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-6.4
"Regime-switching methods and resulting benefits"

DOCUMENT TYPE	Deliverable
DOCUMENT NAME:	swind.deliverable_Dp-6.4.pdf
VERSION:	V1.0
DATE:	2012.10.01
CLASSIFICATION:	R0: General public
STATUS:	Approved

Abstract: This report compiles the series of work carried out in the frame of Task 6.4 of the EU-FP7 project SafeWind, with focus given to regime-switching approaches in wind power forecasting and their potential resulting benefits. Various models and forecasting methods were proposed and validated on sets of test cases in Denmark, Ireland, and Crete. Some of these exercises were for short term forecasting (with lead times between 10 minutes and 6 hours ahead), while others looked at longer lead times, in the order of hours to days. Significant benefits in terms of forecast accuracy were obtained, in both point and probabilistic forecasting set-ups.

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:	Pierre-Julien Trombe (DTU), Miguel Bermejo (UC3M), Ismael Sanchez (UC3M), George Sideratos (NTUA), Nikos Hatzygiriou (NTUA), George Anastasiades (Uni. Oxford), Patrick McSharry (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, or as technical reports, 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, or university library. These papers include:

Trombe P-J, Pinson P (2012). High-resolution forecasting of wind power generation with regime-switching models and off-site observations. Technical report, Technical University of Denmark, Dpt. of Informatics and Mathematical Modelling, Tech. Rep. no. 2012-15.

Bermejo MA, Sánchez I (2012). Short-term prediction of wind energy using Threshold Autoregressive models. Technical Report, Universidad Carlos III Madrid, Dpt. of Statistics, 2012.

McSharry P, Anastasiades G (2012). Quantile forecasting of wind power using variability indices. Working paper, submitted.

Sideratos G, Hatziargyriou N (2012). Wind power forecasting focused on extreme power system events. *IEEE Transactions on Sustainable Energy* **3**(3): 445-454.

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 are commonly evaluated by error measures such as the Mean Absolute Error (MAE) and the Root Mean Square Error (RMSE), which only relate to an evaluation of forecast accuracy for all types of external conditions. This is while in terms of power systems management and participation in electricity markets, extreme events of all sorts may be leading to much higher costs than business-as-usual situations. This is the reason why regime-switching approaches should be envisaged so as to make a difference between situations that could lead to different types of consequences for the decision-makers. Some of the specific works on events eg. ramps Bossavy *et al.* (2012) are focused on very specific events, while here emphasis is placed on more general definitions of regimes, in sense of situations with distinct and characteristic dynamic behaviour of wind power generation and its forecast uncertainty.

In the frame of the EU-FP7 project SafeWind, a Task was defined so as to concentrate on these regime aspects, by developing appropriate methodologies, and also by showing what the benefits of regime-switching forecasting approaches may be. A part of the work there was also focused on putting in the balance these benefits with the corresponding increase in modelling and forecasting complexity. In addition, consideration was given to both point and probabilistic forecasting methodologies, since today both are recognized as important inputs to decision making. This report consists in a collection of work that were published or submitted in scholarly journals or as technical reports, based on the research work carried out in this specific Task of 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. Trombe and Pinson (2012) focuses on short-term forecasting for lead times within the hour ahead, by combining advanced modelling of predictive densities, rival approaches to regime switching, as well as potential use of offsite information. results are given for a case study in Ireland. In parallel, Bermejo and Sánchez (2012) concentrates on a general methodology permitting to identify regime-switching models with an observable regime sequence, based on recursive estimation techniques. This paper is an application of some earlier more methodological work presented in Bermejo *et al.* (2011) to the case of short-term forecasting of wind power generation, with results from Greek wind farms. McSharry and Anastasiades (2012) extends to a probabilistic framework by looking at how nonparametric probabilistic forecasts of wind power generation may e obtained by additionally accounting by some regime information, where regimes are defined by recent variability indices. Application results are given for the case of few wind farms in Denmark. Finally, the work in Sideratos and Hatziargyriou (2012) describes the use of artificial intelligence techniques for the forecasting of wind power generation with focus on extreme events, for longer lead times (in the order of hours to days).

References

Bermejo MA, Pena D, Sánchez I (2012). Identification of TAR models using recursive estimation. *Journal of Forecasting* **30**(1): 31–50.

Bermejo MA, Sánchez I (2012). Short-term prediction of wind energy using Threshold Autoregressive models. 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`].

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]

Sideratos G, Hatziargyriou N (2012). Wind power forecasting focused on extreme power system events. *IEEE Transactions on Sustainable Energy* **3**(3): 445-454.

Trombe P-J, Pinson P (2012). High-resolution forecasting of wind power generation with regime-switching models and off-site observations. Technical report, Technical University of Denmark, Dpt. of Informatics and Mathematical Modelling, Tech. Rep. no. 2012-15.

McSharry P, Anastasiades G (2012). Quantile forecasting of wind power using variability indices. Working paper, submitted.

3 Appendix

IMM – Technical Report – 2012-15

High-resolution forecasting of wind power generation with regime-switching models and off-site observations

Pierre-Julien Trombe[1,*], **Pierre Pinson**[1]

[1] DTU Informatics, Technical University of Denmark, Kgs. Lyngby, Denmark

[*] Corresponding author:
P.-J. Trombe, DTU Informatics, Technical University of Denmark,
Richard Petersens Plads (bg. 305 - room 227), DK-2800 Kgs. Lyngby, Denmark.
Tel: +45 4525 3402, fax: +45 4588 2673, email: pjt@imm.dtu.dk

1 Introduction

With the growing penetration of wind power into power systems, electric utilities are called to revise their operational practices. In particular, experts in energy management recommend to increase the scheduling frequency of electricity generation and delivery from hours to minutes, in order to mitigate the impact of wind power variability on power systems [1]. Transmission System Operators (TSO) expressed concurring views on the integration of large amounts of wind power into power systems [2]. In a few European countries, very short-term wind power forecasts with temporal resolutions from 5 to 15 minutes, and lead times up to 36-48 hours, are already used in a wide range of applications [3]. These include among others optimizing reserve allocation, balancing electricity consumption and production, and controlling wind power fluctuations at large offshore wind farms [4,5]. In particular, one application for which forecasts with specific lead times up to 15-20 minutes are needed is the management of the immediate regulating power reserve. This type of reserve is activated over time intervals up to 15-20 minutes, after the system experiences a sudden and large deviation between scheduled and actual wind power generation [6]. This issue is paramount in countries or regions with limited interconnections, or with no complementary source of energy (e.g., hydro or pumped hydro) that can be both stored and used for fast-acting generation.

Issuing improved wind power forecasts for supporting decision-making in regulating reserve management has the merit of being more cost-effective when compared to other solutions such as increasing backup capacities. For lead times from a few minutes to a few hours, wind power forecasts are best generated with statistical models using historical data. However, developments in wind power forecasting have long been oriented towards energy market applications, placing focus on forecasts at hourly resolutions, as required by the market structure. These approaches heavily rely on the availability of meteorological forecasts of wind speed and direction owing to the strong relation between wind and wind power, the so-called power curve [7]. Employing such a strategy is not realistic when working with lead times of a few minutes. Instead, a number of new modeling and forecasting approaches were recently proposed in view of improving the predictability of wind power fluctuations for very short lead times. These include regime-switching models, off-site predictors and a new type of predictive distribution.

Regime-Switching models – The motivation for applying these models comes from the existence of structural changes in the dynamics of wind power fluctuations at temporal resolutions of a few minutes, hence the term *wind power regime*. Periods of low and high wind power variability alternate, not only modulated by the wind own variability, but also by the power curve that amplifies or dampens wind fluctuations owing to its nonlinear nature. For low or high wind speeds, wind power fluctuations are very small whereas, for moderate wind speeds (i.e., roughly between 7 and 13 $\mathrm{m.s}^{-1}$), wind power fluctuations can become extreme. Originally developed for applications in Econometrics [8], regime-switching models have, since then, also been applied for modeling and forecasting offshore wind power fluctuations in [9–11], improving the accuracy of wind power forecasts when compared to single regime models. Regime-switching models divide into two categories, those for which regimes are observable and determined by expertise, and those for which they are unobservable and estimated jointly with the model. This translates into two classes of time series models, namely Threshold Autoregressive (TAR) and Markov-Switching Autoregressive (MSAR) models [8, 12].

Off-site predictors – Traditional inputs to statistical prediction models consist of on-site observations

(i.e., wind power production, wind speed and direction) and/or meteorological forecasts (wind speed and direction, temperature, atmospheric pressure). However, meteorological forecasts are generated at coarse temporal resolutions, from 1 to 3 hours, and therefore not informative on intra-hour wind fluctuations. Furthermore, wind measurements are rarely available in real-time for applications with lead-times of a few minutes. When wind power data and wind data are not simultaneously available, the difficulty of generating accurate wind power forecasts increases. This is the reason why a number of recent studies explored the potential of off-site observations as new predictors [13–19]. In particular, wind farms and meteorological masts scattered over a region form a net capable of capturing valuable information on the weather conditions. Owing to the synoptic mechanisms in the atmosphere which drive wind variability in space and time, upwind observations can be informative of upcoming changes in weather conditions and be used as extra predictors [20, 21]. Two distinct approaches exist for integrating these off-site predictors into forecasting models, depending on whether (i) the dominant weather conditions are known a priori and the model designed accordingly [13–16], or (ii) there is no a priori information available on weather conditions and it is assumed that the model can capture the associated effects directly from the data [17–19]. Despite their high accuracy, models based on the first type of approach have a clear downside, they tend to be very region or site-dependent, lacking of adaptivity when applied to areas with different weather conditions. In contrast, models based on the second type of approach are more data-driven and require less expert knowledge to capture the spatio-temporal dependencies between sites.

The Generalized Logit-Normal distribution – Wind power generation is a double-bounded process since it can neither be negative nor exceed the wind farm rated capacity. In addition, the distribution of wind power forecast errors changes with respect to the conditional expectation of the forecasts [22]. In particular, heavy skewness near the bounds and a clear heteroscedastic behavior are generally observed. In a parametric framework, a common approach for dealing with these features consists in combining a statistical model that handles the heteroscedasticity (e.g., Generalized Autoregressive Conditional Heteroscedastic (GARCH) models) with a predictive distribution that deals with the effects of the bounds and, potentially, with skewness (e.g., censored and truncated Normal distributions) as in [23]. A generalization of this type of approach was proposed in [24] with the Generalized Logit-Normal (GLN) distribution and applied for forecasting wind power fluctuations at large offshore wind farms.

All three aforementioned approaches yielded substantial gains in wind power predictability, in a wide variety of contexts. However, their predictive performances, yet demonstrated against traditional benchmark models, were not compared against one another. As a result, there seems to be a great deal of confusion on the direction to follow for forecasting wind power fluctuations. In particular, the constraints imposed by short lead time applications (i.e., no wind measurements) offer a difficult test to the robustness of these approaches. For instance, one may wonder whether the relative complexity of regime-switching models is worth the gain in predictability, when compared to more parsimonious models with a single regime and tuned with off-site predictors and the GLN distribution. As a first attempt to clear this point out, we perform a comparative study of the predictive performances of the different approaches and, eventually, explore different combinations of them in order to evaluate whether additional improvements can be obtained. Focus is placed on wind power fluctuations from a single wind farm.

Wind power forecasts and, more generally, forecasts of any continuous quantity are given in the form of either a single-value (i.e., deterministic forecast) or a full probability distribution or density (i.e., probabilistic forecast). As pointed out in [25], forecasts ought to be probabilistic in order to achieve optimal decision-making under uncertainty. This idea found its echoes with a few TSOs which started

using probabilistic information in control rooms [2]. In this work, the accuracy of wind power forecasts is verified with respect to both point and density forecasts even though more importance will be given to the latter ones.

This paper is organized as follows. Section 2 introduces the case study, the data and their characteristics. Section 3 presents the four classes of model considered in this study, namely Autoregressive (AR), AR-GARCH, TAR, MSAR. In section 4, the predictive performances of these models are evaluated both in terms of point and density forecasts. Finally, section 5 delivers concluding remarks.

2 Data and their characteristics

In this section, we present the data and their characteristics. We also perform a number of analysis to introduce some essential principles that motivate modeling assumptions in section 3. In particular, we give a detailed account on the the GLN predictive distribution as proposed in [24], and evaluate spatio-temporal correlations of wind power in view of integrating off-site predictors into time series models.

2.1 Case study

The case study consists of a group of three wind farms located in the South-East of Ireland, the Carnsore wind farm which has a rated capacity (Pn) of 11.9 MW and its two nearest wind farms, Richfield (27 MW) and Ballywater (42 MW), as shown in Figure 1. Ballywater and Richfield are located about 40 km North-East and 17 km West of Carnsore, respectively. The Carnsore wind farm is located at the extreme point of a peninsula, by the sea shore. Richfield and Ballywater are located further away inland but within 5-10 km from the sea, remaining in the zone of influence of the marine weather. In this study, focus is placed on forecasting the wind power generation at the Carnsore wind farm. As aforementioned, no wind measurement is available. Furthermore, available meteorological forecasts have a too coarse temporal resolution to be informative for lead times of a few minutes and thus cannot be used. Our knowledge of weather conditions in Ireland is restricted to the prevalence of southwesterly winds. In addition, passages of low-pressure systems characterized by large wind variability and developments of storms are more frequent over the period from August to January [26].

Ireland and its power system are singular when compared to other countries/regions with high wind power penetrations. Ireland has large wind resource but very limited interconnection capacity with power systems from other countries. More specifically, there exists a single interconnection to Northern Ireland which, in turn, is only connected to the United Kingdom. The target of Ireland is to meet 40% of its energy demand with renewable energy sources by 2020, of which 37% are expected to be covered by the integration of wind power. The small interconnection capacity clearly acts a limiting factor for enabling further wind power into the system since the latter will be unable to spill excess power when needed. Consequently, improved wind power predictability would allow to decrease the frequency of curtailment actions and reduce losses of wind power generation [3].

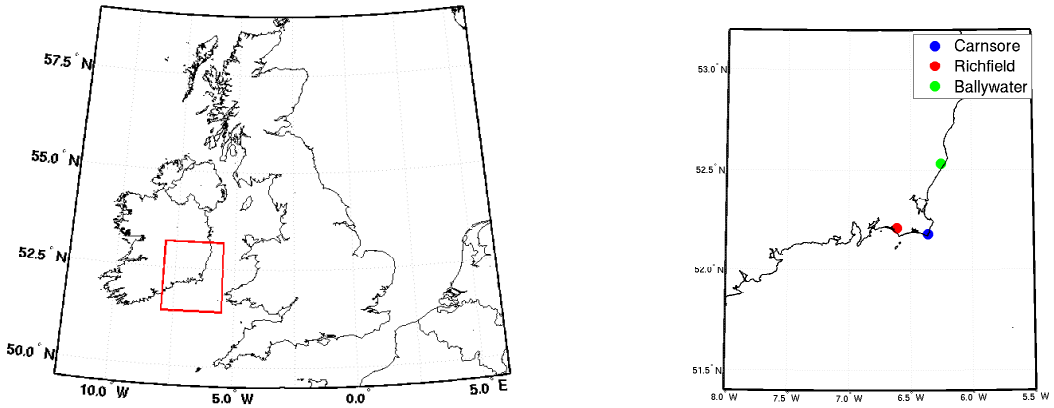


FIGURE 1: *The Carnsore, Richfield and Ballywater wind farms are located in the South-East of Ireland. Carnsore and Richfield are separated by an approximate distance of 17 km, and the distance between Richfield and Ballywater is 40 km.*

2.2 Data quality control

The wind power data used in this study are provided by Eirgrid, the TSO in Ireland. They span the period from December 31, 2006 to June 1, 2009. One time series of wind power production is available for each wind farm, at a temporal resolution of 15 minutes. Following [27], time series are normalized and expressed as a percentage of the wind farm rated capacity. The resulting time series take values on the unit interval $[0, 1]$. The raw data records are complete for Carnsore and Richfield but not for Ballywater for which 3071 values (out of 84864) are reported missing. Since the data consist of output power time series, and not available power, a data quality control is performed. We identify several periods where the output power is curtailed, likely indicating that some wind turbines were temporarily out of order or that an absolute power limitation was imposed. An example is given in Figure 2 which shows the time series of wind power for the Carnsore wind farm. The output power never exceeds 92% of the rated power of Carnsore in the second semester of 2007 and the first semester of 2008. Consequently, we only use the period from July 10, 2008 to 27 March, 2009 in this study, corresponding to more than 25000 data points. This period is shaded in grey in Figure 2.

2.3 The Generalized Logit-Normal predictive distribution

The conversion from wind to power makes that wind power generation is a double-bounded process, with a potentially high concentration of observations near or at the bounds. This feature is illustrated in Figure 3. In addition, the shape of the distribution of the wind power forecast errors evolves with the conditional expectation of the forecasts. Near the bounds, the conditional distribution of wind power forecast errors tends to have a very small standard deviation and to be heavily skewed. Moving away from these bounds, the standard deviation increases and the skewness decreases [22]. When forecasting wind power generation from single wind farms, designing an appropriate strategy for taking these features into account is paramount. In [24], the author proposed the use of the Generalized Logit-Normal

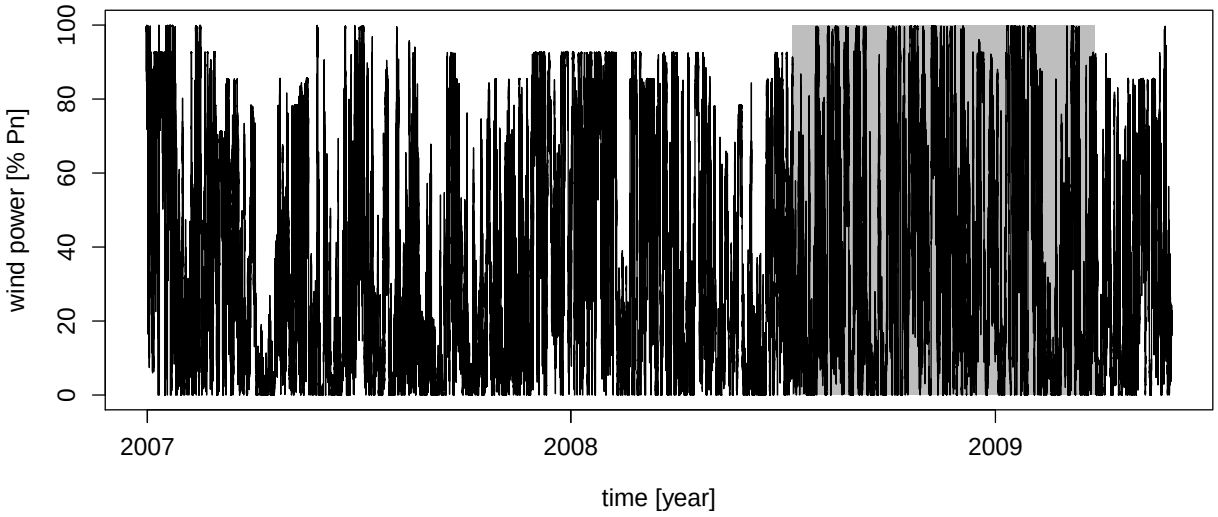


FIGURE 2: *Time series of wind power at Carnsore. The data overlaying the shaded area are considered to be of good quality and used for the experimental part of this study.*

(GLN) distribution. The underlying motivation for using this distribution comes from the work of [28] where it is shown that appropriate data transformations may enhance characteristics such as linearity, homoscedasticity and additivity.

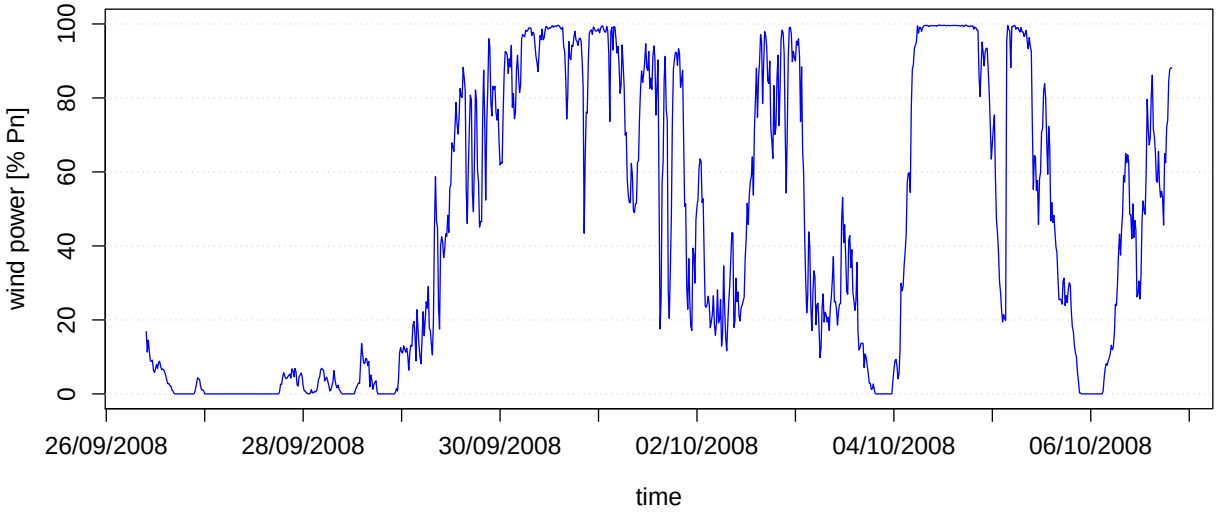


FIGURE 3: *Normalized wind power generation at Carnsore. The temporal resolution of 15 minutes.*

The homoscedasticity of wind power forecast errors can be enhanced by transforming the original time series $\{y_t\}$ as follows:

$$\widetilde{y}_t = \gamma(y_t, \nu) = \log\left(\frac{y_t^\nu}{1 - y_t^\nu}\right), \quad \nu > 0, \quad y_t \in [0, 1] \tag{1}$$

where ν is a shape parameter and the resulting time series $\{\widetilde{y}_t\}$ takes values in $]-\infty, +\infty[$. This transformation, as shown in Figure 4 for a set of different values of ν, aims at outstretching the distribution near the bounds of the interval $[0, 1]$. In the original domain $[0, 1]$, the assumption of homoscedastic

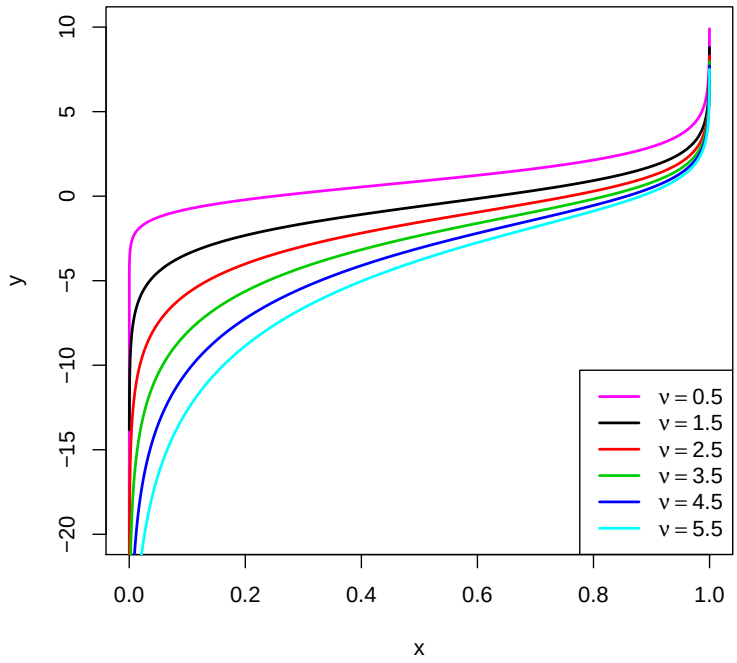


FIGURE 4: *The GLN predictive distribution consists of transforming the original wind power observations in order to deal with the heteroscedasticty near the bounds of the interval* $[0, 1]$

wind power forecast errors does not hold and, even though one may argue that this may still not be the case after transforming the time series, making that such assumption is clearly more appropriate in the transformed domain than in the original one.

However, the concentration of observations at the bounds, in 0 and 1, generates two probability masses that remain in the transformed domain. They are located in $-\infty$ and $+\infty$, respectively. To fix this, the coarsening principle is applied as in [29]. All observations taking values in the open interval $]-\infty, \gamma(\epsilon, \nu)[$ are shifted to $\gamma(\epsilon, \nu)$. Likewise, all observations taking values in $]\gamma(1-\epsilon, \nu), +\infty[$ are shifted to $\gamma(1-\epsilon, \nu)$, with $\epsilon < 0.01$. Two Dirac distributions $\delta_{\gamma(\epsilon,\nu)}$ and $\delta_{\gamma(1-\epsilon,\nu)}$ are introduced so that the one-step ahead predictive distribution in the transformed domain, $Y_{t+1|t}$, is defined as follows:

$$Y_{t+1|t} \sim \omega^0_{t+1|t}\delta_{\gamma(\epsilon,\nu)} + \mathcal{N}(\hat{\mu}_{t+1|t}, \hat{\sigma}^2_{t+1|t})\mathbf{1}_{]\gamma(\epsilon,\nu),\gamma(1-\epsilon,\nu)[} + \omega^1_{t+1|t}\delta_{\gamma(1-\epsilon,\nu)} \quad (2)$$

$$\omega^0_{t+1|t} = \Phi\left(\frac{\gamma(\epsilon,\nu) - \hat{\mu}_{t+1|t}}{\hat{\sigma}_{t+1|t}}\right) \quad (3)$$

$$\omega^1_{t+1|t} = 1 - \Phi\left(\frac{\gamma(1-\epsilon,\nu) - \hat{\mu}_{t+1|t}}{\hat{\sigma}_{t+1|t}}\right) \quad (4)$$

where Φ is the cumulative distribution function of the Normal variable with 0 mean and unit variance.

2.4 Spatio-temporal correlations in wind data

Recent studies showed that it was possible to take advantage of spatio-temporal correlations in wind data at an hourly resolution in order to improve the predictability of wind speed or wind power at regional

scales [15–18]. Nevertheless, for higher temporal resolutions, in the order of a few minutes, the wind variability caused by local effects is magnified and may reduce these correlations. Besides that, other factors which contribute to decrease spatio-temporal correlations of wind data include topographical effects and inter-site distances. When considering wind power data, the potential effects of the power curve cannot be ignored. The power curve is a function of atmospheric variables such as wind speed, wind direction, wind shear and air density. For identical atmospheric conditions at two wind farms, differences in the type, age and size of wind turbines, as well as their geographical spread, may result in large differences in generated power, and thereby decrease spatio-temporal correlations.

For a reasonable number of wind farms, a visual assessment of their respective wind power generation can give clear indications on the potential level of spatio-temporal correlations. Figure 5 shows three time series of normalized wind power from Carnsore, Richfield and Ballywater over a 4-day episode. Wind power fluctuations from Carnsore and Richfield closely follow each other. Still, it appears difficult to identify a clear and recurrent pattern on whether wind fluctuations at Carnsore leads those at Richfield, or whether it is the opposite. This potentially reflects changes in wind direction. Note also that the wind power level at Ballywater is significantly lower than at Carnsore and Richfield.

