Downloaded from orbit.dtu.dk on: Nov 07, 2026

DTU Library

Adaptive modelling and forecasting of offshore wind power fluctuations with Markov-switching autoregressive models

Pinson, Pierre; Madsen, Henrik

Published in:
Journal of Forecasting

Link to article, DOI:
10.1002/for.1194

Publication date:
2012

Document Version
Early version, also known as pre-print

Link back to DTU Orbit

Citation (APA):
Pinson, P., & Madsen, H. (2012). Adaptive modelling and forecasting of offshore wind power fluctuations with Markov-switching autoregressive models. *Journal of Forecasting*, *31*(4), 281–313. https://doi.org/10.1002/for.1194

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

submitted to Journal of Forecasting (October 1, 2009)

Adaptive modelling and forecasting of offshore wind power fluctuations with Markov-switching autoregressive models

Pierre Pinson[*]**, Henrik Madsen**

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

Abstract

Wind power production data at temporal resolutions of a few minutes exhibits successive periods with fluctuations of various dynamic nature and magnitude, which cannot be explained (so far) by the evolution of some explanatory variable. Our proposal is to capture this regime-switching behaviour with an approach relying on Markov-Switching AutoRegressive (MSAR) models. An appropriate parameterization of the model coefficients is introduced, along with an adaptive estimation method allowing to accommodate long-term variations in the process characteristics. The objective criterion to be recursively optimized is based on penalized maximum-likelihood, with exponential forgetting of past observations. MSAR models are then employed for 1-step-ahead point forecasting of 10-minute resolution time-series of wind power at two large offshore wind farms. They are favourably compared against persistence and AutoRegressive (AR) models. It is finally shown that the main interest of MSAR models lies in their ability to generate interval/density forecasts of significantly higher skill.

Keywords: wind power forecasting, regime-switching, adaptive estimation, point forecasting, interval forecasting

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

1 Introduction

Future developments of wind power installations are more likely to take place offshore, owing to space availability, less problems with local population acceptance, and more steady winds. This is especially the case for countries that already experience a high wind power penetration onshore, like Germany and Denmark. This latter country hosts the two largest offshore wind farms worldwide: Nysted and Horns Rev, whose nominal capacities are of 165.5 and 160 MW, respectively. Today, each of these wind farms can supply alone 2% of the whole electricity consumption of Denmark (Mandelbaum 2002). Danish authorities have actually formulated the objectives of having 50% of the electricity consumption met by wind energy at the horizon 2025 (Ea Energy Analyses 2007), mainly by the commissioning of more large offshore wind parks. Such large offshore wind farms concentrate a high wind power capacity at a single location. Onshore, the same level of installed capacity is usually spread over an area of significant size, which yields a smoothing of power fluctuations (Focken *et al.* 2002). This smoothing effect is hardly present offshore. The magnitude of power fluctuations may thus reach very significant levels. Characterizing and modelling the power fluctuations for the specific case of offshore wind farms is a current challenge (Akhmatov 2007; Hendersen *et al.* 2003), for better forecasting offshore wind generation, developing appropriate control strategies, or alternatively for simulating the combination of wind generation with storage or any form of backup generation. A discussion on these aspects is available in Sørensen *et al.* (2007). So far the focus in the wind power forecasting literature has been on forecast horizons in the range of hours to days for energy trading, scheduling and economic dispatch, see (Costa *et al.* 2008; Giebel *et al.* 2003) and references therein. As wind power penetration reaches a certain level, and with large-scale wind farms injecting power at a single point of the electricity network, it becomes crucial to have new types of forecasts with horizons of a few minutes ahead. Two example applications are: *(i)* the optimal operation of reserves for the case of the Transmission System Operator (TSO), for which the relevant horizons is of 10-15 minutes (Sørensen *et al.* 2007); or *(ii)* the use of forecasts as input to the (offshore) wind farms controllers themselves, for which the typical time step is of 5-10 minutes (Kristoffersen and Christiansen 2003).

When inspecting offshore wind power production data averaged at a few-minute rate, one observes sudden changes in the fluctuations characteristics at time scales of a few hours, which so far might be best explained by local atmospheric changes e.g. frontline passages and rain showers (Sørensen *et al.* 2008). Actually, analysis of corresponding wind speed data in the frequency domain reveals that changes in the frequency and magnitude of fluctuations may be the combination of diurnal, mesoscale, and local effects (Vincent *et al.* 2008). These meteorological phenomena add complexity to the modelling of wind power production, which is already non-linear and bounded owing to the characteristics of the wind-to-power conversion function, the so-called power curve. Such succession of periods with power fluctuations of lower and larger magnitudes calls for the use of regime-switching models. Recently, Pinson *et al.* (2008) showed that for the case of the Nysted and

Horns Rev wind farms, Markov-switching approaches, i.e. for which the regime sequence is not directly observable but is assumed to be a first-order Markov chain, were more suitable than regime-switching approaches relying on an observable process e.g. using Smooth Transition AutoRegressive (STAR) models. Note that the idea of employing Markov-Switching AutoRegressive (MSAR) models has also been proposed by e.g. Ailliot *et al.* (2006) for the modelling of the spatio-temporal evolution of wind fields. Despite the final application being somewhat different, the developments of Ailliot *et al.* (2006) also go along the idea that regime switches in wind-related processes may be appropriately modelled with a hidden Markov Chain. Though, a drawback of the Markov-switching approach described by Pinson *et al.* (2008) is that model coefficients are not time-varying, while it is known that wind generation is a process whose dynamic features have smooth long-term variations, at time scales of weeks to months (Costa *et al.* 2008; Giebel *et al.* 2003). Focusing on wind speed first, a number of recent papers have shown that statistical models employed for modeling and forecasting at time-scales in the order of a few hours would clearly benefit from adaptive estimation. If taking the example of the space-time regime-switching model of Gneiting *et al.* (2006), or of the multivariate wind vector model of Hering and Genton (2009), they both concluded that an optimal size for the sliding window used for adaptive estimation is of 45 days. Similarly for the question of (probabilistic) forecasting of wind speed at time scales of hours to days, Sloughter *et al.* (2009) showed that estimation over a sliding window of fairly similar size leads to optimal forecast performance. This is due to some smooth variations in local wind dynamics, which are difficult to account for in a parametric form directly in the models. When considering offshore wind speed at a time resolution of a few minutes, recent works by Vincent *et al.* (2008) (using 4 years of wind speed data recorded at the Horns Rev wind farm) have clearly demonstrated how the spectral characteristics of offshore wind speed exhibit frequent and abrupt changes, in parallel to smooth variations at the time-scales of months and seasons. An important implication of the analysis works of Vincent *et al.* (2008) is that assuming stationarity of wind power fluctuations dynamics does not appear appropriate. These smooth variations in the characteristics of the process dynamics are more complex than a simple periodicity in the signal or in the coefficients of a dynamic model. Furthermore, if focusing then on the specific case of wind power, one should realize that wind power generation is a complex nonlinear function of meteorological conditions, not of wind speed and direction only. Air density and temperature for instance can influence the power curve of a wind farm. There can also be more practical (engineering) reasons to consider that the wind power generation dynamics smoothly evolve with time, related to the 'life' of the wind farm itself. The dynamic behaviour of each individual turbine does change with time, owing to varying shadowing effects, ageing, failures, maintenance, non-optimal facing of the wind, potentially decommissioning and recommissioning, etc. The need for adaptivity of statistical models when considering wind power generation dynamics has already been clearly motivated by e.g. Sanchez (2006) based on example data. A large number of recent works on wind power modeling and forecasting follow, motivate and support the idea such that adaptive estimation is a key feature of statistical models employed, e.g. (Møller *et al.* 2008; Pinson and Madsen 2009; Sanchez 2008).

The main objective of the present paper is to introduce a MSAR model whose coefficients are adaptively and recursively estimated, with application to the modelling and forecasting of offshore wind power fluctuations. The parameterization of the model coefficients employed here is inspired by those initially proposed in (Collings *et al.* 1994; Collings and Rydén 1998) and in (Holst *et al.* 1994). Adaptivity in time is achieved with exponential forgetting of past observations. In addition, the formulation of the objective function to be minimized at each time-step includes a regularization term that permits to increase the generalization ability of estimated models, in addition to improving numerical stability of the recursive estimation procedure. The motivation for introducing a recursive estimation method is to lower computational costs by updating parameter estimates based on newly available observations only. Note that the meaning of recursive estimation is here different with that commonly employed in the context of Markov Chain Monte Carlo (MCMC) techniques (Scott 2002), which solely implies the use of computing-efficient algorithms. Recursivity is in the present paper specifically related to time-adaptivity and potential application to a real-world online environment. This also motivates the choice for a maximum likelihood type of estimation method, since it can fairly easily be made time-adaptive and recursive. In parallel, advantage is taken of the possibility to analytically formulate predictive densities from Markov-switching autoregressive models for associating one-step ahead point forecasts with prediction intervals. Predictive densities are given as a mixture of conditional densities in each regime, the quantiles of which can be obtained by numerical integration methods.

The paper is structured as following. A general formulation of the type of models considered, i.e. MSAR models, is introduced in Section 2. The specific MSAR model parameterization employed is also described. Then, Section 3 focuses on the adaptive estimation of model coefficients, by first introducing the objective function to be minimized at each time-step, and then deriving an appropriate recursive estimation procedure. The issue of forecasting is dealt with in Section 4, by describing how one-step ahead point forecasts and quantile forecasts can be obtained, from a formulation of one-step ahead predictive densities. Simulations in Section 5 allow us to illustrate the tracking properties of the proposed estimation method. MSAR models are then applied for modelling and forecasting power fluctuations at offshore wind farms in Section 6, the coefficients of which are adaptively estimated with the method introduced. Data originates from both Nysted and Horns Rev wind farms, and consists of power averages with a 10-minute temporal resolution. The characteristics of the estimated models are discussed. Our forecast performance evaluation concentrates on point, interval and density forecasts. Section 7 ends the paper with concluding remarks.

2 The Markov-switching autoregressive model and the parameterization of its coefficients

Let $\{y_t\}$, $t = 1, \ldots, n$, be the time-series of measured power production over a period of n time steps. The power production value at a given time t is defined as the average power over the preceding time interval, i.e. between times $t-1$ and t. For the modelling of offshore wind power fluctuations, the temporal resolution of relevant time-series ranges from 1 to 15 minutes. Hereafter, the notation y_t may be used for denoting either the power production random variable at time t or the measured value.

In parallel, consider $\{z_t\}$ a regime sequence taking a finite number of discrete values, $z_t \in \{1, \ldots, r\}$, $\forall t$. It is assumed that $\{y_t\}$ is an AutoRegressive (AR) process governed by the regime sequence $\{z_t\}$ in the following way,

$$y_t \;=\; \boldsymbol{\theta}^{(z_t)\top} \mathbf{x}_t + \varepsilon_t^{(z_t)}, \quad \forall t \tag{1}$$

with

$$\boldsymbol{\theta}^{(z)} \;=\; [\theta_0^{(z)} \;\; \theta_1^{(z)} \;\; \ldots \;\; \theta_p^{(z)}]^\top, \quad z \;=\; 1, \ldots, r \tag{2}$$

$$\mathbf{x}_t \;=\; [1 \;\; y_{t-1} \;\; y_{t-2} \;\; \ldots \;\; y_{t-p}]^\top \tag{3}$$

and where p is the order of the AR process, chosen here to be the same in each regime for simplicity. However, the developed methodology could be extended for having different orders in each regime, since all model parameters will eventually be gathered in a single vector of coefficients, adaptively estimated with the method described in Section 3. This vector, denoted by $\boldsymbol{\Theta}$, is described below. One might also prefer to skip the intercept term of the AR models if not deemed necessary, as it will be the case for one of the case-studies considered, see Section 6. In parallel, $\{\varepsilon_t^{(z)}\}$ is a white noise process in regime z, i.e. a sequence of independent random variables such that $\mathbb{E}[\varepsilon^{(z)}] = 0$ and $\sigma^{(z)} < \infty$. Let us denote by $\eta^{(z)}$ the density function of the innovations in regime z, which we will refer to as conditional density in the following, since being conditional to the regime. For simplicity, it is assumed that innovations in each regime are distributed Gaussian, $\varepsilon_t^{(z)} \sim \mathcal{N}(0, \sigma^{(z)})$, and thus

$$\eta^{(z)}(\varepsilon; \boldsymbol{\Theta}) \;=\; \frac{1}{\sigma^{(z)}\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{\varepsilon}{\sigma^{(z)}}\right)^2\right), \quad \forall z, t \tag{4}$$

Note that other types of distributions could be assumed for the innovations in each regime. This would yield significant differences in the formulas to be derived for estimation of the model coefficients, even though a methodology similar to that introduced hereafter could be employed.

In addition, it is assumed that the regime sequence $\{z_t\}$ follows a first order Markov chain on the

finite space $\{1, \ldots, r\}$: the regime at time k is determined from the regime at time $k-1$ only, in a probabilistic manner,

$$P[z_k = j | z_{k-1} = i, z_{k-2}, \ldots, z_0] \; = \; P[z_k = j | z_{k-1} = i], \quad \forall i, j, k \tag{5}$$

thus explaining the name of Markov-Switching AutoRegressive (MSAR) given to the resulting model.

All the probabilities governing switches from one regime to the others are gathered in the so-called transition matrix $\mathbf{P}(\boldsymbol{\Theta}) = \{p^{ij}\}_{i,j=1,\ldots,r}$, for which the element p^{ij} represents the probability of being in regime j given that the previous regime was i, as formulated in (5). Some constraints need to be imposed on the transition probabilities. Firstly, by definition all the elements on a given row of the transition matrix must sum to 1,

$$\sum_{j=1}^{r} p^{ij} \; = \; 1, \;\; \forall i \tag{6}$$

since the r regimes represent all possible states that can be reached at any time. Secondly, all the elements of the matrix are chosen to be positive: $p^{ij} \geq 0, \forall i, j$, in order to ensure ergodicity, which means that any regime can be reached eventually.

In order for constrainst (6) to be met at any time, the transition probabilities are parameterized on a unit r-dimensional sphere, as initially proposed in (Collings *et al.* 1994; Collings and Rydén 1998). Indeed, by having $p^{ij} = \left(s^{ij}\right)^2$, and for each i, the vector $s^{i\cdot} = [s^{i1} \;\; \ldots \;\; s^{ir}]^\top$ describing a location on a unit r-dimensional sphere, we naturally have

$$\sum_{j=1}^{r} p^{ij} \; = \; ||s^{i\cdot}||_2^2 \; = \; 1, \quad \forall i \tag{7}$$

For recursive estimation of coefficients in Markov-switching autoregression, Holst *et al.* (1994) argue that a more stable algorithm can be derived by considering the logarithms of the standard deviations of the model innovations, i.e.

$$\tilde{\sigma}^{(z)} \; = \; \ln\left(\sigma^{(z)}\right), \quad \forall z \tag{8}$$

In a similar manner, it is also proposed here to consider the logit transform $\tilde{s}^{ij}$ of the s^{ij} coefficients in order to improve the numerical properties of the information matrix to be used in the recursive estimation scheme,

$$\tilde{s}^{ij} \; = \; \ln\left(\frac{s^{ij}}{1 - s^{ij}}\right), \quad \forall i, j \tag{9}$$

which also translates to constraining the s^{ij} coefficients to the open interval $]0, 1[$.

