and finance. *J. Forecasting*, **19**, 231–234.
49. Usaola, J. (2009) Probabilistic load flow with wind production uncertainty using cumulants and Cornish–Fisher expansion. *Int. J. Elec. Power. Syst.*, **31**, 474–481.
50. Vincent, C.L., Giebel, G., Pinson, P. and Madsen, H. (2010) Resolving non-stationary spectral signals in wind speed time-series using the Hilbert-Huang transform. *J. App. Meteorol. Clim.*, **49**, 253–267.
51. Vincent, C.L., Pinson, P. and Giebel, G. (2011) Wind fluctuations in the North Sea. *Int. J. Climatol.*, available online.
52. Wallis, K.F. (1987) Time series analysis and bounded economic variables. *J. Time Ser. Anal.*, **8**, 115–123.

Probabilistic wind power forecasting using radial basis function neural networks

G. Sideratos and N. D. Hatziargyriou, *Fellow, Member, IEEE*

***Abstract*-- A novel methodology for probabilistic wind power forecasting is described. The method is based on artificial intelligence and concentrates on the uncertainty information about the future wind power production predicting a set of quantiles with predefined nominal probabilities. The proposed model uses the point predictions of an existing state-of-art wind power forecasting model and forecasts the prediction uncertainties due to the inaccuracies of the numerical weather predictions (NWP), the weather stability and the deterministic forecasting model. The performance of the proposed model is evaluated on two wind farms that are located in areas with different weather conditions.**

***Index Terms*—Probabilistic wind power forecasting, radial basis function neural network, self-organized map**

I. Introduction

In recent years, significant research efforts have been devoted in wind power forecasting and a number of techniques and tools have been developed. A good overview of these tools is presented in [1-3], illustrating the undisputable progress in the area, but also the limited predictability of wind power in several case studies. Most of the wind power forecasting tools provide wind power forecasts without much information about their uncertainty. Currently, wind power forecasting research focuses on the enhancement of wind power point predictions with their uncertainty expressed by risk indices or by predicted wind power probability densities.

Many studies have shown the benefits obtained by the use of forecasting uncertainty information. In [4], the optimal level of generation reserves is estimated using the uncertainty of wind power predictions. In [5, 6] the optimal coordination of hydro-generation and wind energy production is investigated taking into account the forecasts of a probabilistic prediction model. The increased revenues obtained by basing bidding strategies on predictive wind power densities in day-ahead electricity markets are presented in [7, 8]. Finally, wind power forecasting uncertainty is required for analytical studies based on probabilistic load flow [9], e.g. congestion lines management.

Probabilistic wind power prediction models use meteorological ensembles that are obtained by a high-resolution meteorological model [10, 11] or on traditional wind power and NWP (Numerical Weather Predictions) timeseries. The latter ones apply a statistical method to estimate the predictive distributions in the form of quantiles or intervals. For example, in [12] local linear quantile regression is applied to estimate ten different quantiles. A similar method is used in [13] combined with spline bases to estimate two quantiles of the forecast error of the WPPT [14] forecasting model. In [15], a method to provide the continuous predictive probability density function of wind power is proposed based on kernel density estimators. A similar method is presented in [16] that uses time-adaptive kernels. In [17] prediction intervals are estimated by adaptive resampling using the uncertainty information of an appropriate risk index obtained from on consecutive weather forecasts. Finally, the relations between typical weather situations and forecasting errors are investigated in [18]. An excellent overview of probabilistic wind power forecasting methods is provided in [19].

Some probabilistic models [13, 18] use the point predictions provided by a deterministic wind power prediction model and retrieve the uncertainty information from the NWP timeseries. This paper presents a novel probabilistic wind power forecasting method that uses the point predictions and uncertainty information of a state of the art deterministic model [20], called dWPF. Uncertainty information is obtained from the quality of the NWP, the structure of the dWPF and from the weather stability. The weather stability is defined by the standard deviations of wind speed and wind direction in the NWP grid. The quality of NWP is estimated by training two radial basis function neural networks (RBFNN) with different input variables in order to predict the wind power obtained by the forecasted wind speed and the wind farm manufacturer's power curve. The quality of NWP is estimated comparing the RBFNNs outputs with the dWPF prediction [21]. Finally, the information retrieved from the winning neuron of the dWPF's RBFNNs is used. A similar approach could be conceptually used to exploit point predictions and uncertainty information of other deterministic approaches based on Gaussian functions.

The proposed probabilistic model is designed to predict eight quantiles of the wind power distribution using multiple RBFNNs. An RBFNN is activated depending on how the

The work presented in this paper has been performed within the framework of the EU research project GA no 213740 "SafeWind". The authors gratefully acknowledge the support received from the EU for their research.

G. Sideratos and N. Hatziargyriou are with the National Technical University of Athens, Athens, GR 15773 Greece (phone: 30-210-7723661; fax:30-210-7723968; e-mail: joesider@power.ece.ntua.gr nh@power.ece.ntua.gr)

testing sample fits to the knowledge contained in the dWPF model. The developed RBFNNs are trained with the ordinary orthogonal least square algorithm (OLS) [22] and their performance is improved using a particle swarm optimization (PSO) algorithm [23] to determine the widths of the radial base neurons and the parameters of the input vector [20]. Results obtained from the application of the proposed model on two wind farms, show the superiority of the proposed method over other existing methods.

II. Uncertainty information of wind power predictions

The main idea developed in this paper is to exploit all information about the parameters that are responsible for the point forecast error of the deterministic model in order to estimate the wind power density function. The uncertainty information is obtained from three different sources: the weather stability, the structure of the dWPF model and the NWP accuracy.

A. The weather stability

The weather instability over the area of a wind farm is responsible for the fast and sometimes large variations in wind power production. In order to quantify the extent of the weather stability, the proposed model uses part of the NWP grid that is provided by the meteorological services and covers a large area around the wind farm. Taking a specific case study in Crete as an example, the applied NWP grid covers the area from 21 to 27 longitudes and from 32 to 36 latitudes, while the wind farm of interest is located at 26.2 longitudes and 35.16 latitudes. The area covered is defined by the most correlated points of the NWP grid with the wind power timeseries. The standard deviations of the values of the wind speed and of the wind direction sd_s and sd_w, respectively that correspond to the points of the defined NWP grid are used for the quantification of the weather stability. It should be mentioned that the wind direction values are transformed to polar coordinates.