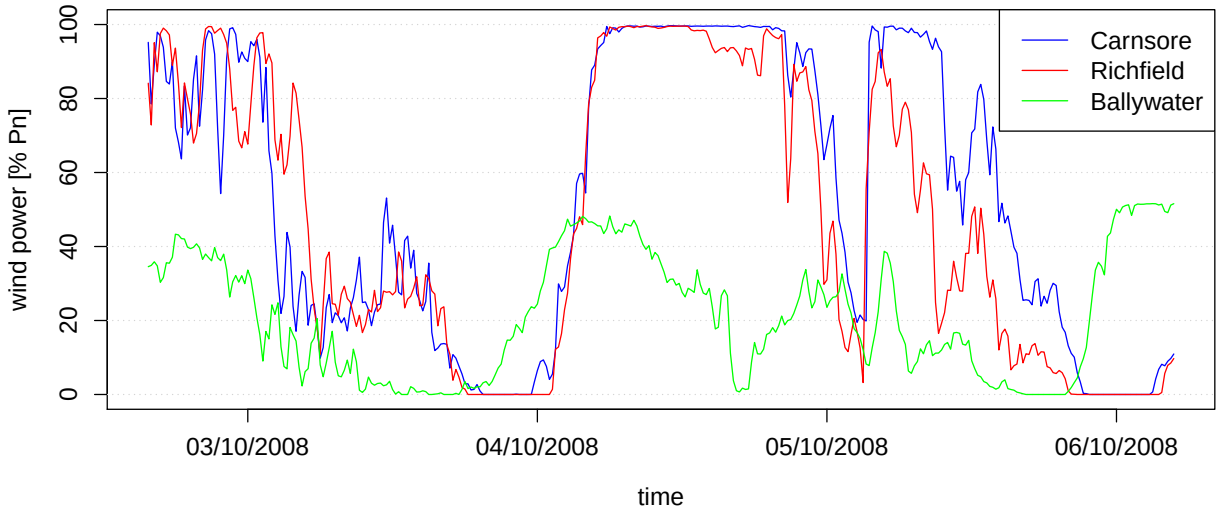


FIGURE 5: *Normalized wind power generation at Carnsore, Richfield and Ballywater with a temporal resolution of 15 minutes.*

Before using off-site observations for prediction applications, it is essential to analyze correlations between wind data from distant sites. Following [21], we assume that these correlations can appropriately be described and quantified by the traditional linear correlation coefficient. In order to evaluate these correlations, we use the pre-whitening technique presented in [30]. Let A and B be two wind farms, with their respective time series of wind power generation $\{y_t^{(A)}\}$ and $\{x_t^{(B)}\}$. $\{x_t^{(B)}\}$ is called the input series and $\{y_t^{(A)}\}$ the output series. The idea is to use the power generation from wind farm B as input for improving the wind power predictability of wind farm A. The procedure is divided into three steps as follows:

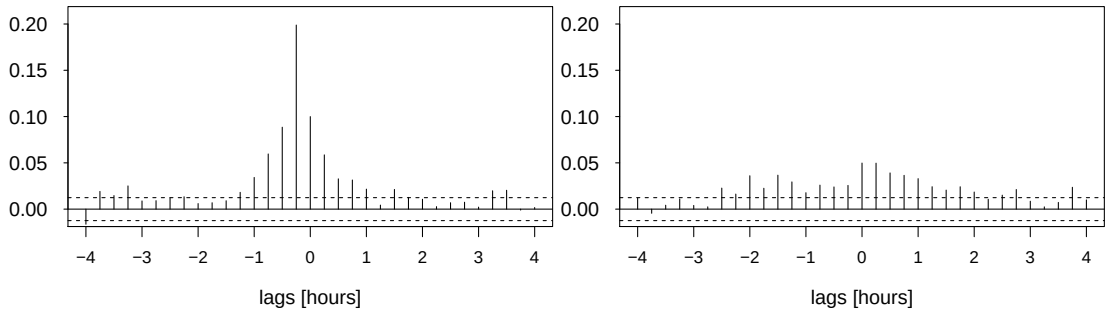
1. An appropriate Autoregressive Moving Average (ARMA) model is fitted to the input series $\{x_t^{(B)}\}$ and a series of residuals $\{e_t^{(B)}\}$ extracted,
2. The output series $\{y_t^{(A)}\}$ is filtered with the same model as in step 1 and a series of residuals $\{e_t^{(A)}\}$

extracted,

3. The cross-correlation function is calculated based on the two series of residuals as follows:

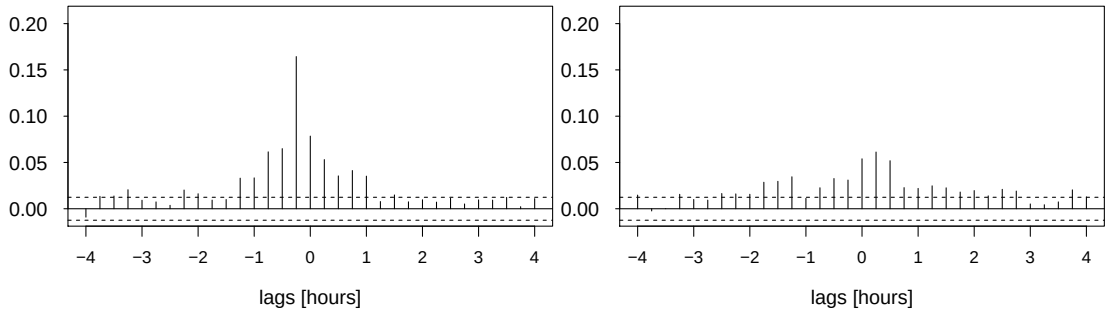
$$\rho_{e^{(A)}e^{(B)}}(\tau) = \frac{\text{cov}(e^{(A)}(t), e^{(B)}(t+\tau))}{\sigma_{e^{(A)}}\sigma_{e^{(B)}}} \tag{5}$$

We repeat the pre-whitening procedure presented hereabove with and without the GLN transformation as given by equation (1) in order to evaluate how this transformation changes the correlation structure between the power generation from two wind farms. The results are reported in Figure 6. Negative lags indicate that wind power fluctuations at Richfield or Ballywater lead those at Carnsore. First, these results reveal larger cross-correlations between Richfield and Carnsore than between Ballywater and Carnsore,



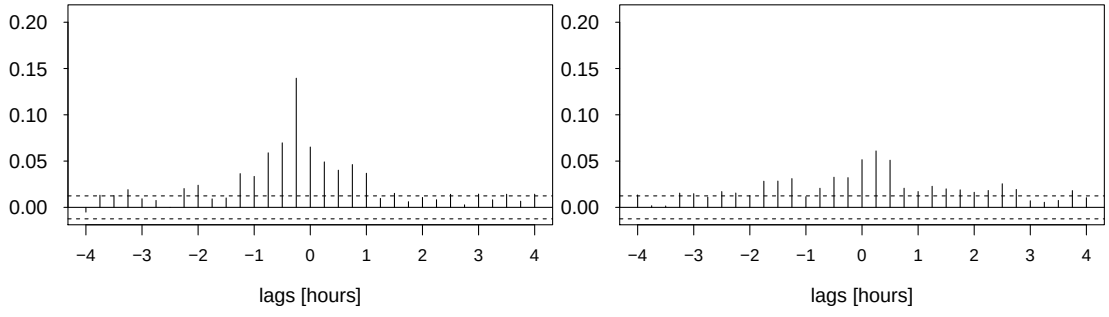
(a) Richfield & Carnsore, No GLN transf.

(b) Ballywater & Carnsore, No GLN transf.



(c) Richfield & Carnsore, GLN transf. with $\nu = 1$

(d) Ballywater & Carnsore, GLN transf. with $\nu = 1$



(e) Richfield & Carnsore, GLN transf. with $\nu = 2$

(f) Ballywater & Carnsore, GLN transf. with $\nu = 2$

FIGURE 6: *Cross-correlations (after pre-whitening) of wind power generation at Carnsore and (left column) Richfield, (right column) Ballywater. Negative lags indicate that wind power fluctuations at Richfield or Ballywater lead those at Carnsore.*

thereby confirming the visual observations made from Figure 5. This result is most likely the consequence of the shorter distance separating Carnsore from Richfield than from Ballywater which would be consistent with the empirical analysis in [21] where spatio-temporal correlations are shown to quickly decrease within a radius of 50 km. Figure 6 also shows that wind power fluctuations at Richfield tend to lead those at Carnsore up to 30 minutes ahead, on average. In contrast, cross-correlations between Ballywater and Richfield are much lower and it appears more difficult to determine a clear tendency on whether wind power fluctuations propagate preferentially from Richfield to Ballywater, or the opposite. A direct extrapolation from these cross-correlations suggests that off-site observations from Richfield have a higher potential for improving wind power predictability at Carnsore than corresponding observations from Ballywater. Finally, one can see that cross-correlations between Carnsore and Richfield are larger without applying the GLN transformation a priori. Cross-correlations tend to decrease with large values of the shape parameter ν. We can think of two potential causes that explain this feature. First, using the GLN distribution may degrade the linear relationship between the two time series, particularly near the bounds where the respective variances may increase. Secondly, using the GLN distribution may enhance the homoscedasticity of the input time series $\{x_t^{(B)}\}$ so that the residuals series $\{e_t^{(B)}\}$ is closer to being a white noise process, and thereby is less informative.

3 Time series modeling

The stochastic nature of wind power generation is described hereafter with time series models. We start by considering linear models (i.e., ARX and ARX-GARCH) before moving on to nonlinear regime-switching models (i.e., TARX and MSARX). For each model, we give the most general formulation, meaning that off-site predictors are included by default, hence the X in model acronyms. Our objective is to estimate models in view of generating, not only accurate point forecasts, but also probabilistic forecasts. All models are thus estimated by Maximum Likelihood Estimation (MLE) rather than Least Squares (LS). Let $\{y_t^{(A)}\}$ (respectively $\{\widetilde{y}_t^{(A)}\}$) be the observed (respectively transformed) time series of wind power generation to be predicted at a given wind farm A. Let $\{x_t^{(WF)}\}$ be a time series of off-site wind power generation observed at a distant wind farm WF, with $WF = B, C, \ldots$. For the sake of simplicity, $y_t^{(A)}$ (respectively $x_t^{(WF)}$) denotes both the random variable and its observed value at time t. Let $\mathbf{\Omega_t} = (y_1^{(A)}, \ldots, y_t^{(A)}, x_1^{(B)}, \ldots, x_t^{(B)}, x_1^{(C)}, \ldots, x_t^{(C)}, \ldots)$ be the set of observations available at time t.

3.1 ARX models

While it is generally acknowledged that wind power generation is a nonlinear process, operational wind power forecasting systems usually rely on linearity assumptions [31]. ARX models are some of the most widely used in practice. There are several reasons for this. First, their formulation is very intuitive and simply consists of a linear combination of lagged variables which leads to fast estimation procedures. Secondly, they stand as very competitive models for generating point forecasts owing to their parsimony (i.e., low number of parameters to be estimated). Thirdly, there exists closed-form formula for generating multi-step ahead forecasts [30].

The linear AR(p)-X(q) model with p autoregressive and q exogenous predictors is given by:

$$y_t^{(A)} = [\boldsymbol{\theta} \quad \boldsymbol{\psi}]\boldsymbol{Y_t} + \sigma\varepsilon_t \tag{6}$$

where

$$\boldsymbol{\theta} = [\theta_0, \theta_1, \ldots, \theta_p] \tag{7}$$
$$\boldsymbol{\psi} = [\psi_{r_B}^{(B)}, \ldots, \psi_{s_B}^{(B)}, \psi_{r_C}^{(C)}, \ldots, \psi_{s_C}^{(C)}, \ldots] \tag{8}$$
$$\boldsymbol{Y_t} = [1, y_{t-1}^{(A)}, \ldots, y_{t-p}^{(A)}, x_{r_B}^{(B)}, \ldots, x_{s_B}^{(B)}, x_{r_C}^{(C)}, \ldots, x_{s_C}^{(C)}, \ldots]^T \tag{9}$$

and $\{\varepsilon_t\}$ is an independent and identically distributed (i.i.d) sequence of random variables with 0 mean and unit variance, and $q = \sum_{WF=(B,C,\ldots)} (s_{WF} - r_{WF} + 1)$.

Let $\boldsymbol{\Theta} = (\boldsymbol{\theta}, \boldsymbol{\psi}, \sigma)$ be the set of parameters to be estimated. For Normally distributed errors, the Maximum Likelihood Estimator (MLE), $\widehat{\boldsymbol{\Theta}}_{MLE}$, is obtained by minimizing the negative log-likelihood function as follows:

$$\widehat{\boldsymbol{\Theta}}_{MLE} = \arg\min_{\boldsymbol{\Theta}} \quad -\log \mathcal{L}(\boldsymbol{\Theta}|\boldsymbol{\Omega_T}) \tag{10}$$

$$\text{where} \quad -\log \mathcal{L}(\boldsymbol{\Theta}|\boldsymbol{\Omega_T}) = \frac{n}{2}\log(2\pi\sigma^2) + \frac{1}{2\sigma^2}\sum_{i=1}^{n}\varepsilon_t^2 \tag{11}$$

$$\text{and} \quad \varepsilon_t = y_t^{(A)} - [\boldsymbol{\theta} \quad \boldsymbol{\psi}]\boldsymbol{Y_t} \tag{12}$$

and $\mathcal{L}$ is the likelihood function.

Two types of predictive density are considered, the censored Normal and the GLN. At time t, given the vector of estimated parameters $\widehat{\boldsymbol{\Theta}}_{MLE}$ and the set of observations $\boldsymbol{\Omega_t}$, the one-step ahead censored Normal density $\hat{f}_{t+1|t}$ is described by the estimated conditional expectation $\hat{\mu}_{t+1|t}$ and standard deviation $\hat{\sigma}$ of the Normal density so that $\hat{f}_{t+1|t}(y^{(A)}|\widehat{\boldsymbol{\Theta}}_{MLE}, \boldsymbol{\Omega_t}) = \mathcal{N}^{[0,1]}(\hat{\mu}_{t+1|t}, \hat{\sigma})$ where $\hat{\mu}_{t+1|t} = [\hat{\boldsymbol{\theta}} \quad \hat{\boldsymbol{\psi}}]\boldsymbol{Y_t}$.

In order to obtain the one-step ahead GLN density, additional steps are needed. First, the transformation given in (1) must be applied for estimating the vector of parameters $\widehat{\boldsymbol{\Theta}}_{MLE}$ in the transformed domain. Second, the one-step ahead predictive density in the transformed domain is obtained by following the formula (2-4). Last, the inverse GLN transformation presented in [24] is applied on a quantile per quantile basis for generating the GLN density in the original domain.

3.2 ARX-GARCH models

ARX-GARCH models are a popular extension of ARX models as they can relax the assumption of constant variance without data transformation. GARCH models were first introduced in Econometrics by [32]. A short review of meteorological applications of GARCH models is available in [11]. This class of model proposes to capture the dynamical structure of the conditional variance, jointly to that of the process conditional expectation $\mathbb{E}(y_t^{(A)}|\boldsymbol{\Omega_t}, \boldsymbol{\Theta})$. The conditional variance h_t^2 is modeled as an ARMA process for the squared errors ε_t^2. It was shown in a number of studies that a GARCH(1,1) structure is

in most cases appropriate to capture the temporal dynamics of h_t^2. The linear AR(p)-X(q)-GARCH(1,1) model with p autoregressive and q exogenous predictors is given by:

$$y_t^{(A)} = [\boldsymbol{\theta} \quad \boldsymbol{\psi}]\boldsymbol{Y_t} + h_t\varepsilon_t \tag{13}$$

$$h_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta h_{t-1}^2 \tag{14}$$

where $\{\varepsilon_t\}$ is an i.i.d sequence of random variables with 0 mean and unit variance. To ensure that the conditional variance is positive, we impose $\omega > 0$ and $\alpha, \beta \geq 0$.

Let $\boldsymbol{\Theta} = (\boldsymbol{\theta}, \boldsymbol{\psi}, \omega, \alpha, \beta)$ be the set of parameters to be estimated. For Normally distributed errors, $\widehat{\boldsymbol{\Theta}}_{MLE}$ is obtained by minimizing the negative log-likelihood function as follows:

$$\widehat{\boldsymbol{\Theta}}_{MLE} = \arg\min_{\boldsymbol{\Theta}} \quad -\log\mathcal{L}(\boldsymbol{\Theta}|\boldsymbol{\Omega_T}) \tag{15}$$

$$\text{where} \quad -\log\mathcal{L}(\boldsymbol{\Theta}|\boldsymbol{\Omega_T}) = \frac{n}{2}\log(2\pi\sigma^2) + \frac{1}{2h_t^2}\sum_{i=1}^{n}\varepsilon_t^2 \tag{16}$$

where ε_t is given by (12) and h_t^2 is given by (14). For the implementation of the model, analytical formula for the first and second order derivatives of the negative log-likelihood function are given in [33].

One-step ahead predictive densities are generated in a similar way as with ARX models, but for a single change. The conditional standard deviation $\hat{\sigma}$ becomes time-varying as follows:

$$\hat{\sigma} = h_t \tag{17}$$

$$\text{with} \quad h_t^2 = \hat{\omega} + \hat{\alpha}\varepsilon_{t-1}^2 + \hat{\beta}h_{t-1}^2 \tag{18}$$

3.3 TARX models

TARX models are the first regime-switching models considered in this study. They are piecewise linear, and the transitions between regimes are governed in a deterministic way by a lagged variable, and are hence observable. See [8] for a more detailed introduction to these models. The TAR($p_1, \ldots, p_R$)-X($q_1, \ldots, q_R$) model with R regimes, p_j autoregressive and q_j exogenous predictors in regime j, with $j = 1, \ldots, R$, is given by:

$$y_t^{(A)} = [\boldsymbol{\theta^{(j)}} \quad \boldsymbol{\psi^{(j)}}]\boldsymbol{Y_t} + \sigma^{(j)}\varepsilon_t \quad \text{if} \quad r_j < z_{t-d} \leq r_{j+1} \tag{19}$$

where

$$\boldsymbol{\theta} = [\theta_0^{(j)}, \theta_1^{(j)}, \ldots, \theta_p^{(j)}] \tag{20}$$

$$\boldsymbol{\psi} = [\psi_{r_B}^{(j,B)}, \ldots, \psi_{s_B}^{(j,B)}, \psi_{r_C}^{(j,C)}, \ldots, \psi_{s_C}^{(j,C)}, \ldots] \tag{21}$$

and $\{\varepsilon_t\}$ is an i.i.d sequence of random variables with 0 mean and unit variance, $\sigma^{(j)}$ the standard deviation in the regime j, z_{t-d} the lagged variable; $d \in N^+$ the delay parameter with usually $d \leq \max(p_1, \ldots, p_R)$, and r_j the threshold values separating the regimes. The regime-switching effect translates into the autoregressive and exogenous coefficients as well as the standard deviation of the error term being state-dependent. Applications of TAR models for forecasting wind power fluctuations can

be found in [9, 10] which alternatively use lagged observations of wind speed, wind direction or wind power for controlling transitions between regimes. A special class of TAR model is the Self-Exciting TAR (SETAR) model which corresponds to the case where the dependent variable is chosen as the lagged variable.

The major issue with TAR models is the joint determination of the delay d and thresholds $r_j, j = 1, \ldots, R$. In particular, the most spread technique for the determination of the r_j is based on the visual assessment of scatter plots of t-ratios (see [8]). In order to fill in the lack of consistency of such approach, an automated procedure for determining the number of regimes and threshold values of TAR models was recently proposed in [34]. It consists of detecting jumps in the values of the estimates of an arranged autoregression by using a recursive least squares (RLS) estimation method. This method can be extended to deal with exogenous predictors without complicating its procedure. Once the threshold values known, the parameters for a given regime can be estimated independently of the parameters of the other regimes by applying the formula given in (10-12) for each regime, and predictive densities can be generated as with ARX models.

3.4 MSARX models

MSARX models are the second type of regime-switching models in this study. Structurally, the major difference between MSARX and TARX models lays in the way the sequence of regimes is determined. With TAR models, this sequence is determined explicitly by a lagged variable, and the transitions between regimes are therefore discontinuous. With MSARX models, the sequence is assumed hidden and inferred directly from the data. More specifically, MSARX models assume that an unobservable Markov process governs the distribution of the observations [12]. This enables smooth transition between regimes.

The MSAR($p_1, \ldots, p_R$)-X($q_1, \ldots, q_R$) model with R regimes, p_j autoregressive and q_j exogenous predictors in regime j, with $j = 1, \ldots, R$, is given by:

$$y_t^{(A)} = [\boldsymbol{\theta}^{(\boldsymbol{z_t})} \quad \boldsymbol{\psi}^{(\boldsymbol{z_t})}]\boldsymbol{Y_t} + \sigma^{(z_t)}\varepsilon_t \tag{22}$$

where

$$\boldsymbol{\theta}^{(\boldsymbol{z})} = [\theta_0^{(z)}, \theta_1^{(z)}, \ldots, \theta_p^{(z)}], \quad z = 1, \ldots, R \tag{23}$$

$$\boldsymbol{\psi}^{(\boldsymbol{z})} = [\psi_{r_B}^{(z,B)}, \ldots, \psi_{s_B}^{(z,B)}, \psi_{r_C}^{(z,C)}, \ldots, \psi_{s_C}^{(z,C)}, \ldots], \quad z = 1, \ldots, R \tag{24}$$

and $\{\varepsilon_t\}$ is an i.i.d sequence of random variables with 0 mean and unit variance, $\{z_t\}$ follows a first order Markov chain with a finite and discrete number of states R and transition probability matrix $\boldsymbol{P}$ of elements $(p_{ij})_{i,j=1,\ldots,R}$:

$$p_{ij} = \Pr(z_t = j | z_{t-1} = i), \quad i, j = 1, \ldots, R \tag{25}$$

$$\sum_{j=1}^{R} p_{ij} = 1, \quad i = 1, \ldots, R \tag{26}$$

Similarly to TARX models, the autoregressive coefficients and standard deviation of the error term are state-dependent. Let $\boldsymbol{\Theta} = (\boldsymbol{\theta^{(1)}}, \ldots, \boldsymbol{\theta^{(R)}}, \boldsymbol{\psi^{(1)}}, \ldots, \boldsymbol{\psi^{(R)}}, \sigma_1, \ldots, \sigma_R, \boldsymbol{P})$ be the set of parameters to estimate. For Normally distributed errors in each regime, $\widehat{\boldsymbol{\Theta}}_{MLE}$ is obtained by minimizing the negative log-likelihood function as follows:

$$\widehat{\boldsymbol{\Theta}}_{MLE} = \arg\min_{\boldsymbol{\Theta}} \quad -\log \mathcal{L}(\boldsymbol{\Theta}|\boldsymbol{\Omega_T}) \tag{27}$$

$$\text{where} \quad \mathcal{L}(\boldsymbol{\Theta}|\boldsymbol{\Omega_T}) = \boldsymbol{\delta}(\prod_{t=1}^{n} \boldsymbol{P}\boldsymbol{D}_t)\mathbf{1}^T \tag{28}$$

$$\boldsymbol{\delta} = \mathbf{1}(\boldsymbol{I}_R - \boldsymbol{P} + \boldsymbol{U}_R)^{-1} \tag{29}$$

$$\boldsymbol{D}_t = \text{diag}(\eta(t,1), \ldots, \eta(t,R)) \tag{30}$$

$$\eta(t,i) = \frac{1}{\sigma^{(i)}}\phi\left(\frac{y_t^{(A)} - [\boldsymbol{\theta^{(i)}} \quad \boldsymbol{\psi^{(i)}}]\boldsymbol{Y_t}}{\sigma^{(i)}}\right), \quad i = 1, \ldots, R \tag{31}$$

where $\boldsymbol{\delta}$ is the stationary distribution of the Markov chain, $\mathbf{1}$ is a unit vector of size R, $\boldsymbol{I}_R$ and $\boldsymbol{U}_R$ Identity and Unity matrices of size $R \times R$, $\boldsymbol{D}_t$ a diagonal matrix and ϕ the probability density function of the Normal distribution. Practical solutions for the implementation of MSARX models are given in [35].

With MSARX models, predictive densities take the form of mixture of densities [12, 35]. For the case where the errors are Normally distributed in each regime, the resulting predictive density is a mixture of R Normal densities that is censored in 0 and 1 later on. At time t, given the vector of estimated parameters $\widehat{\boldsymbol{\Theta}}_{MLE}$ and the set of observations $\boldsymbol{\Omega_t}$, the one-step ahead density can be obtained as follows:

$$\hat{f}^{[0,1]}_{t+1|t}(y^{(A)}|\widehat{\boldsymbol{\Theta}}_{MLE}, \boldsymbol{\Omega_t}) = \sum_{k=1}^{R} \xi_t^{(k)} \phi([\hat{\boldsymbol{\theta}}^{(\boldsymbol{k})} \quad \hat{\boldsymbol{\psi}}^{(\boldsymbol{k})}]\boldsymbol{Y_t}, \hat{\sigma}^{(k)}) \tag{32}$$

$$\text{where} \quad \xi_t = \frac{\boldsymbol{\delta}(\prod_{i=1}^{t} \hat{\boldsymbol{P}}\boldsymbol{D}_i)\hat{\boldsymbol{P}}}{\boldsymbol{\delta}(\prod_{i=1}^{t} \hat{\boldsymbol{P}}\boldsymbol{D}_i)\mathbf{1}^T} \tag{33}$$

and $\xi_t^{(k)}$ is the kth element of the vector of filtered probabilities ξ_t at time t.

In order to obtain predictive densities in a GLN fashion, we can apply the same 3-step procedure as for ARX models that is: (1) data transformation in order to work in the transformed domain, (2) generation of mixture of Normal densities in the transformed domain, and (3) inverse transformation of a set of quantiles of this mixture of Normal densities.

3.5 Estimation procedure

As mentioned in section 2, the data we selected cover the period from July 10, 2008 to 27 March, 2009. This corresponds to about 25000 observations, for each of the three time series (i.e., Carnsore, Ballywater, Richfield). Focus is placed on predicting the wind power generation at the Carnsore wind farm. The first 15000 observations are used for fitting the models. The following 5000 observations

are used for performing a one-fold cross-validation and determining the optimal parametrisation of each model. The last 5000 observations, corresponding to about 63 days, are kept for forecast evaluation.

Cross-validation is jointly performed on the structure of the model (i.e., selection of the optimal AR lags from 1 up to 8, and X lags from 1 to 5, number of regimes R) and a set of values for the shape parameter ν of the GLN distribution (from 0.1 to 3.1 with steps of 0.1). Because of that, and because the likelihood function is unbounded, neither the respective goodness-of-fit nor the predictive power of the models can be compared with respect to likelihood based scores. Instead, the cross-validation procedure is performed by minimizing the Continuous Ranked Probability Score (CRPS) for one-step ahead density forecasts. The CRPS quantifies the accuracy of conditional density forecasts based on two principles: calibration (i.e., the relative position of a forecast with respect to the observed value) and sharpness (i.e., the concentration of the predictive distribution around the observed value) [36].

For each class of models presented in this section, we estimated four different models with: (N) a censored Normal distribution, (X-N) a censored Normal distribution and exogenous regressors, (GLN) a GLN distribution, (GLN-X) a GLN distribution and exogenous regressors. Four different lagged variables z_{t-d} were tried for controlling the regime sequence of TAR models, namely $y_{t-d}^{(Carn)}$, $x_{t-d}^{(Rich)}$, and their respective first order differentiated series. For all four TAR models, $y_{t-1}^{(Carn)}$ was selected as the best lagged variable. The final parametrisation of each model is summarized in Table 1 along with the total number of parameters in order to appreciate their respective cost-complexity. Several observations can be drawn from these results. First, none of the final models includes off-site information from Ballywater. This means that wind power fluctuations from Ballywater are not informative for improving the predictability of wind power fluctuations at Carnsore for the proposed models. On the opposite, all models include two lagged measurements from Richfield, concurring with the early observations in section 2 which indicated that wind power fluctuations at Richfield led those at Carnsore up to 30 minutes ahead. Second, the use of the GLN distribution leads to a reduction of the autoregressive order for AR and MSAR models, while it decreases the optimal regimes number, from four to three, for TAR models. More generally, the use of the GLN distribution yields a reduction in the cost complexity (i.e., the number of parameters to be estimated) of all models but AR-GARCH.

4 Experimental results and forecast evaluation

In this section, we evaluate the predictive performances of the four classes of models presented in the previous section, namely ARX, ARX-GARCH, TARX and MSARX models. The evaluation consists of measuring the accuracy of one-step ahead point and density forecasts, as well as the overall reliability of these forecasts.

4.1 Point forecasts

Electric utilities have a long tradition of using point or deterministic forecasts of wind power [2,7]. In this study, point forecast accuracy is evaluated with respect to the Normalized Mean Absolute Error (NMAE). There is an inverse relationship between point forecast accuracy and the NMAE score: the lower the NMAE, the better. Following [37], we use the median of the predictive densities as the optimal point

TABLE 1: *Summary of model parametrisation after cross-validation. This includes the lagged variables $y_{t-i}^{(Carn)}$, the lagged exogenous variables $x_{t-i}^{(Rich)}$, the number of regimes and total number of parameters.*

Model	$y_{t-i}^{(Carn)}$	$x_{t-i}^{(Rich)}$	Number of regimes	Total number of parameters
AR-N	1:7	-	1	9
AR-X-N	1:7	1:2	1	11
AR-GLN	1:5	-	1	8
AR-X-GLN	1:5	1:2	1	10
AR-GARCH-N	1:5	-	1	9
AR-X-GARCH-N	1:5	1:2	1	11
AR-GARCH-GLN	1:5	-	1	10
AR-X-GARCH-GLN	1:5	1:2	1	12
TAR-N	(1:6, 1:6, 1:5, 1:6)	-	4	31
TAR-X-N	(1:5, 1:5, 1:5, 1:5)	(1:2, 1:2, 1:2, 1:2)	4	36
TAR-GLN	(1:6, 1:3, 1:6)	-	3	22
TAR-X-GLN	(1:6, 1:3, 1:6)	(1:2, 1:2, 1:2)	3	28
MSAR-N	(1:5, 1:5)	-	2	16
MSAR-X-N	(1:5, 1:5)	(1:2, 1:2)	2	20
MSAR-GLN	(1:3, 1:3)	-	2	13
MSAR-X-GLN	(1:3, 1:3)	(1:2, 1:2)	2	17

forecast, due to the nature of the NMAE which is based on a symmetric piecewise linear scoring rule. All models are benchmarked against Persistence since it is one of the most competitive benchmarks for such short lead times. Persistence usually outperforms other common benchmarks such as Climatology, Moving average or Constant forecast (see for instance [19, 24]) which are not included here. It is an Autoregressive model of order 1 with no intercept term and its coefficient value equal to 1. Point forecast results are given in Table 2. It is interesting to note that not all models outperform Persistence and that even the largest improvement does not exceed 3%. Overall, MSARX and ARX-GARCH with a GLN distribution give the best results. When considering each class of models independently of the others, we observe two trends. The first one concerns AR and TAR models for which the use of either off-site information or the GLN distribution yields substantial gains in wind power predictability. These gains are further improved by using both. The second trend regards AR-GARCH and MSAR models for which the use of the GLN distribution alone, without off-site information, leads to negligible gains whereas the opposite (i.e., no GLN distribution and off-site information) leads to appreciable gains.