Finally, the set of coefficients allowing for full characterization of the MSAR model can be summarized as

$$\boldsymbol{\Theta} \;=\; \left[{\boldsymbol{\theta}^{(1)}}^{\top} \;\; \dots \;\; {\boldsymbol{\theta}^{(r)}}^{\top} \;\; \tilde{\boldsymbol{\sigma}}^{\top} \;\; \tilde{\mathrm{s}}^{\top} \right]^{\top} \tag{10}$$

where

$$\boldsymbol{\theta}^{(z)} \;=\; \left[\theta_0^{(z)} \;\; \theta_1^{(z)} \;\; \dots \;\; \theta_p^{(z)} \right]^{\top}, \quad \forall z \tag{11}$$

gives the autoregressive coefficients in regime z, while

$$\tilde{\boldsymbol{\sigma}} \;=\; \left[\tilde{\sigma}^{(1)} \;\; \dots \;\; \tilde{\sigma}^{(r)} \right]^{\top} \tag{12}$$

corresponds to the natural logarithm of the standard deviations of conditional densities in all regimes, and finally

$$\tilde{\mathrm{s}} \;=\; \left[\tilde{s}^{1.\top} \;\; \dots \;\; \tilde{s}^{r.\top} \right]^{\top} \tag{13}$$

is the vector of logit spherical coefficients summarizing the transition probabilities.

3 Adaptive estimation of the model coefficients

There is a large number of papers in the literature dealing with recursive estimation in Hidden Markov Models (HMMs), see e.g. (Collings *et al.* 1994; Collings and Rydén 1998; Holst *et al.* 1994; LeGland and Mevel 1997; Rydén 1997; Stiller and Radons 1999). Most of these estimation methods can be extended to the case of MSAR models, since HMMs are a specific case of the type of models introduced above, for which only an intercept term is used in the definition of the model for each regime. However, it is often considered that the underlying model is stationary and that recursive estimation is motivated by online application and reduction of computational costs only. In contrast here, the model coefficients are allowed to be slowly varying owing to the physical characteristics of the wind power generation process. The time-scales of the regime-switches and of the variations of the MSAR model parameters are assumed to be clearly separable. This assumption comes from meteorological and physical expertise on the problem at hand. The time scale for regime-switching is directly related to the successive passages of meteorological fronts, local rain events, or mesoscale effects, typically being in the order of a few hours. In contrast, the long-term smooth changes in the model parameters are related to time-scales in the order of weeks to months, such changes in the model dynamics being mainly influenced by transitions between various synoptic meteorological regimes, seasonality, changes in the wind farm physical characteristics, etc. This calls for the introduction of an adaptive estimation method permitting to track such long-term changes in the

process characteristics. Since a parametric assumption on these smooth changes would certainly be too rigid, and a parametric model difficult to propose anyways, it is preferred instead to track the MSAR model parameters in a nonparametric framework by employing an exponential forgetting scheme.

Hereafter, it is considered that observations are available up to the current point in time t, and hence that the size of the dataset grows as time increases. The time-dependent objective function to be minimized at each time step is introduced in a first stage, followed by the recursive procedure for updating the set of model coefficients as new observations become available.

3.1 Formulation of the time-dependent objective function

If considering a stationary MSAR model (i.e. for which model coefficients do not change with time), thus not seeking for adaptivity of model coefficients, their estimation can be performed (based on a dataset containing observations up to time t) by maximizing the likelihood of the observations given the model. Equivalently, for a chosen model structure, this translates to minimizing the negative log-likelihood of the observations given the set of model coefficients $\boldsymbol{\Theta}$,

$$S_t(\boldsymbol{\Theta}) \;=\; -\ln\left(P\left[y_1, y_2, \ldots, y_t \mid \boldsymbol{\Theta}\right]\right) \tag{14}$$

which can be rewritten as

$$S_t(\boldsymbol{\Theta}) \;=\; -\sum_{k=1}^{t} \ln\left(u_k(\boldsymbol{\Theta})\right) \tag{15}$$

with

$$u_k(\boldsymbol{\Theta}) \;=\; P\left[y_k \mid y_{k-1}, \ldots, y_1; \boldsymbol{\Theta}\right] \tag{16}$$

In contrast, for the case of maximum-likelihood estimation in MSAR models for which it is assumed that model coefficients can be slowly varying, let us introduce the following time-dependent objective function to be minimized at time t

$$S_t(\boldsymbol{\Theta}) \;=\; -\frac{1}{n_\lambda}\left(\sum_{k=1}^{t} \lambda^{t-k} \ln\left(u_k(\boldsymbol{\Theta})\right)\right) + \frac{\nu}{2}\boldsymbol{\Theta}^\top\boldsymbol{\Theta} \tag{17}$$

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

$$n_\lambda \;=\; \frac{1}{1-\lambda} \tag{18}$$

the effective number of observations is used for normalizing the negative log-likelihood part of the

objective function. The value of the forgetting factor is typically slightly below 1. Note that (17) is a penalized version of what would be a usual maximum-likelihood objective function, with ν the regularization parameter. ν controls the balance between likelihood maximization and minimization of the norm of the model estimates. Such type of regularization is commonly known as Tikhonov regularization (Tikhonov and Arsenin 1977), or alternatively as ridge regression in the statistical literature, following Hoerl (1962). It may allow to increase the generalization ability of the model when used for prediction. In parallel, it permits to increase the numerical stability of the derived recursive estimation scheme, by decreasing the condition number of the information matrix to be inverted at each time step. It is known that this information matrix may be close to singular (Holst *et al.* 1994). This will be further discussed when introducing the recursive updating formulae below. Theoretical and numerical properties of Tikhonov regularization are discussed in (Johansen 1997).

The estimate $\hat{\boldsymbol{\Theta}}_t$ of the model coefficients at time t is finally defined as the set of coefficient values which minimizes (17), i.e.

$$\hat{\boldsymbol{\Theta}}_t \;=\; \arg\min_{\boldsymbol{\Theta}} S_t(\boldsymbol{\Theta}) \tag{19}$$

Note that to our knowledge, there is no literature on the properties of model coefficient estimates for MSAR models when the estimation is performed by minimizing (17). However, Douc *et al.* (2004) show both consistency and asymptotic normality of the maximum likelihood estimator (thus without the regularization term in the objective function) for stationary and non-stationary MSAR models. Those results should then be extended to the case of the penalized likelihood estimator introduced above. We do not aim to perform the necessary theoretical developments here. Instead, a simulation study in Section 5 will allow us to illustrate the nice behaviour of the model estimates.

3.2 Recursive estimation

Imagine being at time t, with the model fully specified by the estimate of model coefficients $\hat{\boldsymbol{\Theta}}_{t-1}$, and a newly available power measure y_t. Our aim in the following is to describe the procedure for updating the model coefficients and thus obtaining $\hat{\boldsymbol{\Theta}}_t$.

Given the definition of the conditional probability u_k in (16), i.e. as the likelihood of the observation y_k given past observations and given the model coefficients (for a chosen model structure), it is straightforward to see that at time t, $u_t(\hat{\boldsymbol{\Theta}}_{t-1})$ can be rewritten as

$$u_t(\hat{\boldsymbol{\Theta}}_{t-1}) \;=\; \boldsymbol{\eta}^\top(\boldsymbol{\varepsilon}_t;\hat{\boldsymbol{\Theta}}_{t-1})\mathbf{P}^\top(\hat{\boldsymbol{\Theta}}_{t-1})\boldsymbol{\xi}_{t-1}(\hat{\boldsymbol{\Theta}}_{t-1}) \tag{20}$$

In the above, $\boldsymbol{\varepsilon}_t$ is the vector of residuals in each regime at time t, thus yielding $\boldsymbol{\eta}(\boldsymbol{\varepsilon}_t;\hat{\boldsymbol{\Theta}}_{t-1})$ the related values of conditional density functions (cf. (4)), given the model coefficients at time $t-1$. In

addition, $\boldsymbol{\xi}_{t-1}(\hat{\boldsymbol{\Theta}}_{t-1})$ is the vector of probabilities of being in such or such regime at time $t-1$, i.e.

$$\boldsymbol{\xi}_{t-1}(\hat{\boldsymbol{\Theta}}_{t-1}) \;=\; \left[\xi_{t-1}^{(1)}(\hat{\boldsymbol{\Theta}}_{t-1}) \;\; \xi_{t-1}^{(2)}(\hat{\boldsymbol{\Theta}}_{t-1}) \;\; \dots \;\; \xi_{t-1}^{(r)}(\hat{\boldsymbol{\Theta}}_{t-1})\right] \tag{21}$$

given the observations up to that time, and given the most recent estimate of model coefficients, that is, $\hat{\boldsymbol{\Theta}}_{t-1}$. In mathematical terms, this writes

$$\xi_{t-1}^{(j)}(\hat{\boldsymbol{\Theta}}_{t-1}) \;=\; P\left[z_{t-1} = j \mid y_{t-1}, y_{t-2}, \dots, y_1; \hat{\boldsymbol{\Theta}}_{t-1}\right] \tag{22}$$

then making $\mathbf{P}^\top(\hat{\boldsymbol{\Theta}}_{t-1})\boldsymbol{\xi}_{t-1}(\hat{\boldsymbol{\Theta}}_{t-1})$ the forecast issued at time $t-1$ of being in such or such regime at time t.

At this same time t, even if the set of true model coefficients $\boldsymbol{\Theta}_{t-1}$ were known, it would not be possible to readily say what the actual regime is. However, one can use statistical inference for estimating the probability $\xi_t^{(j)}$ of being in regime j at time t. This can indeed be achieved by applying the probabilistic inference filter initially introduced by Hamilton (1989). The sequence of regime probabilities for each time index up to time $t-1$ is known, and the regime probabilities at time t (given $\boldsymbol{\Theta}_{t-1}$) are then given by

$$\boldsymbol{\xi}_t(\hat{\boldsymbol{\Theta}}_{t-1}) \;=\; \frac{\boldsymbol{\eta}(\boldsymbol{\varepsilon}_t; \hat{\boldsymbol{\Theta}}_{t-1}) \otimes \mathbf{P}^\top(\hat{\boldsymbol{\Theta}}_{t-1})\boldsymbol{\xi}_{t-1}(\hat{\boldsymbol{\Theta}}_{t-1})}{\boldsymbol{\eta}^\top(\boldsymbol{\varepsilon}_t; \hat{\boldsymbol{\Theta}}_{t-1})\mathbf{P}^\top(\hat{\boldsymbol{\Theta}}_{t-1})\boldsymbol{\xi}_{t-1}(\hat{\boldsymbol{\Theta}}_{t-1})} \tag{23}$$

where $\otimes$ denotes element-by-element multiplication. $\boldsymbol{\xi}_t$ will hence be referred to as the vector of filtered probabilities in the following. Since the parameters of the MSAR models are considered to be varying in time, but only slowly, it is assumed that the small change in parameter estimates from $\hat{\boldsymbol{\Theta}}_{t-1}$ to $\hat{\boldsymbol{\Theta}}_t$ will not affect the vector of filtered probabilities, and thus that $\boldsymbol{\xi}_t(\hat{\boldsymbol{\Theta}}_t) = \boldsymbol{\xi}_t(\hat{\boldsymbol{\Theta}}_{t-1})$. If not making this assumption, one could simply re-estimate the vector of filtered probabilities $\boldsymbol{\xi}_t$ after having obtained $\hat{\boldsymbol{\Theta}}_t$. Note that one does not need to re-estimate the whole sequence of filtered probabilities every time the model coefficients are updated.

In order to derive the recursive estimation procedure, the method employed is based on using a Newton-Raphson step for obtaining the estimate $\hat{\boldsymbol{\Theta}}_t$ as a function of the previous estimate $\hat{\boldsymbol{\Theta}}_{t-1}$, see e.g. (Madsen 2007),

$$\hat{\boldsymbol{\Theta}}_t \;=\; \hat{\boldsymbol{\Theta}}_{t-1} - \frac{\nabla_{\boldsymbol{\Theta}} S_t(\hat{\boldsymbol{\Theta}}_{t-1})}{\nabla^2_{\boldsymbol{\Theta}} S_t(\hat{\boldsymbol{\Theta}}_{t-1})} \tag{24}$$

A little algebra yields the gradient of the objective function calculated at $\hat{\boldsymbol{\Theta}}_{t-1}$,

$$\nabla_{\boldsymbol{\Theta}} S_t(\hat{\boldsymbol{\Theta}}_{t-1}) \;=\; -(1-\lambda)\nabla_{\boldsymbol{\Theta}} \ln(u_t(\hat{\boldsymbol{\Theta}}_{t-1})) \tag{25}$$

as well as its Hessian,

$$\nabla^2_{\boldsymbol{\Theta}} S_t(\hat{\boldsymbol{\Theta}}_{t-1}) \;=\; \lambda \nabla^2_{\boldsymbol{\Theta}} S_{t-1}(\hat{\boldsymbol{\Theta}}_{t-1}) + (1-\lambda)\left(\nu \mathbf{I} - \nabla^2_{\boldsymbol{\Theta}} \ln(u_t(\hat{\boldsymbol{\Theta}}_{t-1}))\right) \tag{26}$$

where $\mathbf{I}$ in (26) is an identity matrix of appropriate dimension.