Another parameter that affects the quality of the point predictions is the vertical profile of the wind speed. The stability of the wind speed vertical profile is measured by the standard deviation of the wind speed values *sd* that correspond to six different horizontal levels of the NWP model to the nearest point to the wind farm.

B. The uncertainty information obtained from the structure of the dWPF model

As described in [20], the dWPF model is a combination of a self-organized map (SOM) and three RBFNNs, as shown in Figure 1.

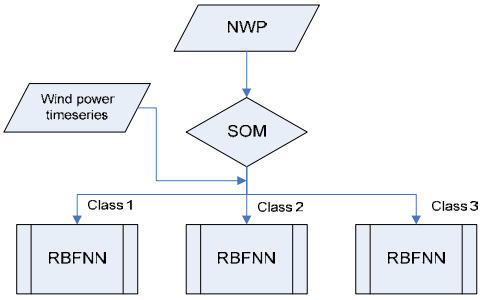


Figure 1. The structure of the dWPF model

Initially, the SOM divides the set of the weather information obtained by NWPs in three classes and activates the corresponding RBFNN that provides the wind power point prediction. Gaussian basis functions are used at the hidden layer and their centers $c_{i,j}$ are estimated by the ordinary OLS algorithm. The number *j* and the width *b* of these functions are defined applying several regularization criteria, as the generalized cross-validation criterion [24], Bayesian interpolation [25] and the Stein unbiased risk estimator [24], while the variables x_i are selected by trial and error.

When an input vector x_i is introduced, the output o_j of the RBFNN hidden layer has the form:

$$o_j = \exp\left(-\left\|\left(c_{i,j} - x_i\right)\right\| \Big/ b^2\right) \qquad (1)$$

In eq. (1), the input vector is compared with each hidden neuron center using the Euclidean norm. The distance of the input vector from every hidden neuron characterizes the degree that the input vector is known to the RBFNN. In this way, the hidden layer output o_j depicts the knowledge that is contained in a RBFNN related to the input vector x_i and this is summarized by the linear function of the second layer to the wind power prediction. Important information about the uncertainty of the dWPF model prediction is contained in the outputs o_j and forms the basis of the proposed probabilistic model.

The neuron that has the maximum contribution to the RBFNN's output is called the *winning neuron n* defined as:

$$n = \arg\max\left\{o_j\right\} \qquad (2)$$

The element of the input vector with the largest distance from the corresponding element of the winning neuron in the Euclidean input space is called the *worst input variable v* with respect to the winning neuron.

$$v = \arg\max\left\{\left\|c_{i,j} - x_i\right\|\right\} \qquad (3)$$

This variable has the largest contribution to the wind power prediction error.

C. The uncertainty information related to the NWP

NWP is the main input to a wind power forecasting model. They are provided by a meteorological weather prediction model that simulates the atmospheric processes over a large area. The stochastic movement of the fronts leads these

models to large errors, namely phase and spatial errors. Phase and spatial errors are due to poor prediction of the front's velocity or the front's track, respectively. In [21] a method to qualify the NWP that are used by the dWPF model is described. In order to compare the wind speed forecasts with the actual values, the wind speed values are transformed to power values through the wind farm manufacturer's power curve. These wind power values are named *theoretical* predictions. The contribution of the NWP to the wind power forecasting error is indicated by the correlation between the errors of the theoretical predictions and those of the dWPF predictions.

For the estimation of the error of the theoretical predictions, two RBFNNs are applied. The input variables of the dWPF model are divided into two groups to feed the two RBFNNs. The two RBFNNs are trained in order to provide the theoretical predictions. The quality of the NWP can be estimated by the differences d_w and d_e between the dWPF prediction and the outputs of these two RBFNNs, as described in [21]. The term d_w denotes the difference between the dWPF prediction and the output of the RBFNN with inputs the wind speed variables and d_e the difference between the dWPF prediction and the output of the RBFNN with inputs the explanatory variables. The correlations of d_w and d_e with the theoretical prediction error outreach 50%. Further details about the method can be found in [21].

III. Description of the Probabilistic Wind Power Forecasting Model

The proposed model runs every hour following the dWPF model execution and provides eight quantiles of the probability density function of wind power for every forecasting time step. The quantiles depict in a different manner the form of the non-parametric probability distribution of the wind power.

Generally, a quantile q_t with nominal proportion $\alpha \in [0,1]$ of the wind power p_t is defined as:

$$\mathrm{P}(p_t < q_t^a) = \alpha \quad \text{or} \quad q_t^a = F_t^{-1}(a) \qquad (4)$$

P is the probability of wind power p_t to be lower than the quantile q_t and F_t is the cumulative density function. The eight quantiles applied, have nominal proportions 0.05, 0.1, 0.25, 0.4, 0.6, 0.75, 0.9 and 0.95 and are selected in order to provide information for the whole wind power distribution. The use of these quantiles in suitable pairs provides the intervals with nominal coverage of 20%, 50%, 80% and 90%. . In this application, a one year data set was used for the formulation of the observed wind power distributions and consequently for the estimation of the observed quantiles and the observed intervals. In order to produce sharper and more realistic intervals, the wind speed and the wind direction timeseries are used conditionally for the formulation of the wind power densities. The observed quantiles in a time period are applied to fit the proposed model, while the performance of the model is evaluated in a different time period. The aim of a well-calibrated probabilistic model is to provide quantile forecasts that have the same cumulative probability with their predefined nominal proportions in the testing period [26].

