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

Trombe, Pierre-Julien ; Pinson, Pierre

Publication date:
2012

Document Version
Publisher's PDF, also known as Version of record

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].

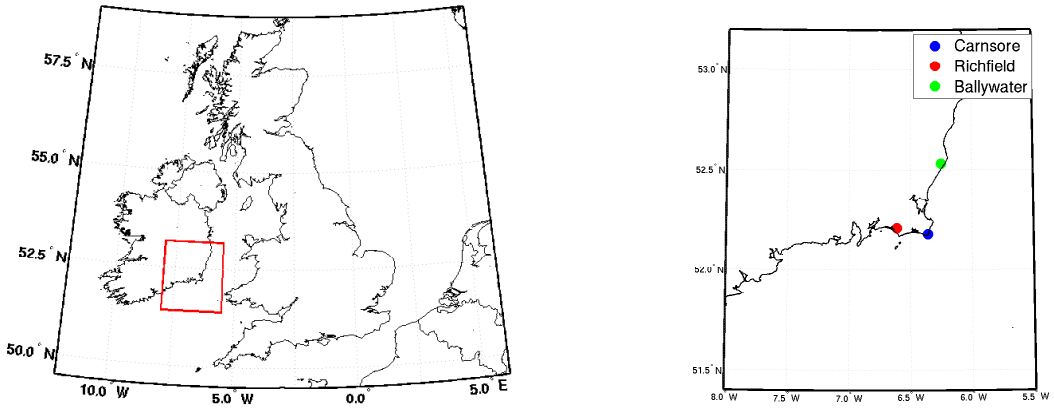


FIGURE 1: *The Carnsore, Richfield and Ballywater wind farms are located in the South-East of Ireland. Carnsore and Richfield are separated by an approximate distance of 17 km, and the distance between Richfield and Ballywater is 40 km.*

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

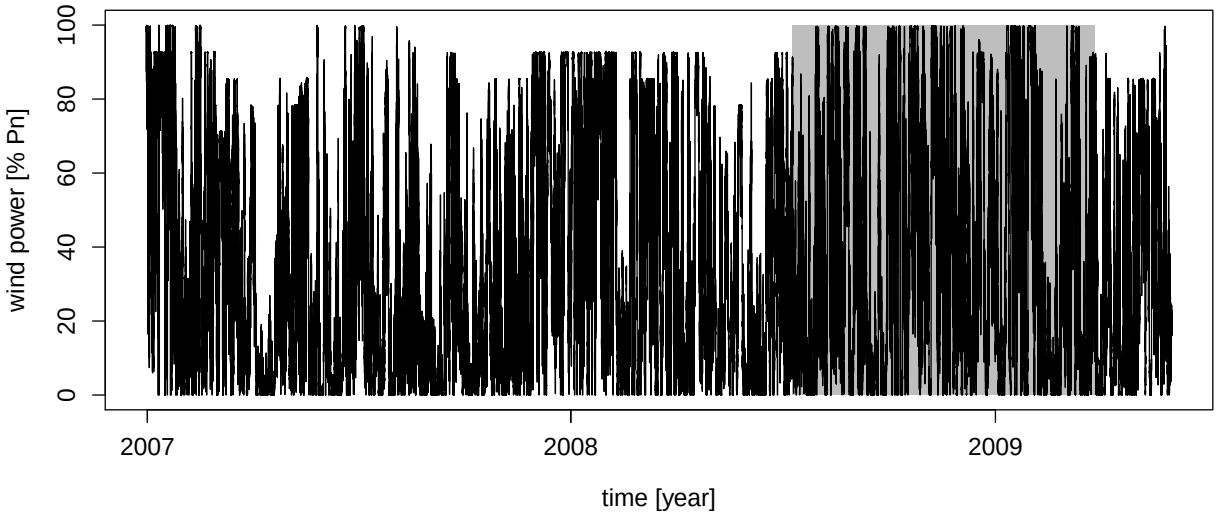


FIGURE 2: *Time series of wind power at Carnsore. The data overlaying the shaded area are considered to be of good quality and used for the experimental part of this study.*

(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.