In parallel, the gradient and Hessian of $\ln(u_t(\hat{\boldsymbol{\Theta}}_{t-1}))$ in (25) and (26) write

$$\nabla_{\boldsymbol{\Theta}} \ln(u_t(\hat{\boldsymbol{\Theta}}_{t-1})) \;=\; \mathbf{h}_t \;=\; \frac{\nabla_{\boldsymbol{\Theta}} u_t(\hat{\boldsymbol{\Theta}}_{t-1})}{u_t(\hat{\boldsymbol{\Theta}}_{t-1})} \tag{27}$$

and

$$\nabla^2_{\boldsymbol{\Theta}} \ln(u_t(\hat{\boldsymbol{\Theta}}_{t-1})) \;=\; \frac{\nabla^2_{\boldsymbol{\Theta}} u_t(\hat{\boldsymbol{\Theta}}_{t-1})}{u_t(\hat{\boldsymbol{\Theta}}_{t-1})} - \frac{\nabla_{\boldsymbol{\Theta}} u_t(\hat{\boldsymbol{\Theta}}_{t-1}) \left(\nabla_{\boldsymbol{\Theta}} u_t(\hat{\boldsymbol{\Theta}}_{t-1})\right)^\top}{u_t^2(\hat{\boldsymbol{\Theta}}_{t-1})} \tag{28}$$

However, by making the assumption such that u_t is (almost) linear in $\boldsymbol{\Theta}$ in the close vicinity of $\hat{\boldsymbol{\Theta}}_{t-1}$, which translates to

$$\frac{\nabla^2_{\boldsymbol{\Theta}} u_t(\hat{\boldsymbol{\Theta}}_{t-1})}{u_t(\hat{\boldsymbol{\Theta}}_{t-1})} \ll \frac{\nabla_{\boldsymbol{\Theta}} u_t(\hat{\boldsymbol{\Theta}}_{t-1}) \left(\nabla_{\boldsymbol{\Theta}} u_t(\hat{\boldsymbol{\Theta}}_{t-1})\right)^\top}{u_t^2(\hat{\boldsymbol{\Theta}}_{t-1})} \tag{29}$$

one then obtains the following approximation of the Hessian of $\ln(u_t)$ evaluated at $\hat{\boldsymbol{\Theta}}_{t-1}$

$$\nabla^2_{\boldsymbol{\Theta}} \ln(u_t(\hat{\boldsymbol{\Theta}}_{t-1})) \;\simeq\; -\frac{\nabla_{\boldsymbol{\Theta}} u_t(\hat{\boldsymbol{\Theta}}_{t-1}) \left(\nabla_{\boldsymbol{\Theta}} u_t(\hat{\boldsymbol{\Theta}}_{t-1})\right)^\top}{u_t^2} \;=\; -\mathbf{h}_t \mathbf{h}_t^\top \tag{30}$$

This assumption is based on the idea that one only makes very small adaptation steps, thus permitting to assume the likelihood function to be locally linear, as if writing a first-order Taylor expansion of the likelihood function at $\hat{\boldsymbol{\Theta}}_{t-1}$. Note that such assumption is similar to that made when applying a recursive Expectation-Maximization (EM) algorithm such as that proposed in (Holst *et al.* 1994). The necessary steps for obtaining $\mathbf{h}_t$ are gathered in the Appendix.

Let us now summarize the above developments and derive the recursive updating scheme to be employed at each time step t. Denote by $\hat{\mathbf{R}}_t$ the matrix corresponding to the estimated inverse covariance matrix of the model coefficients at time t, i.e. $\hat{\mathbf{R}}_t = \nabla^2_{\boldsymbol{\Theta}} S_t(\hat{\boldsymbol{\Theta}}_t)$. We here make a classical assumption for deriving recursive estimation methods for stochastic dynamic systems, see (Ljung and Söderström 1983), which is that the objective criterion S is smooth in the close vicinity of $\hat{\boldsymbol{\Theta}}_t$, and the adaptation step small enough, so that

$$\hat{\mathbf{R}}_t \;=\; \nabla^2_{\boldsymbol{\Theta}} S_t(\hat{\boldsymbol{\Theta}}_t) \simeq \nabla^2_{\boldsymbol{\Theta}} S_t(\hat{\boldsymbol{\Theta}}_{t-1}) \tag{31}$$

Then, using the result (27) in (25), the approximation of (30) in (26), and plugging the gradient

and the Hessian of $S_t(\hat{\boldsymbol{\Theta}}_{t-1})$ into the Newton-Raphson step (24), the 2-step scheme at time t for updating the set of model coefficients can be summarized as

$$\hat{\mathbf{R}}_t = \lambda \hat{\mathbf{R}}_{t-1} + (1-\lambda)\left(\mathbf{h}_t \mathbf{h}_t^\top + \nu \mathbf{I}\right) \tag{32}$$

$$\hat{\boldsymbol{\Theta}}_t = \pi_s \left\{ \hat{\boldsymbol{\Theta}}_{t-1} + (1-\lambda)\hat{\mathbf{R}}_t^{-1}\mathbf{h}_t \right\} \tag{33}$$

where $\mathbf{I}$ is an identity matrix of appropriate dimensions, and π_s a projection operator on the unit spheres defined by the $s^{i\cdot}$ vectors ($i = 1, \ldots, r$). This projection hence concerns transition probabilities only and do not affect autoregressive and standard deviation coefficients. In practice, this projection can be performed by finding the closest point (to the newly updated $s^{i\cdot}$ coefficients) on a unit sphere (of dimension the number of regimes), thus meeting the constraint defined by (7). Note that this 2-step scheme is applied after having calculated the vector of filtered probabilities $\boldsymbol{\xi}_t$. Such a scheme looks somewhat similar to the classical scheme for recursive estimation of regressive parameters, see Madsen (2007).

One observes from (32) and (33) the effects of regularization. It consists of a constant loading on the diagonal of the inverse covariance matrix $\hat{\mathbf{R}}_t$ to be inverted in (33), thus permitting to control its condition number. Then, when $\nu = 0$, (33) simplifies to the classical updating formula for model coefficients tracked with Recursive Least Squares (RLS) or Recursive Maximum Likelihood (RML) methods (again, for more details see Madsen (2007)). It is known that the use of such a type of regularization in recursive estimation schemes may introduce some bias, but globally (if ν is chosen appropriately) permits to have an optimal trade-off between bias and variance of the estimates (Johansen 1997). In addition, it allows for increased numerical stability of the recursive estimation scheme (due to a decrease in the condition number of the matrix $\hat{\mathbf{R}}_t$ to be inverted at each time step), and increased generalization ability of the models estimated.

For initializing the recursive procedure without any prior knowledge on the process considered, one may use equal probabilities of being in the various states, define each AR component as a simple random walk model, put a large load on the diagonal elements of the transition matrix (i.e. large probabilities of staying in the same regime), and have sufficiently large standard deviations of conditional densities in each regime so that conditional density values are not too close to zero while having poor knowledge of the process dynamics. In parallel, the inverse covariance matrix $\hat{\mathbf{R}}_0$ is initialized with $\hat{\mathbf{R}}_0 = \nu \mathbf{I}$, i.e. with a constant load on the diagonal of an identity matrix (of appropriate dimensions) corresponding to the value chosen for the regularization parameter. Then, for the first few steps of the recursive estimation procedure, only (32) is used for gaining information as long as $\hat{\mathbf{R}}_t$ is not invertible. After that, (33) can be used for updating model coefficient estimates.

4 Point and density forecasting

Denote by f_t the density function of wind power values at time t. Given the chosen model structure and the set of true model coefficients $\mathbf{\Theta}_{t-1}$ estimated at time $t-1$, the one-step ahead predictive density of wind generation $\hat{f}_{t|t-1}$ can be expressed as

$$\hat{f}_{t|t-1}(y) \;=\; \sum_{j=1}^{r} \hat{\xi}_{t|t-1}^{(j)} \eta_{t-1}^{(j)} \left(y - {\boldsymbol{\theta}_{t-1}^{(j)}}^{\top} \mathbf{x_t}; \mathbf{\Theta}_{t-1} \right) \tag{34}$$

where $\hat{\xi}_{t|t-1}^{(j)}$ is the one-step ahead forecast probability of being in regime j at time t. The vector $\hat{\boldsymbol{\xi}}_{t|t-1}$ containing such forecast for all regimes is given by

$$\hat{\boldsymbol{\xi}}_{t|t-1} = \mathbf{P}^{\top}(\mathbf{\Theta}_{t-1}) \boldsymbol{\xi}_{t-1}(\mathbf{\Theta}_{t-1}) \tag{35}$$

Since the true model coefficients are obviously not available, they are replaced in the above equations by the estimate $\hat{\mathbf{\Theta}}_{t-1}$ available at that point in time.

Define $\hat{y}_{t|t-1}$ the one-step ahead point prediction of wind power as the conditional expectation of the random variable y_t, given the information set available at time $t-1$. $\hat{y}_{t|t-1}$ can then be derived from the predictive density definition of (34) as

$$\hat{y}_{t|t-1} \;=\; \sum_{j=1}^{r} \hat{\xi}_{t|t-1}^{(j)} {\hat{\boldsymbol{\theta}}_{t-1}^{(j)}}^{\top} \mathbf{x_t} \tag{36}$$

since the distributions of innovations in each regime are all centered.

In parallel, following the definition of conditional densities in (4), the one-step ahead predictive density $\hat{f}_{t|t-1}$ consists of a mixture of Normal densities. This predictive density can hence be explicitly formulated, and quantile forecasts for given proportions calculated with numerical integration methods. Indeed, if denoting by $\hat{F}_{t|t-1}$ the cumulative distribution function related to the predictive density $\hat{f}_{t|t-1}$, the quantile forecast $\hat{q}_{t|t-1}^{(\alpha)}$ for a given proportion α is

$$\hat{q}_{t|t-1}^{(\alpha)} \;=\; {\hat{F}_{t|t-1}}^{-1}(\alpha) \tag{37}$$

The calculation of quantiles for finite mixtures of Normal densities can be found in Rahman *et al.* (2006) for instance. For more details on the question of point and density forecasting with MSAR models, we refer to Krolzig (1997).

Note that the method proposed here disregards the question of uncertainty in parameter estimation, since it gives the exact formulation of the one-step ahead predictive density $\hat{f}_{t|t-1}$ given the true parameters of the MSAR model. Accounting for parameter estimation uncertainty for such type of model is a difficult task which, to our knowledge, has not been treated in the relevant literature. The fact that model coefficients are time-varying and the proposed estimation method

recursive complicates this question even more. However, the use of bootstrap methods may be envisaged, as initially proposed by Craig and Sendi (2002). And, concerning the recursive estimation issue, one may consider adapting the nonparametric block bootstrap method introduced by Corradi and Swanson (2007) to the case of the models considered here.

5 Simulations

In this section, simulations are used for illustrating the convergence and tracking abilities of the proposed adaptive estimation method for MSAR models, for which model coefficients are assumed to be slowly varying. Simulations are based on time-series of 10.500 time steps. The Markov-switching autoregressive process simulated follows the definition given above by (1)-(6), with 2 regimes, and an autoregressive process of order $p = 1$ in each regime. Similar simulations have been carried out with 3 and 4 regimes as well as autoregressive orders up to 4 lags, thus being consistent with the application to wind power data described below. For clarity, it is preferred to show simulations results for this simple case only. In parallel, increasing the number of regimes and the autoregressive order in each regime is not desirable, since it would multiply the number of model coefficients, and then make the estimation problem quite difficult.

It is considered that there is the same probability for the switching between regimes 1 and 2 or between regimes 2 and 1. The MSAR process parameters are linearly varying between values chosen for certain points in time. Between times $t = 1$ and $t = 500$, the set of model parameters is initialized with

$$\begin{aligned}
\boldsymbol{\theta}^{(1)}_{t=1} &= \boldsymbol{\theta}^{(1)}_{t=500} = [0.7 \;\; -0.4]^\top, \quad \sigma^{(1)}_{t=1} = \sigma^{(1)}_{t=500} = 0.1 \\
\boldsymbol{\theta}^{(2)}_{t=1} &= \boldsymbol{\theta}^{(2)}_{t=500} = [0.2 \;\; -0.1]^\top, \quad \sigma^{(2)}_{t=1} = \sigma^{(2)}_{t=500} = 0.2
\end{aligned}$$

with a switching probability of 0.15. This first 500 time-step period will be used for initial batch training of the MSAR models, i.e. by only updating the estimated information matrix $\hat{\mathbf{R}}_t$ at each time step, but not the estimates of the model parameters themselves. Then at time $t = 4500$, the set of true model parameters is given as

$$\begin{aligned}
\boldsymbol{\theta}^{(1)}_{t=4500} &= [0.5 \;\; -0.2]^\top, \quad \sigma^{(1)}_{t=4500} = 0.15 \\
\boldsymbol{\theta}^{(2)}_{t=4500} &= [-0.1 \;\; 0.2]^\top, \quad \sigma^{(2)}_{t=4500} = 0.2
\end{aligned}$$

with a switching probability of 0.2. Finally, from $t = 8500$ and until the end of the dataset, the

Markov-switching autoregressive model is defined by

$$\begin{aligned}
\boldsymbol{\theta}^{(1)}_{t=8500} &= \boldsymbol{\theta}^{(1)}_{t=10500} = [0.45 \;\; -0.05]^\top, \quad \sigma^{(1)}_{t=8500} = \sigma^{(1)}_{t=10500} = 0.2 \\
\boldsymbol{\theta}^{(2)}_{t=8500} &= \boldsymbol{\theta}^{(2)}_{t=10500} = [-0.1 \;\; 0.1]^\top, \quad \sigma^{(2)}_{t=8500} = \sigma^{(2)}_{t=10500} = 0.15
\end{aligned}$$

while the switching probability is set to 0.18.

The MSAR process defined by the above parameters is clearly nonstationary. The rate of change in the parameters is fairly significant in comparison with the smooth long-term changes that are envisaged for the case of the wind power generation dynamics at a temporal resolution of a few minutes.

In order to follow and evaluate the behaviour of the estimation error through time, we simulate 1000 times the 10500 time-step stochastic process described above, and follow the mean and some of the quantiles of the distributions of estimated parameters for each of the model coefficients. The quantiles chosen are those with proportions 0.05 and 0.95, thus only discarding the extreme parameter estimates from our representation of their distributions. These quantiles will then form 90%-confidence resampling intervals informing on model estimation uncertainty on this finite sample.

For each of these simulated processes, the model parameterization described in section 2 and the adaptive estimation method introduced in section 3 are employed. The MSAR model is initialized with the values of the true parameters at time $t = 1$, as defined above. This is in order to see how the recursive estimation method may adapt to changes in the process characteristics from a fairly good prior knowledge. The first 500 time steps then serve for initial recursive estimation of the inverse covariance matrix of model parameters. Note that simulations have also been carried out for the case of an initialization of model parameters far from the true parameter values, and/or with periodic abrupt changes in those true values, though not presented here. Those simulations showed a good ability of the adaptive estimation method to converge towards the true model parameters. The values of both the forgetting factor λ and regularization parameter ν are chosen from our experience of the range of values that may be obtained from application to real-world data. These two parameters are set to $\lambda = 0.995$ (corresponding to an effective number of observations of 200) and $\nu = 0.01$. Other values for these 2 parameters have been tried, yielding similar results in a qualitative sense. Obviously, the value chosen for the forgetting factor has a significant impact on the ability of the recursive estimation method to adapt to changes in the model parameters. When considering a real-world modeling and forecasting exercise, λ and ν should be selected and optimized based on the available data, using e.g. a cross-validation framework.