The wind power quantiles are based on the uncertainty information gathered from the parameters that contribute to the wind power prediction error. In section II, three possible sources, namely the weather stability, the deterministic model's structure, and the NWP quality were identified. Amongst them, the knowledge that is contained in the deterministic wind power prediction model dWPF, provides the basic information about the forecasting uncertainty. The knowledge of the dWPF model is obtained at the neurons of the RBFNNs' hidden layer. The output of the hidden layer o_j of any RBFNN contained in the dWPF model reflects the way a testing sample is known by the RBFNN. In fact, the output o_j is a vector with elements, corresponding each to a neuron and defines its similarity with the input testing case.

In order to quantify the knowledge of an RBFNN about a future case, the output of its hidden layer is classified in multiple classes. These classes are named as the *levels of the RBFNN's knowledge*. The levels of knowledge in the proposed probabilistic model are defined by a self-organized map. In this application, twenty four levels of knowledge are applied for each RBFNN of the dWPF model. This means that the probabilistic model contains three self-organized maps one for each weather class of the dWPF model. The self-organized maps are trained by the hidden layer outputs of the RBFNNs that are obtained by the one-year evaluation of the dWPF model. A self-organized map creates vectors that are statistically the most significant ones inside a training data set [27]. These vectors consist of the proposed levels. The number of these levels is estimated by trial and error in order to capture thoroughly all the outputs of the hidden layer.

Indicative levels of knowledge that correspond to the RBFNN of the first weather class of the dWPF model are illustrated in fig.2.

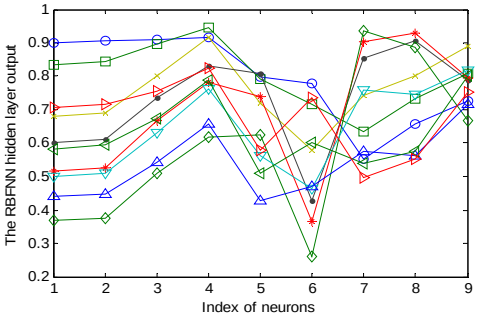


Figure 2. The levels of knowledge of the RBFNN's hidden layer output selected by the self organized map

Depending on the levels of a RBFNN's knowledge, the data set is divided in multiple subsets. For probabilistic wind power forecasting, the data set contains together with the wind power prediction of the dWPF model and the NWP timeseries, the parameters of the weather stability, the uncertainty related to NWP and the uncertainty related to the dWPF model, as described in section II. In detail, these parameters are:

- the standard deviations of the wind speed and the wind direction sd_s and sd_w, respectively, that

correspond to the points of a predefined part of the NWP grid.

- the standard deviation of the wind speed values *sd* that correspond to six different horizontal levels of the NWP model at the point near the wind farm
- the winning neuron *n* and the worst variable *v* of the RBFNNs obtained by the one year dWPF model's evaluation
- the measures d_w and d_e obtained from the NWP quality estimator model [21], as described in section II.

Each subset Ω at time *t* that corresponds to a different level *k* of the knowledge contained in the RBFNN of the weather class *i*, has the form:

$$\Omega_{i,k} = \{\hat{p}^{t+h}, ws^{t+h-1}, ws^{t+h}, ws^{t+h+1}, \ldots \cos(wd^{t+h}/2\pi), \quad \sin(wd^{t+h}/2\pi), \quad \ldots \quad (5) \quad n, \quad v, \quad sd_w, \quad sd_d, \quad sd, \quad h \}$$

$\hat{p}^{t+h}$ is the dWPF's prediction, ws and wd are the wind speed and the wind direction provided by the NWP and *h* is the forecasting time step.

Each subset $\Omega_{i,k}$ that was formed by the self-organized map of the class *i*, is applied for the training of a different RBFNN aiming to predict the eight observed quantiles. This means that one RBFNN is applied for each of the defined levels of knowledge and every RBFNN provides eight output values. The applied RBFNNs are trained by the ordinary OLS algorithm and their performance is optimized by the PSO algorithm described in the following section.

The proposed model is summarized as follows. In the dWPF model, a self-organized map recognizes three different classes related to the weather conditions and activates one of the three RBFNNs that are contained in the model to produce deterministic wind power predictions. During this process, the hidden layer output of the activated RBFNN is used to estimate the uncertainty of the wind power prediction, as described in section II. Depending on the class of the weather condition, as defined by the dWPF model, a separate self-organized map receives at its input the hidden layer output of the activated RBFNN and estimates the level of the knowledge of the dWPF model RBFNN. The RBFNN of the proposed model that corresponds to the estimated level of knowledge is activated and provides the eight quantiles. The structure of the probabilistic wind power forecasting model that corresponds to one class is shown in Fig. 3. The model of the other two classes of weather has a similar structure.

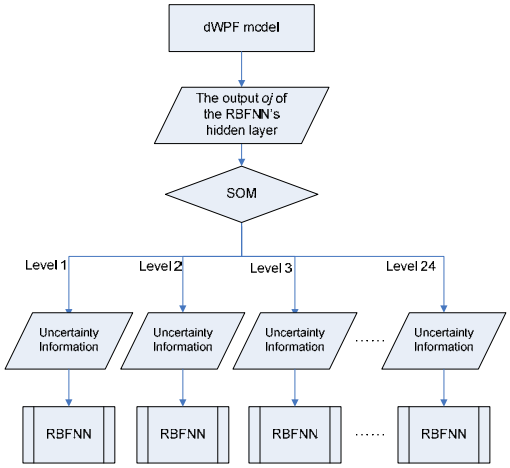


Figure 3. The structure of the proposed probabilistic model

IV. Optimization of the RBFNNs' performance

In most neural network applications [28], it is effective to preprocess the input data, by normalizing the range of all input variables in the same interval like [-1, 1] or [0, 1]. This is the case when all input variables should contribute equally to the RBFNN output. In wind power forecasting however, the wind speed contains the main information, while some input variables are used as explanatory and are less correlated to wind power. In order to improve the performance of the applied RBFNNs, an additional weighted vector $\overline{w}_i$ is applied at the input data that defines the importance of each input variable to the prediction. Thus, eq. (1) is modified as:

$$o_j = \exp\left(-\left\|\overline{w}_i\left(c_{ij} - x_i\right)\right\|/b\right) \qquad (6)$$

The optimal weighted vector $\overline{w}_i$ and the width *b* are estimated by the Particle Swarm Optimization (PSO) algorithm [23]. In order to achieve the best possible fitting of the applied RBFNNs, the cost function f_c that was chosen for the PSO algorithm, contains the values obtained by the Bayesian interpolation *BI* [25], the Stein unbiased risk estimator *SURE* [24] and the predictive densities' bias *BIAS* [26].

$$f_c = a_1 BI + a_2 SURE + a_3 |BIAS| \qquad (9)$$

These criteria are calculated after the number of the RBFNN's neurons has been defined and can ensure the generalization ability of the neural networks. The a_1, a_2 and a_3 coefficients are chosen in order to normalize the weight of each criterion.