4.2 Density Forecasts

Forecasts of any quantity contain an inherent part of uncertainty. Supplying information on this uncertainty is paramount for developing efficient decision-making strategies, as shown in the context of wind power trading by [38]. Here, information on this uncertainty is provided in the form of full predictive densities of wind power, for all four classes of models. The accuracy of these densities is assessed with respect to the Normalized CRPS (NCRPS). This score is a generalization of the NMAE score for proba-

TABLE 2: *One-step ahead forecast performances. Results are given in terms of Normalized Mean Absolute Error (NMAE) and Normalized Continuous Ranked Probability Score (NCRPS). Point (respectively probabilistic) forecast improvements are given with respect to Persistence (respectively a AR-N model).*

Model	NMAE	NCRPS
Persistence	3.77	-
AR-N	3.87 (-2.7%)	3.38
AR-X-N	3.80 (-0.7%)	3.28 (2.9%)
AR-GLN	3.77 (0.2%)	2.99 (11.7%)
AR-X-GLN	3.70 (1.9%)	2.90 (14.1%)
AR-GARCH-N	3.76 (0.4%)	3.04 (10.2%)
AR-X-GARCH-N	3.73 (1.1%)	2.97 (12.1%)
AR-GARCH-GLN	3.76 (0.3%)	2.82 (16.8%)
AR-X-GARCH-GLN	**3.67 (2.8%)**	2.75 (18.7%)
TAR-N	3.84 (-1.9%)	3.05 (9.8%)
TAR-X-N	3.73 (1.0%)	2.96 (12.4%)
TAR-GLN	3.77 (0.1%)	2.88 (16.6%)
TAR-X-GLN	3.70 (1.9%)	2.81(16.9%)
MSAR-N	3.77 (0.1%)	3.01 (11.1%)
MSAR-X-N	3.67 (2.7%)	2.93 (13.4%)
MSAR-GLN	3.76 (0.3%)	2.79 (17.7%)
MSAR-X-GLN	**3.67 (2.8%)**	**2.71 (19.8%)**

bilistic forecasts and measures the difference between the observed cumulative distribution functions and those predicted [36]. It can be interpreted in a similar way as the NMAE, meaning the lower the NCRPS the better. All models are benchmarked against an AR model with a censored Normal distribution (AR-N). Results for one-step ahead densities are reported in Table 2. The best result is given by the MSAR model with off-site information and the use of the GLN distribution (MSAR-X-GLN), with a relative improvement of almost 20% when compared to an AR-N model. In addition, we observe a common trend across all four classes of models when considered independently of the others. Their ranking is dominated by models including both off-site observations and the GLN distribution (X-GLN), then come models specified with the GLN distribution and no off-site predictors (GLN), then models with off-site predictors but without GLN distribution (X), and finally models with neither the GLN distribution nor off-site predictor (N).

Figures 7 and 8 give an illustration of these predictive densities over two arbitrary examples of 100 observations each. Densities are depicted as prediction intervals with nominal coverage rates ranging from 10 to 90%. Point forecasts corresponding to the median of these densities are also presented. Prediction intervals generated with the best two models (i.e., ARX-GARCH-GLN and MSAR-X-GLN) are compared. In particular, in Figure 7, large forecast errors result in wider prediction intervals for the ARX-GARCH-GLN model than for the MSAR-X-GLN model.

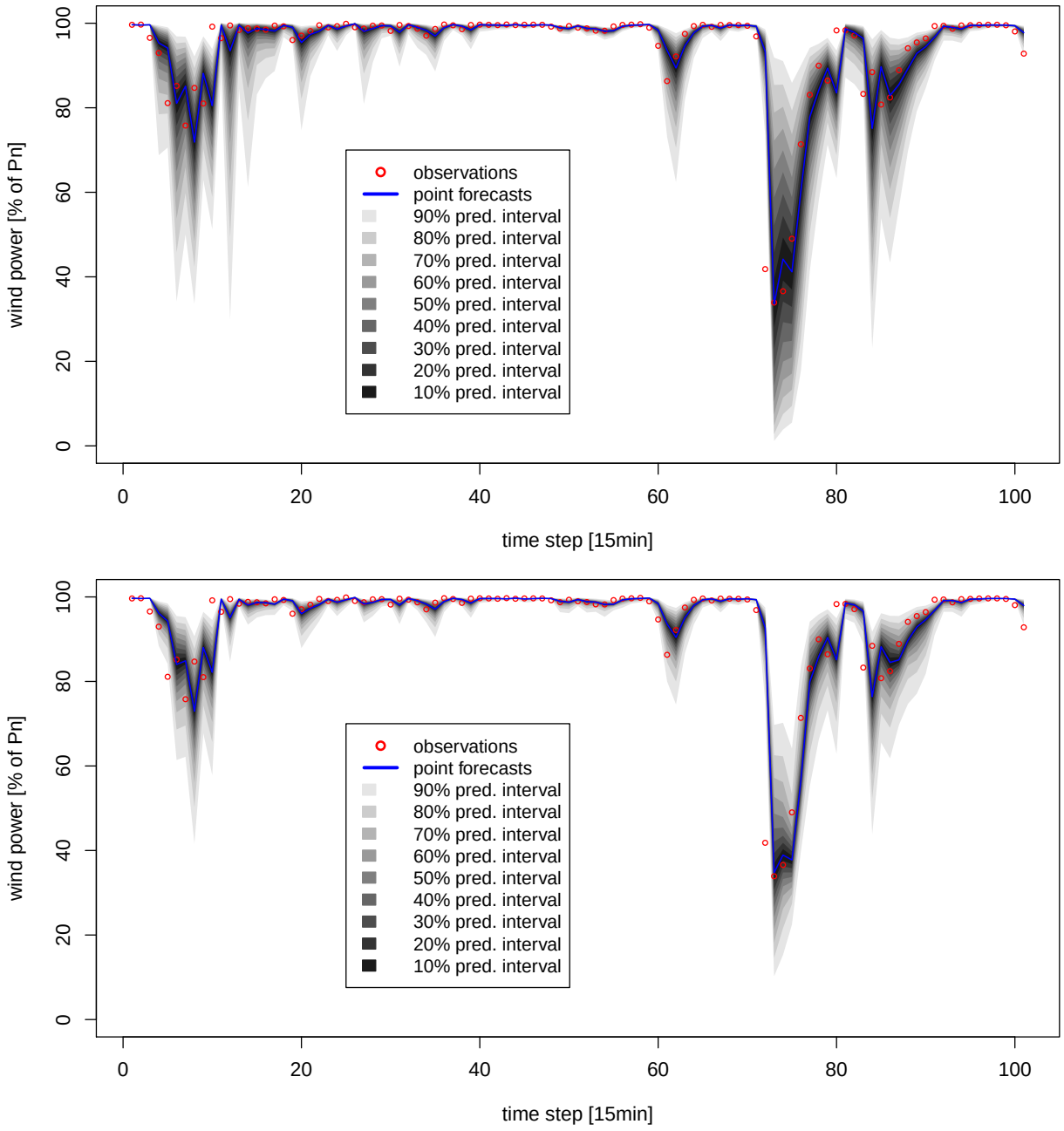


FIGURE 7: *Example 1. Time series of normalized wind power generation at Carnsore and one-step ahead point forecasts and prediction intervals with nominal coverage from 10 to 90%. ARX-GARCH-GLN model (Top panel), MSAR-X-GLN model (Bottom panel).*

4.3 Forecast reliability

The CRPS is a global score that averages the predictive accuracy of conditional densities based on their calibration and associated sharpness. However, it is not informative on the behavior of these densities in terms of probabilistic reliability. Reliability measures how well the predicted probabilities of an event correspond to their observed frequencies. For instance, one may want to measure the proportion of observations actually lower than the 5^{th} percent quantile or larger then the 95^{th} percent quantile for

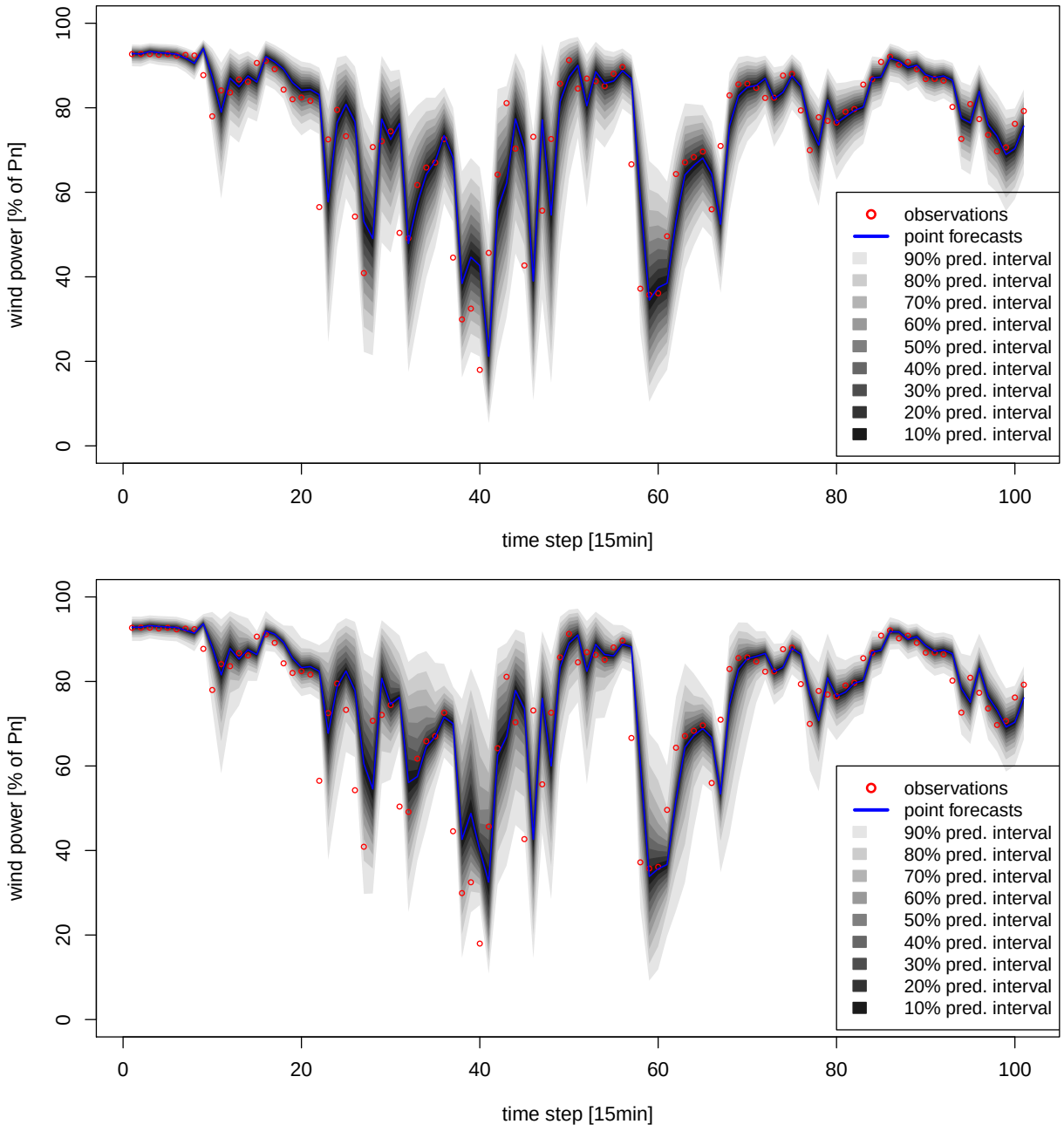


FIGURE 8: *Example 2. Time series of normalized wind power generation at Carnsore and one-step ahead point forecasts and prediction intervals with nominal coverage from 10 to 90%. ARX-GARCH-GLN (Top panel) model, MSARX-GLN model (Bottom panel).*

evaluating the ability of the predictive density tails in predicting extreme or rare events. In this study, the reliability of the predictive densities of wind power is evaluated with four reliability diagrams as shown in Figure 9. These diagrams are generated for each of the four classes of models by comparing the nominal (i.e., theoretical) proportions of a set of quantiles with the observed proportions of the same set. Here, we used 19 quantiles, from the 5^{th} percent quantile to the 95^{th} percent quantile with a step of 5^{th} percent. The best reliability is given by the model whose diagram is closer to the ideal case in Figure 9, that is the MSAR-X-GLN model.

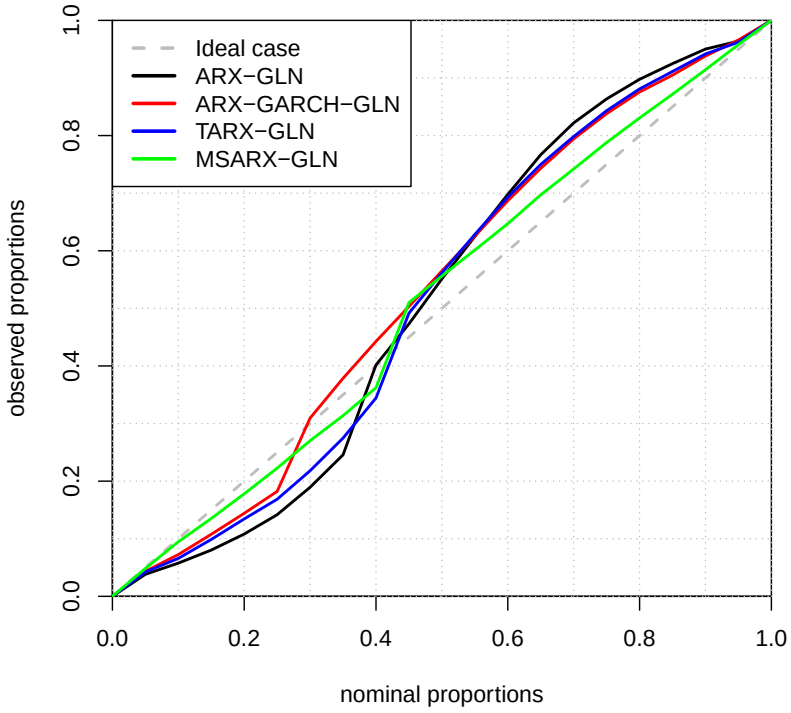


FIGURE 9: *Reliability diagram of predictive densities of wind power.*

4.4 Discussion

The results presented in this section highlight a number of interesting points but also raise a few questions. Let us summarize some of our comments herebelow:

1. In the Irish case study chosen for this work, the variability of wind power fluctuations can be considered as extreme. For instance, the NMAE value of the Persistence is about 50% larger than that at the Horns Rev 1 wind farm where wind power fluctuations are known to be characterized by a high variability [9, 11]. In that sense, this case study offered a difficult test to all models, enhancing the impact of the results obtained.

2. Irrespectively of the availability of off-site measurements, the use of the GLN distribution is recommended for very short-term forecasts. In particular, it enables an improved modeling of the heteroscedastic behavior of wind power time series, which translates to substantial gains in predictability even for models already explicitly accounting for heteroscedasticity in their formulation (i.e., MSARX and ARX-GARCH) . However, it calls for further research on its potential for multi-step ahead forecasts. This issue was not addressed here but will be investigated in the future. In addition, focus should be placed on developing a more consistent framework than cross-validation for estimating the optimal value of the shape parameter ν of the GLN distribution. For instance, the estimation of ν could be performed jointly with the estimation of the model via the Expectation-Maximization (EM) algorithm (see [39]).

3. The results obtained with TARX models are relatively disappointing, particularly, when analyzed from a perspective including the cost complexity of these models and the level of expertise re-

quired to tune them. It is also worth noting that TAR models are outperformed by linear in mean ARX-GARCH models. It could be expected that TARX models perform much better for point forecasting especially in combination with the GLN distribution since the introduction of regimes via the thresholds could reduce the strong influence of the probability masses in $\gamma(\epsilon, \nu)$ and $\gamma(1-\epsilon, \nu)$ on the autoregressive coefficient estimates.

4. Density forecasts of wind power generated with Markov-Switching models have superior calibration and sharpness when compared to those generated with other models in this study. Beyond this result, it is important to stress the underlying assumption in MSAR models which leads to such result, that is the existence of an unobservable regime sequence which governs the wind power generation. As of today, our knowledge is limited and we can only assume that the estimated regime sequence is linked to some weather regime. Therefore, it would be useful to investigate the use of data (e.g., quick scan satellite images, weather radar images) that can describe weather conditions over large spatial areas and high temporal resolutions for improving the characterization of this regime sequence.

5 Conclusion

This work considered the probabilistic forecasting of wind power generation from a single wind farm, over very short lead times (i.e., 15 minutes). Realistic assumptions were made regarding the online availability of wind data in the current wind power context, meaning that neither wind measurements nor wind forecasts are available for the temporal resolution of interest. The sole data that are used consist of on-site observations of wind power generation, along with corresponding observations from the two nearest wind farms located in a radius of 50 km. Focus is placed on the most recent approaches from the wind power forecasting literature, including regime-switching models, the use of off-site predictors and a new predictive distribution. The predictive performances of these approaches and their associated models are compared against one another to assess their respective merits. Eventually, combinations of these approaches are proposed and proved to generate improved wind power forecasts.

Through an application with three wind farms in Ireland, we show that regime-switching models for which the sequence of regime is unobservable (i.e., Markov-Switching) generate more accurate point forecasts, better calibrated and sharper conditional densities, than single regime or other regime-switching models for which the regimes are observable. Furthermore, gains in wind power predictability can be increased by taking advantage of off-site information when available or using a more appropriate predictive distribution such as the GLN distribution, as introduced in [24]. The highest gains were obtained by using simultaneously off-site observation and the GLN distribution.

The superior predictive power of Markov-Switching models is interesting in two aspects. First, because this type of models is rather generic and thus non site-dependent, requiring very little expert knowledge to be tuned. It confirms the potential shown for offshore applications [9, 11]. Second, because Markov-Switching models assume the existence of an unobservable regime sequence that can be interpreted as a hidden weather regime. This indicates that substantial gains in wind power predictability could be obtained by integrating more meteorological data at high spatio-temporal resolution such as satellite images, weather radar images, or meteorological forecasts. In particular, this a prerequisite for extending

regime-switching approaches to multi-step ahead wind power forecasts.

Acknowledgments

This work was partly supported by the European Commission under the SafeWind project (ENK7-CT2008-213740) and by the Danish Public Service Obligation (PSO) program through the project "Radar@Sea" (PSO-2009-1-0226). Eirgrid is acknowledged for providing the wind power data from the Carnsore, Richfield and Ballywater wind farms. The authors also express their gratitude to Sven Creutz Thomsen for preparing the data.

References

[1] GE Energy. Western wind and solar integration study. Technical report, National Renewable Energy Laboratory (NREL), Golden, CO., 2010.

[2] L. Jones and C. Clark. Wind integration - A survey of global views of grid operators. In *Proceedings of the 10^{th} International Workshop on Large-Scale Integration of Wind Power into Power Systems, Aarhus, Denmark*, 2011.

[3] H. Holttinen, A.G. Orths, P. Eriksen, J. Hidalgo, A. Estanqueiro, F. Groome, Y. Coughlan, H. Neumann, B. Lange, F. Hulle, and I. Dudurych. Currents of change. *IEEE Power and Energy Magazine*, 9:47–59, 2011.

[4] V. Akhmatov, C. Rasmussen, P. B. Eriksen, and J. Pedersen. Technical aspects of status and expected future trends for wind power in Denmark. *Wind Energy*, 10:31–49, 2007.

[5] J.R. Kristoffersen and P. Christiansen. Horns Rev offshore wind farm: its main controller and remote control system. *Wind Engineering*, 27:351–359, 2003.

[6] V. Akhmatov. Influence of wind direction on intense power fluctuations in large offshore wind farms in the North Sea. *Wind Engineering*, 31:59–64, 2007.

[7] G. Giebel, R. Brownsword, G. Kariniotakis, M. Denhard, and C. Draxl. The state-of-the-art in short-term prediction of wind power: A literature overview. Technical report, ANEMOS.plus, 2011.

[8] H. Tong. *Non-linear time series: a dynamical system approach*. Oxford University Press, 1990.

[9] P. Pinson, L.E.A. Christensen, H. Madsen, P.E. Sørensen, M.H. Donovan, and L.E. Jensen. Regime-switching modelling of the fluctuations of offshore wind generation. *Journal of Wind Engineering and Industrial Aerodynamics*, 96:2327–2347, 2008.

[10] C. Gallego, P. Pinson, H. Madsen, A. Costa, and A. Cuerva. Influence of local wind speed and direction on wind power dynamics - Application to offshore very short-term forecasting. *Applied Energy*, 88:4087–4096, 2011.

[11] P-J. Trombe, P. Pinson, and H. Madsen. A general probabilistic forecasting framework for offshore wind power fluctuations. *Energies*, 5:621–657, 2012.

[12] S. Fruhwirth-Schnatter. *Finite mixture and Markov-Switching models*. Springer, 2006.

[13] M.C. Alexiadis, P.S. Dokopoulos, and H.S. Sahsamanoglou. Wind speed and power forecasting based on spatial correlation models. *IEEE Transactions on Energy Conversion*, 14:836–842, 1999.

[14] I.G. Damousis, M.C. Alexiadis, J.B. Theocharis, and P.S. Dokopoulos. A fuzzy model for wind speed prediction and power generation in wind parks using spatial correlation. *IEEE Transactions on Energy Conversion*, 19:352–361, 2004.

[15] T. Gneiting, K. Larson, K. Westrick, M.G. Genton, and E. Aldrich. Calibrated probabilistic forecasting at the Stateline wind energy center: The regime-switching space-time method. *Journal of the American Statistical Association*, 101:968–979, 2006.

[16] K.A. Larson and K. Westrick. Short-term wind forecasting using off-site observations. *Wind Energy*, 9:55–62, 2006.

[17] A.S. Hering and M.G. Genton. Powering up with space-time wind forecasting. *Journal of the American Statistical Association*, 105:92–104, 2010.

[18] J. Tastu, P. Pinson, and H. Madsen. Multivariate conditional parametric models for a spatio-temporal analysis of short-term wind power forecast errors. In *Proceedings of the European Wind Energy Conference, Warsaw, Poland*, 2010.

[19] A. Lau. Probabilistic wind power forecasts: from aggregated approach to spatio-temporal models. *PhD Thesis, University of Oxford*, 2011.

[20] J. Tastu, P. Pinson, E. Kotwa, H. Madsen, and H.A. Nielsen. Spatio-temporal analysis and modeling of short-term wind power forecast errors. *Wind Energy*, 14:43–60, 2011.

[21] R. Girard and D. Allard. Spatio-temporal propagation of wind power prediction errors. *Wind Energy*, 2012. In Press.

[22] M. Lange. On the uncertainty of wind power predictions - analysis of the forecast accuracy and statistical distribution of errors. *Journal of Solar Energy Engineering*, 127:177, 2005.

[23] A. Lau and P. McSharry. Approaches for multi-step density forecasts with application to aggregated wind power. *The Annals of Applied Statistics*, 4:1311–1341, 2010.

[24] P. Pinson. Very short-term probabilistic forecasting of wind power with Generalized Logit-Normal distributions. *Journal of the Royal Statistical Society, Series C*, 61:555–576, 2012.

[25] T. Gneiting. Editorial: Probabilistic forecasting. *Journal of the Royal Statistical Society*, 171:319–321, 2008.

[26] Met Eireann, the Irish National Meteorological Service. [http://www.met.ie].

[27] H. Madsen, P. Pinson, T.S Nielsen, H.Aa. Nielsen, and G. Kariniotakis. Standardizing the performance evaluation of short-term wind power prediction models. *Wind Engineering*, 29:475–489, 2005.

[28] G.E.P. Box and D.R. Cox. An analysis of transformations. *Journal of the Royal Statistical Society. Series B*, 26:211–252, 1964.

[29] E. Lesaffre, D. Rizopoulos, and R. Tsonaka. The logistic transform for bounded outcome scores. *Biostatistics*, 8:72–85, 2007.

[30] H. Madsen. *Time series analysis*. CRC Press, 2008.

[31] H.A. Nielsen, P. Pinson, T.S. Nielsen, L.E. Christiansen, H. Madsen, G. Giebel, J. Badger, X.G. Larsén, H.V. Ravn, J. Tøfting, et al. Intelligent wind power prediction systems: Final report. Technical report, Informatics and Mathematical Modelling, Technical University of Denmark, DTU, 2007.

[32] T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31:307–327, 1986.

[33] G. Fiorentini, G. Calzolari, and L. Panattoni. Analytic derivatives and the computation of GARCH estimates. *Journal of Applied Econometrics*, 11:399–417, 1998.

[34] M.A. Bermejo, D. Peña, and I. Sánchez. Identification of TAR models using recursive estimation. *Journal of Forecasting*, 30:31–50, 2011.

[35] W. Zucchini and I.L. MacDonald. *Hidden Markov models for time series: An introduction using R*. Chapman & Hall/CRC, 2009.

[36] T. Gneiting, F. Balabdaoui, and Raftery A.E. Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society B*, 69:243–268, 2007.

[37] T. Gneiting. Quantiles as optimal point forecasts. *International Journal of Forecasting*, 27:197–207, 2011.

[38] P. Pinson, C. Chevallier, and G. Kariniotakis. Trading wind generation with short-term probabilistic forecasts of wind power. *IEEE Transactions on Power Systems*, 22:1148–1156, 2007.

[39] A.P. Dempster, N.M. Laird, and D.B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 1–38, 1977.

Short-term prediction of wind energy using Threshold Autoregressive models

Bermejo, M.A. Sánchez, I.

Department of Statistics, UC3M

January 24, 2012

1 Introduction

The main goal in this paper is the calculation of wind power short-term predictions (until 6 hours) by using univariate time series models. The short-term predictions are necessary for the implementation of wind energy in the intraday energy market, and for management operations in wind farms.

There are an extense literature about forecast methods in wind power. Costa et al. (2008) and Giebel et al. (2011) give a review of the different methods proposed. In this paper, we focus in the calculation of predictions by using statistical models. Especially, we focus in a class of regime-switching models, the self-exciting threshold autoregressive model (SETAR). Figure 1 displays an example of wind power generation. The figure shows the possible existence of different regimes in different levels of wind power.

The SETAR models allow the existence of different linear regimes. The process changes among the different regimes according to an observed variable, called threshold variable. In this paper, the increment of the wind power ($\Delta p_{t-1} = p_{t-1} - p_{t-2}$) is used as threshold variable. The existence of different regimes caused by Δp_{t-1} implies a different performance when the wind power is increasing and decreasing. This effect happens due to the inertia of the mechanical component of the wind generator (see, e.g., Burton et al., 2001).

The identification of the SETAR model is done by using the automatic tool proposed in Bermejo et al. (2011).

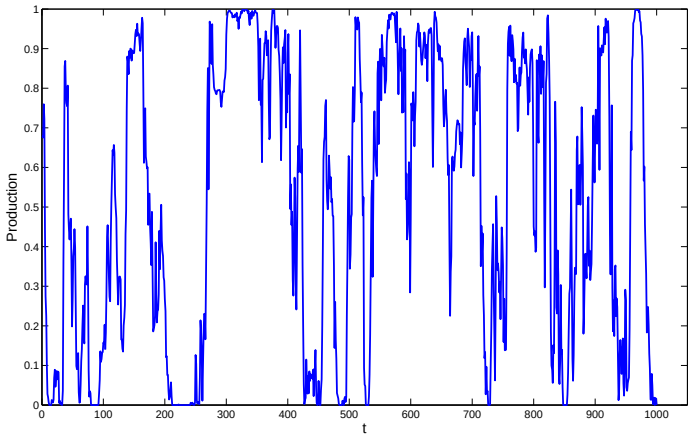


Figure 1: Hourly wind power generation.