The evolution of the distributions of estimated model parameters is depicted in Figure 1 for the autoregressive coefficients, in Figure 2 for the standard deviations of conditional densities in each

regime, and in Figure 3 for the transition probabilities. For each of these Figures, the first row shows the evolution of model coefficients in the first regime, while the second row gathers that for the case of the second regime. Note that due to the property such that transition probabilities on a given row of $\mathbf{P}_t$ sum to 1, the estimation error in p_t^{11} (respectively p_t^{21}) is the opposite of that in p_t^{12} (respectively p_t^{22}).

Over the first 500 time-step period, the estimated model coefficients are not updated, since this period serves for initial estimation of the inverse covariance matrix $\hat{\mathbf{R}}_t$. At $t = 500$ it is verified that $\hat{\mathbf{R}}_t$ is invertible, and from that point in time both $\hat{\mathbf{R}}_t$ and $\hat{\mathbf{\Theta}}_t$ are recursively updated. Invertibility of $\hat{\mathbf{R}}_t$ is facilitated by the regularization scheme employed, since insuring a constant loading on its diagonal.

Over the following 8000 time steps, the mean of the distributions of estimated model coefficients follows fairly well the changes in true model coefficients, indicating a general ability of the estimation method introduced to track long-term and smooth changes in MSAR model coefficients. This comment is valid for AR parameters, standard deviations of conditional densities, as well as transition probabilities. The variations in the rate of change of the model coefficients introduced at time step 4500 also does not seem to influence the tracking ability of the method. Even the change of direction in the evolution of transition probabilities is accommodated (see Figure 3). In a general manner, one may notice that the model parameter estimates may be slightly lagging behind the true ones. This comes from the exponential forgetting scheme employed, which still gives some weight to fairly old observations, thus introducing inertia in the adaptation of estimates of model coefficients.

Finally for the case of the final period of the dataset, for which the true model coefficients are kept at the same values for 2000 time steps, the estimated model coefficients stabilizes at values very close to the true ones. The small bias observed, especially for the case of AR coefficients and for standard deviations of conditional densities, may originate from the regularization scheme employed. Indeed as mentioned above, it is known that Tikhonov regularization may introduce some bias, though it globally provides an optimal trade-off between bias and variance of the model estimates. An interesting point is that estimates of transition probabilities are close to being unbiased, while it is known that they are the most difficult coefficients to estimate in MSAR models, see discussion by Holst *et al.* (1994). One might argue that the estimation problem is easier here since it only consists in tracking changes in model coefficients. It has however also been noticed that when initializing transition probabilities with values far from the true ones, the convergence of model estimates was still satisfactory.

While the mean of the distributions of model estimates permits to show the general tracking ability of the estimation method introduced, the resampling 90%-confidence intervals inform us about the uncertainty in model coefficient estimation. Such uncertainty is fairly high for all model coefficients, especially for the transition probabilities. Even though estimates are close to unbiased,

it still remains difficult to accurately track MSAR model coefficients. Note that the size of the 90%-confidence intervals evolve with time and with the level of the various coefficients, as they all are inter-related. Globally, the set of simulations performed with various number of regimes and orders of autoregressive processes (in each regime) has shown a good ability of the proposed estimation model to track model parameters. This thus gives us confidence in the potential applicability of the proposed models and estimation techniques to real-world wind power data, as it is done below.

6 Results on offshore wind power data

In order to analyze the performance of the proposed MSAR models and related adaptive estimation method for the modelling and forecasting of offshore wind power fluctuations, they are applied to real-world case studies. The exercise consists of one-step ahead forecasting of time-series of wind power production. Firstly, the data for the offshore wind farm is described. Then, the configuration of the various models and the setup used for estimation purposes are presented. Finally, a collection of results is shown and commented, related to point, interval and density forecasting.

6.1 Case studies

The two offshore wind farms are located at Horns Rev and Nysted, off the west coast of Jutland and off the south coast of Zealand in Denmark, respectively. The former has a nominal capacity of 160 MW, while that of the latter reaches 165.5 MW. The annual energy yield for each of these wind farms is of around 600 GWh. Today, they represent the two largest offshore wind farms worldwide.

For both wind farms, the original power measurement data includes raw one-second measurements for each wind turbine. Focus is given to the total power output at Horns Rev and Nysted. For each wind farm, time series of power production are normalized by their rated capacity P_n. Hence, power values or error measures are all expressed in percentage of P_n. A sampling procedure has been developed in order to obtain time-series of 10-minute power averages. This averaging rate makes that the very fast fluctuations related to the turbulent nature of the wind disappear and reveal slower fluctuations at the minute scale. Because there may be some erroneous or suspicious data in the raw measurements, the sampling procedure has a threshold parameter τ_v, which corresponds to the minimum percentage of data that need to be considered as valid within a given time interval, so that the related power average is considered as valid too. The threshold chosen is $\tau_v = 75\%$. At Horns Rev, the available raw data is from 16$^{\text{th}}$ February 2005 to 25$^{\text{th}}$ January 2006. And, for Nysted, this data has been gathered for the period ranging from 1$^{\text{st}}$ January to 30$^{\text{th}}$ September 2005. As a result of this procedure, the wind power production datasets at these two wind farms consist of 26650 and 27350 valid data points for Nysted and Horns Rev, respectively.

6.2 Model configuration, estimation setup and benchmarking

In a first stage, it is necessary to define periods that are used for identification (and initial training) of the statistical models and periods that are used for evaluating what the performance of these models may be in operational conditions. These two parts of the dataset are referred to as learning and testing sets. For both wind farms, the first 6000 data points are used as a training set, and the remainder for out-of-sample evaluation of the 1-step ahead forecast performance of the Markov-switching autoregressive models. These evaluation sets contain $N_n = 20650$ and $N_h = 21350$ data points for Nysted and Horns Rev, respectively. Over the learning period, a part of the data is used for one-fold cross validation (the last 2000 points) in order to select optimal values of the forgetting factor and regularization parameter, as well as an optimal number of regimes and an optimal autoregressive order. This is done in a trial-and-error manner, by evaluating the results obtained from a set of different setups for the model structure and hyper-parameters λ and μ. For more information on cross validation, we refer to (Stone 1974). The error measure that is to be minimized over the cross validation set is the Normalized Root Mean Square Error (NRMSE) of 1-step ahead point forecasts, since it is aimed at having 1-step ahead point forecast that would minimize such criterion over the evaluation set. The number of regimes is restricted to vary between 2 and 4. Regarding the autoregressive order, models with up to 4 lags have been checked, with and without an intercept term. It is considered that the forgetting factor cannot be less than $\lambda = 0.98$, since lower values would correspond to an effective number of observations (cf. (18)) smaller than 50 data points (thus covering the previous 8-9 hours only, reaching time-scales close to those of regime switches). Such low value of the forgetting factor would then not allow for adaptation with respect to slow variations in the process characteristics, but would serve more to compensate for model misspecification. No restriction is imposed on the potential range of values for the regularization parameter ν.

For the fitting of various MSAR models, the AR coefficients in each regime are initialized such that the intercept term (if employed) is 0, and with autoregressive patterns that are close to being a simple random walk model. In practice, this translates to initializing the autoregressive model in regime j with a parameter $\theta_1^{(j)}$ for the first lag as

$$\theta_1^{(j)} \;=\; 1 + e_j, \quad e_j \sim \mathcal{N}(0, \tau) \tag{38}$$

with τ small (in the order of 10^{-2}), and with the other parameters $\theta_k^{(j)}$ (if relevant) as

$$\theta_k^{(j)} \;=\; e_j, \quad e_j \sim \mathcal{N}(0, \tau), \quad k > 1 \tag{39}$$

It is verified that the AR models used as initialization in each regime is stationary. For stationarity conditions of an autoregressive process, see e.g. (Chatfield 2004; Madsen 2007). If not, another random drawing of potential parameters following the above equations is performed. In parallel,

matrices of transition probabilities are initialized with 0.9 on the diagonal, corresponding to the probability of staying in the same regime. The off-diagonal elements are set to $(1/(r-1))$, simulating an equal probability of switching from the current regime to any other one. Finally, the standard deviation for the Gaussian conditional densities in each regime is set to 0.2. Such a large value permits to have widespread conditional densities at the beginning of the estimation process, limiting potential numerical issues for the recursive estimation scheme, which in turn would be due to very low conditional density values used e.g. in the denominator of the probabilistic inference filter (23).

The results from the cross-validation procedure, i.e. the number of regimes and autoregressive lags, the use of an intercept term or not, as well as values of the forgetting factor λ and regularization parameter ν, are gathered in Table 1. For both test cases, the forgetting factor takes a value very close to 1, seemingly indicating that changes in process characteristics are indeed slow. The value of 0.999 in the Table corresponds to a number of effective observations of 1000, or seen differently to a period covering the previous 7 days. Sudden and abrupt changes in power fluctuations characteristics are in contrast dealt with thanks to the Markov-switching mechanism. One may then think that λ is so close to 1 that there is no benefit from time-adaptivity of the model paramater. An effective number of observations of around 7 days, however, actually translates to fairly rapid variations in the MSAR model parameters. Values greater than 0.999 have been tried, but for both case-studies the value of 0.999 has been deemed optimal. In comparison, works of Gneiting *et al.* (2006), or Hering and Genton (2009), both focusing on 2-hour ahead forecasting of wind speed, conclude that an optimal sliding window for estimation of the parameters of their dynamic models is of around 45 days. Here data with a finer temporal resolution is available, the local wind climate and time scales involved different, thus potentially yielding different dynamics. This may explain the faster adaptation of our model parameters. In parallel, optimal regularization parameter values are not equal to zero, illustrating the interest of our proposal. The values selected for ν are those that permitted minimizing the 1-step ahead NRMSE on the cross-validation exercise. These values could be increased if interested in dampening variations in model estimates, though this would certainly affect forecasting performance. In contrast, one could also lower the ν values, which may affect both forecast performance and numerical properties of the recursive estimation scheme. Regarding the structure of the MSAR models, a similar number of lags is selected for both cases (with an additional intercept term at Horns Rev), while the number of regimes is larger for Horns Rev than for Nysted. This may indicate a more complex structure of the power fluctuations at the former wind farm.

Two benchmarks are considered in order to compare the performance of the MSAR models on the 1-step ahead forecasting exercise performed below. One the one hand, persistence (which is another name for forecasting based on a random walk model) is a common and serious benchmark for such short look-ahead times. On the other hand, an obvious other benchmark is the AR model with adaptive estimation of its parameters, in order to see the benefits of introducing Markov-switching. In the present case, the adaptive estimation method employed in that described as Recursive Pre-

diction Error (RPE) method in Madsen (2007). This estimation method is equivalent to Recursive Maximum Likelihood (RML) estimation if assuming Gaussianity of model residuals. Similarly to the adaptive estimation method introduced here, adaptivity is based on exponential forgetting of past observations. The level of forgetting is controlled by a forgetting factor, also denoted λ. For an optimal choice of the order of the AR model and of the forgetting factor, the same cross-validation procedure is employed. Models with up to 4 lags (with and without an intercept term) are considered. They are initialized with parameter values close to that of a random walk model, as it is done in (38)-(39) in each regime for the case of the AR models of the Markov-switching approach. The results of the cross validation procedure are collated in Table 2. The structure of AR models obtained is fairly similar to that for the AR parts in each regime of the MSAR models identified above. One notices lower values of the optimal forgetting factors obtained, which is certainly due to the periodic change in the autoregressive dynamics of the process considered.

6.3 Point forecasting results

For evaluation of out-of-sample forecast accuracy, we follow the approach presented in (Madsen *et al.* 2005) for the evaluation of short-term wind power forecasts. Focus is given to the use of error measures such as NRMSE, Normalized Mean Absolute Error (NMAE), and Normalized bias (Nbias). NRMSE is the most relevant error criterion since point forecasts have been defined as conditional expectations. As already mentioned above, forecasts from the proposed Markov-switching autoregressive models are also benchmarked against those obtained from persistence and adaptive AR models. Even though persistence is the most simple way of generating forecasts — the 1-step ahead persistence forecast is equal to the last power measurement — this benchmark is actually difficult to ourperform for look-ahead times such as those considered in the present paper.

The forecast performance assessment over the evaluation set is summarized in Table 3. It has been observed that forecasts where unbiased for all models. Nbias values are therefore not shown and discussed here. NMAE and NRMSE criteria have lower values when employing Markov-switching models. This is satisfactory as it was expected that predictions would be hardly better than those from persistence. The reduction in NRMSE and NMAE when going from more basic to more advanced approaches is larger for the Nysted wind farm than for the Horns Rev wind farm. In addition, the level of prediction error is in general higher for the latter wind farm. This confirms the findings in (Pinson *et al.* 2008), where it is shown that for this temporal resolution and whatever the chosen modelling approach, the level of forecast accuracy is higher at Nysted. The Horns Rev wind farm is located in the North Sea (while Nysted is in the Baltic sea, south of Zealand in Denmark). It may be more exposed to stronger fronts causing fluctuations with larger magnitude, and that are less predictable. If comparing the various types of models employed, it appears that the difference in forecast accuracy (either in terms of NMAE or NRMSE) is fairly small. For instance if focusing on the RMSE criterion for the Nysted wind farm, the decrease observed when going from

persistence to MSAR model is of 7.5%, while it is of 2.6% when going from AR to MSAR models. In view of the large size of the datasets used for forecast evaluation, such decrease in the various criterion can still be seen as a significant increase in forecast accuracy. Point forecasting accuracy may not be the main benefit from the application of MSAR models to wind power fluctuations data at temporal resolutions of a few minutes. Benefits coming from the characterization of the regime sequences or from higher skill of probabilistic forecasts (see below) may be more evident. One should realize however that improvements in terms of point forecast accuracy indeed exist, while for such look-ahead times, outperforming persistence or simple autoregressive models may be very difficult. Since the costs related to power system control and management are very high, even small improvements in the range of a few percents translate in practice to significant monetary savings.

In parallel to the question of forecast accuracy, an expected major interest of the Markov-switching approach is that one can better appraise the characteristics of short-term fluctuations of wind generation offshore by studying the estimated model coefficients, standard deviations of conditional densities, as well as transition probabilities. Autoregressive coefficients may inform on how the persistent nature of power generation may evolve depending on the regime, while standard deviations of conditional densities may tell on the potential amplitude of wind power fluctuations depending on the regime. Finally, the transition probabilities may tell if such or such regime is more dominant, or if some fast transitions may be expected from certain regimes to the others.