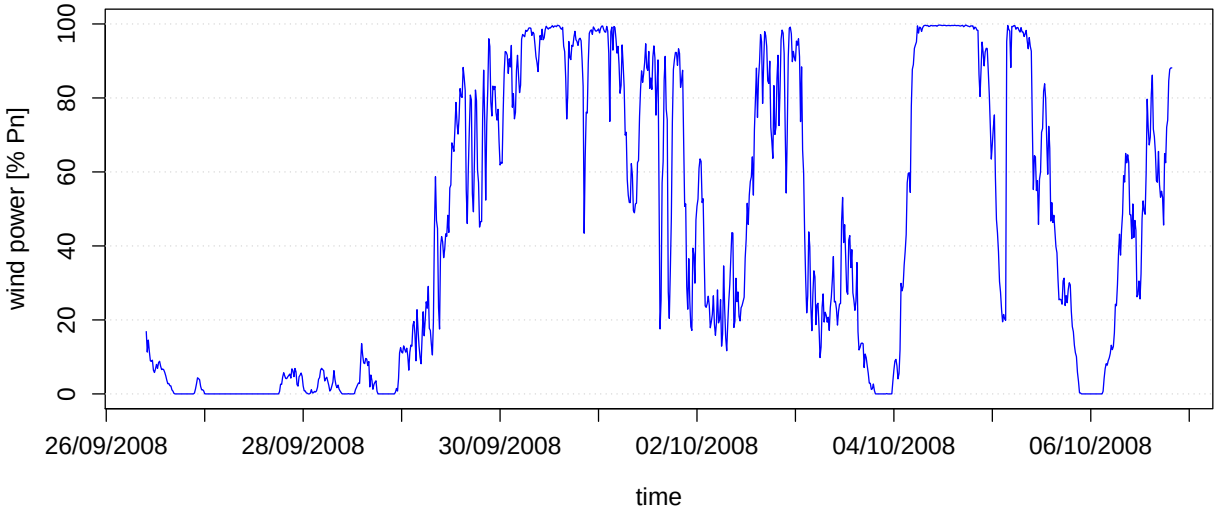


FIGURE 3: *Normalized wind power generation at Carnsore. The temporal resolution of 15 minutes.*

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

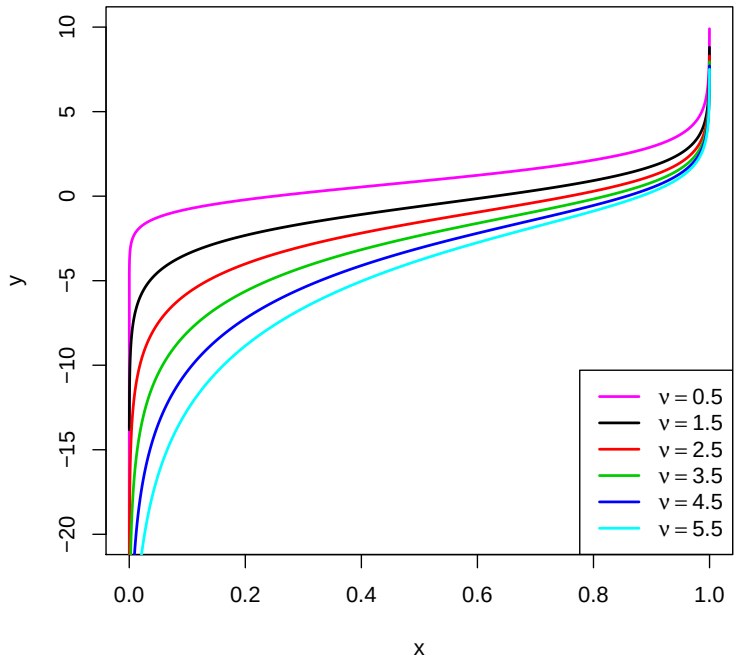


FIGURE 4: *The GLN predictive distribution consists of transforming the original wind power observations in order to deal with the heteroscedasticty near the bounds of the interval* $[0, 1]$

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.