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

$$e_{t+h|t} = p_{t+h} - \widehat{p}_{t+h|t}, \tag{1}$$

and its first conditional moment will be

$$m^{(0)}_{t+h|t} = \mathrm{E}_t \left[e_{t+h|t} \right]. \tag{2}$$

Finally, the conditional central moments of the forecast error $e_{t+h|t}$ will be

$$m^{(j)}_{t+h|t} = \mathrm{E}_t \left[\left(e_{t+h|t} - m^{(0)}_{t+h|t} \right)^j \right]; \text{ with } j = 2, 3, ..., M. \tag{3}$$

2 Self-Exciting Threshold Autoregressive models

A time series p_t is a SETAR(k;p,d) process if it follows

$$p_t = \phi_0^{(j)} + \sum_{i=1}^{p} \phi_i^{(j)} p_{t-i} + e_t^{(j)}; \text{ if } r_{j-1} \leq y_{t-d} < r_j, \tag{4}$$

where $j = 1, \ldots, k$. The integer k is the number of regimes, y_{t-d} is the threshold variable and the values of the thresholds are $-\infty = r_0 < r_1 < \ldots < r_k = \infty$; d is called the delay parameter. In each

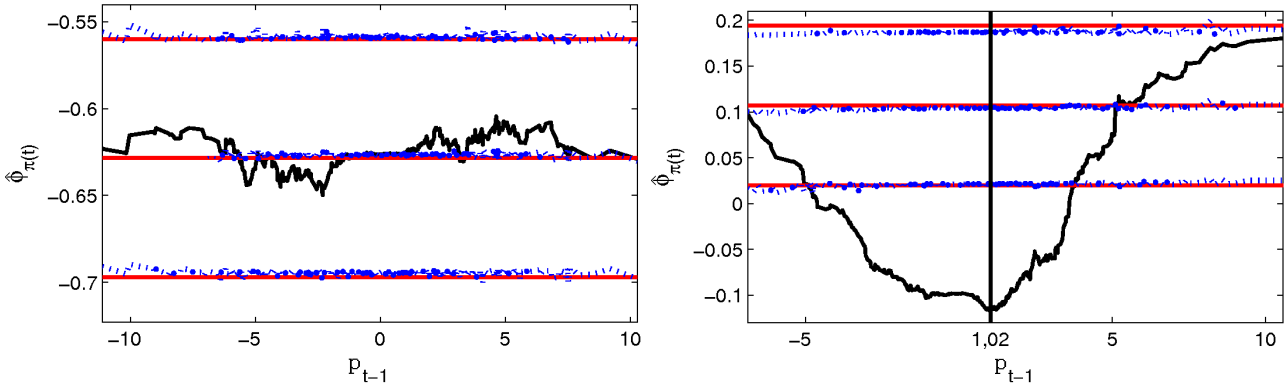
regime, $e_t^{(j)}$ is a sequence of independent and identically distributed random variables with zero mean and finite and constant standard deviation $\sigma^{(j)}$. The SETAR models were introduced by Tong (1978, 1983) and Tong and Lim (1980).

2.1 Identification of SETAR models

There are two main approaches to identify a SETAR model. The first is based on likelihood ratio (LR) tests. Chan (1990) and Chan and Tong (1990) developed the null distribution of the LR test using a Gaussian process and found it to be non-standard. Hansen (1999) used asymptotic and bootstrap distributions to overcome this problem. If the thresholds $r_1, \ldots, r_{k-1}$ were known, LR tests would supply the most powerful tests. However, that is not the case in a practical situation. In practice, the threshold is a nuisance parameter which is not identified under the null hypothesis. This problem has a negative impact on the efficiency of the procedures. To circumvent this problem, LR tests need to assume certain ranges of possible thresholds. As a result, LR tests need both intensive computational methods and non-standard reference distributions.

The second main approach to detecting a SETAR model is by means of portmanteau tests based on the predictive residuals of some arranged autoregressions. If the model is linear, the sequence of predictive residuals of the arranged autoregression have known properties. Petruccelli and Davis (1986) proposed a CUSUM-type test using these predictive residuals that is sensitive to the presence of a SETAR structure. Tsay (1989) considered a variant of this idea that is based on a standard F test. Recently, Bermejo et al. (2011) developed a graphical procedure based on recursive and time-varying estimation of the parameters. The proposed procedure has a superior identification performance than previous proposals.

In this paper, the automatic procedure Aut-ARLS proposed by Bermejo et al. (2011) is used to detect SETAR structure in wind power data. The main idea of the Aut-ARLS is the time-varying recursive estimation of the parameters of an arranged autoregressive model. Then, if the observations are arranged according the variable y_{t-d} and the sequence of estimates tend to show a structural change around the value r, a SETAR model with threshold variable y_{t-d} and threshold value r is identified. Figure 2 shows an example of the performance of the method. Figure 2a displays an example when the data is generated from a linear AR model. Figure 2b shows an example when the data is generated from a SETAR model with threshold variable y_{t-1} and threshold value 0.



(a) Linear model

(b) SETAR model

Figure 2: Example of Aut-ARLS method. Sequence of estimates of an arranged autoregressive model, which data is arranged according to p_{t-1}.

2.2 Forecasting with SETAR models

The calculation of predictions by using SETAR models is apparently easier than the calculation by using another nonlinear models, because SETAR models is formed by several linear models. This advantage is true when the forecast horizon h is lower than the delay parameter d, because the estimation of the one-step-ahead prediction will be calculated by means of

$$\hat{p}_{t+h|t} = \hat{\phi}_0^{(j)} + \sum_{i=1}^{p} \hat{\phi}_i^{(j)} p_{t-i}, \tag{5}$$

where j is the regime which the observation p_{t+h-d} belongs. That is, the regime in which the process will be is known and the predictions can be calculated by using a linear AR model. The problem arise when $h > d$, because the threshold variable y_{t+h-d} will be unobserved, then we do not know the regime in which the process will be. For this reason, the uncertainty about regimes must be taken into consideration when the predictions are calculated.

Tong (1990) proposed a numerical integration method to estimate the conditional mean. The numerical integration can be complex with SETAR models with several regimes. Clements and Smith (1997) made a review about the ability of different methods to calculate predictions by using SETAR models. They showed that, in general, the best results are obtained by using a Monte Carlo method. This method allows to calculate predictions of complex SETAR models and to avoid the assumption of normality.

3 SETAR structure in wind power data

The identification of the SETAR structure needs to choose a set of possible threshold variables. Firstly, the existence of possible differences in the process when the wind power is increasing or decreasing is studied. Then, the procedure Aut-ARLS is applied on wind power data arranged according by Δp_{t-1}. Figure 6 displays an obtained example in Rokas wind farm. The threshold value estimated is 0, that is, the performance of wind power is different when the process increase or decrease.

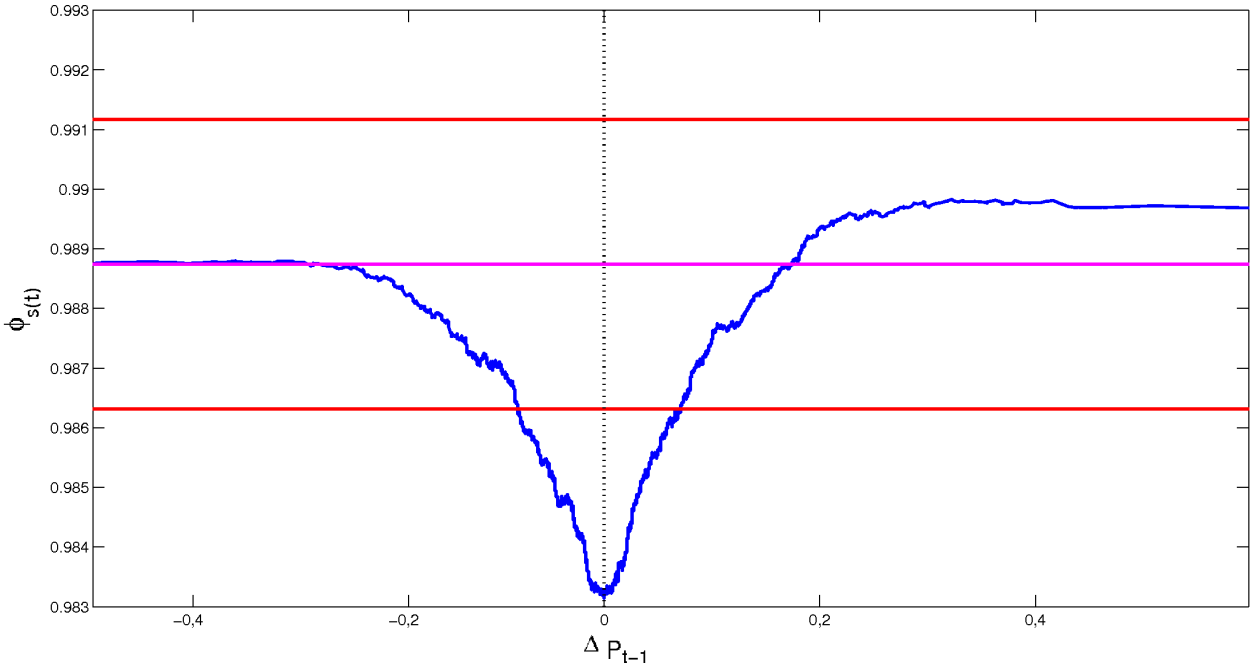
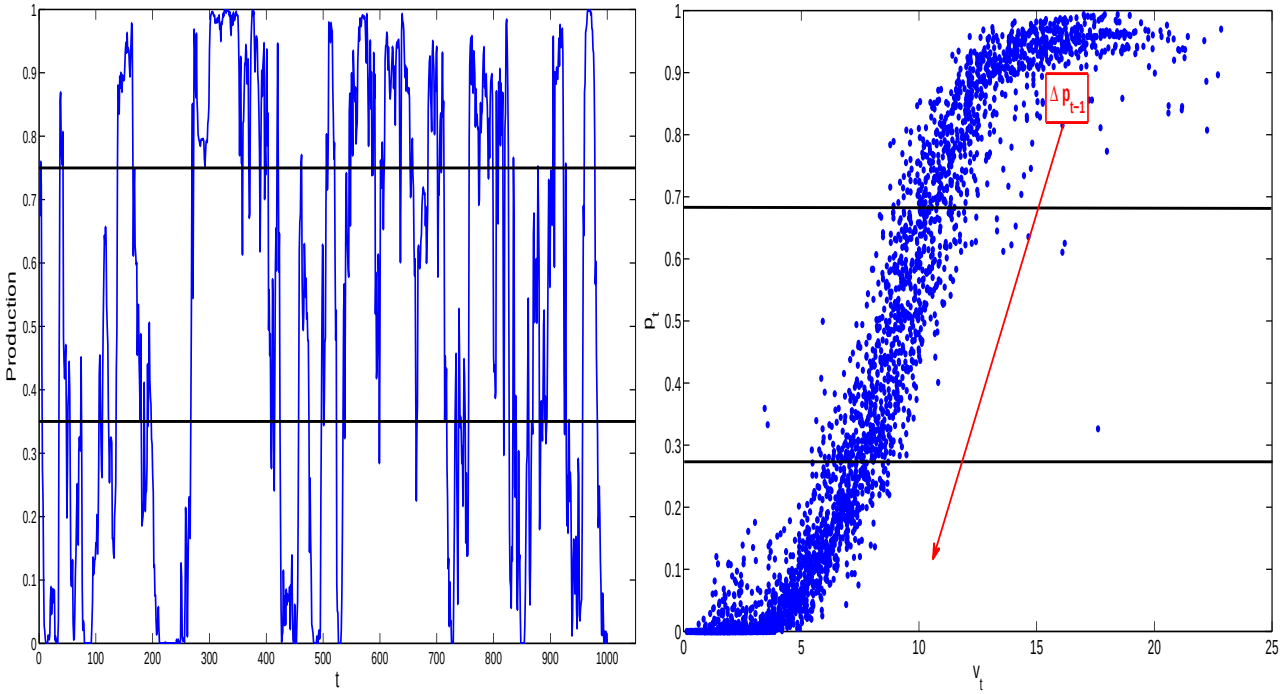


Figure 3: Sequence of estimates of an arranged autoregressive model, which data is arranged according to the variable Δp_{t-1}.

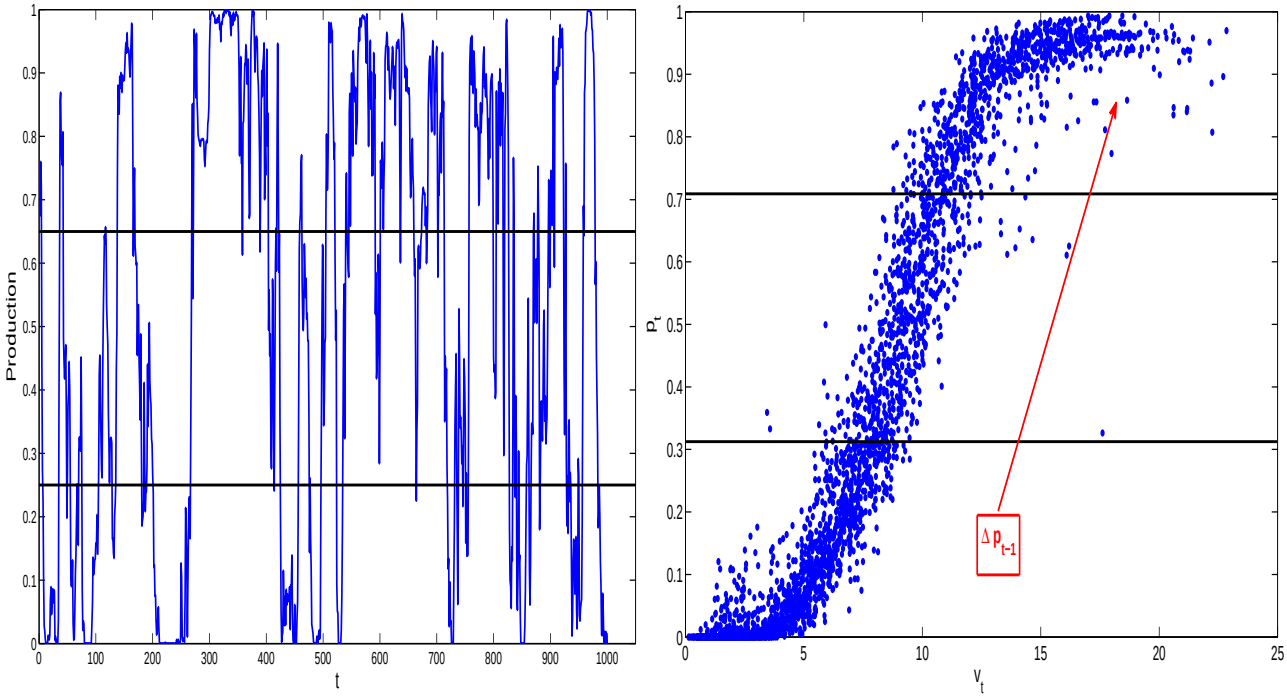
Once the existence of two main regimes is detected, the existence of different regimes depending on the levels of wind power is studied. Then, Aut-ARLS is applied by using p_{t-1} as threshold variable, when the process increases and decreases. Figure 4 displays the structure detected when the process decreases. Three regimes are detected corresponding to high, medium and low levels of wind power. The same results are obtained when the process increases, as can we seen in Figure 5

An important conclusion is that the same SETAR structure is founded in all available Greek wind farms. In figure 6 we can see the box-plot of the estimated threshold values for all the analyzed wind farms.



(a) Wind power (b) Power vs Speed

Figure 4: SETAR structure identified when observations are arranged according to p_{t-1}, if $\Delta p_{t-1} \leq 0$.



(a) Wind power (b) Power vs Speed

Figure 5: SETAR structure identified when observations are data arranged according to p_{t-1}, if $\Delta p_{t-1} > 0$.

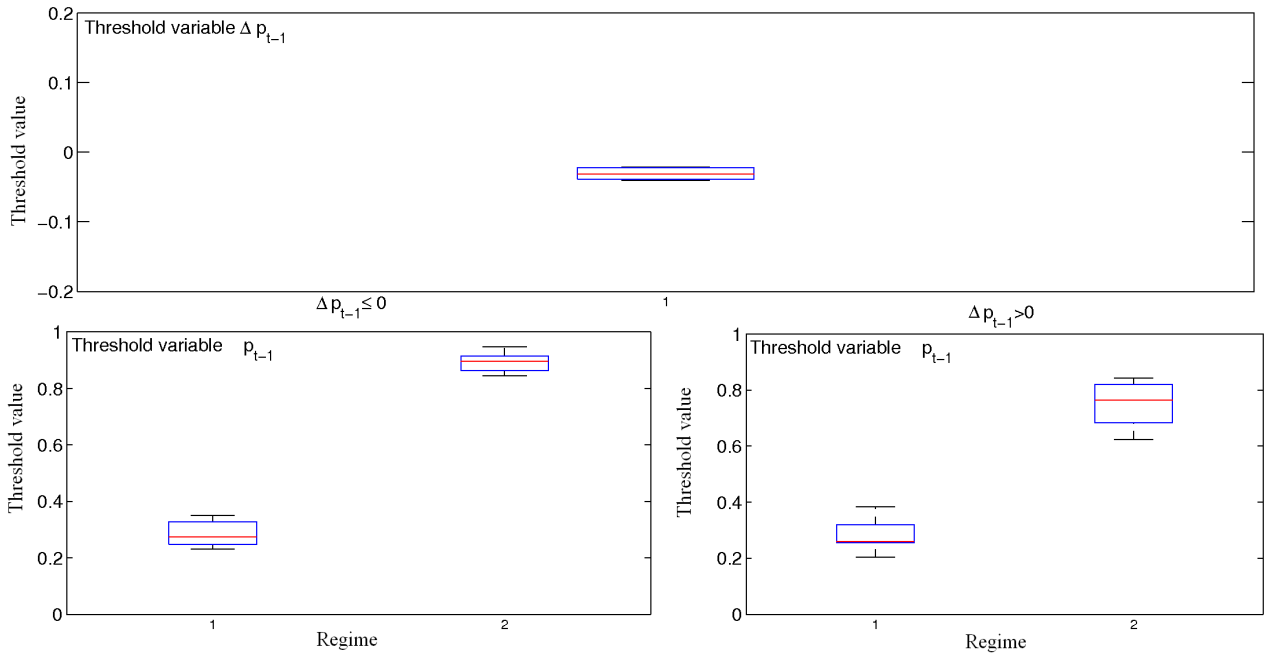


Figure 6: Box-plot of the estimated threshold values for all wind farms.

4 Short-term wind power predictions

The estimation of the predictions by using SETAR models are based on Monte Carlo simulation of future regimes. Then, N different trajectories $\hat{p}^N_{t+h|t}$ are generated according to an estimated predictive density $f_{t+h|t}$. Usually, the estimation of the sequence of predictive densities is carried out by using sample moments. In this paper, we propose to use the conditional moments.

The estimation of the conditional moments is done by means of regression models, with the independent variables being the powers of the point forecasts of wind power generation. This SETAR model with conditional moments is called SETAR-TVCD. The proposed model to estimate the conditional moments has the following form

$$m^{(j)}_{t+h|t} = \hat{\alpha}^{(j)}_{0t|t-h} + \hat{\alpha}^{(j)}_{1t|t-h}\hat{p}_{t+h|t} + \hat{\alpha}^{(j)}_{2t|t-h}\hat{p}^2_{t+h|t} + \hat{\alpha}^{(j)}_{3t|t-h}\hat{p}^3_{t+h|t} \tag{6}$$

The estimation of model (6) 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).

Mean of the Relative Root Mean Square Error in 6 wind farms										
			SETAR				SETAR-TVCD			
h	ARMA	TV-AR	TN	CN	BETA	MEP	TN	CN	BETA	MEP
1	1,18	1,08	1	1	1	1	1	1	1	1
2	1,26	1,12	1,18	1,18	1,19	1,09	1,15	1,11	1,09	1
3	1,42	1,17	1,29	1,22	1,26	1,10	1,21	1,17	1,13	1
4	1,65	1,22	1,35	1,33	1,34	1,11	1,26	1,21	1,16	1
5	1,87	1,25	1,57	1,44	1,45	1,12	1,35	1,27	1,26	1
6	1,96	1,28	1,66	1,56	1,50	1,17	1,41	1,30	1,32	1

Table 1: Results in 6 hourly wind farms

The predictions are calculated by using different models:

- **ARMA**. Best linear model is chosen by using AIC. The same model is used to predict at all horizons.
- **TV-AR**. Several AR(1) model with time-varying parameters are estimated by using a RLS method with several forgetting factors. The best model is chosen.
- **SETAR**. The predictive density $f_{t+h|t}$ is estimated by means of sample moments. Different parametric distributions are considered: truncated normal (TN), censored normal (CN), beta and maximum entropy (MEP).
- **SETAR-TVCD**. Conditional moments are used. Different parametric distributions are considered: truncated normal (TN), censored normal (CN), beta and maximum entropy (MEP).

The root mean square error (RMSE) is calculated in all available wind farms. Table 1 show the average of the relative RMSE assuming SETAR-TVCD with MEP distribution is the best model. The results confirms the assumption, because there is not any relative value lower than 1.

References

[1] Bermejo, M.A., Peña, D. y Sánchez, I. (2011). Identification of TAR models using recursive estimation. *Journal of Forecasting,* **30:** 31-50.

[2] Burton, T., Sharpe, D., Jenkins, N. y Bossanyi, E. (2001). *Wind Energy Handbook.* John Wiley & Sons: West Sussex.

[3] Chan, K.S. (1990). Testing for threshold autoregression. *The Annals of Statistics,* **18:** 1886-1894.

[4] Chan, K.S. y Tong, H. (1990). On likelihood ratio tests for threshold autoregression. *Journal of the Royal Statistical Society, Series B,* **52:** 469-476.

[5] Clements, M.P. y Smith, J. (1997). The performance of alternative forecasting methods for SETAR models. *International Journal of Forecasting,* **13:** 463-475.

[6] Costa A. et al. 2008. A review on the young history of the wind power short-term prediction. *Renewable and Sustainable Energy Reviews,* **12:** 1725-1744.

[7] Giebel, G., Brownsword, R. y Kariniotakis, G. (2011). The State-of-the-Art in Short-Term Prediction of Wind Power, Deliverable Report D1.1, Project Anemos, disponible en http://anemos.cma.fr/download/ANEMOS_D1.1_StateOfTheArt_v1.1.pdf.

[8] Hansen, B.E. (1999). Threshold effects in non-dynamic panels: estimation, testing, and inference. *Journal of Econometrics*, **93**: 345–368.

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

[10] Petruccelli, J. y Davies, N. (1986). A portmanteau test for Self-Exciting Threshold Autoregressive-Type nonlinearity in time series. *Biometrika,* **73:** 687-694.

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

[12] Tong, H. (1978). Discussion of a paper by A.J. Lawrance and N.T. Kottegoda. *Journal of the Royal Statistical Society, Series B,* **140:** 34-35.

[13] Tong, H. (1983). *Threshold models in non-linear time series analysis.* Springer: New York.

[14] Tong, H. (1990). *Non-linear Time Series: A Dynamical System Approach.* Oxford University Press: Oxford.

[15] Tong, H. y Lim, K.S. (1980). Threshold autoregression, limit cycles and cyclical data (con discusión). *Journal of Royal Statistical Society, Series B (Methodological),* **42:** 245-292.

[16] Tsay, R.S. (1989). Testing and modeling threshold autoregressive processes. *Journal of the American Statistical Association,* **84:** 231-240.

Quantile forecasting of wind power using variability indices

Patrick McSharry[a,b], Georgios Anastasiades[a,b]

[a] *Oxford Centre of Industrial and Applied Mathematics, Mathematical Institute, University of Oxford, Oxford, United Kingdom*

[b] *Smith School of Enterprise and the Environment, University of Oxford, Oxford, United Kingdom*

Abstract

Wind power forecasting techniques have received substantial attention recently due to the increasing penetration of wind energy in national power systems. While the initial focus has been on point forecasts, the need to quantify forecast uncertainty and communicate the risk of extreme ramp events has led to an interest in producing probabilistic forecasts. Using four years of wind power data from three wind farms in Denmark, we develop quantile regression models to generate short-term probabilistic forecasts from 15 minutes up to six hours ahead. More specifically, we investigate the potential of using various variability indices as explanatory variables in order to include the influence of changing weather regimes. These indices are extracted from the same univariate wind power series and optimized specifically for each quantile. The forecasting performance of this approach is compared with that of various benchmark models. Our results demonstrate that variability indices can increase the overall skill of the forecasts and that the level of improvement depends on the specific quantile.

Keywords: wind power forecasting, wind power variability, quantile forecasting, density forecasting, quantile regression, continuous ranked probability score, quantile loss function

1. Introduction

Wind power is one of the fastest growing renewable energy sources (Barton & Infield (2004)). However, due to the large variability of wind speed, and the unpredictable and dynamic nature of the earth's atmosphere,

Email addresses: patrick.mcsharry@smithschool.ox.ac.uk (Patrick McSharry), georgios.anastasiades@exeter.ox.ac.uk (Georgios Anastasiades)

there is great variability in wind power production. This inherent variability of wind speed is the main cause of the uncertainty observed in wind power generation. Recently, scientists have been attempting to, directly or indirectly, model this uncertainty and produce accurate forecasts of wind power production.

Since there is no efficient way to store wind energy, the wind power production decreases to zero if wind speed drops below a certain level. On the other hand, excessively strong winds can cause serious damage to the wind turbines, and hence they are automatically shut down at the 'disconnection speed', leading to an abrupt and quite unexpected drop of power generation. In addition, the wind power generated is limited by the capacity of each turbine. Therefore, it is quite important to produce accurate wind power forecasts for an efficient operation of wind turbines and reliable integration of wind power into the national grid.

The length of the relevant forecast horizons usually depends on the required application. For example, in order to schedule power generation, forecast horizons of several hours are usually enough, but for maintenance planning we need forecast horizons of several days or even weeks (McSharry, Pinson, & Gerard (2009)).

The literature of wind power forecasting starts with the work of Brown, Katz, & Murphy (1984) where they used autoregressive processes to model and simulate the wind speed, and then estimate the wind power by applying suitable transformations to values of wind speed. Most of the early and recent literature focuses on producing wind power point forecasts, directly, or indirectly in the sense that the focus is on modeling the wind speed and then transforming the forecasts through a power curve (Sanchez (2006), Taylor, McSharry, Member, IEEE, & Buizza (2009)).

Sanchez (2008) proposed an adaptive forecast combination procedure (AEC) that tends to be similar to the use of the best available predictor in a time varying environment. The proposed procedure was applied to two wind farms where alternative forecasts were available, showing the advantage of the proposed method. Jursa & Rohrig (2008) introduced a new short-term prediction method based on the application of evolutionary optimization algorithms for the automated specification of two well-known time series prediction models, namely, neural networks and the nearest neighbour search. The proposed method was used to produce wind power forecasts for a certain wind farm.

One of the most popular methods for producing point forecasts, is the use of time series models and especially the autoregressive integrated moving average (ARIMA) model, applied in the past by Nielsen (1999), Kariniotakis, Nogaret, & Stavrakakis (1997) and Makarov, Hawkins, Leuze, & Vidov (2002). Recent stud-

ies have also used the generalized autoregressive conditional heteroskedasticiry (GARCH) models for the conditional variance (Tol (1997), Taylor et al. (2009), Lau & McSharry (2010)). Tol (1997) fitted an autoregressive (AR)-GARCH model to daily Canadian mean wind speed series, and Taylor et al. (2009) an autoregressive moving average (ARMA)-GARCH model to the wind speed time series in the United Kingdom. Lau & McSharry (2010) used an ARIMA-GARCH process to model the aggregated wind power time series in Ireland.

Recent research has focused on producing probabilistic or density forecasts, because the point forecast methods are not able to quantify the uncertainty related to the prediction. Up to now, the number of studies on multi-step quantile/density forecasting is relatively small compared to point forecasting. Moeanaddin & Tong (1990) estimated densities using recursive numerical methods, which are quite computationally intensive. Pinson (2010), by applying a suitable transformation, managed to produce ten minutes ahead density forecasts at the Horns Rev wind farm in Denmark.

Taylor et al. (2009) used sophisticated time series models and weather ensemble[1] predictions to produce density forecasts for five wind farms in the United Kingdom. Also Lau & McSharry (2010) produced multi-step density forecast for the aggregated wind power series in Ireland, using ARIMA-GARCH models and exponential smoothing methods.

The quantile regression method, introduced by Koenker & Bassett (1978), has been extensively used to produce wind power quantile forecasts, using a variety of explanatory variables, among which, wind speed, wind direction, temperature and atmospheric pressure. Recent literature includes papers by Bremnes (2004), Nielsen, Madsen, & Nielsen (2006) and Moller, Nielsen, & Madsen (2008). Davy, Milton, Russell, & Coppin (2010), proposed a new variability index that is designed to detect rapid fluctuations of wind speed or power that are sustained for a length of time, and used it as an explanatory variable in the quantile regression model they constructed.