The set of model coefficients at the end of the evaluation set for Nysted (i.e. at time $t = N_n$) can be summarized by the AR coefficients and standard deviations of related conditional densities in the two regimes,

$$\begin{aligned}
\boldsymbol{\theta}^{(1)}_{t=N_n} &= [1.546 \quad -0.843 \;\; 0.291]^\top, \quad \sigma^{(1)}_{t=N_n} = 0.042 \\
\boldsymbol{\theta}^{(2)}_{t=N_n} &= [1.478 \quad -0.549 \;\; 0.071]^\top, \quad \sigma^{(2)}_{t=N_n} = 0.008
\end{aligned} \tag{40}$$

while the final matrix of transition probabilities reads

$$\mathbf{P}_{t=N_n} = \begin{bmatrix} 0.946 & 0.054 \\ 0.026 & 0.974 \end{bmatrix}$$

In parallel for Horns Rev, the AR coefficients and related standard deviations (in each of the three regimes) at the end of the dataset (i.e. at $t = N_h$) are

$$\begin{aligned}
\boldsymbol{\theta}^{(1)}_{t=N_h} &= [0.002 \;\; 1.421 \quad -0.486 \;\; 0.061]^\top, \quad \sigma^{(1)}_{t=N_h} = 0.017 \\
\boldsymbol{\theta}^{(2)}_{t=N_h} &= [0.039 \;\; 1.154 \quad -0.373 \;\; 0.159]^\top, \quad \sigma^{(2)}_{t=N_h} = 0.068 \\
\boldsymbol{\theta}^{(3)}_{t=N_h} &= [0.0 \quad 1.316 \quad -0.377 \;\; 0.061]^\top, \quad \sigma^{(3)}_{t=N_h} = 4.10^{-4}
\end{aligned}$$

while the final matrix of transition probabilities is

$$\mathbf{P}_{t=N_h} = \begin{bmatrix} 0.911 & 0.064 & 0.025 \\ 0.082 & 0.913 & 0.005 \\ 0.043 & 0.003 & 0.954 \end{bmatrix}$$

The characteristics of the hidden Markov chain can be analysed by determining the corresponding stationary distributions and expected return times for the various regimes. They are gathered in Table 4. At Nysted, the last regime may be seen as dominant since having the largest marginal probability or, equivalently, the shortest expected return time. This last regime is that with the smallest standard deviation, thus telling us that the most dominant regime at this wind farm is that with smoother fluctuations around the mean dynamics. This not the case at Horns Rev, since the dominant regime is instead the first one, the standard deviation of which is significantly larger than than for the third regime, and respectively smaller that that for the second regime. One also verifies from the return periods of Table 4 and the optimal forgetting factors that, as mentioned in Section 3, the time scales for regime-switching and long-term variations in the model parameters are clearly separable, in the order of few time-steps (ca. 15-40 minutes) for the former one, and of approximately a week for the latter one.

Let us focus then on the autoregressive dynamics in each regime at both wind farms. Studying the roots of the transfer operator $\Phi_j(B) = 1 + \sum_{i=1}^{p} \theta_p^{(j)} B^p$ related to each AR component composing the MSAR models estimated ($j = 1, \ldots, r$), permits evaluating the stationarity of each of these components. The AR model in regime j is stationary if the roots of the related transfer operator $\Phi_j(B)$ are within the unit circle, see e.g. Madsen (2007, pp. 119-121). In the present cases, some of the AR components of estimated MSAR models clearly are nonstationary. It is known however that this does not translate to the resulting MSAR models themselves being nonstationary, since being weighted combinations of (potentially nonstationary) AR components. This important point has been introduced, studied from simulations, and thoroughly discussed by Holst *et al.* (1994). As we do not aim at employing MSAR models for multi-step ahead forecasting, this issue of the stationarity of estimated MSAR models is not investigated further here.

For both test cases, either the dynamics in each regime are different, or if the dynamics are fairly similar (e.g. at Horns Rev for regimes 1 and 3) the standard deviation of the conditional densities then are significantly different. In a general manner, very large differences in standard deviations of conditional densities among the regimes can be observed (up to a factor of 170 for the regimes with larger and smaller standard deviations of conditional densities at Horns Rev). This actually shows that the interest of the proposed Markov-switching autoregressive models may lie more in their ability to resolve among various potential magnitude of power fluctuations than in their ability to capture their autoregressive pattern. A clear benefit will then come from the providing of prediction intervals along with the point forecasts, as they will inform of the possibility of power

fluctutations of smaller or larger magnitude to come. This will be highlighted and discussed in the following Section.

We finally study an arbitrarily chosen episode of power generation at the Nysted wind farm. For confidentiality reason, the dates defining beginning and end of this period cannot be given. The episode consists of 500 successive time-steps with power measurements and corresponding one-step ahead forecasts as obtained from the fitted MSAR model. These 500 time steps represent a period of 83 hours (3 days approximately). In parallel to the time-series of measurements and forecasts in Figure 4, Figure 5 depicts the evolution of the filtered probabilities, i.e. the estimated probabilities of being in such or such regime at each time step.

Remember that the two regimes correspond to very different standard deviations of conditional densities (by a factor of more than 5). The regimes have then clearly been separated by the Markov-switching modelling approach. If focusing on the power time-series of Figure 4, one observes successive periods with fluctuations of lower and larger magnitude. Then, by comparison with the evolution of filtered probabilities in Figure 5, one sees that periods with smoother behaviour of power generation are associated with very high probability of being in the second regime. This is valid for time steps between 220 and 280 or between 360 and 430 for instance. This also shows that regimes are not obviously related to a certain level of power generation, as it would be the case if using e.g. Self-Exciting Transition Auto-Regressive (SETAR) or Smooth Transition Auto-Regressive (STAR) models (Pinson *et al.* 2008). Such a discrimination between regimes allows us to imagine that the study of obtained series of filtered probabilities along with the evolution of some meteorological variables will give useful information for better understanding the meteorological phenomena that govern such behaviour. This would then permit to develop prediction methods taking advantage of additional explanatory variables.

6.4 Evaluation of interval and density forecasts

In a second stage, focus is given to the probabilistic information provided by the MSAR models. Indeed, even if point predictions in the form of conditional expectations are expected to be relevant for power management purposes, the whole information on fluctuations will actually be given by prediction intervals giving the potential range of power production in the next time step, with a given probability i.e. the so-called nominal coverage rate. Ideally, full predictive density would provide the full information about expected wind power fluctuations. The possibility of associating point predictions with central prediction intervals is therefore considered here. Central prediction intervals are intervals that are centred in probability around the median. For instance, a central prediction interval with a nominal coverage rate of 90% has its bounds consisting of the quantile forecasts with nominal proportions 0.05 and 0.95. Note that for real-world applications, prediction intervals with large nominal coverage rates e.g. 90% would be preferred, since covering a large range of possibilities and discarding the very tails only (which potentially include outliers and

extreme events). For the present test cases, prediction intervals are generated over the evaluation set for both Horns Rev and Nysted. The nominal coverage of these intervals range from 10% to 90%, with a 10% increment. This translates to numerically calculating 18 quantiles of the predictive densities obtained from (34).

Figures 6 and 7 illustrate the interest of the Markov-switching approach employed for interval forecasting on the test case of the Horns Rev wind farm. It depicts time-series of one-step ahead point and interval forecasts (with 90% nominal coverage) over 100 time-steps (16.7 hours) in Figure 6, along with the evolution of associated filtered probabilities in Figure 7. The width of the prediction intervals clearly varies during this 100 time-step period, with their size being directly influenced by forecasts of filtered probabilities and standard deviations of conditional densities in each regime (cf. (34)). In this episode, prediction intervals are wider during periods with power fluctuations of larger magnitude. It is composed by successive periods with main contributions from the first and second regimes, which for Horns Rev correspond to the regimes with standard deviations of conditional densities of medium and large magnitude. Even though point predictions may be less accurate (in a mean squared error sense) during these periods of larger fluctuations, MSAR models can provide this valuable information about their potential magnitude. In addition, prediction intervals are not symmetric, as even if conditional densities are assumed to be Gaussian in each regime, the resulting one-step ahead predictive densities are not.

For the evaluation of the probabilistic forecasts (both interval and density forecasts) generated with the MSAR models as described in Section 4, we follow the verification framework introduced by Gneiting *et al.* (2007). A similar framework developed for the specific case of the wind power application can be found in (Pinson *et al.* 2007). Focus is given first to the overall skill of probabilistic forecasts, by employing a proper skill score on density forecasts. Overall skill encompasses both reliability (i.e. probabilistic correctness) and sharpness of interval and density forecasts. The Continuous Ranked Probability Score (CRPS) is chosen here. The value of the CRPS at time t, i.e. for a given forecast density $\hat{f}_{t|t-1}(y)$ with cumulative distribution function $\hat{F}_{t|t-1}(y)$, and corresponding measurement y_t, is calculated as

$$\text{CRPS}_t \;=\; \int_y \Big(\hat{F}_{t|t-1}(y) - H(y-y_t)\Big)^2 \, dy \tag{41}$$

where $H(.)$ is the Heaviside step function, taking the value 1 for $y \geq y_t$ and 0 otherwise. The CRPS value over an entire evaluation set is calculated as the average of CRPS_t values, and expressed in percentage of the nominal capacity P_n of the wind farm considered. CRPS is a negatively rewarding score: the lower the better.

Since AR models seem to be serious competitors to MSAR models for point forecasting, we also consider them as a benchmark in this probabilistic forecasting exercise. Interval and density forecasts from AR models can be fairly easily generated, since the model residuals are assumed to be distributed Gaussian, and their variance tracked by the adaptive estimation method employed. A

1-step ahead density forecast then takes the form of a Gaussian density centered on the 1-step ahead point forecast, and with variance the current estimate of the variance of the model residuals.

The CRPS values of 1-step ahead density forecasts from AR and MSAR models on the evaluation set at Horns Rev and Nysted are gathered in Table 5. For both test cases MSAR models have a significantly higher probabilistic forecast skill, with a decrease in CRPS of 7.14% and 7.62% if compared with AR models. This clearly demonstrates the ability of MSAR models to capture successive periods with various levels in the magnitude of power fluctuations, since the increased probabilistic forecasting skill only comes from the introduction of the Markov-switching mechanism. It also confirms a comment made above, such that the main interest of MSAR models may not obviously be in the switching autoregressive dynamics for the conditional expectation of the process, but in the switching between different levels of fluctuation around this conditional expectation.

In parallel, an important property of probabilistic forecasts relates to their reliability, which corresponds to their probabilistic correctness. For evaluating such a property, one may employ a frequentist approach and verify the observed coverage of prediction intervals against their nominal ones. The observed coverage for the various prediction intervals are gathered in Table 6. The agreement between nominal coverage rates and observed ones is fairly good, even though deviations from perfect reliability of up to 6% can be observed for certain nominal coverage rates. There seem to be in general a trend of having prediction intervals that are too wide for Nysted (except for the 90% nominal coverage rate) while these intervals are generally too narrow for the Horns Rev case.

Since it was explained above that prediction intervals are defined by two quantile forecasts for given nominal proportions, it appears relevant to also verify the observed proportions of quantiles composing the bounds of the prediction intervals. This also allows to see if the interval forecasts are properly centred. Such an evaluation is performed in Figure 8 by the use of reliability diagrams, which gives the observed proportions of the quantiles against the nominal ones. The closer to the diagonal the better. For both wind farms, the reliability curve lies below the diagonal, indicating that all quantiles are underestimated (in probability). This underestimation is more significant for the central part of predictive densities, and in general for the Horns Rev test case. It may be that the Gaussian assumption should be re-thought, in order to have conditional densities that would be better representative of the process considered. Ideally, using nonparametric density estimation in each regime may allow to correct for that.

This lack of reliability of MSAR predictive densities is not a crucial problem. It is known that the reliability of probabilistic forecast can be fairly easily enhanced without affecting their resolution, i.e. their inherent ability to resolve among situations with different levels of uncertainty (Toth *et al.* 2003). MSAR probabilistic forecasts could indeed be recalibrated with e.g. conditional parametric models (Nielsen *et al.* 2006) or smoothed bootstrap (Hall and Rieck 2001). Here instead of calibration of probabilistic forecasts, our focus has principally been to generate one-step ahead forecasts with minimum mean squared error, and to illustrate the possibility of accompanying them with

prediction intervals. One should note though that the use of a maximum-likelihood estimation method (as it is the case here) for MSAR models permits to maximize the sharpness of resulting probabilistic forecasts. If further work should be concentrated on the providing of reliable interval forecasts (and still having maximum sharpness), one should then directly formulate the estimation method for MSAR model coefficients in that sense, using an appropriate skill criterion, as done by Gneiting *et al.* (2006) for instance.

7 Concluding remarks

An appealing approach to the regime-switching modelling of short-term wind power fluctuations based on MSAR models has been introduced. Such a modeling approach can be employed for data analysis and forecasting purposes. The temporal resolution considered is in the order of a few minutes. MSAR models have been allowed to have time-varying coefficients, though slowly varying, in order to follow the long-term variations in the wind generation process characteristics. An appropriate recursive estimation method using penalized maximum likelihood, where penalization consists of Tikhonov regularization, has been described. The proposal for inclusion of a regularization term comes from the aim of application to noisy wind power data, for which the use of non-regularized estimation may result in ill-conditioned numerical problems. The tracking abilities of the method have been illustrated from simulations on artificial data. MSAR models have then been employed for modelling and forecasting wind power fluctuations at Horns Rev and Nysted, the two largest offshore wind farms worldwide. These models and related estimation method have been evaluated on a one-step ahead forecasting exercise on time-series with a 10-minute resolution. They have been favourably compared against persistence and AR-model-based forecasts. MSAR models have the additional benefit of informing of the characteristics of such fluctuations. Indeed, it has been possible to identify regimes with different autoregressive dynamic behaviours, and more importantly with different variances of conditional densities. This shows the ability of the proposed approach to characterize periods with smaller or larger magnitudes of power fluctuations. In the future, the series of regime sequences may be compared with the time-series of some meteorological variables over the same period, in order to reveal if power fluctuations characteristics can indeed be explained by the dynamics of these meteorological variables.

In addition to generating point predictions of wind generation, it has been shown that the interest of the approach proposed also and mainly lied in the possibility of associating point forecasts with prediction intervals or even full predictive densities. Indeed when focusing on power fluctuations, even if point forecasts give useful information, one is mainly interested in the magnitude of potential deviations from these point predictions. The prediction intervals derived here have a higher skill and resolution than those that would be obtained from AR models since resulting from a mixture of conditional densities in each regime. The reliability of prediction intervals obtained from the

MSAR models has been discussed. Even though this reliability is not perfect, prediction intervals could then be recalibrated with existing techniques, without altering their resolution. In parallel, it is known that for the wind generation process, noise distributions are not Gaussian, and that the shape of these distributions is influenced by the level of some explanatory variables (Pinson 2006). Therefore, in order to better shape predictive densities, the Gaussian assumption employed for conditional densities in each regime should be relaxed in the future. Nonparametric density estimation may be achieved with kernel density estimators, as in e.g. (Dorea and Zhao 2002), though this may introduce some problems in a recursive maximum likelihood estimation framework e.g. multimodality of conditional densities. These open questions related to conditional densities and estimation of more general MSAR models will be the focus of further research.