The generalized cross-validation criterion [24] is used for the selection of the optimal number of the hidden layer's neurons.

V. Evaluation of the proposed model

A. The evaluation framework

The evaluation of deterministic models is based on the deviation between the predicted and the actual value of the wind power. The validation of a probabilistic model is more complicated and should consider the whole testing data set.

The verification of probabilistic forecasts is determined by the reliability and the sharpness of the predictive densities.

Reliability is the ability of the model to provide quantile forecasts $\hat{q}^{a}_{t/t+h}$ that satisfy eq. (4) on the whole data set, namely the predicted wind power values $\hat{q}^{a}_{t/t+h}$ should have, as much as possible, similar probabilities with the respective nominal ones α. To compute the reliability of a model, first, the *indicator variable* should be calculated for every look-ahead time, separately. The indicator variable $I^{(a)}_{t/t+h}$ of the quantile prediction $\hat{q}^{a}_{t/t+h}$ at the forecasting step h with nominal probability α is defined as the *hit* or the *miss* of eq. (4). This means:

$$I^{(a)}_{t/t+h} = \begin{cases} 1, & \text{if } p_{t+h} < \hat{q}^{a}_{t/t+h} \\ 0, & otherwise \end{cases} \qquad (10)$$

p_{t+h} is the wind power measurement at time $t+h$. From the reliability definition, the proportion $\hat{a}^{a}_{h}$ of the hits in the evaluation set should be as close as possible to the nominal probability of the quantile forecasts. Then, the reliability of a model can be defined as the deviation of the empirical from the nominal probabilities or can be illustrated by diagrams that show the nominal probabilities versus the empirical ones. In this paper, the reliability is presented as the bias of the probabilities [30] or the deviation of the nominal proportions.

$$b^{a}_{h} = a - \hat{a}^{a}_{h} \qquad (11)$$

The sharpness of the predictive densities is defined as the average of the distance that covers the prediction intervals. The sharpness expresses the ability of the model to concentrate the uncertainty information of the wind power prediction. In the meteorological literature, the sharpness is referred to as the model's deviation from the climatological mean probabilities [31]. The sharpness $\overline{\delta}^{(1-2a)}_{h}$ is evaluated on the prediction intervals, namely on the symmetrical pairs of the quantiles, without considering the wind power observations. It has the following form:

$$\overline{\delta}^{(1-2a)}_{h} = \frac{1}{N}\sum_{t=1}^{N}(\hat{q}^{(1-a)}_{t/t+h} - \hat{q}^{(a)}_{t/t+h}) \quad \forall \alpha < 0.5 \quad (12)$$

The evaluation of the proposed model is based on the Continuous Ranked Probability Score $CRPS_h$ [32]. This score combines the reliability and the sharpness to one criterion and evaluates the whole predictive density as follows:

$$CRPS_h = \frac{1}{N}\sum_{t=1}^{N}\int_{-\infty}^{+\infty}(\hat{F}_{t/t+h}(p) - I(p \geq p_{t+h})^2 dp \qquad (13)$$

$\hat{F}_{t/t+h}(p)$ is the cumulative predictive distribution and N is the number of the testing samples. Since the number of the forecasted quantiles is limited to eight, the cumulative predictive distribution is a discrete function and the wind power values between two consecutive quantiles $\hat{q}^{a_1}_{t/t+h}$ and $\hat{q}^{a_2}_{t/t+h}$ are considered to have the same probability. The integral of eq. (13) is computed for the range of the nominal capacity of the wind farm with resolution 0.5% of the nominal capacity. The $CRPS_h$ is equivalent to the mean absolute error (MAE) criterion of the deterministic evaluation [31] and it is expressed as percentage of the wind farm nominal power P_n.

Applying the above evaluation criteria, the proposed model is compared with three naïve models that are used for benchmarking in probabilistic forecasting applications [26]. The persistence method that is used in the deterministic wind power prediction [32] corresponds to the persistence distribution. The persistence distribution is formed by the observations of the most recent hours. For this wind power forecasting application, the last 12 hours observations are used to form the persistence distribution. The second reference method is the unconditional distribution that is based on all available wind power values. This method does not take into account any knowledge about the past when it provides a probabilistic prediction. Finally, the uniform distribution of wind power is used. The uniform distribution assumes any wind power value probable to occur with the same probability at any time step.

B. The case studies

In this paper, results from the application of the proposed model to two wind farms are presented; the Lasithi wind farm and the Klim wind farm. The Lasithi wind farm is located at the eastern part of Crete in Greece with coordinates 26.2 longitudes and 35.16 latitudes. The wind farm consists of 30 wind turbines BONUS 600 MK4 and its nominal capacity is 18MW. The wind farm area is semi-complex with low predictability, since the prevailing winds are influenced by a combination of sea breeze and valley-hill breeze. The wind power timeseries cover the period from 1st January of 2007 to 31st August 2008. The timeseries of the first eight months of the year 2008 are used for the evaluation of the proposed model performance, while the data set that corresponds to the year of 2007 is used for model training. The training and the testing sets of the dWPF model use the timeseries of 2006.