Bossavy, Girard, & Kariniotakis (2010) extracted two new indices that are able to recognize and predict ramp events in the wind power series, and used them to produce quantile estimates with the quantile regression forest method as their basic model. Finally, Gneiting (2011) studied the behavior of quantiles as optimal predictors and illustrated the relevance of decision theoretic guidance in the transition from a

[1] A relatively new approach for wind power forecasting is to use ensemble forecasts produced from numerical weather prediction (NWP) methods (Taylor et al. (2009) and Pinson & Madsen (2009)). Using this approach meteorological forecasts are transformed to wind power forecasts.

predictive distribution to a point forecast using the Bank of Englands density forecasts of United Kingdom inflation rates, and probabilistic predictions of wind energy resources in the Pacific Northwest.

In this paper we use univariate wind power series from three different wind farms in Denmark, to produce very short-term quantile forecasts, from 15 minutes up to six hours ahead. In order to produce quantile forecasts, we will use a quantile regression model, with explanatory variables extracted form the same wind power time series. More specifically, four new variability indices will be produced (extracted form the original wind power time series), which serve to capture the volatile nature of the wind power series. These indices, together with some lagged versions of the wind power series, will be used as explanatory variables in the quantile regression model. As for any regression model, we need predictions (point forecasts) for the future values of the explanatory variables, in order to produce future estimates. For the point forecasts we will use time-series models that are able to model both the mean and the variance of the underlying series.

The three Danish wind farms were chosen according to their monthly wind power capacity and standard deviation. We choose one high, one low, and one average variability wind farm in order to better understand the ability of each model to produce probabilistic forecasts under different circumstances.

The indices used will be independently optimized for each of the three wind farms, using a one-fold cross validation technique. In fact, two different optimizations will take place for each wind farm: The first one will aim to minimize the check function value produced by a 1-step ahead quantile regression forecast, for each of nineteen different quantiles. The second one will aim to minimize the average check function value, produced by taking the average over all 24 forecast horizons (equal to six hours), for each quantile. The final forecast results will be compared with those of some widely used benchmark models.

The remainder of the paper is presented as follows. In Section 2 we will introduce the wind power data, and in Section 3 the new variability indices will be derived. Section 4 will present the methodology behind quantile regression models with ways to evaluate the resulting quantile forecasts. In Section 5 we will optimize the variability indices using the in-sample data, and in Section 6 the out-of-sample forecast performance of the quantile regression models will be assessed. Section 7 will conclude the paper.

2. Wind power time series

We use wind power data recorded at three wind farms in Denmark summarized in Table 1. These wind farms were chosen to have different amounts of wind power variability, and have the smallest percentage of

missing values[2] compared to the one of other available wind farms.

Wind farm station name	Region	Wind power variability	No of turbines	Wind farm rated capacity (MW)
DØR	6	Low	1	1000
ALB	3	Medium	23	25500
VES	2	High	8	2195

Table 1: The Danish wind farms used in this study.

Our data sets contain wind power measurements recorded every 15 minutes for four years, from 01-Jan-2007 until 31-Dec-2010. The data of each wind farm is bounded between zero and the maximum capacity of the turbines. The zero value is attained in the case of excessively strong wind, where the turbines shut down in order to prevent them from damage. In order to facilitate comparisons between the data sets of different capacities, we normalize the wind power data of each wind farm by dividing by the total (rated) capacity. Hence, the data is now bounded within the interval [0,1].

We dissect the data of each farm into a set of exactly two years (2007 and 2008) for in-sample model training and calibration, and an out-of-sample testing set (the remaining two years) for out-of-sample testing and model evaluation. The in-sample set is dissected again into two sub-sets, a training set and a testing set. For the in-sample training set we use the first 1.5 years and for the in-sample testing set the remaining half year. This way we can use a *one-fold cross validation technique* to optimize the indices introduced in Section 3, and test the performance of our final chosen model using the out-of-sample testing set.

The time series plots for the year 2010, together with the monthly capacity and standard deviation indices are shown in Figure 1. The monthly capacity and variability indices were generated by taking the mean and standard deviation of wind power, respectively, for each month over the entire four year period. As we can clearly observe, the three wind farms have different wind power variability. More specifically, the first two wind farms have the highest and lowest possible wind power variability for all four years (from all the available wind farms in Denmark), without having any significant changes[3] in the capacity form year to year. The third wind farm was chosen to have an average (medium) variability compared to the other two farms, but with the same stationarity criterion for capacity over the four years.

[2]The missing values in the data sets were imputed using linear interpolation.

[3]Wind power variability may change form year to year by addition of new turbines or removal (maybe for maintenance) of existing ones.

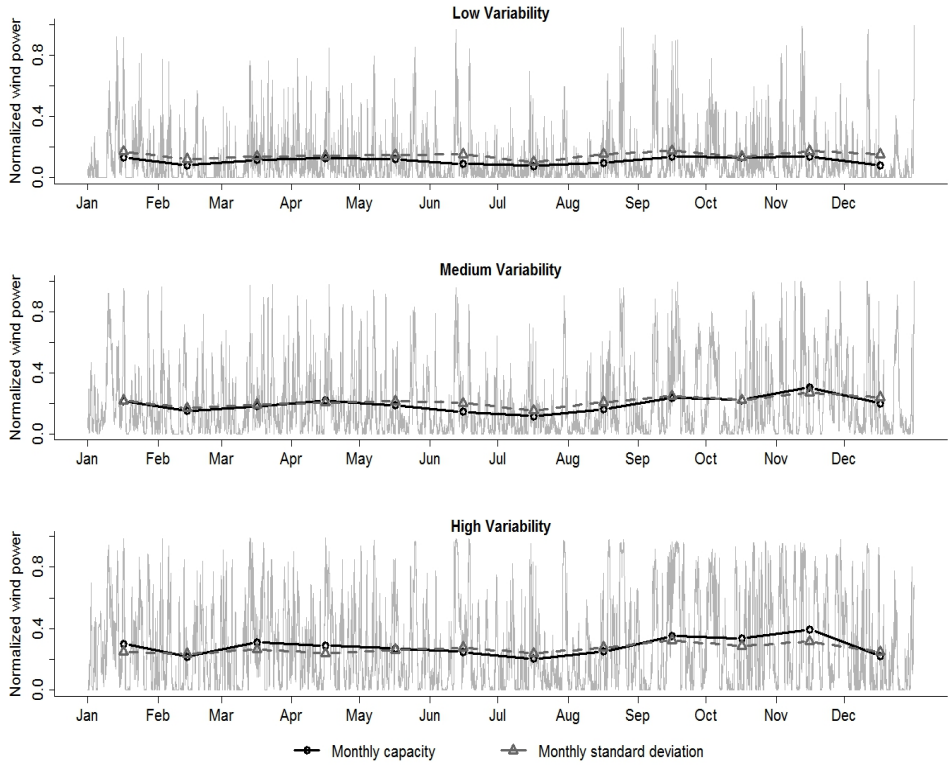


Figure 1: Time series plots of normalized power data for the three chosen Danish wind farms, for year 2010.

3. Indices of wind power variability

Davy et al. (2010), proposed a variability index that is design to detect rapid fluctuations of wind speed or power that are sustained for a length of time. They defined this variability index as the standard deviation of a band-limited signal in a moving window, and they constructed such an index for a wind speed time series. This variability index depends on four parameters: the order of the filter (integer greater than one), the upper and lower frequencies of the extracted signal, and the width of the moving window. We would like to use such an index as an explanatory variable in our quantile regression, but a proper optimization of this is too time consuming because of the number of the parameters involved.

Instead, we propose a parsimonious variability index, which depends on only two parameters, and is constructed as follows. Firstly we smooth our original wind power series using an averaging window of size m, in order to obtain the smoothed wind power series,

$$r_t = \frac{1}{m} \sum_{i=0}^{m-1} y_{t-i}, \tag{1}$$

starting from $t = m$. Note that this series behaves in a fully retrospective way, in the sense that each point of the series depends only on the historical values of the original series. Since the smoothed series is $m-1$ points smaller than the original series, we set $r_t = r_m$, for $t = 1, 2, ..., m-1$.

Finally, the new variability index is just the standard deviation of the extracted smoothed wind power series in a moving window of width n. So, if r_t is a given point of the smoothed series, we define the new index as

$$SD_t = \sqrt{Var(r_{t-n+1}, ..., r_{t-1}, r_t)}, \tag{2}$$

for $t \geq n$. Again, we impute the first $n-1$ points of the series by setting $SD_t = SD_n$, for $t = 1, 2, ..., n-1$. This index can be optimized much easier than the one proposed by Davy et al. (2010), since it has only two parameters: the smoothing parameter m, and the variability parameter n .

By similar reasoning, we create another three variability indices. We smooth the original wind power series as described above, and then instead of finding the standard deviation, we find the interquartile range (IQR), the 5% and the 95% quantiles of the smoothed series over a moving window (different for each series). These three new indices can be defined as:

$$IQR_t = IQR(r_{t-n+1}, ..., r_{t-1}, r_t), \tag{3}$$

$$\mathbb{P}\left((r_{t-n+1}, ..., r_{t-1}, r_t) < Q05_t\right) = 0.05, \tag{4}$$

$$\mathbb{P}\left((r_{t-n+1}, ..., r_{t-1}, r_t) < Q95_t\right) = 0.95, \tag{5}$$

for $t \geq n$. We also impute their values for $t = 1, ..., n-1$ as described above. An example of the construction of the three variability wind power indices is shown in Figure 2. These indices will be properly optimized, and together with some lagged values of the original power series will be used as explanatory variables in the quantile regression introduced in the next section.

4. Quantile regression, forecasting, and evaluation methodology

4.1. Quantile regression

Given a random variable[4], y_t, and a strictly increasing CDF, $F_t(y)$, the α_i-quantile, $q_t^{(\alpha_i)}(y)$, with proportion $\alpha_i \in [0, 1]$ is defined as the value for which the probability of obtaining values of y_t below $q_t^{(\alpha_i)}$ is α_i:

$$\mathbb{P}\left(y_t < q_t^{(\alpha_i)}\right) = \alpha_i \quad or \quad q_t^{(\alpha_i)} = F_t^{-1}(\alpha_i) \tag{6}$$

[4]The notation y_t, is used for denoting both the stochastic state of the random variable at time t, and the measured value at that time.

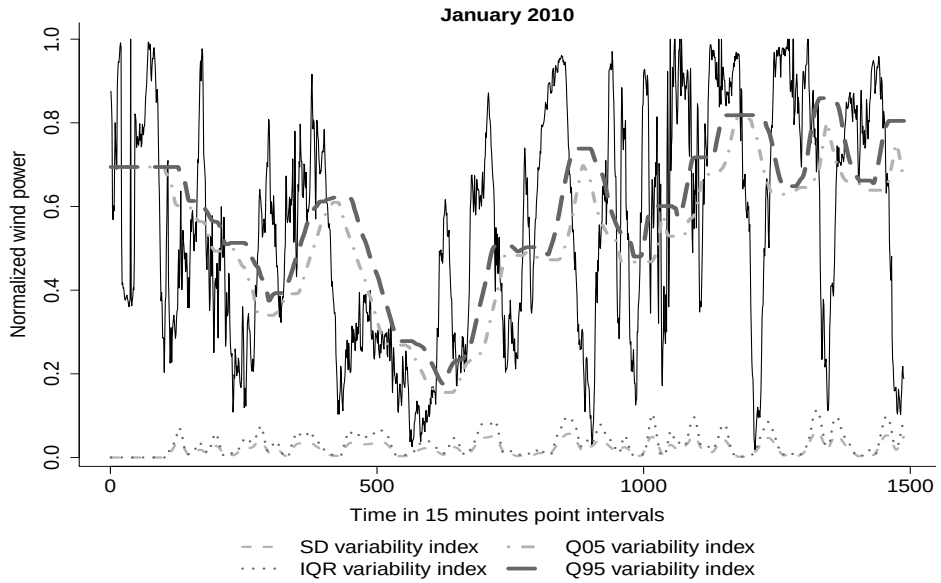


Figure 2: Wind power time series plot of the medium variability farm, together with the four variability indices. The parameters are chosen to be the same for all indices just for comparison ($m = 100$ and $n = 30$).

The quantile regression, introduced by Koenker & Bassett (1978), models $q_t^{(\alpha_i)}$ for $\alpha_i \in [0, 1]$, as a linear combination of some given explanatory variables. This is quite similar to multiple linear regression where the mean is modeled in exactly the same way. So, the α_i-quantile is modeled as:

$$\begin{aligned} q_t^{(\alpha_i)} &= \gamma_0^{(\alpha_i)} + \gamma_1^{(\alpha_i)} x_{t,1} + ... + \gamma_p^{(\alpha_i)} x_{t,p} \\ &= \gamma_0^{(\alpha_i)} + \sum_{j=1}^{p} \gamma_j^{(\alpha_i)} x_{t,j}, \end{aligned} \tag{7}$$

where $\gamma_j^{(\alpha_i)}$ are unknown coefficients depending on α_i, and $x_{t,j}$ are the p known regressors. In quantile regression, a regression coefficient estimates the change in a specified quantile of the response variable produced by a one unit change in the corresponding predictor variable.

We define the *quantile loss function*, also known as the *check function*, for a given proportion $\alpha_i \in [0, 1]$ as

$$\begin{aligned} \rho_{\alpha_i}(u) &= \big(\alpha_i - \mathbb{1}_{\{u<0\}}\big)u \\ &= \begin{cases} \alpha_i u, & u \geq 0, \\ (\alpha_i - 1)u, & u < 0, \end{cases} \end{aligned} \tag{8}$$

where u is a given function. Then, the sample α_i-quantile can be estimated by minimizing $\sum_{t=1}^{T} \rho_{\alpha_i}(y_t - q)$ with respect to q. Hence, we can estimate the unknown coefficients, $\gamma_j^{(\alpha_i)}$, by replacing q with the right-hand side of equation (7):

$$\hat{\gamma}^{(\alpha_i)} = \underset{\gamma}{\operatorname{argmin}} \sum_{t=1}^{T} \rho_{\alpha_i}\big\{y_t - (\gamma_0 + \gamma_1 x_{t,1} + ... + \gamma_p x_{t,p})\big\}, \tag{9}$$

where $\hat{\gamma}^{(\alpha_i)}$ is a vector containing the unknown coefficients. Usually, these estimates are calculated using linear programming techniques as in Koenker & D'Orey (1987).

In this paper we will use quantile regression to forecast the values of quantiles with nominal proportion $\alpha_i = \{0.05, 0.01, ..., 0.95\}$, for forecast horizons[5] $k = 1, 2, ..., 24$. We denote the forecast for the quantile with nominal proportion α_i issued at time t for forecast time $t + k$, by $\hat{q}_{t+k|t}^{(\alpha_i)}(y)$. In order to produce these forecasts, we use equation (7), and the estimated coefficients, $\hat{\gamma}^{(\alpha_i)}$:

$$\begin{aligned} \hat{q}_{t+k|t}^{(\alpha_i)}(y) &= \gamma_0^{(\alpha_i)} + \gamma_1^{(\alpha_i)} \hat{x}_{t+k|t,1} + ... + \gamma_p^{(\alpha_i)} \hat{x}_{t+k|t,p} \\ &= \gamma_0^{(\alpha_i)} + \sum_{j=1}^{p} \gamma_j^{(\alpha_i)} \hat{x}_{t+k|t,j}, \end{aligned} \tag{10}$$

where $\hat{x}_{t+k|t,j}$ for $j = 1, ..., p$ denote the forecasts of the explanatory variables $\mathrm{x}_{t,j}$, issued at time t with lead time $t + k$.

The random variable y_t will represent the normalized wind power time series, (y_t), and the explanatory variables will be represented by time series, $(x_{t,j})$, extracted from the normalized wind power series. So, in order to produce the forecasts, $\hat{x}_{t+k|t,j}$, we will fit suitable time series models to the variables $(x_{t,j})$, and then predict from these models up to t+k values ahead.

4.2. Quantile forecast evaluation

The evaluation of the quantile forecasts, for each quantile, $\alpha_i = \{0.05, 0.01, ..., 0.95\}$, will be done using the quantile loss function:

The ***quantile loss function*** (Koenker & Bassett (1978)), also known as the ***check function*** is used to define a specific quantile of the distribution and was defined in Section 4, equation (8) .

[5]The forecast horizon, k, is measured in time steps of 15 minutes.

Hence, we can estimate a particular quantile, $\hat{q}^{(\alpha_i)}$, with proportion α_i, using

$$\hat{q}^{(\alpha_i)} = \min_q \sum_{t=1}^{N} \rho_{\alpha_i}(y_t - q), \tag{11}$$

and therefore we can evaluate a series of quantile forecasts, $\hat{q}^{(\alpha_i)}_{t+k|t}$, issued at time t with lead time $t+k$ and nominal proportion α_i, using:

$$QL(k, \alpha_i) = \frac{1}{N} \sum_{t=1}^{N} \rho_{\alpha_i}(y_{t+k} - \hat{q}^{(\alpha_i)}_{t+k|t}). \tag{12}$$

.

Using the different quantile forecasts we can also reconstruct the whole probability / cumulative forecasted distribution. We use the ***Continuous Ranked Probability Score (CRPS)*** in order to evaluate the the density forecasts for each forecast horizon:

The CRPS (Matheson & Winkler (1976)) is computed by taking the integral of the Brier scores for the associated probability forecasts at all real valued thresholds,

$$crps(\hat{F}_{t+k|t}(y), y_{t+k}) = \int_{-\infty}^{+\infty} \big(\hat{F}_{t+k|t}(y) - \mathbb{1}_{\{y \geq y_{t+k}\}}\big)^2 \, \mathrm{d}y \tag{13}$$

$$= \int_0^1 QS_{\alpha_i}\big(\hat{F}^{-1}_{t+k|t}(\alpha_i), y_{t+k}\big) \, \mathrm{d}\alpha_i, \tag{14}$$

where $\hat{F}_{t+k|t}(y)$ corresponds to the CDF forecast, and y_{t+k} to the corresponding verification. $\mathbb{1}_{\{y \geq y_{t+k}\}}$ is an indicator function that equals one if $y \geq y_{t+k}$ and zero otherwise. The quantile score, QS_{α_i}, (Gneiting & Raftery (2007)) is defined by

$$QS_{\alpha_i}(q, y) = 2\big(\alpha_i - \mathbb{1}_{\{y<q\}}\big)\big(y - q\big). \tag{15}$$

Hence, the average of these $crps$ values over each forecast-verification pair gives the CRPS for each forecast horizon k:

$$\text{CRPS}(k) = \frac{1}{N} \sum_{t=1}^{N} crps(\hat{F}_{t+k|t}(y), y_{t+k}) \tag{16}$$

$$= 2 \int_0^1 QL(k, \alpha_i) \, \mathrm{d}\alpha_i, \tag{17}$$

where $QL(k, \alpha_i)$ is the mean quantile loss function defined in equation (12). Representation (17) will be particularly useful because we will derive the CRPS using the loss function value for each quantile.

5. Optimization of the variability indices

Our main goal is to check whether or not the four variability indices (introduced in Section 3) can help to provide trustworthy quantile forecasts of wind power, when used as explanatory variables in the quantile regression model (7). For this purpose, we have to optimize the two parameters (m, n) of these indices using the following procedure.

For each index, we sample different combinatios of parameters from the range $m, n = \{0, 8, 16, ..., 192\}$, in order to produce 625 different realizations of each index, for each wind farm. A preliminary analysis showed that creating a moving window larger than 192 time-points wide (2880 minutes) did not increase the performance of the indices.

Then, for each set of parameters, we fit the following four different quantile regression models on the in-sample training set (of each wind farm), with the normalized wind power series, (y_t), as the response, for each of the 19 quantiles $\alpha_i = \{0.05, 0.1, ..., 0.95\}$:

$$\begin{aligned}
\text{SD model:} \quad & q_t = \gamma_{01} + \gamma_{11} y_{t-1} + \gamma_{21} y_{t-2} + \gamma_{31} y_{t-3} + \gamma_{41} SD_t \\
\text{IQR model:} \quad & q_t = \gamma_{02} + \gamma_{12} y_{t-1} + \gamma_{22} y_{t-2} + \gamma_{32} y_{t-3} + \gamma_{42} IQR_t \\
\text{Q05 model:} \quad & q_t = \gamma_{03} + \gamma_{13} y_{t-1} + \gamma_{23} y_{t-2} + \gamma_{33} y_{t-3} + \gamma_{43} Q05_t \\
\text{Q95 model:} \quad & q_t = \gamma_{04} + \gamma_{14} y_{t-1} + \gamma_{24} y_{t-2} + \gamma_{34} y_{t-3} + \gamma_{44} Q95_t,
\end{aligned} \tag{18}$$

where $q_t \equiv q_t^{(\alpha_i)}$ is defined in (6), $\gamma_{hl} \equiv \gamma_{hl}^{(\alpha_i)}$ are the regression coefficients, and y_{t-j} are lagged wind power series. The choice of using three lagged series of the response as explanatory variables was taken after many trials of different numbers of lags for all three wind farms.

Next, we produce point forecasts from 15 minutes up to six hours ahead[6], from each point of the in-sample testing set, by fitting ARIMA$(1, 1, 1)$ models to each realization of the four variability indices of the above regressions. The point forecasts will be used as predictors to produce the quantile forecasts for the normalized wind power time series, as explained in Section 4.1. Our choice of ARIMA$(1, 1, 1)$ model was made mainly for simplicity, after exploring the forecast performances of various time series models. Fitting a different model for each set of parameters is very computationally expensive, and in this case the sacrificed accuracy is very small.

[6] i.e. with forecast horizon $k = 1, 2, ..., 24$.

Moreover, modeling the variance of the indices using ARCH/GARCH models (in combination with an ARIMA model for the mean), does not provide a consistent and significant improvement of the RMSE[7] of the point forecasts. This is mainly because of the very small forecast horizon we have, and hence it suffices to use a simple ARIMA model.

In order to produce point forecasts of the lagged wind power series, model solution using BIC identified an ARIMA$(0, 1, 2)$ - GARCH$(1, 1)$ model for the low variability farm, an ARIMA$(1, 1, 3)$ - GARCH$(1, 1)$ for the medium variability farm, and an ARIMA$(2, 1, 1)$ - GARCH$(1, 1)$ for the high variability farm. These models have the ability to capture the heteroskedastic effects that the wind power series have, taking into account the non-linear nature of the variations. Also, these forecasts are calculated only once for all different realizations of the quantile regression models, and hence there is no point in this case to sacrifice the (small) accuracy gain for simplicity and computational efficiency.

After producing quantile forecasts for 24 different forecast horizons, we evaluate them (i) using the check function value averaged over all forecast horizons, and (ii) using the check function value of only the first step ahead forecasts. The results justify our inspection of better forecast performance for the models with small moving windows. We repeat the above procedure by restricting our parameters' range even more for each variability index, and sample every different combination of parameters from the range $m, n = \{0, 1, 2, ...50\}$.

We end up with distinct sets of parameters (for each model and wind farm) that minimize the averaged and 1-step ahead check function value of each different quantile. The check function minimization results are shown in Tables 3 and 4 of the Appendix. In general, we can not distinguish any particular parameter pattern, but there are some features that are worth mentioning. For all the models, it is more common to have the smoothing window width (m) smaller than the variability window width (n), especially for quantiles less or equal to the median. This pattern changes for the upper quantiles (larger than the median) were we do not observe a clear pattern. Also, on average, the parameters for the 24-steps optimization are smaller than the corresponding ones of the 1-step optimization.

Finally, using (17), we gather two more sets of parameters (for each model and wind farm) that minimize (i) the total CRPS averaged over all forecast horizons, and (ii) the total CRPS of the first step ahead forecasts. Using these we would like to check whether or not there is an improvement in the overall skill of the models

[7] We used the Root Mean Square Error to evaluate the point forecast performance of various time series models.

when treating each quantile separately. The CRPS minimization results are shown in Table 2 of the Appendix. Again, on average, the parameters for the 24-steps optimization are smaller than the corresponding ones of the 1-step optimization, and the smoothing window width is smaller than the variability window.

6. Out-of-sample forecast performance results

In this section we will fit the optimized models of (18) to the whole two years in-sample set (of each farm), produce quantile forecasts from 15 minutes up to six hours ahead from each point of the out-of-sample set, and assess their forecast performance using the CRPS, check function and reliability diagrams.

Firstly, it will be useful to check the R-squared value of the four fitted models for each quantile. Essentially, by using quantile regression we fit a linear regression line for each quantile, and hence the R-squared value will tell us the proportion of total variance explained by each model, for each quantile. Of course this is an in-sample calculation, but we expect to get roughly an idea of how good the models are performing. Figure 3 shows the R-squared values for all models and wind farms when we optimize the models using the check function value of each quantile. A general observation is that we have relatively high R-squared values (> 0.5) for all quantiles, confirming the ability of the models to explain a large proportion of the total variability. Also the variance explained is much larger for the upper quantiles than the lower quantiles, mainly because of the asymmetric shape of the wind power distribution.

Some models for some specific wind farms and quantiles close to the median, have their R-squared values close to one. If we also examine these models' optimal parameters (Tables 3, 4), we observe that when the R-squared is close to one, both parameters are chosen to be zero. In other words, the variability indices of these models are identical to their original wind power series. When this happens the regression coefficients of the lagged series are almost zero, and the coefficient of the corresponding variability index is almost equal to one (see Figure 8 of the Appendix). This is perfectly acceptable since for a quantile regression model, the response is a specific quantile of a series (or data points) and not the series itself. Intuitively, if the specific quantile is close to the mean value of the series, then the check function is minimized by this mean value. Hence, the optimization forces the parameters to be zero, and consequently forecasts this quantile by predicting the mean of the series itself.

A final observation is that the R-squared values of the low variability wind farms are significantly smaller than the corresponding ones of the other farms. This is reasonable as the variability indices are used to

capture the different regime changes of the series, and hence we expect our models to explain a larger proportion of the total variance when the variability is higher.

Figure 4 shows the analogous R-squared values when the models are optimized based on the CRPS (1-step or averaged 24-steps). The observations are very similar to the ones stated above, with the difference that now the plots are much smoother (since we don't treat each quantile separately), without any values close to 1.

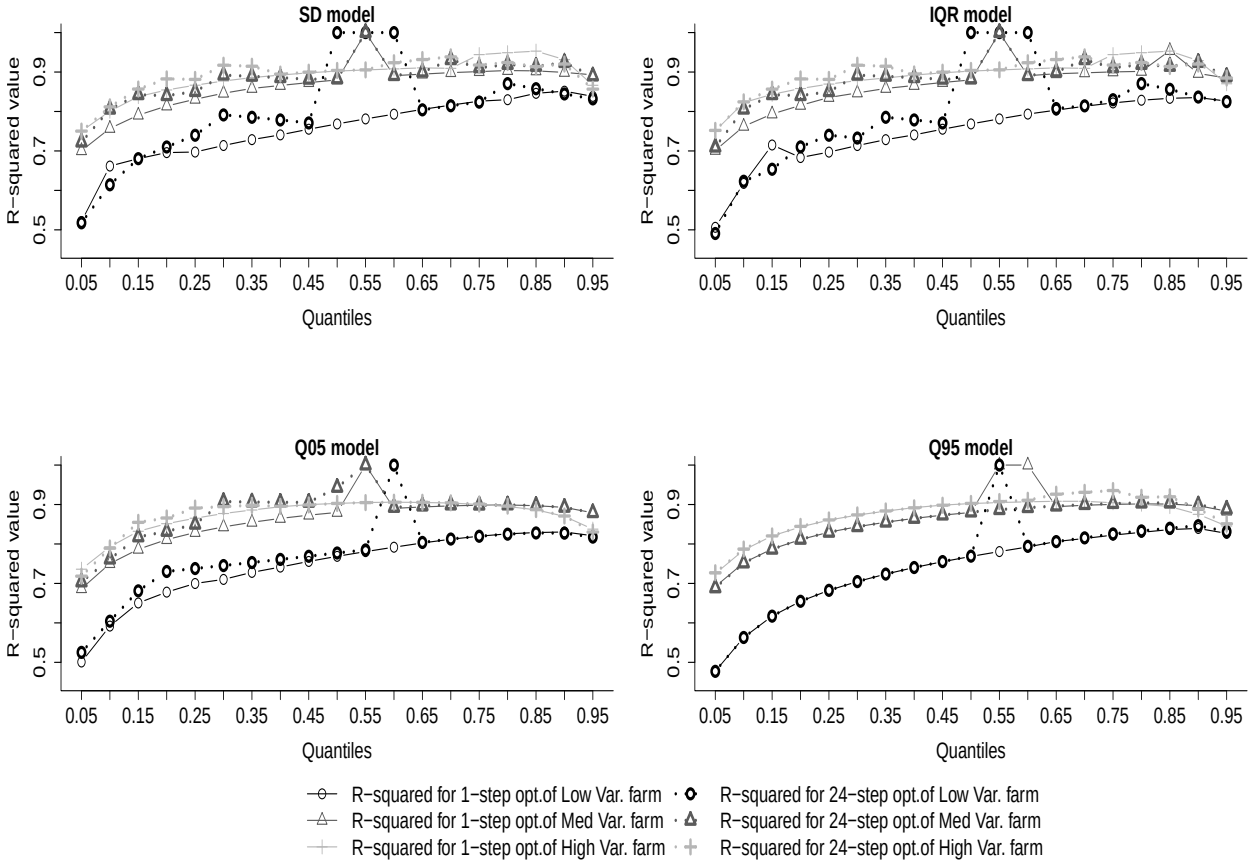


Figure 3: R-squared values for models with 1-step check function optimization and averaged over 24-steps check function optimization.