One-step ahead predictive densities of power generation have been explicitly formulated. Such densities consist of finite mixtures of conditional densities in each regime. However, it has been explained that the issue of parameter uncertainty was not considered, and that this may also affect the quality of derived prediction intervals (and probabilistic forecasts in general), especially in an adaptive estimation framework where the quality of parameter estimation may also vary with time. Novel approaches accounting for such parameter uncertainty should hence be proposed. The derivation of analytical formula might be difficult. In contrast, one may think of employing nonparametric block bootstrap techniques similar to that proposed in (Corradi and Swanson 2007).

Acknowledgments

The work presented has been carried out in the frame of the projects 'Power Fluctuations in Large Offshore Wind Farms' and 'Mesoscale atmospheric variability and the variation of wind and production for offshore wind farms', sponsored by the Danish Public Service Obligation (PSO) fund (PSO 4104/6506 and PSO 7141), which is hereby greatly acknowledged. The first author has been partly supported by the Danish Ministry for Science, Technology and Innovation (grant no. FTP-274-08-0573). DONG Energy Generation A/S and Vattenfall Danmark A/S are acknowledged for providing the wind power data for the Nysted and Horns Rev offshore wind farms. The authors are grateful to Tobias Rydén and Tilmann Gneiting for fruitful discussions.

References

Ailliot P, Monbet V, Prevosto M. 2006. An autoregressive model with time-varying coefficients for wind fields. *Environmetrics* **17**: 107–117.

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

Chatfield C. 2004. *The Analysis of Time-Series: An Introduction (6^{th} ed.)*. Chapman & Hall/CRC: London.

Collings IB, Krishnamurthy V, Moore JB. 1994. On-line identification of hidden Markov models via prediction error techniques. *IEEE Transaction on Signal Processing* **42**: 3535–3539.

Collings IB, Rydén T. 1998. A new maximum likelihood gradient technique algorithm for on-line hidden Markov model identification. Proc. IEEE International Conference on Acoustics, Speech, Signal Processing, Seattle (USA).

Corradi V, Swanson NR. 2007. Nonparametric bootstrap procedures for predictive inference based on recursive estimation schemes. *International Economic Review* **48**: 67–109.

Costa A, Crespo A, Navarro J, Lizcano G, Madsen H, Feitosa E. 2008. A review on the young history of the wind power short-term prediction. *Renewable and Sustainable Energy Reviews* **12**: 1725–1744

Craig BA, Sendi PP. 2002. Estimation of the transition matrix of a discrete-time Markov chain. *Health Economics* **11**: 33–42.

Dorea CCY, Zhao LC. 2002. Nonparametric density estimation in hidden Markov models. *Statistical Inference for Stochastic Processes* **5**: 55–64.

Douc R, Moulines E, Rydén T. 2004. Asymptotic properties of the maximum likelihood estimator in autoregressive models with Markov regimes. *Annals of Statistics* **32**: 2254–2304.

Ea Energy Analyses. 50% wind power in Denmark by 2025 - English summary. 2007. Technical Report, Ea Energy Analyses, Copenhagen, Denmark. [Available at: www.windpower.org]

Focken U, Lange M, Monnich M, Waldl H-P, Beyer H-G, Luig A. 2002. Short-term prediction of the aggregated power output of wind farms - A statistical analysis of the reduction of the prediction error by spatial smoothing effects. *Journal of Wind Engineering and Industrial Aerodynamics* **90**: 231–246.

Giebel G, Kariniotakis G, Brownsword R. 2003. The state of the art in short-term prediction of wind power - A literature overview. Technical report, ANEMOS EU project, deliverable report D1.1. [Available online: http://www.anemos-project.eu]

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

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

Hall P, Rieck A. 2001. Improving coverage accuracy of nonparametric prediction intervals. *Journal of the Royal Statistical Society B* **63**: 717–725.

Hamilton JD. 1989. A new approach to the economic analysis of nonstationay time-series and business cycles. *Econometrica* **57**: 357–384.

Hendersen AR, Morgan C, Smith B, Sørensen HC, Barthelmie RJ, Boesmans B. 2003. Offshore wind energy in Europe - A review of the state-of-the-art. *Wind Energy* **6**: 35–52.

Hering AS, Genton MG. 2009. Powering up with space-time wind forecasting. *Journal of the American Statistical Association*, forthcoming.

Hoerl AE. 1962. Application of ridge analysis to regression problems. *Chemical Engineering Progress* **58**: 54–59.

Holst U, Lindgren G, Holst J, Thuvesholmen M. 1994. Recursive estimation in switching autoregressions with a Markov regime. *Journal of Time Series Analysis* **15**: 489–506.

Johansen TA. 1997. On Tikhonov regularization, bias and variance in nonlinear system identification. *Auto-*

matica **33**: 441–446.

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

Krolzig H-M. 1997. *Markov-Switching Autoregressions. Modeling, Statistical Inference and Application to Business Cycle Analysis*. Springer: Berlin (Lecture Notes in Economics and Mathematical Systems, vol. 454).

LeGland F, Mevel L. 1997. Recursive estimation in hidden Markov models. Proc. IEEE Conference on Decision and Control, San Diego.

Ljung L, Söderström T. 1983. *Theory and Practice of Recursive Estimation*. MIT Press: Boston.

Madsen H. 2007. *Time-Series Analysis*. Chapman & Hall/CRC: London.

Madsen H, Pinson P, Nielsen TS, Nielsen HAa, Kariniotakis G. Standardizing the performance evaluation of short-term wind power prediction models. 2005. *Wind Engineering* **29**: 475–489.

Mandelbaum R. 2002. Reap the wild wind. *IEEE Spectrum* **39**: 34–39.

Møller JK, Nielsen HAa, Madsen H. 2008. Time-adaptive quantile regression. *Computational Statistics and Data Analysis* **52**: 1292-1303.

Nielsen HAa, Nielsen TS, Madsen H, Badger J, Giebel G, Landberg L, Sattler K, Voulund L, Tøfting J. 2006. From wind ensembles to probabilistic information about future wind power production - Results from an actual application. Proc. IEEE PMAPS Conference, Probabilistic Methods Applied to Power Systems, Stockholm, Sweden.

Pinson P. 2006. *Estimation of the uncertainty in wind power forecasting*. PhD Thesis, Ecole des Mines de Paris, Paris, France. [Available online: www.pastel.paristech.org]

Pinson P, Christensen LEA, Madsen H, Sørensen PE, Donovan MH, Jensen LE. 2008. Regime-switching modelling of the fluctuations of offshore wind generation. *Journal of Wind Engineering and Industrial Aerodynamics* **96**: 2327–2347.

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

Pinson P, Nielsen HAa, Møller JK, Madsen H, Kariniotakis G. 2007. Nonparametric probabilistic forecasts of wind power: required properties and evaluation. *Wind Energy* **10**: 497–516.

Rahman M, Rahman R, Pearson LR. 2006. Quantiles for finite mixtures of Normal distributions. *International Journal of Mathematical Education in Science and Technology* **37**: 352–357.

Rydén T. 1997. On recursive estimation for hidden Markov models. *Stochastic Processes and their Applications* **66**: 79–96.

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* **22**: 679–693.

Scott SL. 2002. Bayesian methods for hidden Markov models: Recursive computing in the 21th century. *Journal of the American Statistical Association* **9**:, 337–351.

Sloughter JM, Gneiting T, Raftery AE. 2009. Probabilistic wind speed forecasting using ensembles and Bayesian model averaging. *Journal of the American Statistical Association*, forthcoming.

Stiller JC, Radons G. 1999. Online estimation of hidden Markov models. 1999. *IEEE Signal Processing Letters* **6**: 213–215.

Stone M. 1974. Cross-validation and assessment of statistical predictions (with discussion). *Journal of the*

Royal Statistical Society B **36**: 111–147.

Sørensen PE, Cutululis NA, Vigueras-Rodriguez A, Jensen LE, Hjerrild J, Donovan MH, Madsen H. 2007. Power fluctuations from large wind farms. *IEEE Transactions on Power Systems* **22**: 958–965.

Sørensen PE, Cutululis NA, Vigueras-Rodriguez A, Madsen H, Pinson P, Jensen LE, Hjerrild J, Donovan MH. 2008. Modelling of power fluctuations from large offshore wind farms. *Wind Energy* **11**: 29–43.

Tikhonov AN, Arsenin VY. 1977. *Solutions of Ill-posed Problems*. Wiscon: Washington.

Toth Z, Tallagrand O, Candille G, Zhu Y. 2003. Probability and ensemble forecasts. In: Jolliffe IT, Stephenson DB, editors. *Forecast verification: a practitioner's guide in atmospheric science*, John Wiley & Sons: New York; 137–163.

Vincent CL, Giebel G, Pinson P, Madsen H. 2008. Resolving non-stationary spectral information in wind speed time-series using the Hilbert-Huang transform. *Journal of Applied Meteorology and Climatology*, forthcoming.

Appendix: Derivatives of relevant variables

The mathematical developments necessary for the application of the recursive estimation procedure at each time step t are gathered in the Appendix. More precisely, it gathers all necessary derivatives for obtaining $\mathbf{h}_t$ as defined in (27). For that purpose, it is necessary to obtain the value of the gradient $\nabla_{\boldsymbol{\Theta}} u_t$ at $\hat{\boldsymbol{\Theta}}_{t-1}$. This is achieved by derivation of (20) as

$$\begin{aligned}\nabla_{\boldsymbol{\Theta}} u_t(\hat{\boldsymbol{\Theta}}_{t-1}) = & \quad \left(\nabla_{\boldsymbol{\Theta}} \boldsymbol{\eta}(\varepsilon_t; \hat{\boldsymbol{\Theta}}_{t-1})\right)^\top \mathbf{P}^\top(\hat{\boldsymbol{\Theta}}_{t-1}) \boldsymbol{\xi}_{t-1}(\hat{\boldsymbol{\Theta}}_{t-1}) \\ & + \quad \boldsymbol{\eta}^\top(\varepsilon_t; \hat{\boldsymbol{\Theta}}_{t-1}) \left(\nabla_{\boldsymbol{\Theta}} \mathbf{P}^\top(\hat{\boldsymbol{\Theta}}_{t-1})\right) \boldsymbol{\xi}_{t-1}^\top(\hat{\boldsymbol{\Theta}}_{t-1}) \\ & + \quad \boldsymbol{\eta}^\top(\varepsilon_t; \hat{\boldsymbol{\Theta}}_{t-1}) \mathbf{P}^\top(\hat{\boldsymbol{\Theta}}_{t-1}) \left(\nabla_{\boldsymbol{\Theta}} \boldsymbol{\xi}_{t-1}(\hat{\boldsymbol{\Theta}}_{t-1})\right) \end{aligned} \tag{42}$$

In the above, the gradients of three different quantities need to be calculated: they correspond to the gradients of the conditional densities, of the transition probability matrix, and finally of the filtered probabilities. They are given below. Note that these gradients are a function of the true model parameters, which in practice are replaced by available estimates.

Gradient of the conditional densities η

It is clear that $\boldsymbol{\eta}$ is independent of the $\tilde{s}^{ij}$ coefficients related to transition probabilities, and this whatever t. Therefore for a given state j, $\nabla_{\boldsymbol{\Theta}} \eta^{(j)}(\varepsilon_t; \hat{\boldsymbol{\Theta}}_{t-1})$ is composed by the vector of derivatives of $\eta^{(j)}$ with respect to the AR coefficients and log-variances of conditional densisties, calculated at ε_t, and by a vector of zeros of length r^2.

The derivative of $\eta^{(j)}$ with respect to $\theta_i^{(j)}$, calculated at $\varepsilon_t^{(j)}$ given $\hat{\mathbf{\Theta}}_{t-1}$, writes

$$\partial_{\theta_i^{(j)}} \eta^{(j)}(\varepsilon_t^{(j)}; \hat{\mathbf{\Theta}}_{t-1}) \;=\; \frac{x_{t,j}}{(\sigma_{t-1}^{(j)})^2} \varepsilon_t^{(j)} \eta^{(j)}(\varepsilon_t^{(j)}; \hat{\mathbf{\Theta}}_{t-1}) \delta_{ij} \tag{43}$$

where $x_{t,j}$ is the j^{th} element of $\mathbf{x}_t$, and where δ_{ij} is an indicator variable, i.e. such that

$$\delta_{ij} \;=\; \begin{cases} 1, & \text{if } i = j \\ 0, & \text{otherwise} \end{cases} \tag{44}$$

In parallel, the derivative of $\eta^{(j)}$ with respect to $\tilde{\sigma}^{(j)}$, and calculated at ε_t, is similarly given by

$$\partial_{\tilde{\sigma}^{(j)}} \eta^{(j)}(\varepsilon_t^{(j)}; \hat{\mathbf{\Theta}}_{t-1}) \;=\; \left(\frac{{\varepsilon_t^{(j)}}^2}{{\sigma_{t-1}^{(j)}}^2} - 1 \right) \eta^{(j)}(\varepsilon_t^{(j)}; \hat{\mathbf{\Theta}}_{t-1}) \delta_{ij} \tag{45}$$

In practice, this means that it is only needed to compute $r(p+1)$ derivatives for building $\nabla_{\mathbf{\Theta}} \eta^{(j)}(\boldsymbol{\varepsilon}_t; \hat{\mathbf{\Theta}}_{t-1})$ at each time step.

Gradient of the transition probability matrix $\mathbf{P}^\top$

For the case of the transition probability matrix, things are just the other way around. That is, derivatives of $\mathbf{P}^\top(\hat{\mathbf{\Theta}}_{t-1})$ with respect to AR coefficients or to standard deviations of conditional densities are zeros, so that it is only necessary to compute the r^2 derivatives related to $\tilde{s}^{ij}$ coefficients.

As a starting point, the derivative of $\mathbf{P}^\top$ with respect to s^{ij} in the direction tangent to the constraint spherical surface defined by (7) is given in (Collings and Rydén 1998) and writes

$$\partial_{s^{ij}} \mathbf{P}^\top(\hat{\mathbf{\Theta}}_{t-1}) \;=\; 2 s_{t-1}^{ij} \left(\mathbf{e}_j - {s_{t-1}^{i\cdot}}^\top \otimes {s_{t-1}^{i\cdot}}^\top \right) \mathbf{e}_i^\top \tag{46}$$

with $\mathbf{e}_j$ a vector of size r whose elements are zeros except for the j^{th}, which equals 1.