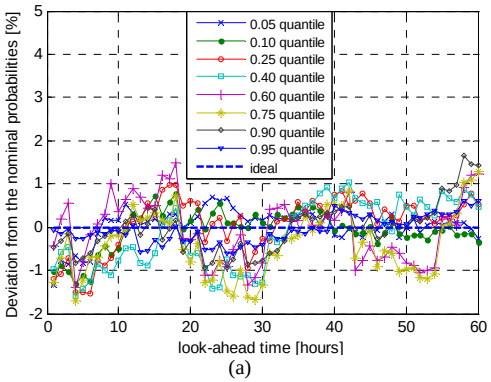
The NWP provided by a high-resolution meteorological model are in the form of multiple horizontal grids with 15Km resolution that cover an extended area, like the Mediterranean area. Every horizontal grid corresponds to a different height over the land surface. Usually the meteorological services provide the NWP that correspond to the nearest grid point to the wind farm. In this application, six NWP grids that correspond to different heights and cover a large area around the wind farm are used. The predictions of the meteorological variables are calculated for every point of the horizontal grids. These predictions are updated two times per day and their forecast horizon is three days. The time resolution of the NWP is three hours. This means there are 24 values for each meteorological variable and for each point of the horizontal NWP grids per update. On the other hand, the dWPF model and the proposed probabilistic model run every hour, namely

when a new wind power measurement is available, with hourly time resolution. Due to the difference of the runs frequency between the wind power prediction models and the meteorological model, the forecasting horizon of the proposed model is limited to 60 hours.

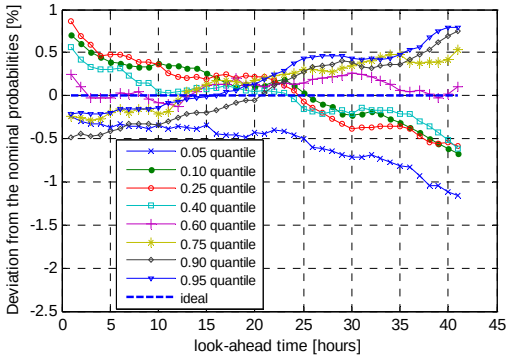
The Klim wind farm is located at the north-western part of Jutland, 8 km from the north coast and 50 km west of the city Aalborg. The farm contains 35 wind turbines of 600 kW and its total rated capacity is 21 MW [1]. The terrain of the wind farm is considered flat. The power production timeseries cover the period from 1st January 2000 to 31st December 2002. The data of the period of the first year were used to derive the dWPF model, the proposed model was trained with the timeseries of the second year, and the evaluation data set covers the period of 2002. The NWP data comes from the Danish HIRLAM model and is limited to the point close to wind farm. Also, in this case-study, the NWP are provided for 48 hours ahead with a time resolution of one hour every six hours and therefore, the forecasting horizon of the proposed model is limited to 41 hours. The weather predictions are provided for three different horizontal levels. In this case study, the weather stability is expressed only with the variable *sd*.

C. Results

As discussed in paragraph A, the most essential feature of a probabilistic forecasting model is the reliability of the quantile forecasts. First, the indicator variable is estimated for each quantile forecast, for each nominal proportion and for each forecasting time step, separately. Then, the proportion of the hits on the entire testing set is calculated and is compared with the nominal proportion of the respective quantile. Finally, the average bias of eq. (11) for every nominal proportion is computed.



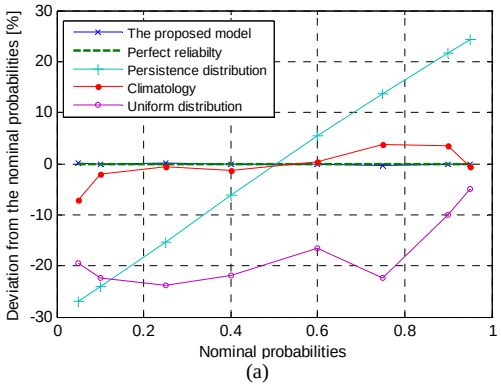
(a)



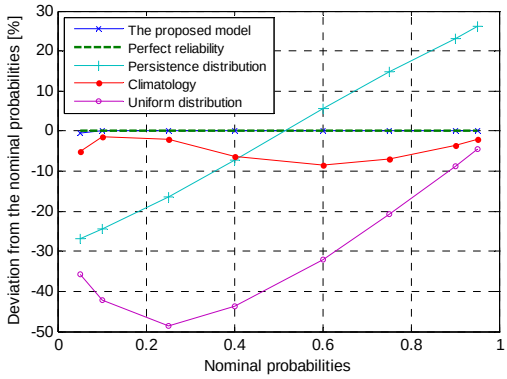
(b)

Figure 4. Reliability diagrams of the proposed model as a function of the forecasting horizon: (a) for Lasithi case study and (b) for Klim case study

Figures 4(a) and 4(b) show the deviation of the reliability of the proposed model from the perfect reliability, as a function of the look-ahead time for the two case studies. The perfect reliability corresponds to the ideal case when the observed proportions of the quantile forecasts inside the testing set are the same with the nominal probabilities. It is clear that for both case studies, the reliability of the proposed model is very close to the perfect reliability. The comparison of the proposed model with the reference methods is illustrated in figures 5(a) and 5(b), where the reliability is expressed as a function of the nominal probabilities. It is shown that the deviation from the perfect reliability is less than 1% for every nominal proportion. In addition, the proposed model outperforms all benchmarks for both case studies.