Next, we produce quantile forecasts from each point of the out-of-sample set (final two years), from 15 minutes up to six hours ahead. This is achieved for all the models and for every wind farm. The forecast performance of each model is assessed using the evaluation methods of Section 4.2. In order to facilitate the comparison of forecast performance across different models, we will introduce two widely used *probabilistic benchmarks*:

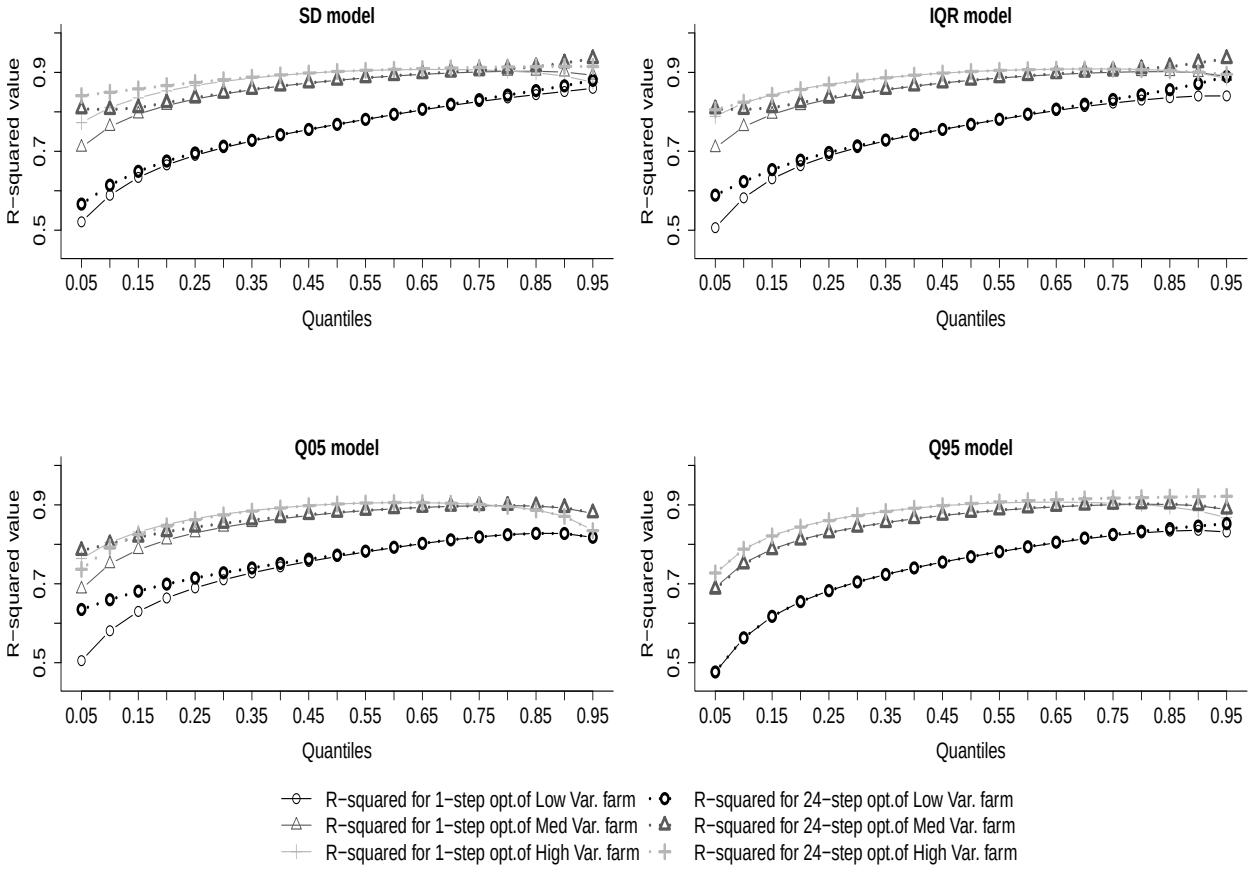


Figure 4: R-squared values for models with 1-step CRPS optimization and averaged over 24-steps CRPS optimization.

- *Persistence distribution*: It is defined as the distribution of the last n observations. The persistence benchmark is independently optimized for each wind farm, by using the same optimization methods as for the variability indices: 1-step check function, average of 24-steps check function, 1-step CRPS and average of 24-steps CRPS minimization.
- *Unconditional distribution*: We construct this benchmark by using all the past observations of the time series. This benchmark assumes that the time ordering of the observations is not relevant when attempting to predict the distribution the response. It is also referred to as *climatology*.

The third benchmark used in this paper is the quantile regression model with only the three lagged series as explanatory variables. This benchmark will help us to identify the gain in forecast performance acquired by using the four variability indices. In the next figures we will present the evaluation results of the four models, and include only the best model(s) for each evaluation criterion. In case that the results are quite similar

across wind farms we will present the averaged results[8] over the three wind farms. Also, it will be necessary to quantify the gain of some forecasting models to a chosen reference system. Following McSharry et al. (2009), this gain, denoted as an improvement with respect to the considered reference forecasts system, is called a *Skill Score* and is defined as:

$$SkillScore(k) = \frac{SCORE_{\text{ref}}(k) - SCORE(k)}{SCORE_{\text{ref}}(k)} = 1 - \frac{SCORE(k)}{SCORE_{\text{ref}}(k)}, \tag{19}$$

where k is the lead time of the forecast and $SCORE$ is considered the evaluation criterion (such as CRPS or quantile loss function score).

6.1. Out-of-sample model comparison and evaluation

Firstly we would like to check the density forecasting performance of the competing models. Figures 5 (a), (c) and (e) show the out-of-sample normalized CRPS for the three wind farms, where the competing models were optimized by minimizing the averaged over 24-steps check function. The plots show only the best two models, namely, the SD and IQR models, together with the three benchmarks. Th subplots show the Skill CRPS of the four competing models, with the three lagged series as the reference model.

A first observation is that the CRPS values are getting larger as the variability of the series increase, which is reasonable because intuitively it is easier to forecast a low variability series than a high one. On the other hand, the two best models perform better for the high variability wind farm. They perform better in the sense that they manage to outperform their competing benchmarks (especially climatology), with greater score difference for each forecast horizon, than the lower variability farms' models. The subplots inside each figure help us to compare the four models. The SD and IQR models do not produce very promising density forecasts for the first lead time, but their performance increases and outperform all the other models after approximately two time steps, for all wind farms.

Figures 5 (b), (d), (e) show the Skill CRPS of the 1-step check function optimization models, with the corresponding (averaged over) 24-steps optimization models as reference, for each wind farm. These Skill CRPS plots can help us to compare the two ways of optimizations carried out in this paper. The SD and IQR models produced by the 1-step optimization procedure perform much better when forecasting

[8]The averaging of the results is reasonable if the relative performances of the methods are similar for each of the three locations (Taylor et al. (2009)).

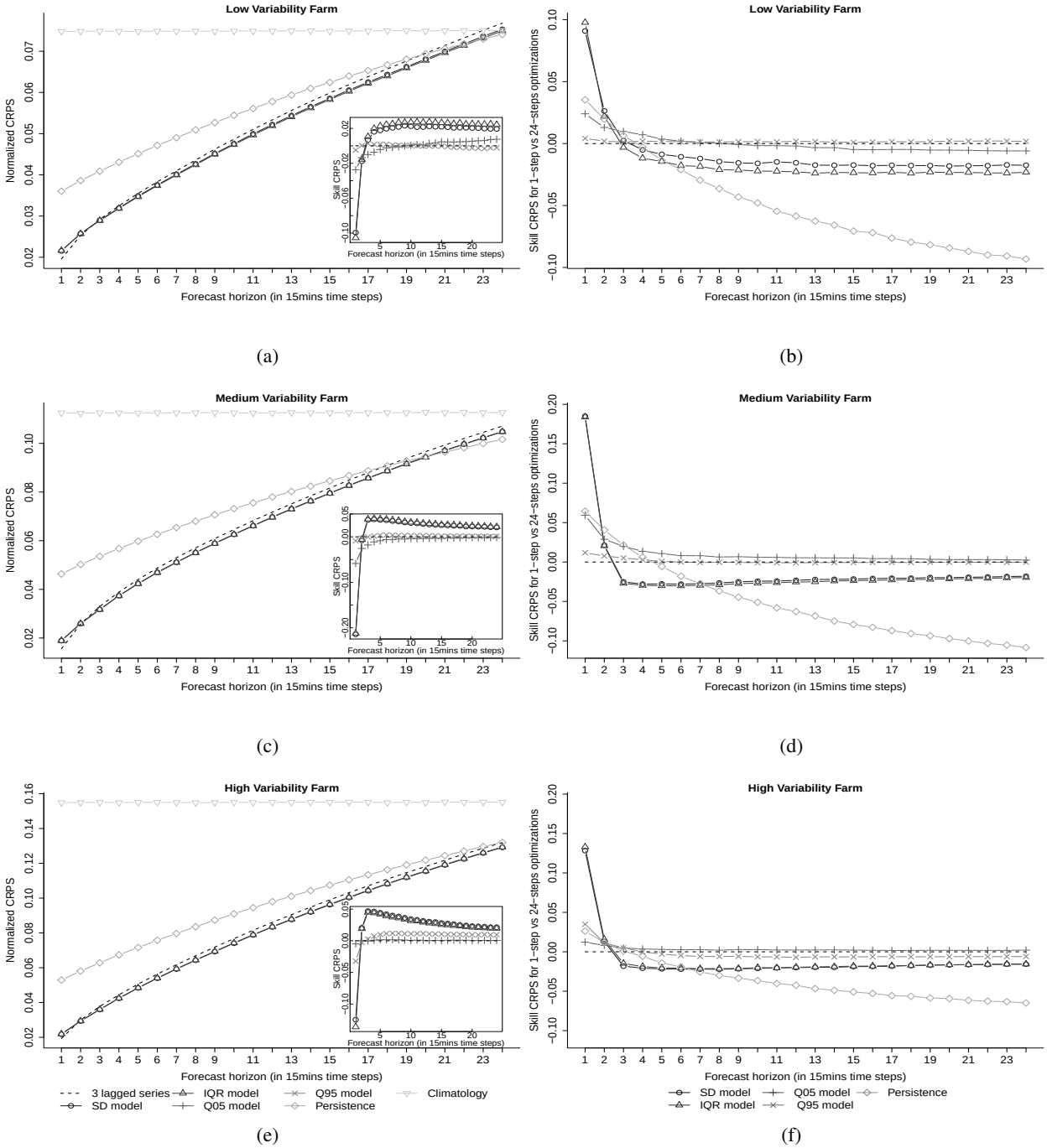


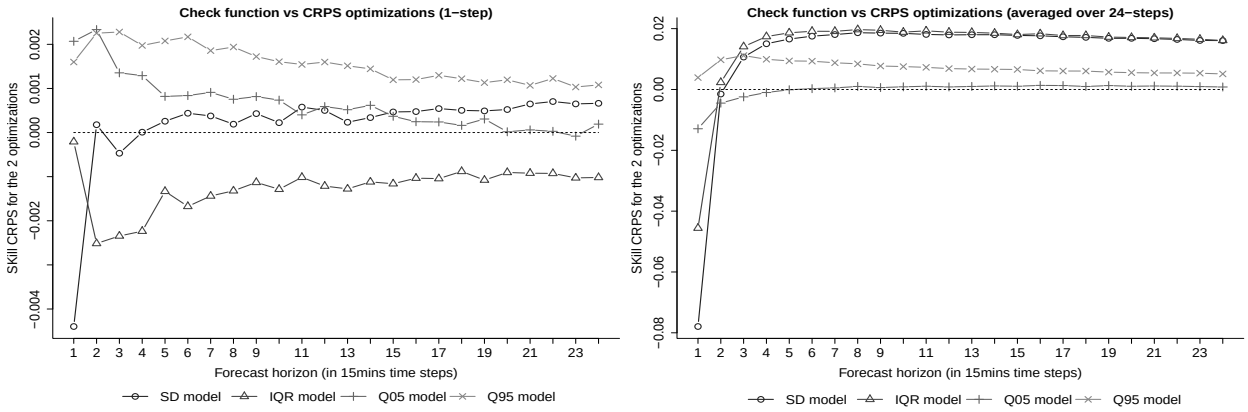
Figure 5: (a), (c), (e): Normalized CRPS for all three wind farms showing the three benchmarks and the best model(s). The subplots show the Skill CRPS of our competing models, with the three lagged series as the reference model. These results are outcome of the 24-steps check function optimization procedure, for each quantile. (b), (d), (f): Skill CRPS of the 1-step optimization models, with the corresponding 24-steps optimization models as reference.

the first two lead times. They can perform up to 17% better than the corresponding models of the 24-steps optimization. The Q05 and Q95 models also perform better when they are optimized using the 1-step optimization procedure.

For forecast horizons greater than 30 minutes (two time steps) the averaged over 24-steps SD and IQR models, produce density forecasts with skill gain up to 3%, compared to the analogous models of the 1-step optimization. On the other hand the Q05 and Q95 models behave almost identically with the corresponding 24-steps optimized models. What is also noteworthy is the huge gain in forecast performance of the persistence benchmark for large forecast horizons, when optimized using the 24-steps optimization method. The 1-step ahead optimized persistence benchmark may outperform the corresponding 24-steps optimized one for up to (maximum) three forecasted time steps, but then it is massively outperformed by the second one.

Finally we would like to check if there is any gain in the models' density forecasting performance when we treat each quantile separately. In other words, we want to check whether or not the resulting models of the (1-step or 24-steps) check function optimization outperform the corresponding ones of the CRPS optimization. Figure 6 (a) shows the Skill CRPS of the optimized models using the 1-step ahead check function optimization, with the corresponding 1-step CRPS optimized models as reference. The scores shown are averaged over all wind farms, because the relative performance of the methods are similar for each location. The performance gain (or loss) is maximum 0.2%, and hence the effort spent to optimize each models' quantile individually does not seem to be worthwhile. On the other hand, the SD and IQR models resulting from a 24-steps check function optimization, seem to outperform the corresponding CRPS optimized models by roughly 2% for forecast horizons greater than 30 minutes, and be outperformed by maximum 8% for the first lead time (Figure 6 (b)). The Q05 and Q95 models again have identical density forecasting performance for both optimization criteria.

Next, we briefly summarize the results found using the CRPS criterion. The best density forecasting models are the SD and IQR models, without any distinguishable -in terms of performance- difference between them. Both models can outperform all used benchmarks for forecast horizons up to (at least) 20 time steps ahead, for all three different variability wind farms. There are two more questions that we need to answer: what type of optimization is the best (1-step or 24-steps), and is it worthy to treat each quantile separately (by using check function instead of CRPS minimization)? The results show that if somebody is interested in forecasting only (up to) two time steps ahead ($\leq$ 30 minutes), then it is recommended to use one of the



(a) (b)

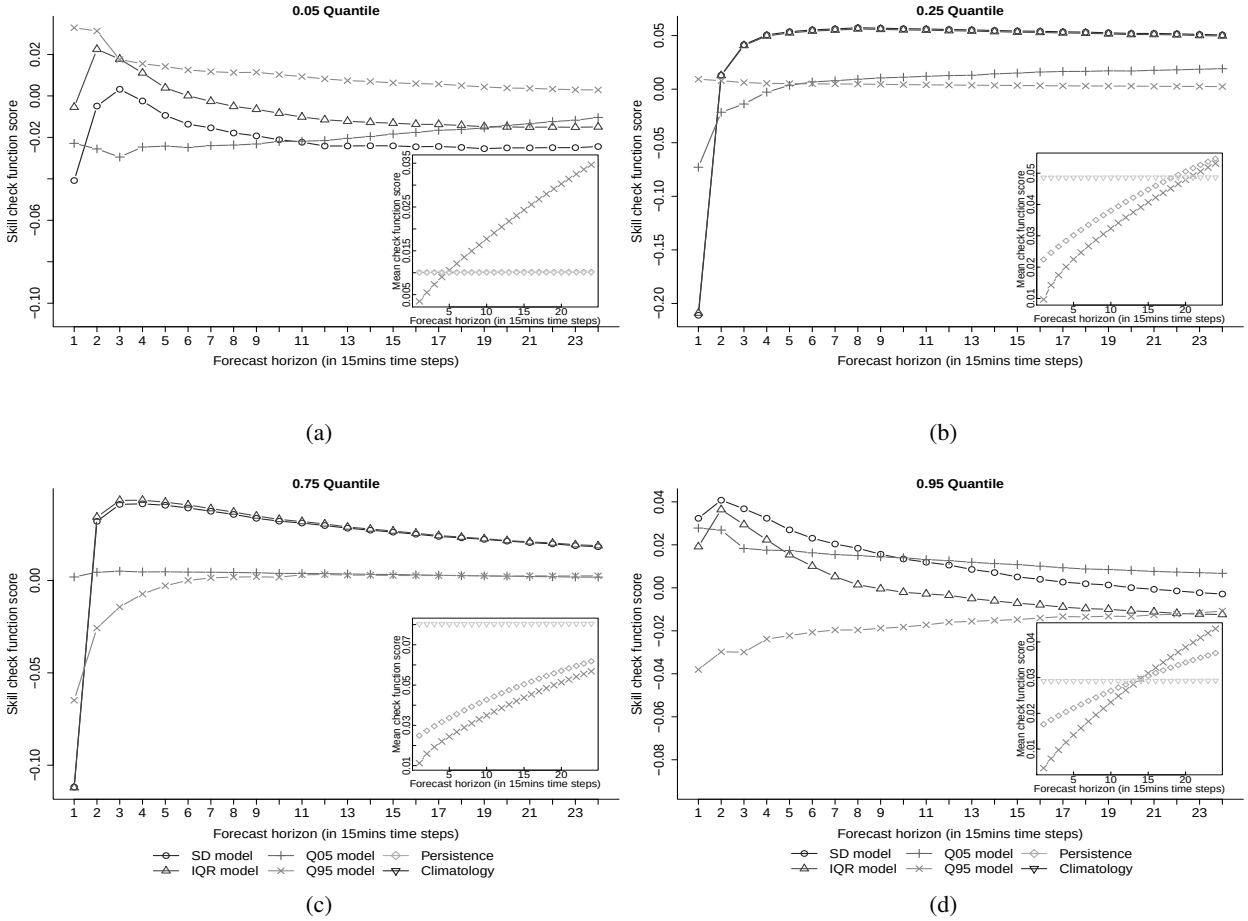
Figure 6: (a) Skill CRPS of the 1-step check function optimization models, with the corresponding 1-step CRPS optimization models as reference. (b) Skill CRPS of the 24-steps check function optimization models, with the corresponding (averaged over) 24-steps CRPS optimization models as reference. Note: Results are averaged over all wind farms.

SD or IQR models, optimized using the 1-step CRPS minimization. This is a pretty fast optimization as all parameters are the same for all the quantiles. On the other hand, in order to predict lead times larger than two time steps, it is recommended to optimize the SD or IQR models using an averaged over 24-steps check function minimization, for each quantile.

Now, we would like to compare the forecasting performance of each model, for some specific quantiles. For the purpose of these evaluations we will use the quantile loss function. Since all the models have quite similar performance for all wind farms, we will present the averaged results over the three wind farms. Moreover, for forecasting a specific quantile there is no advantage in using the CRPS optimization procedure, and hence the results presented will be only for the check function optimized models, for the specific quantiles. The chosen quantiles are the 0.05, 0.25, 0.75 and 0.95 quantiles. The 0.05 and 0.95 quantiles form the two tails of the forecasted density, and represent the rare events (such as ramps, cut offs) of a wind power series.

Figure 7 shows the Skill check function scores of the four models, with the three lagged series as the reference model. The subplots show the normalized check function score of the best model(s), together with persistence and climatology benchmarks. Note that these plots are outcomes of the averaged over 24-steps check function optimization. A first noteworthy observation is about the performances of the Q05 and Q95 models, for the lower and upper quantiles. We see that regardless of the Q95 model's structure, it is the

most appropriate for producing quantile forecasts for the 0.05 quantile for all forecasts horizons (Figure 7 (a)). In contrast, the Q05 model is the most appropriate for producing quantile forecasts (for $k > 9$) of the upper quantile, with the Q95 model performing the worst (Figure 7 (d)).



(a) (b) (c) (d)

Figure 7: Skill check function scores, averaged over the three wind farms, with the three lagged series as the reference model. The subplots show the mean check function score of the best model(s), together with persistence and climatology benchmarks. These results are outcome of the (averaged over) 24-steps check function optimization procedure, for each quantile.

For forecasting all the other quantiles (including the ones whose plots are not included), the SD and IQR seem to be the most appropriate, without again any distinguishable -in terms of performance- difference between them. Figures 7 (b) and (c) show that the 1-step ahead forecast performances of these two models are not as good as the other models' (similarly for the density forecasting evaluation before). By using a 1-step check function minimization, we manage to increase their forecasting performance for the first lead

time, but produce higher check function scores for all the other forecast horizons. Since the CRPS arises from the check function scores of the different quantiles, this result was expected (and therefore there is no need to present any quantile forecast results for the 1-step ahead optimization).

Finally, by having a look at the different subplots of Figure 7 we observe that the models outperform the persistence and climatology benchmarks only for the middle quantiles. For the lower quantile, the Q05 model outperforms these two benchmarks only for the first four lead times (up to one hour). Due to the asymmetric shape of the wind power distribution, our best models for the upper quantile outperforms the two benchmarks for up to 14 time steps ahead. For the 0.30 up to 0.85 quantiles, the SD and IQR models outperform the benchmarks for all 24 forecast horizons.

7. Conclusions and discussions

In this paper we showed how to produce wind power quantile and density forecasts, for lead times from 15 minutes up to six hours ahead, using a single univariate wind power series. In order to do that we introduced some useful variability indices, which are able to capture the volatile behavior of the wind power series.

We used linear quantile regression as our main tool for producing quantile forecasts for 19 different quantiles, with three lagged versions of the wind power series as the main explanatory variables. Four models were created, each one having as a fourth explanatory variable one of the four extracted variability indices.

In order for the final results to be consistent, we used data from three wind farms in Denmark, each one chosen to have different wind power variability (low, medium and high). We used in total four years of data, with a 15 minutes resolution, for each wind farm. The first two years were used for in-sample model calibration and training, and the final two years for out-of-sample forecasting and evaluation.

All four quantile regression models were optimized in the in-sample set, in order to find their specific set of parameters which minimize the check function score at each quantile. Each optimization run that produces a check function score for a specific quantile is quite computationally expensive, so we chose to find the global minimum in a two-dimensional grid of parameter values. In fact this approach gave us the opportunity to try four different optimizations without any extra cost: 1-step ahead check function minimization, average over 24 forecast horizons check function minimization, 1-step ahead CRPS minimization, average over 24 forecast horizons CRPS optimization. The first two give a unique set of parameters for each quantile, and the last two a set of common parameters for all quantiles.

Our main goal was to check how well these models performing compared to the persistence and climatology probabilistic benchmarks. It is worth mentioning that persistence is a very strong and simple benchmark for very short forecast horizons, and was optimized using the same cost (optimization) functions as the four regression models. The density forecasts of the models were evaluated using the CRPS. The best two models found, are the SD and IQR models. They managed to outperform the benchmarks for most of the forecasts horizons. The models of the high variability wind farm, outperform the benchmarks for all forecast horizons and with greater score differences compared to the other two farms.

The above best two models, optimized using the 1-step ahead check function minimization, outperform the models of the 24-steps check function minimization only for the first lead time. Moreover, the 1-step CRPS optimized models do not seem to have different forecasting performance for the corresponding check function optimized models. If one is interested only in the first lead time, we recommend using the 1-step CRPS optimization, as it produces models with a single set of parameters for all quantiles. On the other hand, for larger forecast horizons, the 24-steps check function optimization is more appropriate.

Acknowledgments

The authors would like to thank Energinet.dk for data provision and support. This work has been partly supported by the European Commission under the SafeWind project (ENK7 - CT2008 - 213740), Her Majestys Government and an IBM Innovation Award.

References

Barton, J. & Infield, D. (2004). Energy storage and its use with intermittent renewable energy. *Energy Conversion, IEEE Transactions on*, *19*(2), 441 – 448.

Bossavy, A., Girard, R., & Kariniotakis, G. (2010). Forecasting uncertainty related to ramps of wind power production.

Bremnes, J. (2004). Probabilistic wind power forecasts using local quantile regression. *Wind energy*, *7*, 47–54.

Brown, B. G., Katz, R. W., & Murphy, A. H. (1984). Time series models to simulate and forecast wind speed and wind power. *Journal of Applied Meteorology*, *23*, 11841195.

Davy, R., Milton, J., Russell, C., & Coppin, P. (2010). Statistical downscaling of wind variability from meteorological fields. *Boundary-Layer Meteorol*, *135*, 161–175.

Gneiting, T. (2011). Quantiles as optimal point forecasts. *International Journal of Forecasting*, *27*(2), 197 – 207.

Gneiting, T. & Raftery, A. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, *102*, 359–378.

Jursa, R. & Rohrig, K. (2008). Short-term wind power forecasting using evolutionary algorithms for the automated specification of artificial intelligence models. *International Journal of Forecasting*, *24*(4), 694 – 709.

Kariniotakis, G., Nogaret, N., & Stavrakakis, G. (1997). Advance short term forecasting of wind power production. Energy Conference, Dublin Castle, Ireland, EWEA.

Koenker, R. & Bassett, G. (1978). Regression quantiles. *Econometrica*, *46(10)*, 33–50.

Koenker, R. & D'Orey, V. (1987). Computing regression quantiles. *Applied Statistics*, *36*, 383–393.

Lau, A. & McSharry, P. (2010). Approaches for multi-step density forecasts with application to aggregated wind power. *Annals of Applied Statistics*, *4*, 1311–1341.

Makarov, Y., Hawkins, D., Leuze, E., & Vidov, J. (2002). California iso wind generation forecasting service design and experience. Windpower, Portland, OR, AWEA.

Matheson, J. & Winkler, R. (1976). Scoring rules for continuous probability distributions. *Management Science*, *22*, 1087–1096.

McSharry, P., Pinson, P., & Gerard, R. (2009). Methodology for the evaluation of probabilistic forecasts. SafeWind report.

Moeanaddin, R. & Tong, H. (1990). Numerical evaluation of distributions in non-linear autoregression. *Journal of time series analysis*, *11*, 33–48.

Moller, J., Nielsen, H., & Madsen, H. (2008). Time-adaptive quantile regression. *Computational Statistics and Data Analysis*, *52(3)*, 12921303.

Nielsen, H., Madsen, H., & Nielsen, T. (2006). Using quantile regression to extend an existing wind power forecasting system with probabilistic forecasts. *Wind energy*, *9*, 95–108.

Nielsen, T. (1999). Experiences with statistical methods for wind power prediction. European Wind Energy Conference, Nice, France, EWEA.

Pinson, P. (2010). On probabilistic forecasting of wind power time-series. *Wind Energy*.

Pinson, P. & Madsen, H. (2009). Ensemble-based probabilistic forecasting at horns rev. *Wind energy*, *12*, 137–155.

Sanchez, I. (2006). Short term prediction of wind energy production. *International Journal of Forecasting*, *22*, 43–56.

Sanchez, I. (2008). Adaptive combination of forecasts with application to wind energy. *International Journal of Forecasting*, *24*(4), 679 – 693.

Taylor, J., McSharry, P., Member, S., IEEE, & Buizza, R. (2009). Wind power density forecasting using ensemble predictions and time series models. *IEEE Transactions on Energy Conversion*, *24*, 775–782.

Tol, R. (1997). Autoregressive conditional heteroskedasticity in daily wind speed measurements. *Theoretical Applied Climatology*, *56*, 113–122.