Then, by noticing that

$$\partial_{\tilde{s}^{ij}} \mathbf{P}^\top(\hat{\mathbf{\Theta}}_{t-1}) \;=\; \partial_{s^{ij}} \mathbf{P}^\top(\hat{\mathbf{\Theta}}_{t-1}) \partial_{\tilde{s}^{ij}} s_{t-1}^{ij} \tag{47}$$

and that

$$\partial_{\tilde{s}^{ij}} s_{t-1}^{ij} \;=\; s_{t-1}^{ij}(1 - s_{t-1}^{ij}) \tag{48}$$

one finally obtains

$$\partial_{\tilde{s}^{ij}} \mathbf{P}^\top(\hat{\mathbf{\Theta}}_{t-1}) \;=\; 2 \left(s_{t-1}^{ij} \right)^2 (1 - s_{t-1}^{ij}) \left(\mathbf{e}_j - {s_{t-1}^{i\cdot}}^\top \otimes {s_{t-1}^{i\cdot}}^\top \right) \mathbf{e}_i^\top \tag{49}$$

Gradient of the filtered probabilities $\boldsymbol{\xi}_{t-1}$

The probabilistic inference filter given in (23) permits to recursively estimate the vector of filtered probabilities. Therefore, by derivation of (23), one obtains a similar filter permitting to recursively estimate $\nabla_{\boldsymbol{\Theta}}\boldsymbol{\xi}_{t-1}$ from $\nabla_{\boldsymbol{\Theta}}\boldsymbol{\xi}_{t-2}$, as well as $\nabla_{\boldsymbol{\Theta}}\mathbf{P}^{\top}$ and $\nabla_{\boldsymbol{\Theta}}\boldsymbol{\eta}$. First, note that one has

$$\nabla_{\boldsymbol{\Theta}}\boldsymbol{\xi}_{t-1}(\hat{\boldsymbol{\Theta}}_{t-1}) \;=\; \begin{bmatrix} \nabla_{\boldsymbol{\theta}^{(1)}}\boldsymbol{\xi}_{t-1}(\hat{\boldsymbol{\Theta}}_{t-1}) \\ \vdots \\ \nabla_{\boldsymbol{\theta}^{(j)}}\boldsymbol{\xi}_{t-1}(\hat{\boldsymbol{\Theta}}_{t-1}) \\ \vdots \\ \nabla_{\boldsymbol{\theta}^{(r)}}\boldsymbol{\xi}_{t-1}(\hat{\boldsymbol{\Theta}}_{t-1}) \\ \nabla_{\tilde{\boldsymbol{\sigma}}}\boldsymbol{\xi}_{t-1}(\hat{\boldsymbol{\Theta}}_{t-1}) \\ \nabla_{\tilde{\mathbf{s}}}\boldsymbol{\xi}_{t-1}(\hat{\boldsymbol{\Theta}}_{t-1}) \end{bmatrix} \tag{50}$$

The various elements composing $\nabla_{\boldsymbol{\Theta}}\boldsymbol{\xi}_{t-1}(\hat{\boldsymbol{\Theta}}_{t-1})$ are detailed below. Regarding derivatives with respect to the AR parameters, one has

$$\partial_{\boldsymbol{\theta}_i^{(j)}}\boldsymbol{\xi}_{t-1}(\hat{\boldsymbol{\Theta}}_{t-1}) \;=\; \boldsymbol{\tau}_{t-1,i}^{(j)} - \frac{\boldsymbol{\xi}_{t-1}(\hat{\boldsymbol{\Theta}}_{t-1})}{u_{t-1}(\hat{\boldsymbol{\Theta}}_{t-1})}|\boldsymbol{\tau}_{t-1,i}^{(j)}| \tag{51}$$

where

$$\begin{aligned} \boldsymbol{\tau}_{t-1,i}^{(j)} \;=\; & \frac{1}{u_{t-1}(\hat{\boldsymbol{\Theta}}_{t-1})}\Big[\Big(\partial_{\boldsymbol{\theta}_i^{(j)}}\eta^{(j)}(\varepsilon_{t-1}^{(j)};\hat{\boldsymbol{\Theta}}_{t-1})\Big)\,\mathbf{e}_j \otimes \Big(\mathbf{P}^{\top}(\hat{\boldsymbol{\Theta}}_{t-1})\boldsymbol{\xi}_{t-2}(\boldsymbol{\Theta}_{t-2})\Big) \\ + \;\; & \boldsymbol{\eta}(\boldsymbol{\varepsilon}_{t-1};\hat{\boldsymbol{\Theta}}_{t-1}) \otimes \Big(\mathbf{P}^{\top}(\hat{\boldsymbol{\Theta}}_{t-1})\partial_{\boldsymbol{\theta}_i^{(j)}}\boldsymbol{\xi}_{t-2}(\boldsymbol{\Theta}_{t-2})\Big)\Big] \end{aligned} \tag{52}$$

and $|.|$ denotes the sum of the vector elements. In the above, $\mathbf{e}_j$ is a unit vector as introduced in (46). $\otimes$ stands for the element-by-element multiplication.

The derivatives with respect to the standard deviation of conditional densities in each regime are similar to those for the AR parameters given in (51)-(52), that is,

$$\partial_{\tilde{\sigma}^{(j)}}\boldsymbol{\xi}_{t-1}(\hat{\boldsymbol{\Theta}}_{t-1}) \;=\; \boldsymbol{\chi}_{t-1}^{(j)} - \frac{\boldsymbol{\xi}_{t-1}(\hat{\boldsymbol{\Theta}}_{t-1})}{u_{t-1}(\hat{\boldsymbol{\Theta}}_{t-1})}|\boldsymbol{\chi}_{t-1}^{(j)}| \tag{53}$$

where

$$\begin{aligned} \boldsymbol{\chi}_{t-1}^{(j)} \;=\; & \frac{1}{u_{t-1}(\hat{\boldsymbol{\Theta}}_{t-1})}\Big[\Big(\partial_{\tilde{\sigma}^{(j)}}\boldsymbol{\eta}(\varepsilon_{t-1}^{(j)};\hat{\boldsymbol{\Theta}}_{t-1})\Big)\,\mathbf{e}_j \otimes \Big(\mathbf{P}^{\top}(\hat{\boldsymbol{\Theta}}_{t-1})\boldsymbol{\xi}_{t-2}(\boldsymbol{\Theta}_{t-2})\Big) \\ + \;\; & \boldsymbol{\eta}(\boldsymbol{\varepsilon}_{t-1};\hat{\boldsymbol{\Theta}}_{t-1}) \otimes \Big(\mathbf{P}^{\top}(\hat{\boldsymbol{\Theta}}_{t-1})\partial_{\tilde{\sigma}^{(j)}}\boldsymbol{\xi}_{t-2}(\boldsymbol{\Theta}_{t-2})\Big)\Big] \end{aligned} \tag{54}$$

In a similar manner, one can obtain the derivatives with respect to the $\tilde{s}^{ij}$ parameters calculated at $\hat{\boldsymbol{\Theta}}_{t-1}$. They are given by

$$\partial_{\tilde{s}^{ij}}\boldsymbol{\xi}_{t-1}(\hat{\boldsymbol{\Theta}}_{t-1}) \;=\; \boldsymbol{\psi}^{ij}_{t-1} - \frac{\boldsymbol{\xi}_{t-1}(\hat{\boldsymbol{\Theta}}_{t-1})}{u_{t-1}(\hat{\boldsymbol{\Theta}}_{t-1})}|\boldsymbol{\psi}^{ij}_{t-1}| \tag{55}$$

where

$$\begin{aligned}\boldsymbol{\psi}^{ij}_{t-1} \;=\;& \frac{1}{u_{t-1}(\hat{\boldsymbol{\Theta}}_{t-1})}\Big[\boldsymbol{\eta}(\boldsymbol{\varepsilon}_t;\hat{\boldsymbol{\Theta}}_{t-1})\left(\partial_{\tilde{s}^{ij}_{t-1}}\mathbf{P}^{\top}(\hat{\boldsymbol{\Theta}}_{t-1})\right)\boldsymbol{\xi}_{t-2}(\boldsymbol{\Theta}_{t-2}) \\ &+\; \boldsymbol{\eta}(\boldsymbol{\varepsilon}_t;\hat{\boldsymbol{\Theta}}_{t-1})\mathbf{P}^{\top}(\hat{\boldsymbol{\Theta}}_{t-1})\partial_{\tilde{s}^{ij}}\boldsymbol{\xi}_{t-2}(\boldsymbol{\Theta}_{t-2})\Big]\end{aligned} \tag{56}$$

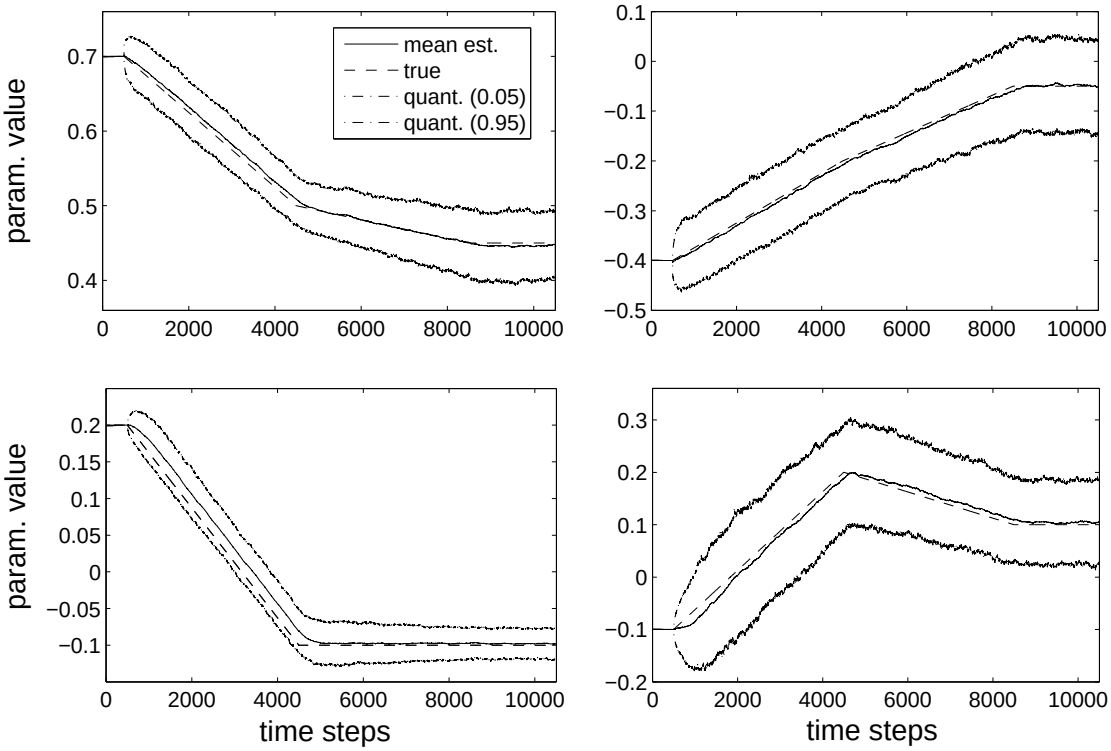

FIGURE 1: *Evolution of the autoregressive parameters (top left: $\theta_0^{(1)}$, top right: $\theta_1^{(1)}$, bottom left: $\theta_0^{(2)}$, bottom right: $\theta_1^{(2)}$). True values are given by dashed lines, the mean estimates are depicted with solid lines, while the dotted lines indicate the quantiles with proportions 0.05 and 0.95 of the distributions of model estimates over the 1000 simulations.*

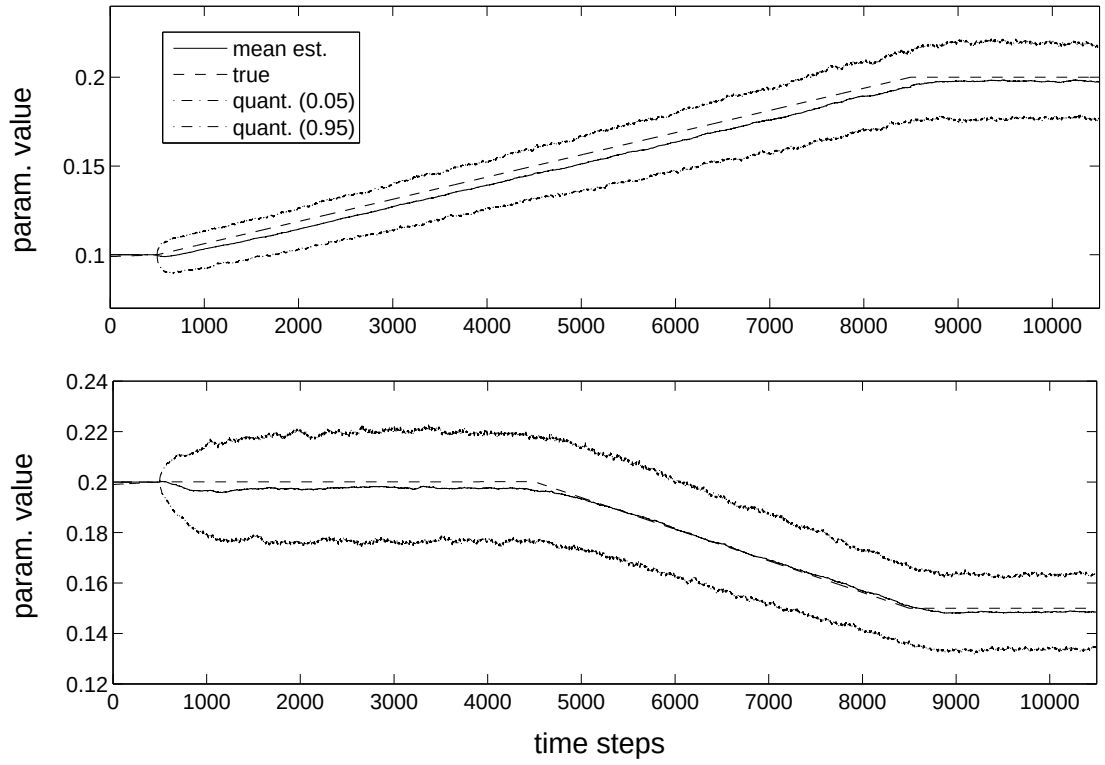

FIGURE 2: *Evolution of the standard deviations of conditional densities in each regime (top: $\sigma^{(1)}$, bottom: $\sigma^{(2)}$).True values are given by dashed lines, the mean estimates are depicted with solid lines, while the dotted lines indicate the quantiles with proportions 0.05 and 0.95 of the distributions of model estimates over the 1000 simulations. and 0.95.*