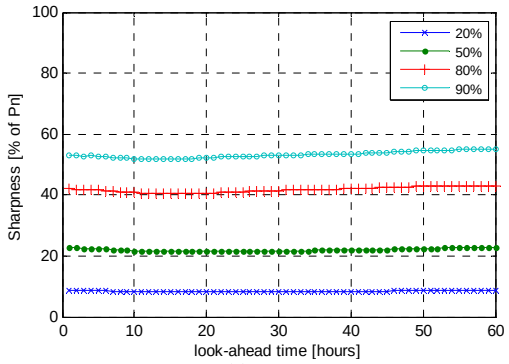
(a)



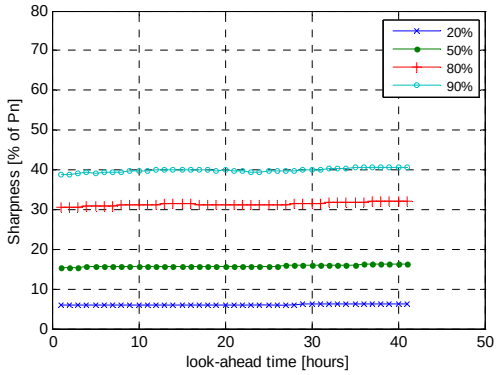
(b)

Figure 5. Reliability diagrams of the proposed model as a function of the nominal proportions: (a) for Lasithi case study and (b) for Klim case study

Figures 4(a) and 4(b) show that the deviation of the nominal probabilities does not outreach ±2% for the case of Lasithi wind farm and ±1.5% for the Klim wind farm for every quantile and every look-ahead time. This means that, if the nominal proportion of a quantile is 40% and the deviation is equal to -1%, then the observed proportion of this quantile in the testing set would be equal to 41%. The high reliability of the proposed model is also proved in figures 5 (a) and (b), where its performance is compared with the three reference methods. For the Lasithi case study, the maximum deviation of the model reliability is equal to 0.4% and appears for the quantile forecasts with nominal proportion 75%, while for the Klim case study, it is less than 0.55% at the 5% quantile forecasts.



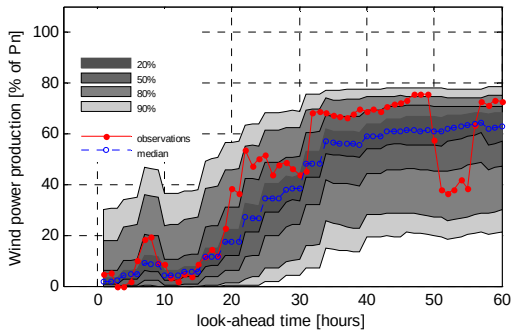
(a)



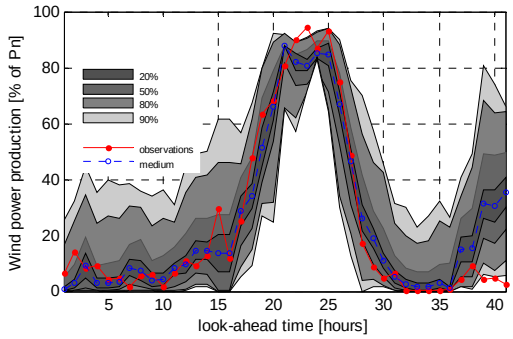
(b)

Figure 6. The sharpness of the proposed model: (a) for the Lasithi case study and (b) for the Klim case study.

Figures 6(a) and 6(b) show the sharpness of the forecasting intervals as a function of the forecasting horizon. In these studies, the sharpness of each interval is quite constant, since the observed quantiles that are used to train the proposed model have been formed based on the timeseries of the wind speed and of the wind direction, instead of the wind power point prediction error timeseries. Thus, the knowledge about the future wind power production is concentrated in a small percentage of the rated capacity of each wind farm, especially for time horizons beyond six hours ahead. For example, the sharpness of the 20% interval does not exceed 10% of the nominal capacity for every look-ahead time for both case studies. The sharpness of the 90% interval is approximately equal to 40% for the Klim wind farm. For the Lasithi wind farm, the sharpness is higher because the prediction intervals tend to cover the high fluctuations that occur due to the rejection of wind power [33], which is an unknown parameter. An example of this situation is presented in fig. 7(a). The good-shaped prediction densities for the Klim case study are shown in the example of figure 7(b).



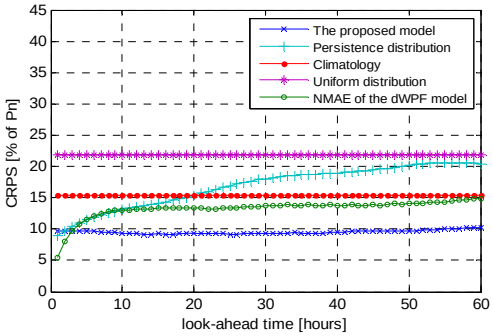
(a)



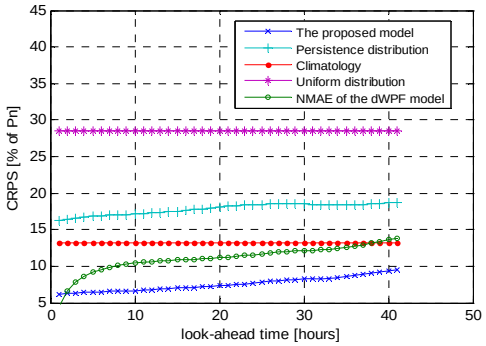
(b)

Figure 7. Two examples from the evaluation of the proposed model: (a) for the Lasithi case study, (b) for the Klim case study.

The good performance of the proposed probabilistic model is also proven by applying the *CRPS* criterion. Figures 8(a) and (b) show the *CRPS* of the model together with the normalized mean absolute error of the dWPF model for each look-ahead time. For the Lasithi case study, the proposed model performs better than the persistence method for every look-ahead hour, except the next first hour. The improvement of the model with respect to persistence reaches 52% at the 51st time step, while the *CRPS* of the model is lower than 10% at the first 53 hours-ahead. At the Klim wind farm case, the *CRPS* of the proposed model is always lower than the CRPS of the reference methods and less that 10%.



(a)



(b)

Figure 8. The *CRPS* as a function of the look ahead time of the proposed model and of the benchmarks and the NMAE of the dWPF model: (a) for the Lasithi wind farm (b) for the Klim wind farm.

Since wind power production is dependent on several factors, it is necessary to evaluate the model conditional to some of these factors. Figure 9 shows the average deviation from the perfect reliability and figure 10 shows the CRPS of the model conditional to the level of wind power for the Lasithi case study. In particular, the range of the wind power is divided to five equal intervals of 3.6 MW and the CRPS is computed for each interval, separately. The reliability for the first two levels is very close to the ideal reliability for most prediction steps. For higher wind power levels, the proposed model provides very reliable densities for horizons shorter than 30 hours ahead, while for longer horizons its performance is lower. Similar conclusions are obtained from figure 10. The CRPS is low at levels 1, 2 and 3, about 13% at level 4 and always higher than 16% at level 5. This is due to the high prediction uncertainty of these cases.