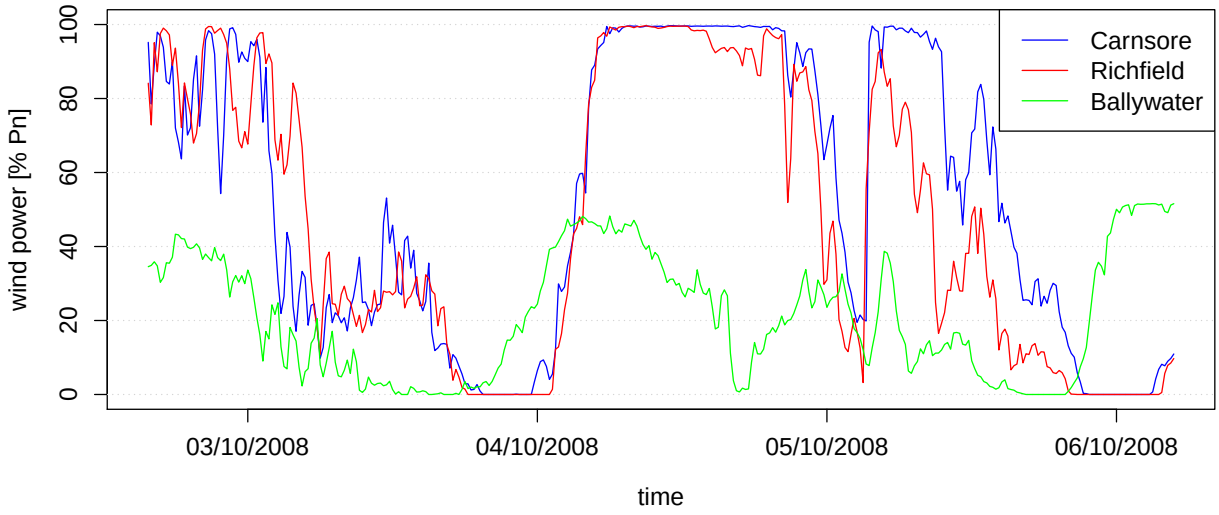


FIGURE 5: *Normalized wind power generation at Carnsore, Richfield and Ballywater with a temporal resolution of 15 minutes.*

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:

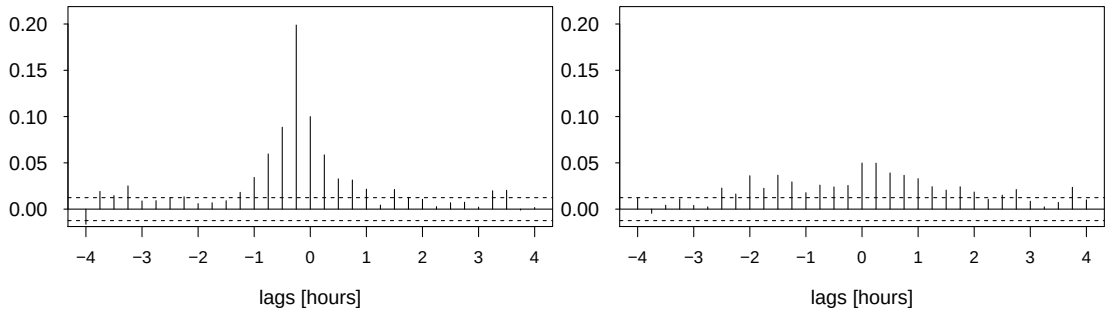
1. An appropriate Autoregressive Moving Average (ARMA) model is fitted to the input series $\{x_t^{(B)}\}$ and a series of residuals $\{e_t^{(B)}\}$ extracted,
2. The output series $\{y_t^{(A)}\}$ is filtered with the same model as in step 1 and a series of residuals $\{e_t^{(A)}\}$