Appendix

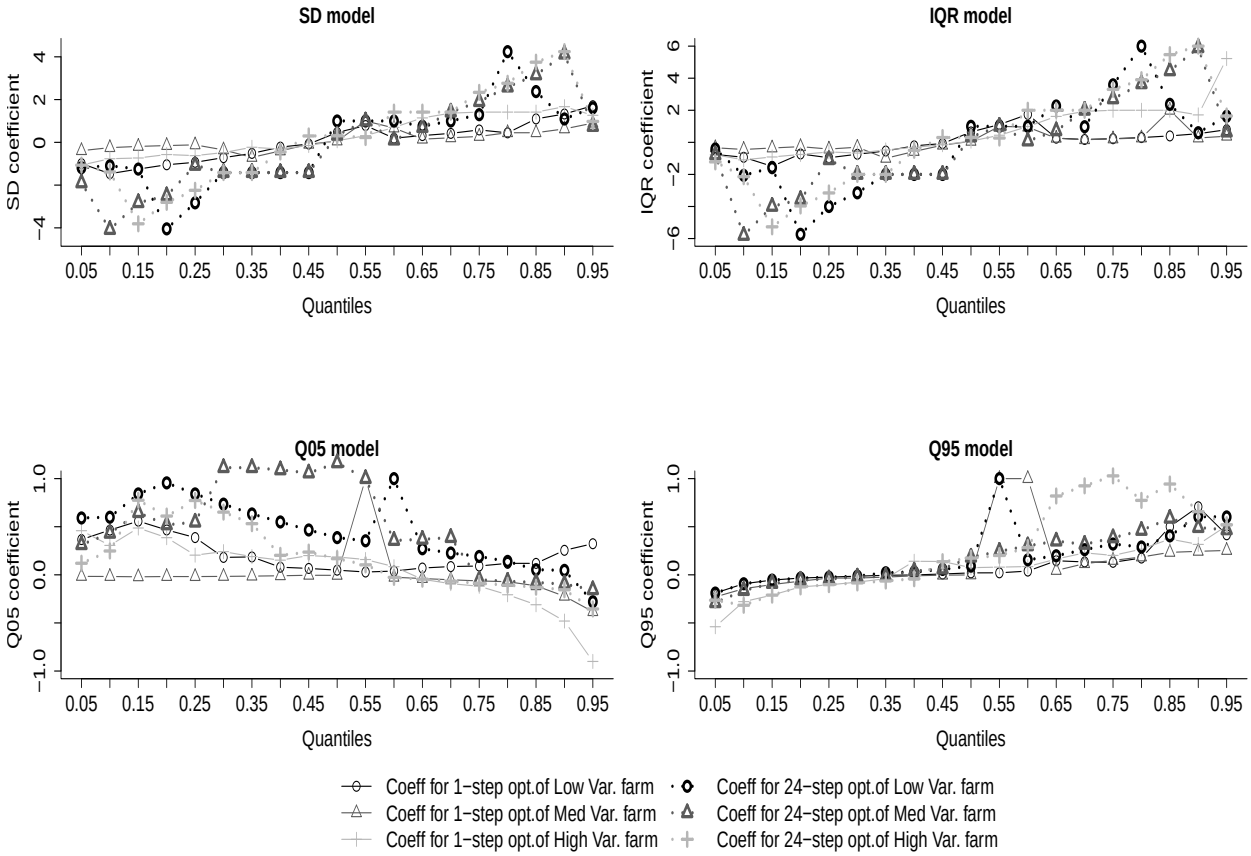


Figure 8: Variability indices' coefficients for models with 1-step check function optimization and averaged over 24-steps check function optimization.

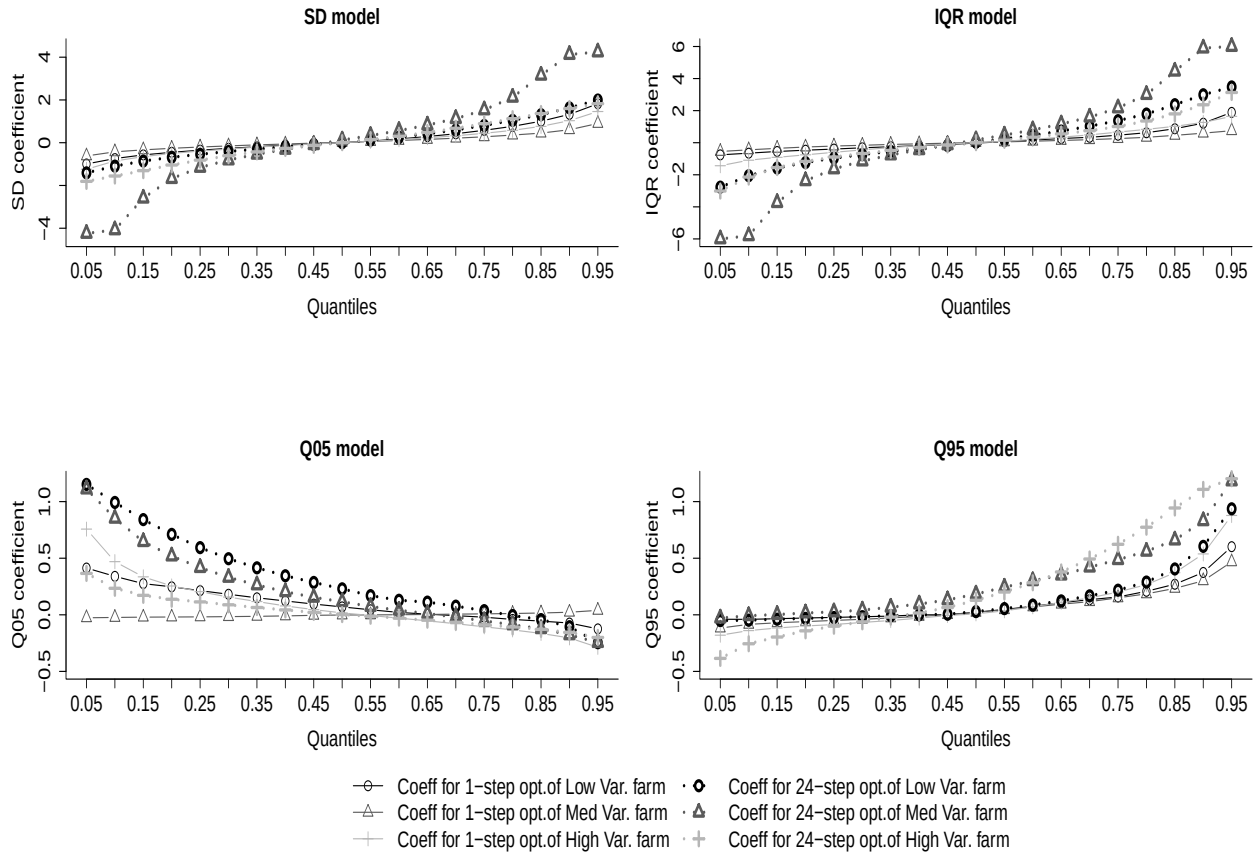


Figure 9: Variability indices' coefficients for models with 1-step CRPS optimization and averaged over 24-steps CRPS optimization.

	Low Var.		Med Var.		High Var.		Low Var.		Med Var.		High Var.	
	m	n	m	n	m	n	m	n	m	n	m	n
Model	1-step CRPS optimization						24-step CRPS optimization					
SD	0	9	0	12	0	8	0	6	3	2	0	4
IQR	2	4	0	5	0	4	2	3	3	2	2	3
Q05	0	17	24	3	0	11	0	8	0	7	0	15
Q95	2	11	2	9	0	11	0	12	5	3	0	6

Table 2: 1-step and 24-steps CRPS optimization results for all models of (18).

	Low Var.		Med Var.		High Var.		Low Var.		Med Var.		High Var.	
	m	n	m	n	m	n	m	n	m	n	m	n
Quantile	1-step optimization of **SD** model						24-steps optimization of **SD** model					
0.05	0	9	0	18	2	7	2	5	3	3	2	7
0.10	0	4	0	20	2	7	0	6	3	2	2	4
0.15	0	4	0	21	0	7	0	4	2	2	3	2
0.20	0	4	0	21	0	7	3	2	2	2	2	2
0.25	2	3	0	21	2	4	2	2	0	3	2	2
0.30	2	3	2	4	2	4	0	2	0	2	0	2
0.35	2	3	2	2	0	7	0	2	0	2	0	2
0.40	5	3	2	2	5	3	0	2	0	2	0	3
0.45	5	3	0	4	6	3	0	2	0	2	7	0
0.50	28	2	6	3	9	0	0	0	7	0	7	0
0.55	14	2	0	0	2	2	0	0	0	0	0	3
0.60	0	8	13	2	2	2	0	0	0	12	0	2
0.65	0	8	2	11	2	2	5	3	0	3	0	2
0.70	0	9	0	12	3	2	5	3	0	2	0	2
0.75	0	9	0	12	0	2	5	3	2	2	2	2
0.80	0	15	0	9	0	2	3	2	2	2	2	2
0.85	0	8	0	12	0	2	2	3	3	2	3	2
0.90	0	9	2	9	0	3	0	12	3	2	3	2
0.95	2	8	0	12	0	10	3	7	0	15	0	14
Quantile	1-step optimization of **IQR** model						24-steps optimization of **IQR** model					
0.05	2	4	0	9	0	5	2	6	2	4	2	4
0.10	0	4	0	5	0	4	2	3	3	2	2	3
0.15	0	3	0	5	0	4	2	3	2	2	3	2
0.20	0	4	0	5	0	4	3	2	2	2	2	2
0.25	2	3	0	4	0	4	2	2	0	3	2	2
0.30	2	3	0	4	2	3	2	2	0	2	0	2
0.35	2	3	2	2	5	3	0	2	0	2	0	2
0.40	5	3	2	2	5	3	0	2	0	2	0	3
0.45	5	3	0	3	6	3	0	2	0	2	7	0
0.50	28	2	6	3	9	0	0	0	7	0	7	0
0.55	14	2	0	0	2	2	0	0	0	0	0	3
0.60	11	2	13	2	2	2	0	0	0	7	0	2
0.65	2	4	0	4	2	2	5	2	0	3	0	2
0.70	0	11	0	7	3	2	5	3	0	2	0	2
0.75	0	11	0	7	0	2	5	2	2	2	2	2
0.80	0	11	0	7	0	2	3	2	2	2	2	2
0.85	0	11	0	2	0	2	2	3	3	2	3	2
0.90	0	11	0	12	0	3	0	9	3	2	3	2
0.95	2	7	0	12	4	2	5	4	0	7	0	4

Table 3: 1-step and 24-steps check function optimization results for the SD and IQR models of (18)

	Low Var.		Med Var.		High Var.		Low Var.		Med Var.		High Var.	
	m	n	m	n	m	n	m	n	m	n	m	n
Quantile	1-step optimization of **Q05** model						24-steps optimization of **Q05** model					
0.05	0	18	48	48	2	10	0	14	0	14	7	4
0.10	0	14	15	4	2	10	0	12	3	6	3	9
0.15	0	11	25	3	2	7	0	8	0	7	0	7
0.20	0	11	35	6	2	7	0	6	0	7	0	7
0.25	0	11	24	3	0	11	0	6	0	6	0	5
0.30	0	17	24	2	2	7	0	6	0	3	0	5
0.35	2	11	24	3	2	7	0	6	0	3	0	5
0.40	2	17	25	2	0	8	0	6	0	3	0	7
0.45	2	16	24	3	2	5	0	6	0	3	5	3
0.50	2	16	13	12	6	2	0	6	0	2	5	3
0.55	21	0	0	0	6	2	2	5	0	0	5	3
0.60	21	0	2	14	9	0	0	0	7	0	0	18
0.65	17	0	2	14	0	7	2	5	7	0	0	16
0.70	16	0	2	11	0	7	2	5	7	0	0	15
0.75	16	0	2	11	0	5	2	5	0	18	0	15
0.80	15	0	2	12	0	5	2	5	0	14	0	15
0.85	16	0	0	5	0	5	14	5	0	14	0	15
0.90	12	0	0	5	0	4	14	7	0	14	0	15
0.95	12	0	0	4	0	4	0	4	0	11	0	3
Quantile	1-step optimization of **Q95** model						24-steps optimization of **Q95** model					
0.05	0	4	0	6	0	4	0	4	0	4	0	8
0.10	0	6	0	6	0	4	0	6	0	5	0	5
0.15	0	6	0	6	0	5	0	6	0	5	0	5
0.20	0	8	0	6	0	5	0	10	0	10	0	8
0.25	0	9	2	18	0	7	0	10	0	10	0	8
0.30	0	10	2	13	0	7	0	10	5	19	2	8
0.35	0	45	2	13	0	7	17	0	3	22	2	8
0.40	0	45	12	14	8	0	4	6	6	3	3	7
0.45	21	0	16	11	8	0	4	6	6	3	6	2
0.50	21	0	32	3	9	0	4	6	5	3	6	2
0.55	0	18	0	0	9	0	0	0	5	3	0	6
0.60	0	18	0	0	0	10	3	7	5	3	0	6
0.65	0	11	0	16	0	9	2	8	5	3	0	4
0.70	2	11	2	9	0	9	3	7	3	5	0	4
0.75	0	15	2	9	0	11	2	8	3	5	0	4
0.80	0	15	2	9	0	11	0	12	3	5	0	6
0.85	3	7	2	9	0	11	0	12	3	5	0	6
0.90	3	7	2	10	2	11	0	12	3	6	0	10
0.95	2	13	3	10	2	11	2	11	2	9	0	14

Table 4: 1-step and 24-steps check function optimization results for the Q05 and Q95 models of (18).

Wind Power Forecasting Focused on Extreme Power System Events

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

***Abstract*— Any small improvement of the wind power forecasting performance can provide additional benefits to the end-users (TSOs, wind farm operators etc.). Several regimes can be defined based on the different wind power profiles that lead to large forecasting errors and related to specific meteorological events. The regime-switching approach gives the opportunity to predict wind power with a different predictor for each regime, reducing essentially the forecasting error. In this paper, the regime sequence is estimated by a modified ARTMAP and RBFNNs are applied as predictors. A novel adaptive learning method has been developed for the on-line learning of the applied RBFNNs. The proposed model was tested on a real wind farm and was compared with a state-of-art forecasting model.**

***Index Terms*— Wind power forecasting, regime switching, ARTMAP, RBFNN, extreme events**

I. Introduction

The large fluctuations of wind power generation due to the intermittent nature of wind are one of the main concerns of Transmission System Operators (TSOs) of power systems with high penetration of wind energy. Especially, in isolated island systems with high wind penetration, the TSO is obliged to maintain large power reserves in gas or oil generators [1]. For TSOs wind power forecasting is therefore of great importance, both for operational and financial reasons. The same applies to Independent Power Producers (IPPs) who usually participate in short term electricity markets providing bids or making bilateral contracts for their wind power production. The deviations between actual production and contracted energy are penalized depending on the energy regulations [2]. The largest penalties come from steep wind power variations that are difficult to anticipate [3].

Nowadays, wind power forecasting tools have increased revenues for both TSOs and IPPs, optimizing reserves and minimizing regulation costs. Several studies have quantified the benefits of accurate forecasts for the economic and secure operation of power systems [4-9]. For example, 12% higher average prices in the NordPool electricity market due to insufficient predictions have been estimated in [3, 9], while significant profits are estimated by the Danish TSO [10] with 1% improvement of the quality of wind power predictions. In [11] a significant income gain of more than 600 k£ for a 50 MW wind farm IPP by increasing his forecast accuracy by 1.2% is calculated.

So far, most wind power forecasting models aim to produce the best forecasts for every possible case, without any specific consideration for critical events [10]. These events, characterized as extreme power system events, are straightforward related with the low predictability of the wind power and critically affect the overall performance of the wind power forecasting models [12]. Wind power predictions are based mainly on Numerical Weather Predictions (NWP) obtained by a high resolution meteorological model. In [12] extreme power system events are defined as cases, where NWP fail to capture the temporal or spatial track of an incoming front, cases with severe variability of wind power production, high wind speeds near the cut-off limit of the wind turbine leading to a large drop of wind power and ramp events characterized by a significant wind power production increase or decrease in one hour [13]. Ramp events related to a large-scale atmospheric process can be identified by NWP quite well, however when they are related to local phenomena, like valley-hill winds, forecasting tools fail to anticipate them [14]. Extreme power system events lead to high wind power variations and consequently to high prediction errors that can be very costly for the end-users [12]. It should be noted that extreme power system events are not necessarily related to extreme meteorological events. They are considered extreme from the power system point of view, i.e. they focus on the consequences of large forecast errors in meteorological predictions, thus directly translating to large forecast errors for wind power generation.

In this paper, a novel regime switching method based on artificial intelligence is proposed in order to provide forecasts for more than one day-ahead predictions with special consideration of extreme events. Regime switching techniques have been applied in [15] for modeling wind power fluctuations in a very short time scale. The proposed model manages to improve wind power predictability considering the extreme events as a separate regime associated with the uncertainty of the NWP. For the estimation of the regimes, a neural network based on the adaptive resonance theory [16] is applied. The so called RBF-pARTMAP estimates 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)

probability of the regimes' occurrence. The wind power predictions are produced by radial basis function neural networks (RBFNN), each trained with data that correspond to a different regime. The final output of the proposed model is obtained combining the RBFNNs predictions with the probabilities of the regimes estimated by the RBF-pARTMAP network. A novel hybrid adaptive method is applied to enhance the RBFNNs with new information, based on the combined use of MRAN (Minimal Resource Allocation Network) algorithm [17] and Genetic Algorithms [18], called GA-mMRAN.

II. Definition of Regimes

In this Section, a technique to group extreme events to regimes depending on the underlying meteorological conditions is presented. A different wind power predictor for each regime is designed to improve wind power predictability, especially in these cases. For the identification of extreme events, two different power curve models are used together with a continuous wavelet analysis [19].

The first power curve model, named mPC corresponds to the wind farm manufacturer's power curve and uses the wind speed obtained by NWP to estimate the future wind generation. The second power curve model, named PC, consists of a set of multiple polynomial equations. Each equation corresponds to a different interval of the wind direction and is constructed using the corresponding wind speed and wind power values with the least square method. The wind speed and the wind direction obtained from the NWP are used.

The under-estimation or over-estimation of wind power by the power curve models indicates intensity errors and phase errors related mainly to ramp events. These errors are divided in three different regimes, as follows. When both power curve models over-estimate wind generation, the first regime is defined. This regime corresponds to phase errors that occur when the NWP predicts a front, before it reaches the wind farm location and to intensity errors that occur when the track of a front does not include the wind farm location. In addition, this regime includes errors at cut-off speeds, since wind power is always over-estimated in these cases. The second regime occurs when both power curve models under-estimate wind power. This regime corresponds to cases of phase and intensity errors that are caused by temporal and spatial estimation of a front with opposite characteristics than the previous regime. The third regime corresponds to cases where either the mPC prediction is over-estimated or the PC prediction is under-estimated. The third regime does not contain extreme events and the corresponding NWPs are considered reliable.

The three regimes described above are divided to six regimes depending on the wind power timeseries variability. For this purpose, the continuous Haar wavelet transform (CWT) [19] is applied to the wind power timeseries *p(t)*. The CWT measures the similarity between the timeseries and the wavelet function. The coefficients of CWT are calculated by the following equation.

$$C(a,b,f(t),\psi(t)) = \int_{-\infty}^{+\infty} p(t)\frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right)dt \qquad (1)$$

Parameter α is the scale that defines the spread of the wavelet ψ(t) and b is the translation parameter that determines the central position of the wavelet.

Setting a threshold to the coefficients of the CWT divides the timeseries in periods with high variability and low variability.

The classification of the events to more regimes could be achieved by dividing the coefficients of the wavelet decomposition by more thresholds and by defining periods with different scales of variability. In this case, the estimation of the future regime would be more difficult and the model complexity would be increased reducing its performance. On the other hand, the implementation of the wavelet decomposition could be avoided making the model more flexible, but also more naïve. Experiments have shown that the adopted definition of regimes is the most effective.

III. The Regime Switching Model

The proposed regime switching model is designed to improve the performance of dWPF, an existing state of the art wind power prediction model [20, 21]. In brief, the dWPF model is a combination of a self-organized map (SOM) and three RBFNN, as shown in Figure 1. It receives as input the numerical weather predictions (NWP) that correspond to the wind farm area and the most recent measurement of the wind power. The NWP comprise wind speed and wind direction that correspond to the time stamp of the forecasting hour and the previous and next time stamps.

Initially, the SOM divides the set of the weather information obtained by NWPs in three subsets, depending on the wind speed and the wind direction. Each of the above subsets is used to train a separate RBFNN, so that the model can learn how to handle unpredictable events more effectively. The RBFNNs receive as input the previous value of wind power, data from NWPs, such as wind speed and wind direction, and the hour of prediction.

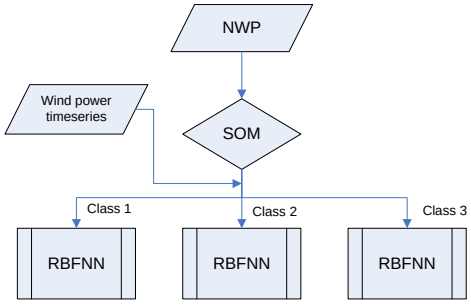


Figure 1. The structure of the dWPF model

The weather conditions related or not to a specific regime are identified using NWPs that correspond to the prediction hour and come from different NWP's updates. Usually, NWP are updated every 6 hours. The weather stability is computed from the forecasts of three updates [22]. So, the weather stability is defined by the standard deviations sd_s and sd_w of the

wind speed and of the wind direction that come from these updates and correspond to the prediction hour. Also, for the weather stability at the vertical level, forecasted wind speeds that correspond to different heights from the land surface are used. The third measure is the standard deviation *sd* of these wind speed forecasts.

Recognition of the regimes requires also estimation of the quality of the NWP. In [21], a method is described that qualifies NWPs. Accordingly, wind speed values are transformed to power values, named *theoretical* predictions, by the mPC model. Comparison of the performance of the dWPF model with the mPC model shows high correlation as expected, since the forecasting ability is dependent on the NWP accuracy. Two RBFNNs with different input variables are trained next, in order to learn the theoretical predictions. The input vector of the first RBFNN consists of the wind power and three wind speed forecasts that correspond to different time stamps close to the prediction hour. The input vector of the second RBFNN comprises the wind power, the wind direction and the forecasting hour. The quality of the NWP can be estimated by the deviations d_w and d_e between the dWPF prediction and the outputs of the two RBFNNs, respectively. Further information can be found in [21].

Based on information about weather stability and NWP quality and using the wind power timeseries and the NWP, the input vector $X_{t/t+h}$ of the regime switching model is formed (2). t is the time that the information becomes available and time $t+h$ corresponds to the prediction hour.

$$\begin{aligned} X_{t/t+h} = \{ & \hat{p}^{t+h}, d_s, d_w, \hat{ws}^{t+h-1}, \hat{ws}^{t+h}, \hat{ws}^{t+h+1}, \ldots \\ & \cos(\hat{wd}^{t+h-1}/2\pi), \sin(\hat{wd}^{t+h-1}/2\pi), \cos(\hat{wd}^{t+h}/2\pi), \ldots \\ & \sin(\hat{wd}^{t+h}/2\pi), \cos(\hat{wd}^{t+h+1}/2\pi), \sin(\hat{wd}^{t+h+1}/2\pi), \ldots \\ & sd_w, sd_d, sd, h, \mathrm{H} \} \end{aligned} \quad (2)$$

$\hat{p}^{t+h}$ is the prediction obtained by the dWPF model, h is the prediction step and H is the prediction hour. $\hat{ws}$ and $\hat{wd}$ are the wind speed and the wind direction, respectively.

Assuming that state s is a sequence of the regimes that have been defined in Section II taking values in the interval [1..6], at time $t+h$, the probability $\xi_{k,t+h}$ of the regime k, is dependent on the state s_{t+h-1} of the previous hour and the explanatory variables of the vector $X_{t/t+h}$. So, the conditional probabilities $\xi_{k,t+h}$ of each regime k can be computed by:

$$\xi_{k,t+h} = \Pr(s_{t+h} = k \mid s_{t+h-1} = l, X_{t/t+h}) \quad \forall k \in [1..6] \quad (3)$$

The state s_{t+h-1} belongs to the regime l and the state s_{t+h} at the time $t+h$ is produced by the following equation:

$$s_{t+h} = \arg\max\{\Pr(s_{t+h} = k \mid s_{t+h-1} = l, X_{t/t+h})\} \quad \forall k \in [1..6] \quad (4)$$

The state s_{t+h} and the conditional probabilities $\xi_{k,t+h}$ for each regime k are estimated by inference using RBF-pARTMAP. The training of the RBF-pARTMAP networks is described in section IV. The RBF-pARTMAP solves (3) applying Bayes' rule, as follows:

$$\xi_{k,t+h} = \frac{\Pr(s_{t+h} = k)\Pr(s_{t+h-1} = l, X_{t/t+h} \mid s_{t+h} = k)}{\Pr(s_{t+h-1} = l, X_{t/t+h})} \quad \forall k \in [1..6] \quad (5)$$

After the estimation of the regime sequence, a different RBFNN is applied to each of the 6 regimes.

The six RBFNNs are enhanced by the input parameterization [20] and the adaptive learning algorithm GA-mMRAN. In most neural network applications [23], it is effective to preprocess the input data, in order to normalize the range of all input variables in the same interval, e.g. [-1, 1] or [0, 1]. This is the case when all input variables contribute equally to the RBFNN output. In wind power forecasting however, wind speed provides 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 $\bar{w}$ is used as input that defines the relative importance of each input variable. Accordingly, the outputs $\varphi_j(\bar{x})$ of the hidden layer of a RBFNN have the following form:

$$\varphi_j(\bar{x}) = \exp\left(-\left\|\bar{w}\left(c_j - \bar{x}\right)\right\| / b\right) \quad (6)$$

$\bar{x}$ is the input vector and c_j and b are the centres and the width of the radial basis functions, respectively. The PSO algorithm [24] is used to compute the width b of the radial basis function of the applied RBFNNs. In order to achieve the best possible fitting of the RBFNNs, the cost function f_c contains performance criteria, i.e. the values obtained by the Bayesian interpolation *BI* [25], the Stein unbiased risk estimator *SURE* [26] and the mean absolute error NMAE normalized by the nominal capacity of the wind farm.

$$f_c = a_1 BI + a_2 SURE + a_3 NMAE \quad (7)$$

α_1, α_2 and α_3 coefficients are chosen in order to normalize the value of each criterion.

The applied RBFNNs are trained off-line with the following procedure. At each iteration of the PSO, the RBFNN is trained by the OLS algorithm [27] that incrementally estimates the centers of the hidden units. Then, the generalized cross-validation criterion defines the optimal number of the hidden layer's neurons. After the number of the RBFNN's neurons is defined, the performance criteria are calculated in order to estimate the generalization ability of the neural networks and are substituted in (7).

The predictions $\hat{y}^k_{t/t+h}$ obtained by each RBFNN are combined with the probabilities $\xi_{k,t+h}$ (8) providing the final prediction of the proposed regime switching model.

$$\hat{p}_{t/t+h} = \xi_{k,t+h} * \hat{y}^k_{t/t+h} \quad (8)$$

In summary, the proposed model estimates the probabilities of the six regimes that are related to the three different types of the NWP error, further divided by two levels according to the wind power variability. Each regime is handled by a different RBFNN. Independently from the regime predicted, the final

prediction of the model combines all six RBFNNs predictions weighted by the regimes probabilities. In this way, the proposed model prediction is based on the extreme power system events related straightforward to the NWP errors, while the dWPF model forecasts do not take into account these errors. This is the reason why, the proposed model forecasts better extreme events, as defined in the introduction, while the dWPF model provides better predictions of normal events.

Figure 2 shows the two steps implementation of the regime switching model. At the first step, the RBF-pARTMAP and the RBFNNs of the model are trained using the new wind power observation and the past information about the current time provided by previous predictions. Once the networks are updated, the model prediction is obtained at step 2.

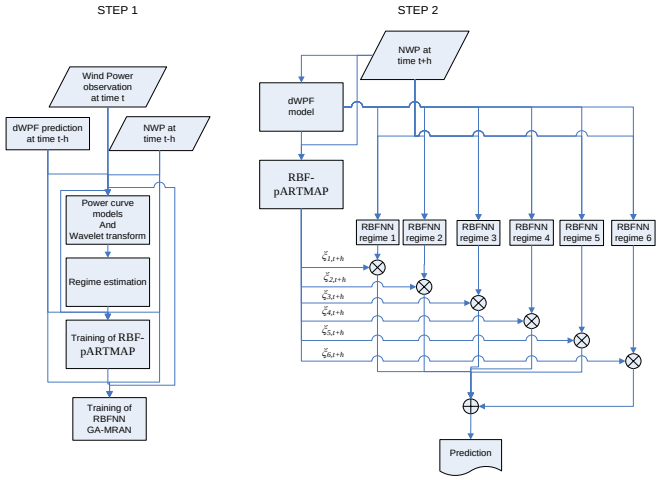


Figure 2. The implementation procedure of the proposed model

IV. Training of the RBF-pARTMAP network

ARTMAP (Adaptive Resonance Theory Map) neural networks have been used in many applications. In [16], a self-constructed network is developed, known as fuzzy ARTMAP that creates categories of the input data and the desired outputs, named hidden units and classes, respectively. More specifically, an input pattern is compared with the existing hidden units by a matching function. A threshold known as *vigilance* is set as a matching criterion that determines if the network will learn the problem finely or coarsely. The hidden units are connected with the classes through a map field. The fuzzy ARTMAP has been applied for regression or pattern recognition, but has been proven inadequate under highly noisy conditions.