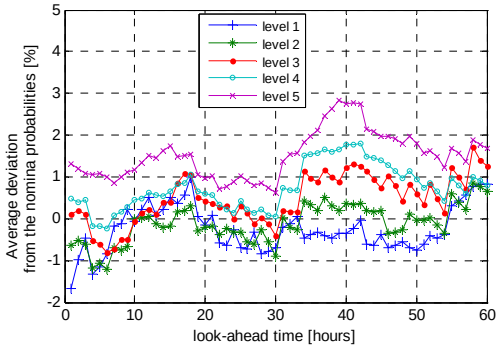


Figure 9. The average deviation of the observed probabilities from the nominal probabilities conditional to the level of wind power for the Lasithi case study.

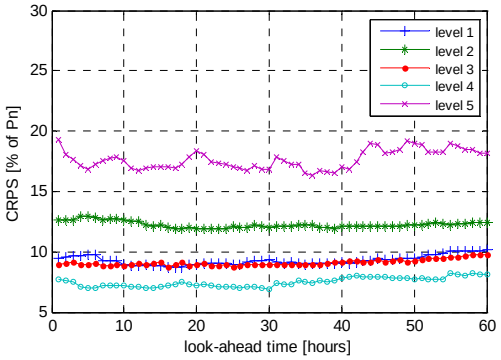


Figure 10. The continuous ranked probability score conditional to the level of wind power for the Lasithi case study

Finally, the usefulness of the input variables, namely the uncertainty indices d_w and d_e, the weather stability measures sd_s and sd_w and the RBFNN uncertainty indices (the winning neuron n, and the worst input variable v) is validated by training and evaluating the proposed model without these variables. The results for the Lasithi case study are illustrated in figure 11. It is shown that all the parameters used have a significant impact on the wind power density estimation. The uncertainty indices d_w and d_e have the largest impact on the proposed model performance.

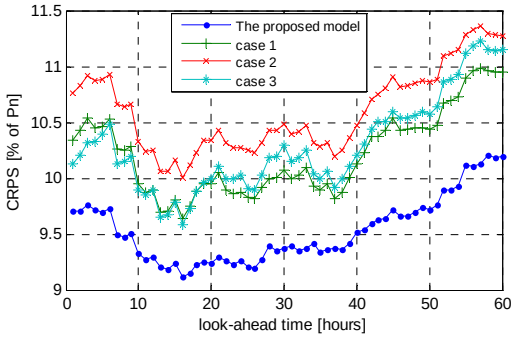


Fig. 11. The performance of the proposed model trained without the RBFNNs uncertainty indices (case 1), without the uncertainty indices (case 2) and without the weather stability measures (case 3).

VI. Conclusions

In this paper, a novel probabilistic wind power prediction model is presented. The model receives as input the point predictions and the uncertainty information of a state-of-art deterministic prediction model based on the use of RBFNNs. Information about the prediction uncertainty of this model is obtained from the hidden layer output of the RBFNNs. Self-organized maps classify the uncertainty knowledge in multiple levels and RBFNNs are applied for each level in order to predict eight quantiles of the future wind power distribution. A novel feature of the method is that additional uncertainty information is obtained by the weather stability that is estimated from several points of the NWP grid. Finally, the PSO algorithm is used to estimate the best parameters of the RBFNNs optimizing the bias of the quantile forecasts together with several generalization criteria. The evaluation of the proposed model on two different wind farms shows that the

model can perform very satisfactory and is very robust in different weather conditions and in different terrains.

VII. References

[1] G. Kariniotakis et al. “What performance can be expected by short term wind power prediction models depending on site characteristics?”, Proc. of EWEC 2004. London 2004.

[2] G. Kariniotakis et al. “Evaluation of Advanced Wind Power Forecasting Models –Results of the Anemos Project”, EWEC, Athens 2006.

[3] V. Guénard, G., Kariniotakis, I. Martí, , “ANEMOS Advanced Wind Power Forecasting. Operational Challenges and On-line Performance”. Proc. of EWEC'07, Milan, Italy, 7-10 May 2007.

[4] R. Doherty, M. O’Malley, “A new approach to quantify reserve demand in systems with significant installed wind capacity”, IEEE Trans. on Power Systems, vol. 20, pp. 587–5952005.

[5] J. Usaola, O. Ravelo, G. Gonzalez, F. Soto, C. Davila, B. Diaz-Guerra, “Benefits for wind energy in electricity markets form using short term wind power prediction tools: a simulation study”, Wind Engineering, vol. 28, no. 1, pp. 119-128, 2004

[6] E.D. Castronuovo, J.A. Pecas Lopes, “On the optimization of the daily operation of a wind-hydro power plant”, IEEE Trans. on Power Systems, vol. 19, pp. 1599–1606, 2004.

[7] G. N. Bathurst, J.Weatherhill, and G. Strbac, “Trading wind generation in short-term energy markets,” IEEE Trans. Power System, vol. 17, no. 3, pp. 782–789, Aug. 2002.

[8] P. Pinson, C. Chevallier, and G. Kariniotakis, “Trading Wind Generation From Short-Term Probabilistic Forecasts of Wind Power,” IEEE to Power System, vol. 22, no. 3, pp. 1148–1156, Aug. 2007

[9] T.S.Karakatsanis,N.D.Hatziargyriou, “Probabilistic Constrained Load Flow based on Sensitivity Analysis” ,*IEEE Trans on Power Systems*, Vol 9,No 4, November 1994,pp 1853-1860

[10] G. Giebel, L. Landberg, J. Badger, K. Sattler, H. Feddersen, T.S. Nielsen, H.Aa. Nielsen, H. Madsen, “Using Ensemble Forecasting for Wind Power”, EWEC’03, Madrid, 16-20 June, 2003

[11] J.W. Taylor, P.E. McSharry, and R. Buizza, “Wind Power Density Forecasting Using Ensemble Predictions and Time Series Models”, IEEE Trans. on Energy Conversion, vol. 24, pp. 775-782, 2009

[12] J. B. Bremnes, “Probabilistic wind power forecasts using local quantile regression”, Wind Energy, vol. 7, no 1, pp. 47–54, 2004.