extracted,

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

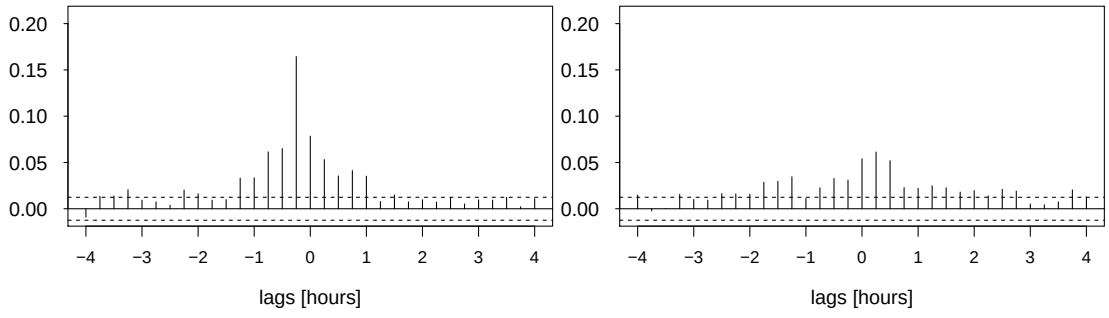
$$\rho_{e^{(A)}e^{(B)}}(\tau) = \frac{\text{cov}(e^{(A)}(t), e^{(B)}(t+\tau))}{\sigma_{e^{(A)}}\sigma_{e^{(B)}}} \tag{5}$$

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,



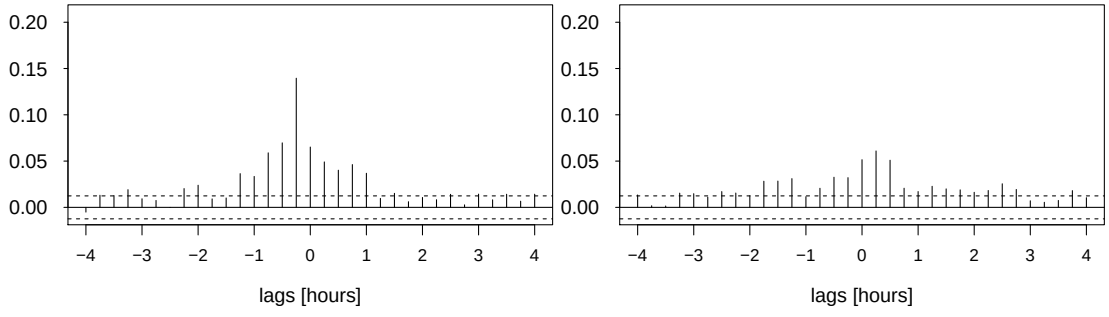
(a) Richfield & Carnsore, No GLN transf.

(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.