In [28], an ARTMAP network, called HS-ARTMAP, is proposed that uses hyper-spherical basis functions instead of the hyper-rectangles of the ordinary fuzzy ARTMAP for the modeling of the hidden units. The HS-ARTMAP uses straightforward the vigilance criterion for the selection of the hidden units. The hyper-spherical basis functions are selected for applications in which the desired outputs have consecutive constant values. For wind power forecasting applications, it was proven [16] that NNs with Gaussian radial basis functions $\varphi_j(x)$ (RBF) have a better performance.

$$\phi_j(x) = \exp\left(-\left\|\left(w_j - x\right)\right\| / b_j^2\right) \qquad (9)$$

w_j and b_j are the mean vector and the radius of the jth Gaussian basis function, respectively The RBF-pARTMAP network introduced in this paper consists of a RBF-like module, an ART-like module and the probability estimation module. The RBF-like module is located at the input of the network and receives the input patterns. The ART-like module receives the desired output, namely the wind power. A map field consists of the synapses of the hidden units of the RBF-like module and the classes of the ART-like module. The map field provides the network knowledge to the probability estimation module. The following figure shows the structure of the RBF-pARTMAP network.

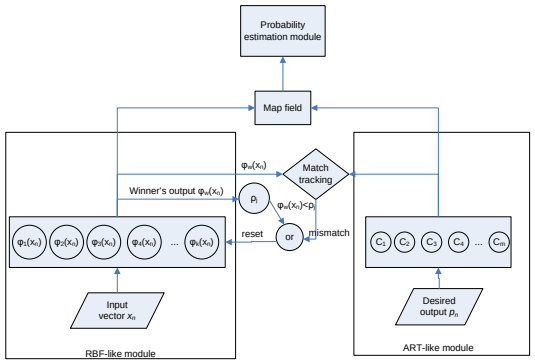


Figure 3. The structure of the RBF-pARTMAP network

The parameters of a hidden unit in the RBF-like module, i.e. the mean vector w in (9), the vigilance ρ and the radius b should be defined before training the RBF-pARTMAP,. All required parameters for the training are determined by the PSO algorithm described in Section III.

The first input pattern $x_1 \in X_{t/t+h}$ that the network receives is saved in the RBF-like module as the first hidden unit φ_1 with vigilance ρ_1 equal to ρ, radius b_1 equal to b and a match number $N_{c,1}$ equal to 1 [29]. At the same time, the corresponding desired output d_1 comprises the first class u_1 of the ART-like module. Then, the element of the map field $cw_{1,1}$ that corresponds to the first hidden unit and to the first class is set to 1. Next, when a pair of $\{x_n, d_n\}$ is received at the network input, the following procedures are executed.

a) Training of the ART-like module

In the ART-like module, the Euclidean distances between the desired output d_n and the existing classes u_i are estimated. Then, the winning neuron k^* is the index of the class with the minimum distance.

$$k^* = \arg\min_{i=1..K}\{\left\|d_n - u_i\right\|\} \qquad (10)$$

K is the number of the existing classes.

A new class in the ART-like module is created when the distance of the k^*th class from the pattern d_n is larger than a threshold value θ_d. As a result, the matching criterion of the ART-like module has the following form:

$$\left\|d_n - u_{k^*}\right\| < \theta_d \qquad (11)$$

and the winning neuron would be the new k^*+1 class.

If the matching criterion of (11) is not met, then the class u_{k^*} is updated by the following equation.

$$u_{k^*}^{new} = \frac{N_{k^*}}{N_{k^*}+1} u_{k^*}^{old} + \frac{1}{N_{k^*}+1} d_n \quad (12)$$

N_{k^*} is the number of the patterns d_n for every n that had been assigned to the k^*th class in the past.

b) Training of the RBF-like module

Since the assigned class k^* with the desired output d_n is estimated, the RBF outputs $\varphi_j(x_n)$ are computed by (9). For each input pattern x_n the hidden unit with the larger output $\varphi_{j^*}(x_n)$ is denoted as the winning neuron j^*. Then, the following conditions are checked

$$\varphi_{j^*}(x_n) \geq \rho_{j^*} \quad (13)$$

$$\left\| w_{j^*} - x_n \right\| \leq b_{j^*} \quad (14)$$

If they are not satisfied, a new hidden unit is generated and initialized as follows:

$$\begin{aligned} \rho_{J+1} &= \rho \\ w_{J+1} &= x_n \\ b_{J+1} &= b \\ cw_{J+1,k} &= 1 \\ J &= J+1 \end{aligned} \quad (15)$$

J is the number of the hidden units of the network before it is updated by the pattern $\{x_n, d_n\}$. After a new hidden unit is generated, the training algorithm is repeated with the next pattern $\{x_{n+1}, d_{n+1}\}$.

If (13) and (14) are satisfied, and the element $cw_{j^*,dn}$ of the map field is equal to 1, the algorithm terminates increasing by 1 the matching number N_{c,j^*} associated with the winning neuron [23]. If the element $cw_{j^*,dn}$ is equal to zero, then the winning neuron is disabled, another winning neuron is chosen and the algorithm starts from the beginning.

Finally, if (13) is satisfied and (14) is not, then $cw_{j^*,dn}$ is checked. If it is equal to zero, the vigilance ρ_{j^*} of the winning neuron is slightly increased by a parameter dr, defined before the training starts and the winning neuron is disabled. Another winning neuron is chosen and the algorithm starts from the beginning. If $cw_{j^*,dn}$ is equal to 1, the matching number N_{c,j^*} associated with the winning neuron is increased by 1 and the winning neuron is updated as follows:

$$b_{j^*}^{new} = \frac{b_{j^*}^{old} + \left\| x_n - w_{j^*}^{old} \right\|}{2} \quad (16)$$

$$w_{j^*}^{new} = x_n + \frac{b_{j^*}^{new}}{b_{j^*}^{old}} (w_{j^*}^{old} - x_n) \quad (17)$$

Further information on the training algorithm can be found in [22].

c) The probability estimation module

This module is activated only in the evaluation mode. During testing of the RBF-pARTMAP network, the outputs $\varphi_j(x_n)$ of the basis functions in the RBF-like module are calculated. Next, only the outputs $\varphi_j(x_n)$ of the RBF that satisfy (18) are considered, in order to estimate the conditional probabilities $\xi_{k,t+h}$.

$$\varphi_j(x_n) \geq \theta_c \quad (18)$$

The parameter θ_c is predefined as described in the next section. If subscript j´ denotes indices of the hidden units where the above condition is met and subscript j´´ denotes indices of the hidden units associated with the class k, the probabilities of (5) can be computed by:

$$\Pr(s_{t+h} = k) = \frac{N_k}{\sum_{\forall k} N_k} \quad (19)$$

$$\Pr(s_{t+h-1} = l, X_{t/t+h} \mid s_{t+h} = k) = \frac{\sum_{\forall j"} N_{c,j"}}{N_k} = 1 \quad (20)$$

$$\Pr(s_{t+h-1} = l, X_{t/t+h}) = \frac{\sum_{\forall j'} N_{c,j'}}{\sum_{\forall j} N_{c,j}} \quad (21)$$

N_k is the number of the output patterns d_n for every n that was assigned to the kth class in the past. Substituting (19)-(21) in (5), provides the conditional probabilities $\xi_{k,t+h}$:

$$\xi_{k,t+h} = \frac{N_k}{\sum_{\forall j'} N_{c,j}} \quad (22)$$

d) Estimation of the RBF-pARTMAP's parameters

The same input parameterization described in section III can be applied to the RBF-pARTMAP, since the outputs of its RBF-like module have the same form with the RBFNN hidden layer.

The optimal weighting vector $\overline{w}$ of the RBF-pARTMAP and its predefined parameters are estimated by the PSO algorithm [24]. These parameters are the initial values of the vigilance ρ and the radius b, the adjustment of the vigilance dr and the parameter θ_c of the probability estimation module. The cost function here is the mean of all failures of the network to predict the regimes in two independent data sets formed randomly.

V. On-line Training of the RBFNN

During the on-line operation of the regime switching model, an adaptive learning method is applied in order to enhance the RBFNNs with information that was not captured during off-line training. This method is applied to the RBFNN that corresponds to the regime estimated by the RBF-pARTMAP. The on-line learning method is a modification of the MRAN algorithm [17]. The MRAN algorithm creates a neural network by adding neurons depending on its performance and the distance of the input vector from its nearest neuron. Also, it

implements a pruning strategy that removes neurons with low contribution to the network's output.

Accordingly, a neuron is added to a RBFNN when:

$$\text{If } e_t > e_{\min} \text{ AND } \left\| \bar{x} - \bar{c}_{nr} \right\| > d_{thr} \qquad (23)$$

Where e_t is the absolute error of the RBFNN, $\bar{x}$ is the input vector and $\bar{c}_{nr}$ is the center of the neuron with the smaller Euclidian distance from the input vector. e_{min} and d_{thr} are constant values.

The parameters of a new neuron are set as:

$$\begin{aligned} a_{K+1} &= e_t, \\ \bar{c}_{K+1} &= \bar{x} \text{ and} \\ b_{K+1} &= \kappa \left\| \bar{x} - \bar{c}_{nr} \right\| \end{aligned} \qquad (24)$$

a is the weight of the connection of the new hidden unit to the RBFNN's output, $\bar{c}$ is the center and b is the width of the Gaussian function that represents the new hidden unit and finally, κ is the scale width factor.

If (23) is not satisfied, the input vector $\bar{x}$ and the corresponding observation y are stored in the data set belonging to the corresponding RBFNN. This data set consists of the most recent cases. So, when a new case is added in the data set, the oldest one is removed. The neurons that have been added during the on-line operation are checked next, in order to evaluate the above data set. More specifically, the sum squared error of the RBFNN is computed removing each time a different neuron or a different combination of neurons. The network with the lowest sum squared error is kept after this pruning procedure.

The neurons that have been added to the initial network and have survived the pruning procedure are tuned by the genetic algorithm. The genetic algorithm estimates the parameters of the above neurons that minimize the sum squared error of the data set. It uses the previous parameters of the neurons as initial population, it keeps one parent as elite, it selects the parent with the 'roulette' method to reproduce the children and it applies two points' crossovers to the 60% of the parents and adaptive mutation to the 40%.

For each forecasting step, the adaptive learning algorithm is implemented to the RBFNN of the regime switching model that corresponds to the regime predicted by the RBF-pARTMAP network exhibiting acceptable execution times. More specifically, the execution time of the pruning procedure is less than 30 seconds, since the maximum number of neurons that can be added is set to 21, while the maximum execution time of the genetic algorithm is set to 2 minutes. So, even if all RBFNNs of the model need to be adapted, the maximum execution time does not exceed fifteen minutes, which is clearly satisfactory for hourly predictions within a 36 h time horizon.

VI. Application to the Klim wind farm

The proposed regime switching model is evaluated at the Klim case study. 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 in Denmark. The farm contains 35 wind turbines of 600 kW and its total rated capacity is 21 MW. 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 are used to train the dWPF model while the parameters of the proposed model are estimated using the data of the second year. The length of the training period for both models is selected in order to capture all the wind attributes of the site. However, the models could be trained with data of an eight months period (four months for each one) retaining their acceptable predictability [20]. They could also be retrained automatically every four months by the PSO algorithm, as described in Section III. The evaluation data set of the proposed model covers the period of the third year 2002.

The NWP data comes from the Danish 0.15o HIRLAM model. NWP are provided for 48 hours ahead with a time resolution of one hour every six hours. The forecasting horizon of the proposed model is limited to 36 hours, because the proposed model runs every hour and uses NWP from older updates.

The evaluation framework follows the evaluation protocol for the wind power prediction models described in [30]. It consists of the normalized mean absolute error (NMAE), the normalized root mean squared error (NRMSE) and the improvements with respect to the persistence method and the simple power curve model mPC described in section II. The persistence is a naïve forecasting method that considers that the wind generation is constant for the whole forecasting horizon and equal to the most recent wind power measurement. Since the dWPF model has already been compared with other state-of-art models in [31], it is considered as the model reference in this paper.

Figure 4 shows the NMAE of the proposed model, of the dWPF model, of the persistence and of the mPC model. For the NMAE criterion, the proposed model is always better than the persistence method and the dWPF model, except from the first prediction step. More specifically, the NMAE of the proposed model does not exceed 10% for the day-ahead predictions and is always lower than 11.6%. The NMAE reduction achieved by the proposed model compared to the dWPF model is higher than 1% for forecasting horizons longer than three hours ahead. Especially, for the 8 hours ahead this reduction outreaches 2.3%. The better performance of dWPF model in the first step is due to the fact that the proposed model is trained with samples from the whole forecasting horizon, while the dWPF model is trained for one hour ahead only using the last wind power observation.

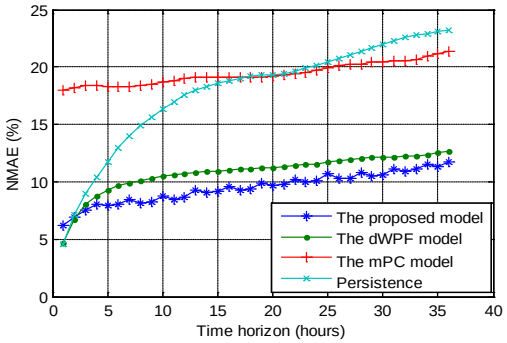


Figure 4. The NMAE of the proposed model, of the dWPF model and of the persistence for every look-ahead time

Similar results are obtained for the NRMSE criterion (figure 5). The proposed model always outperforms dWPF, except from the one-hour ahead predictions. Similar to the dWPF model, the NRMSE criterion is lower than 15% until the 25th prediction step and does not exceed 16.5%. The reduction of the NRMSE is higher than 1% for all time steps except from the first two steps. It reaches a maximum of 2.9% for the ninth prediction step.

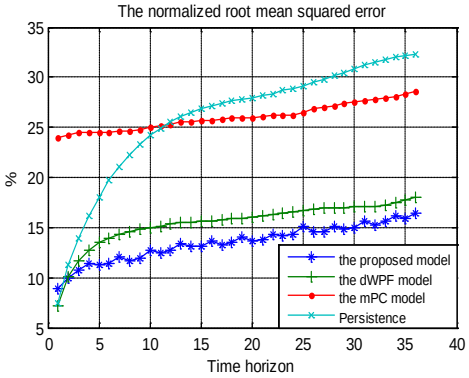


Figure 5. The normalized root mean squared error of the proposed model, of the dWPF model and of the persistence for every look-ahead time.

The overall improvement of the NMAE achieved by the proposed model with respect to the dWPF and the persistence method is illustrated in figure 6. The same figure shows the improvement of the dWPF model with respect to persistence. The improvement of the regime switching model w.r.t. persistence is always higher than 40% beyond the fifth prediction step and it ranges around 50% for horizons longer than 8 hours-ahead. The improvement w.r.t the dWPF model exceeds 22% at the ninth time step. Furthermore, at the horizons between 4 hours ahead and 30 hours ahead, it is higher than 10%.

In Table I, the average performance of the two models is presented. Regimes 1 and 2 correspond to extreme events, when wind generation is under-estimated, regimes 5 and 6 to extreme events when wind power is over-estimated and regimes 3 and 4 correspond to normal events. As expected, the proposed model outperforms the dWPF model in all regimes except regime 3 that consists of normal events with low variability.

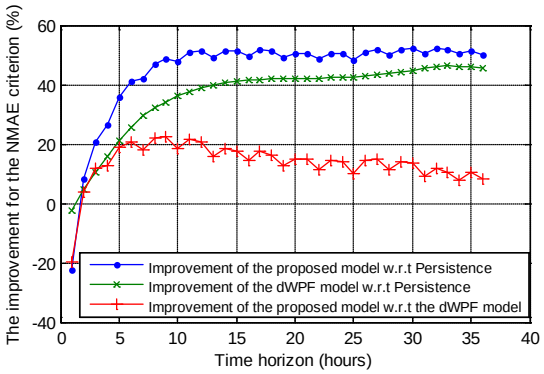


Figure 6. The improvement of the proposed model with respect to the dWPF model and to the Persistence method for every look-ahead time

TABLE I.
COMPARISON OF THE PROPOSED MODEL AND THE DWPF MODEL PERFORMANCE

	PROPOSED MODEL		dWPF MODEL	
	NMAE	NRMSE	NMAE	NRMSE
REGIME 1	**12.43%**	**16.25%**	17.01%	21.08%
REGIME 2	**15.48%**	**19.44%**	18.28%	23.65%
REGIME 3	12.47%	18.73%	**10.32%**	**15.68%**
REGIME 4	**6.33%**	**9.77%**	7.60%	11.36%
REGIME 5	**10.53%**	**13.78%**	11.94%	15.32%
REGIME 6	**10.95%**	**14.28%**	12.81%	16.34%

The superiority of the proposed model compared to the dWPF model is clearly proven in cases characterized as extreme events. Figures 7 show the predictions of the proposed model and of the dWPF model against measurements during several extreme events. All wind power values, i.e. all wind power observations and wind power predictions are expressed as percentage of the WF nominal capacity of the wind farm P_n.

Figure 7a shows an example of an up-ramp event. The wind generation is over-estimated from both mPC and dWPF models. At the first 18 time steps, the state belongs to regime 2 and then it changes to regime 1. At the 23rd time step it belongs again to regime 2, changing back at the 26th time step. It returns to regime 2, from 28th to 30th and from 35th to 36th steps.

An intensity error is illustrated in figure 7b. A wrong estimation of the track of the front by the NWP has led to a huge prediction error. At the first 6 time steps, the state is fluctuates between regime 3 and regime 4. At the 7th time step, it changes to regime 6 and returns to regime 3 at the 11th step. At the 17th step, it changes to the regime 1 and remains until the end of the forecasting horizon.

In figure 7c, a phase error is presented. It is clear that the proposed model handles the NWP better than the dWPF model. At the beginning, the state fluctuates between the regimes 1 and 4. At the 9th step it is stabilized at regime 2 until the 20th time step and then it changes to regime 1. Finally, at the 32nd time step, it turns to regime 6.

Similarly, in the case of a down-ramp event shown in figure 7d, the state takes values between 5 and 6. Finally, in figure 7e the state is between the regimes 5 and 6 at the beginning, turns to the regimes 1 and 2 and at the end fluctuates between the regimes 3 and 4.

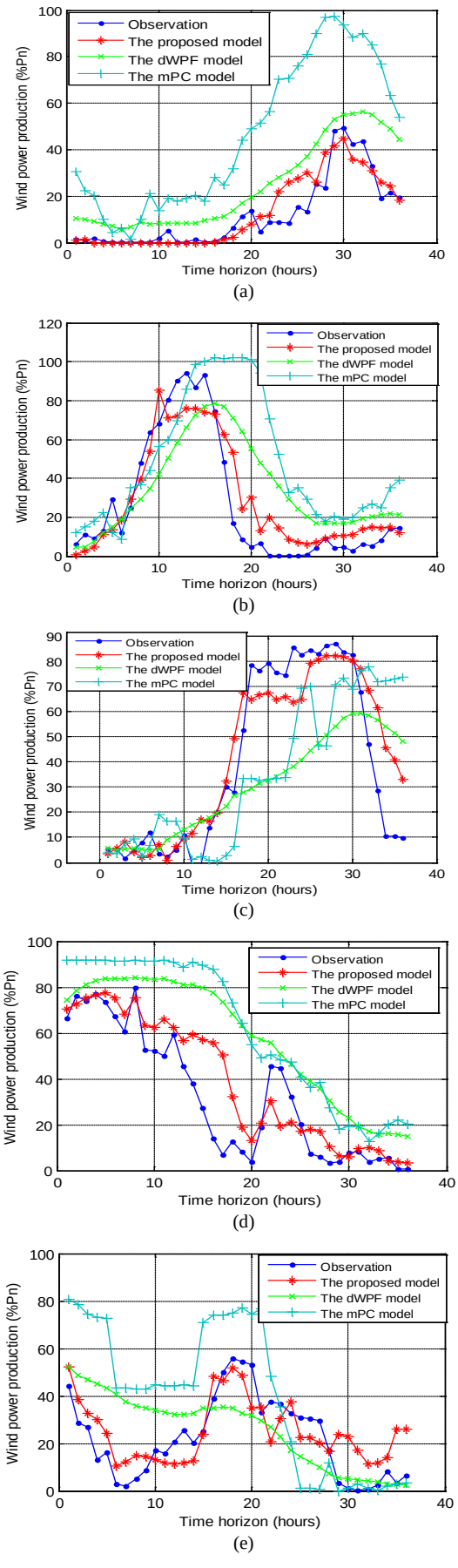


Figure 7. The predictions of the proposed model, of the dWPF model and the mPC model against measurements during several extreme power system events: a) on up-ramp event, b) on intensity error event, c) on a phase error event, d) on a down-ramp event and e) on a high wind power fluctuation event.

VII. Application to Lasithi Case Study

The Lasithi wind farm is located at the eastern part of Crete in Greece. The wind farm consists of 30 wind turbines BONUS 600 MK4 and its nominal capacity P_n is equal to 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 available wind power timeseries cover the period from 1st January of 2006 to 31st March 2008. The timeseries of the first three 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 NWP come from the European Centre of Medium-range Weather Forecasts (ECMWF). They are provided two times per day with ten days forecasting horizon.

Figure 8 shows the performance of the proposed regime switching model and the dWPF model. Due the longer horizon, the dWPF model outperforms the proposed model at the first nine time steps. However, the performance of the proposed model is better than the dWPF model at the horizons longer than 10 hours ahead. The improvement outreaches the 1% for the NMAE criterion at the most forecasting steps.

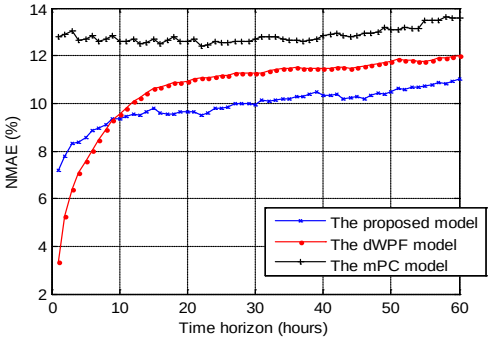


Figure 8. The NMAE of the proposed model, of the dWPF model and of the mPC model for every look-ahead time

VIII. Conclusions

The paper presents a novel regime switching model that improves the predictability of a state-of-art wind power prediction model, especially at extreme events. The extreme events are grouped to six different regimes by combining the predictions of two power curve models with a continuous wavelet decomposition of the wind power timeseries. Next, the Bayes rule and the radial basis function models are integrated in a HS-ARTMAP network forming the RBF-pARTMAP network. The RBF-pARTMAP is used to estimate the probability occurrence of each regime. The final prediction is obtained from the combination of the regimes probabilities with the predictions of the six RBFNNs. For on-line operation, a novel adaptive learning algorithm is applied that enhances the RBFNNs performance using the new observations. The

evaluation of the regime switching model at the Klim and Lasithi wind farms shows significant improvement of wind power forecasting over state of the art models, especially in cases of extreme events. The proposed model can be applied not only for wind power forecasting of individual wind farms but also, of larger regions and can be used effectively by TSOs for mid-term operational planning.

IX. References

[1] Hatziargyriou N., et al, "Energy Management and Control of Island Power Systems with Increased Penetration from Renewable Energy Sources", IEEE PES 2002 Winter Power Meeting, New York, January 2002.

[2] Parsons B., Milligan M., Zavadil R., Brooks D., Kirby B., Dragoon K., Caldwell J., "Grid Impacts of Wind Power: A Summary of Recent Studies in the United States," EWEC Wind Energy Journal, June, 2003

[3] Sørensen, B. and Meibom P., 'Can Wind Power be Sold in a Deregulated Electricity Market?' Proc. of the European Wind Energy Conference, Nice, France, 1-5 March 1999, pp. 375-378.

[4] Hatziargyriou N.., et al, '"MORE CARE" Control Advice for Secure Operation of Isolated Power Systems with Increased Renewable Energy Penetration & Storage', Proc. of the European Wind Energy Conference, Copenhagen, Denmark, 2-6 June 2001, pp. 1142-1145.

[5] ANEMOS Project. *'Wind prediction in electricity markets'*. ANEMOS Project, Deliverable D8.2. [Online]. Available: http://anemos.cma.fr/.

[6] Usaola J, Ravelo O, González G, Soto F, Carmen Dávila M, Díaz–Guerra B. 'Benefits for wind energy in electricity markets from using short term wind power prediction tools; a simulation study'. *Wind Engineering* 2004; 28: 119–128.

[7] Tsikalakis A.G., Hatziargyriou N. D., Katsigiannis Y. A., Georgilakis P. S., 'Impact of wind power forecasting error bias on the economic operation of autonomous power systems', *Wind Energy*, vol. 12, issue 4, pp. 315-331.

[8] Barthelmie, R.J., Murray F., and Pryor S.C. 'The economic benefit of short-term forecasting for wind energy in the UK electricity market'. *Energy Policy*, 36(5), pp. 1687-1696, 2008.

[9] Nielsen, L.H., et al, 'Wind Power and a Liberalised North European Electricity Exchange', Proc. of the European Wind Energy Conference, Nice, France, 1-5 March 1999, pp. 379-382,

[10] ANEMOS Deliverable 1.1 '*The State-Of-The-Art in Short-Term Prediction of Wind Power: A Literature Overview*', 2003

[11] 1 Parkes, J., Wasey J., Tindal A., Munoz L. 'Wind Energy Trading Benefits Through Short Term Forecasting'. *Proc. of the European Wind Energy Conference and Exhibition*, Athens (GR), 27 Feb-2 Mar 2006

[12] Pinson, P. and the SafeWind project consortium, '*Defining a catalogue of extreme events*'. EWEC'10, Warsaw, Poland, 20-23 April 2010.

[13] Zack, J.W., Young S., Cote M., Nocera J., Aymami J. and Vidal J., 'Development and Testing of an Innovative Short-Term Large Wind Ramp Forecasting System'. *Proc. of the European Wind Energy Conference and Exhibition*, Warsaw (PL), 20-23 Apr 2010

[14] Möhrlen, C., Jørgensen J., Sattler K., McKeogh E., 'On the accuracy of land cover data in NWP forecasts for high resolution wind energy prediction', *Proc. of the European Wind Energy Conference*, Copenhagen, Denmark, 2-6 June 2001, pp. 854-857

[15] P. Pinson et al, "Regime-switching modelling of fluctuations of offshore wind generation", *J. Wind Eng. Ind. Aerodyn*, vol 96, pp. 2327– 2347, 2008

[16] Carpenter G. A., Grossberg S., & Reynolds J. H. "ARTMAP: Supervised real-time learning and classification of nonstationary data by a self-organizing neural network", *Neural Networks*, vol. 4, pp. 565–588, 1991

[17] Sundararajan N., Saratchandran P., and Yingwei L., '*Radial Basis Function Neural Networks with Sequential Learning: MRAN and Its Applications*', Singapore: World Scientific, 1999

[18] Goldberg DE, '*Genetic algorithms in search optimization and machine learning'*, Addison-Wesley, Reading MA, 1989

[19] Chui, C.K., '*Wavelets: a tutorial in theory and applications'*, Academic Press, 1992.

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

[23] Simon Haykins, *'Neural networks'*, Pearson-Prentice Hall, second edition, 2005

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

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

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

[27] 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.

[28] M. Su, J. Lee, and K. Hsieh, "A new ARTMAP-based neural network for incremental learning," *Neurocomputing*, Vol.69, pp. 2284-2300, 2006

[29] S.J. Verzi, G.L. Heileman, M. Georgiopoulos, M.J. Healy, ''Boosted ARTMAP,'' *IEEE International Joint Conference on Neural Networks Proceedings*, IEEE World Congress on Computational Intelligence, New York, USA, 1998, pp. 396–401.

[30] 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, issue 6, 475–489, 2005

[31] G. Kariniotakis, I. Martí, D. Casas, P. Pinson, T. S. Nielsen, H. Madsen, J. Giebel, G. and Usaola, I. Sanchez, A. M. Palomares, R. Brownsword, J. Tambke, U. Focken, P. Lange, M. and Louka, G. Kallos, C. Lac, G. Sideratos, and G. Descombes, 'What performance can be expected by short-term wind power prediction models depending on site characteristics?'. in Proc. of EWEC'04, London, UK., Nov. 2004

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 Dipl.-Electr. Mech. Eng. degree 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.