[13] H. A. Nielsen, H. Madsen, and T. S. Nielsen, “Using quantile regression to extend an existing wind power forecasting system with probabilistic forecasts”, Wind Energy, vol 9, no 1-2, pp. 95–108, 2006.

[14] T. S.Nielsen and H.Madsen, ‘Statistical methods for predicting wind power’, in Proceedings of the 1997 EuropeanWind Energy Conference, EWEC’97, Dublin, Ireland, pp. 755–758, 1997

[15] J. Juban, L. Fugon, and G. N. Kariniotakis, “Probabilistic short-term wind power forecasting based on kernel density estimators,” Proc. of the EWEC’07, Milan, Italy, May 2007.

[16] R.J. Bessa, V. Miranda, A. Botterud, Z. Zhou, J. Wang, “Time-Adaptive Quantile-Copula for Wind Power Probabilistic Forecasting,” Renewable Energy, Vol. 40, No. 1, pp. 29-39, 2012.

[17] P. Pinson, G. Kariniotakis, “On-line assessment of prediction risk for wind power production forecasts” Wind Energy, vol 7, pp. 119–132, 2004.

[18] M. Lange, “Analysis of the Uncertainty of Wind Power Predictions” PhD dissertation, Carl von Ossietzky Oldenburg University, 2003.

[19] Giebel G., “The State of the Art in Short-Term Prediction of Wind Power - A Literature Overview, 2nd Edition” Deliverable 1.2b of the ANEMOS.plus project. Available: Anemos-plus.eu.

[20] G. Sideratos and N. Hatziargyriou, “An Advanced Radial Base Structure for Wind Power Forecasting”, International journal on Power and Energy Systems, ACTA Press, Vol 12, November 2008.

[21] G. Sideratos, N. Hatziargyriou, ” An Advanced Statistical Method for Wind Power Forecasting”, IEEE Transactions on Power System, Vol. 22, Issue 1, pp. 258-265, February 2007.

[22] S. Chen, C.F.N. Cowan, and P.M. Grant, “Orthogonal least squares learning for radial basis function networks”, IEEE Transactions on Neural Networks, vol. 2, pp. 302-309.

[23] J. Kennedy, R. Eberhart, "Particle Swarm Optimization". Proc. of IEEE International Conference on Neural Networks, vol IV. pp. 1942–1948, 1995.

[24] M.J.L Orr, "Regularization in the selection of radial basis function centers", Neural Computation, vol.7, pp. 606-623.

[25] D. J. C. MacKay, “Bayesian interpolation,” Neural Computation., vol. 4, no.3, pp. 415–447, May 1992.

[26] P. Pinson, H.Aa. Nielsen, J.K. M_ller, H. Madsen, G. Kariniotakis, “Nonparametric probabilistic forecasts of wind power: required properties and evaluation”, Wind Energy, vol. 10, no 6, pp. 497-516, 2007

[27] T. Kohonen, ‘Self-Organizing Maps’, Springer, 2001 Berlin

[28] S. Haykins, ’Neural networks’, Pearson-Prentice Hall, second edition, 2005

[29] P. Pinson, P. McSharry, H. Madsen, “Reliability diagrams for density forecasts of continuous variables: accounting for serial correlation”. Quarterly Journal of the Royal Meteorological Society, vol. 136, issue 646, pp. 77-90c, 2010

[30] T. Gneiting, F. Balabdaoui, and A. E. Raftery, “Probabilistic forecasts, calibration and sharpness”. Journal of the Royal Statistical Society: Series B (Statistical Methodology), Vol. 69, pp. 243-268, 2007.

[31] J. E. Matheson, and R. L. Winkler, “Scoring Rules for Continuous Probability Distributions,” Management Science, vol. 22, pp. 1087–1096, 1976.

[32] H. Madsen, P. Pinson, H. Aa. Nielsen, T. S. Nielsen, and G. Kariniotakis, ‘Standardizing the performance evaluation of short-term wind power prediction models’, Wind Engineering, vol. 29, no 6, pp. 475–489, 2005

[33] P. Georgilakis, “Technical challenges associated with the integration of the wind power into power systems”, Renewable and Sustainable Energy reviews, vol 12, pp. 852-863, 2008

VI. Biographies

George Sideratos was born in Chios, Greece in 1976. He received Electrical and Computer Engineering degree from the National Technical University of Athens (NTUA), Athens, in 2002, the Ph.D. degree in electrical engineering from NTUA, in 2010. He is currently a senior researcher in the Power System laboratory of NTUA. His research interests include wind power and load forecasting and artificial intelligence techniques.

Nikos D. Hatziargyriou (S’80–M’82–SM’90-FM’09) was born in Athens, Greece, in 1954. He received the Diploma in Electrical and Mechanical Engineering from the National Technical University of Athens (NTUA), Athens, in 1976, and the M.Sc. and Ph.D. degrees in electrical engineering from the University of Manchester Institute of Science and Technology (UMIST), Manchester, U.K., in 1979 and 1982, respectively. He is professor at the Power Division of the Electrical and Computer Engineering Department of NTUA. His research interests include dispersed and renewable generation, artificial intelligence techniques in power systems, and dynamic security assessment of power systems. Since 2007 he is Deputy CEO of the Public Power Corporation (PPC) in Greece, responsible for Transmission and Distribution Networks, island DNO and the Center of Testing, Research and Prototyping. He is Fellow Member of IEEE, immediate past Chair of the Power System Dynamic Performance Committee, member of CIGRE, Convener of SCC6, member of the BoD of EURELECTRIC and member of the EU Advisory Council of the Technology Platform on SmartGrids.