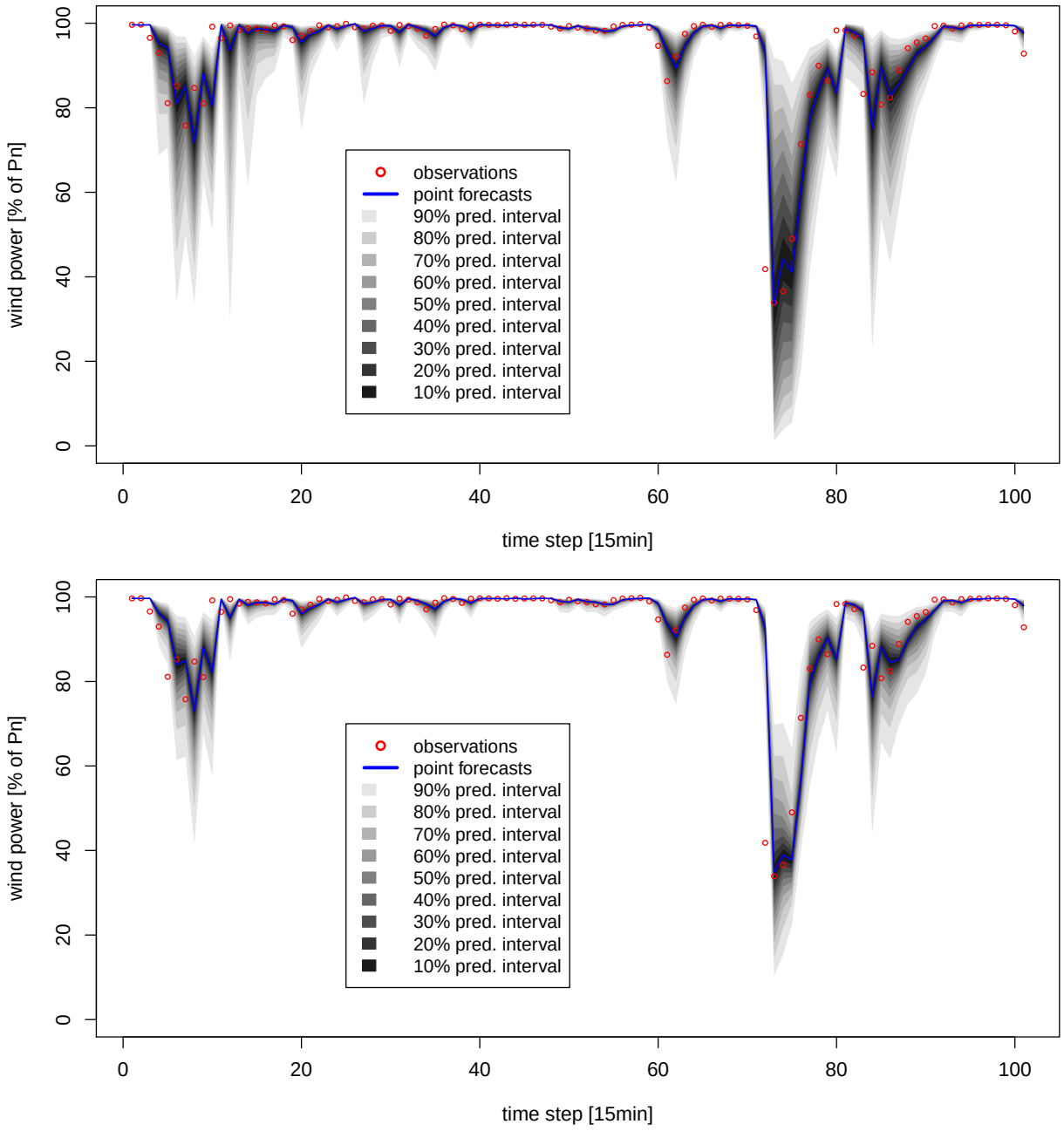


FIGURE 7: *Example 1. Time series of normalized wind power generation at Carnsore and one-step ahead point forecasts and prediction intervals with nominal coverage from 10 to 90%. ARX-GARCH-GLN model (Top panel), MSAR-X-GLN model (Bottom panel).*

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

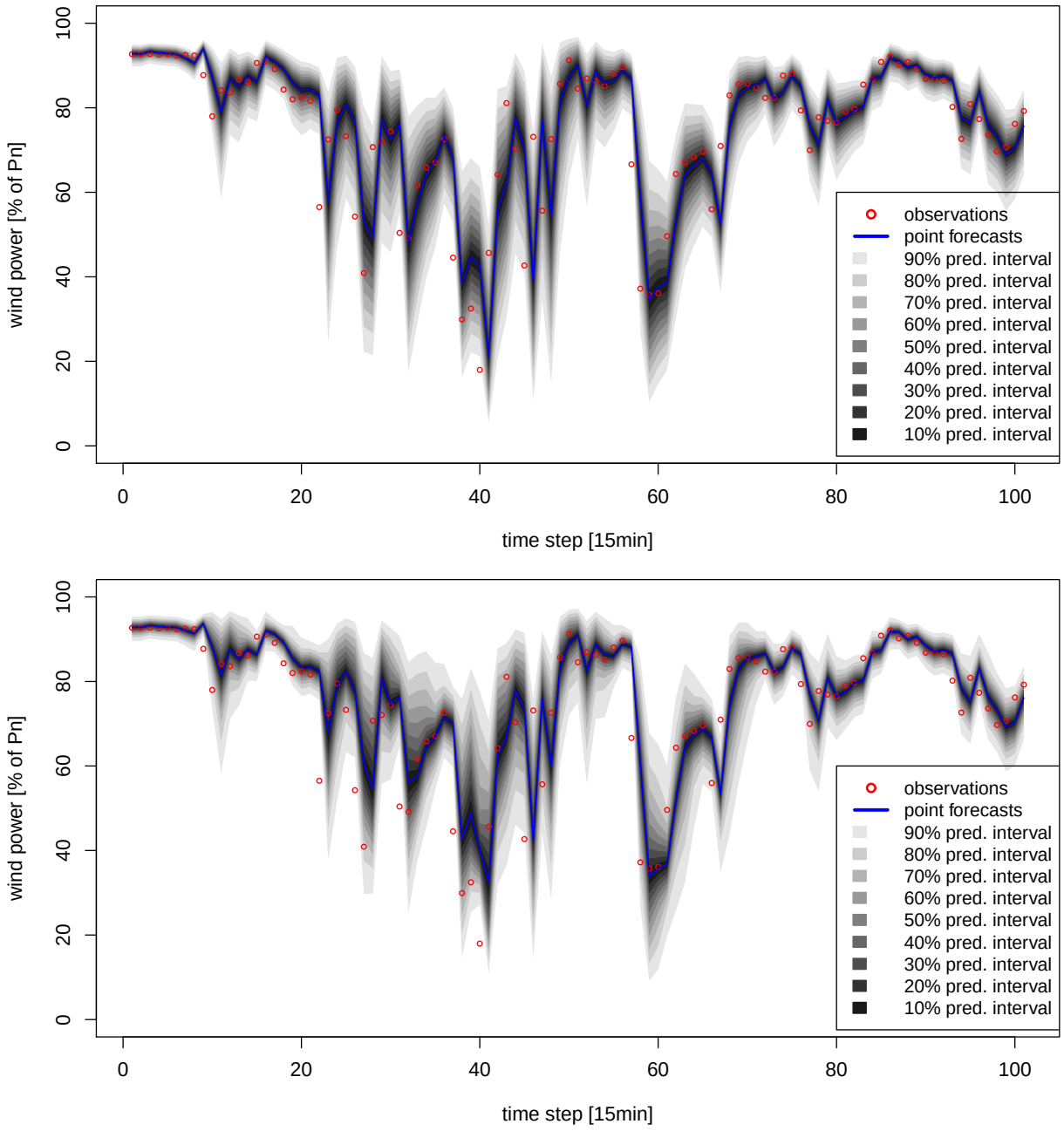


FIGURE 8: *Example 2. Time series of normalized wind power generation at Carnsore and one-step ahead point forecasts and prediction intervals with nominal coverage from 10 to 90%. ARX-GARCH-GLN (Top panel) model, MSARX-GLN model (Bottom panel).*

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.

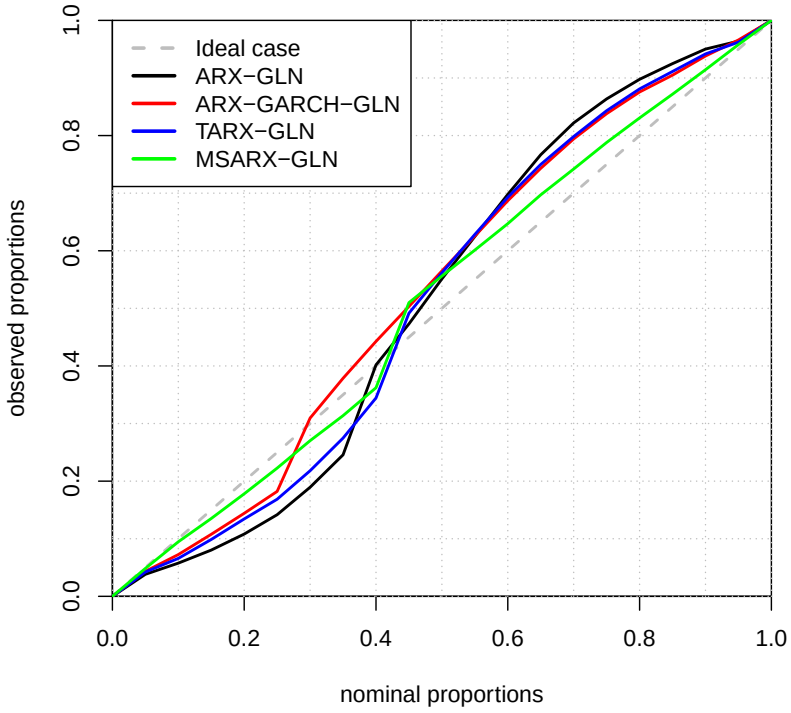


FIGURE 9: *Reliability diagram of predictive densities of wind power.*

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.