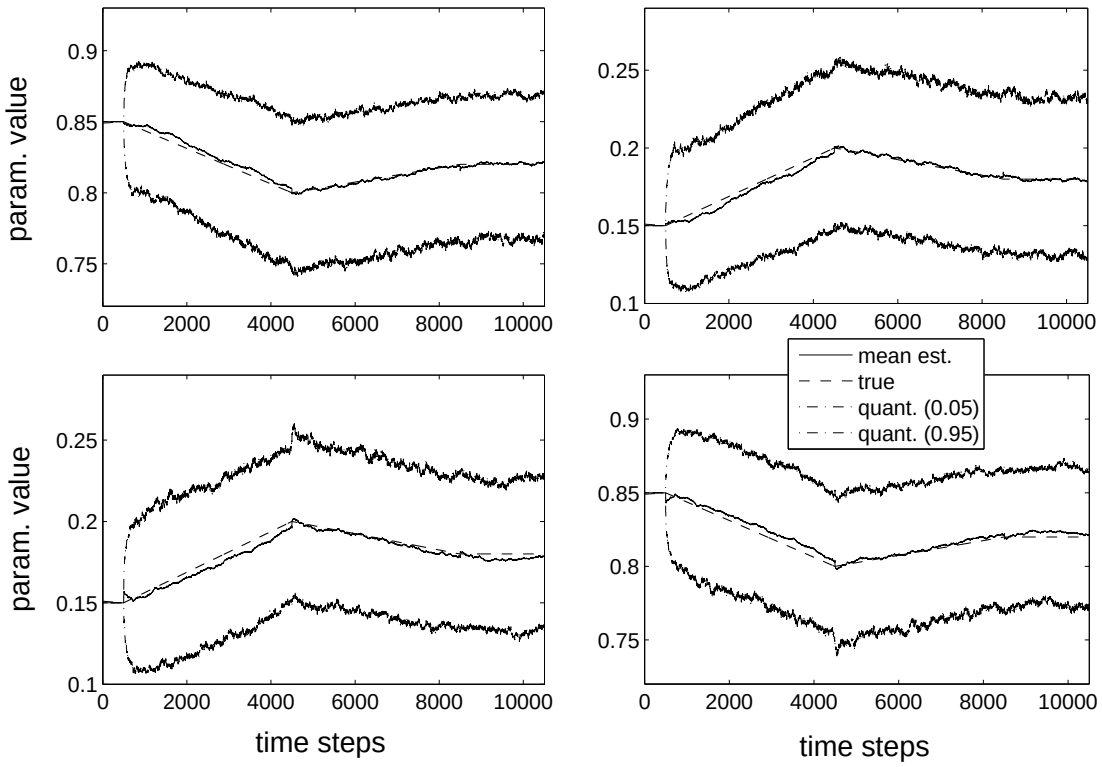

FIGURE 3: *Evolution of the transition probabilities of the Markov-switching autoregressive model (top left:* p^{11}*, top right:* p^{12}*, bottom left:* p^{21}*, bottom right:* p^{22} *). True values are given by dashed lines, the mean estimates are depicted with solid lines, while the dotted lines indicate the quantiles with proportions 0.05 and 0.95 of the distributions of model estimates over the 1000 simulations.*

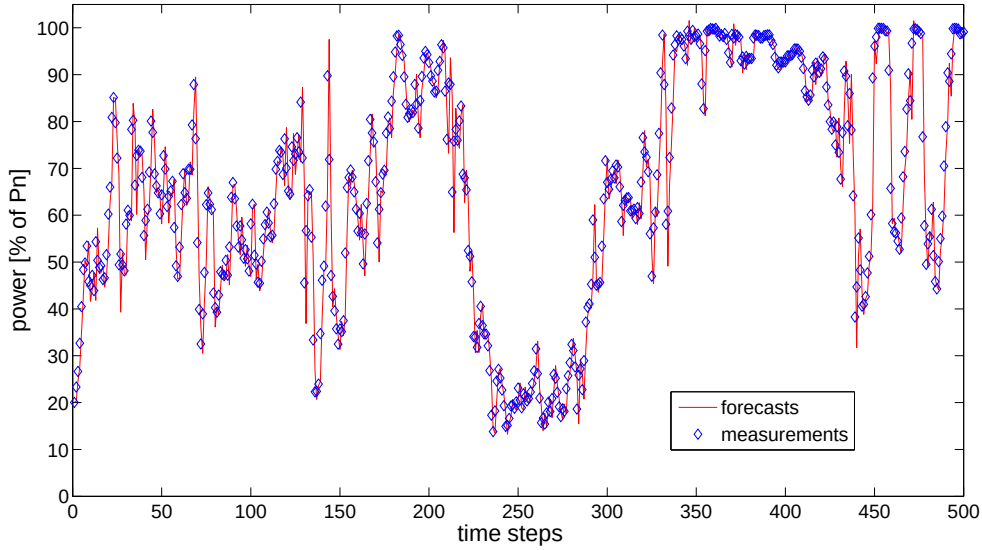

FIGURE 4: *Time-series of normalized power generation at Nysted (both measurements and one-step ahead predictions) over an arbitrarily chosen episode.*

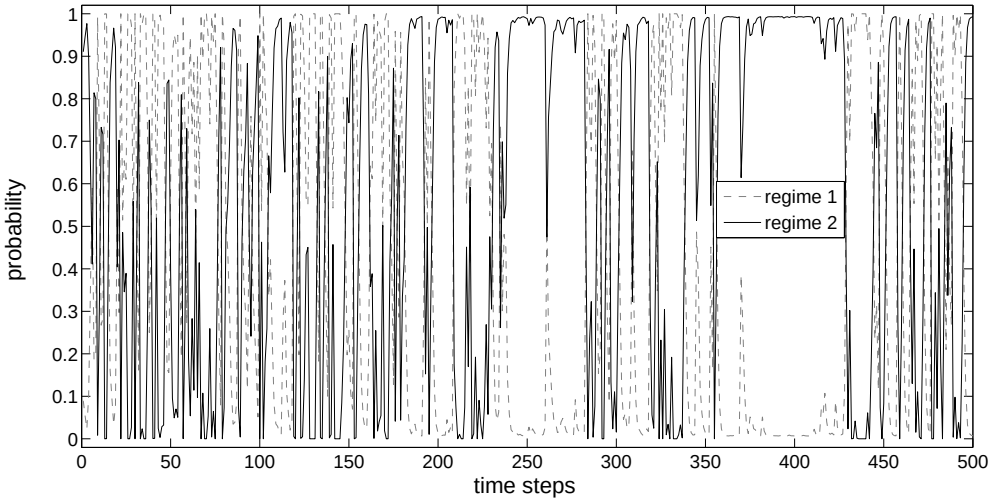

FIGURE 5: *Evolution of filtered probabilities given by the Markov-switching model over the same period.*

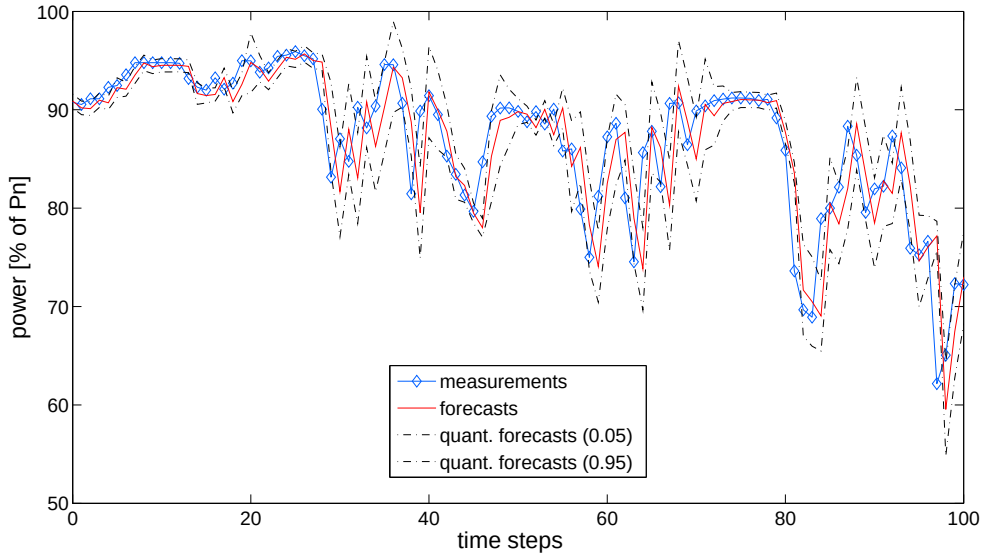

FIGURE 6: *Time-series of normalized power generation at Horns Rev (both measurements and one-step ahead predictions) over an arbitrarily chosen episode, accompanied with prediction intervals with a nominal coverage rate of 90%.*

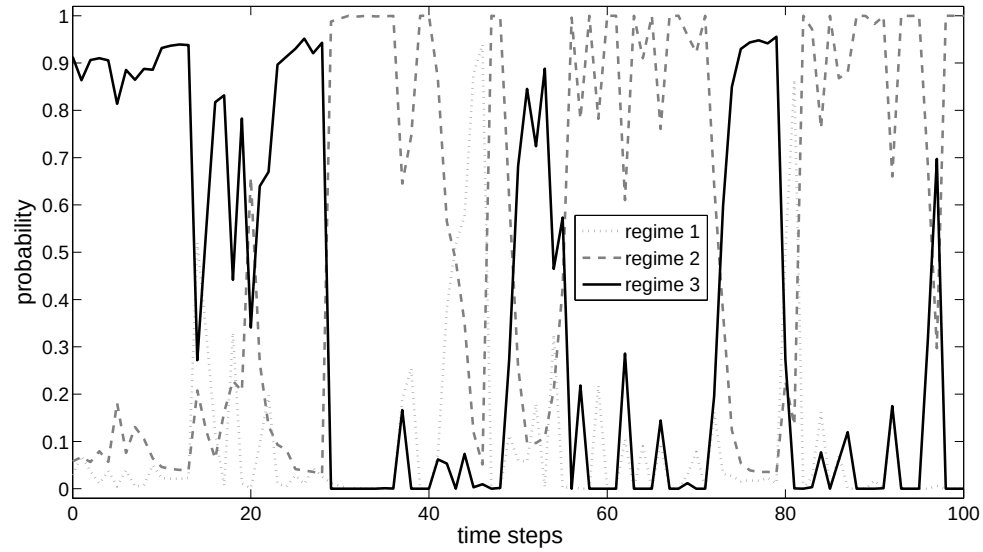

FIGURE 7: *Evolution of filtered probabilities given by the Markov-switching model over the same period.*

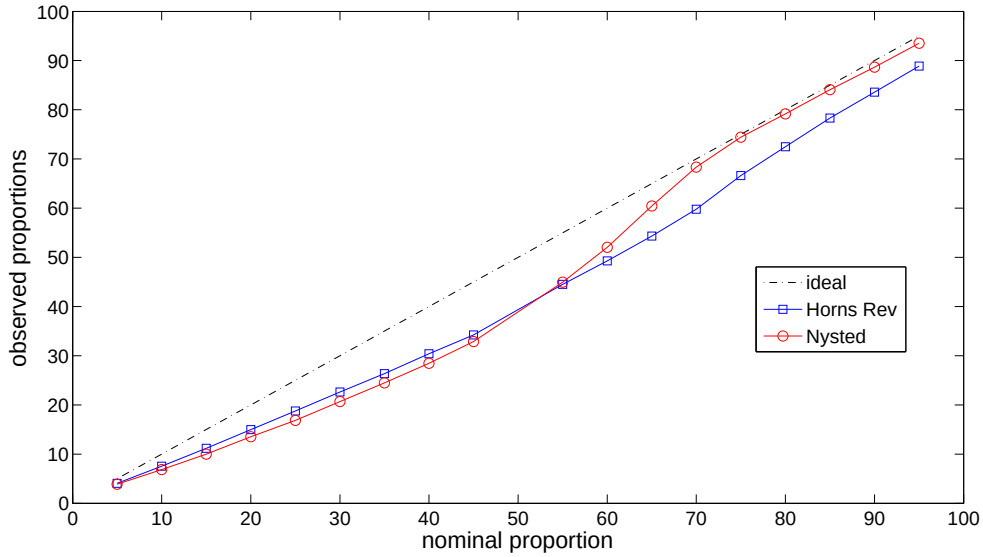

FIGURE 8: *Reliability evaluation of quantile forecasts obtained from the Markov-switching autoregressive models for both Horns Rev and Nysted. Such reliability diagram compare nominal and observed quantile proportions.*

TABLE 1: *Results of the cross validation procedure for MSAR models for the Nysted and Horns Rev wind farms.*

	regimes	lags	intercept	λ	ν
Nysted	2	3	no	0.999	0.02
Horns Rev	3	3	yes	0.999	0.06

TABLE 2: *Results of the cross validation procedure for AR models, for the Nysted and Horns Rev wind farms.*

	lags	intercept	λ
Nysted	2	no	0.995
Horns Rev	3	yes	0.997

TABLE 3: *One-step ahead forecast performance over the evaluation set for Nysted and Horns Rev. Results are for persistence, AR and MSAR models. Performance is summarized with NMAE, and NRMSE criteria, given in percentage of the nominal capacity P_n of the wind farm considered.*

	persistence		AR model		MSAR model	
	NMAE	NRMSE	NMAE	NRMSE	NMAE	NRMSE
Nysted	2.37	4.11	2.21	3.90	2.18	3.80
Horns Rev	2.52	4.77	2.48	4.62	2.43	4.57

TABLE 4: *Stationary distributions and expected return times for the various regimes, for both Nysted and Horns Rev test cases. Stationary distributions are given in terms of marginal probabilities of each regime, while the expected return times are given in number of time-steps.*

	Nysted		Horns Rev		
Regime	1	2	1	2	3
Marginal probability	0.32	0.68	0.42	0.32	0.26
Expected return time	3.08	1.48	2.38	3.15	3.81

TABLE 5: *One-step ahead skill of density forecasts over the evaluation set for Nysted and Horns Rev. Results are for AR and MSAR models. Skill is evaluated with a CRPS criterion, given in percentage of the nominal capacity P_n of the wind farm considered.*

	AR model	MSAR model	Improvement [%]
Nysted	1.82	1.69	7.14
Horns Rev	2.10	1.94	7.62

TABLE 6: *Empirical coverage of the interval forecasts produced from the Markov-switching autoregressive models for Horns Rev and Nysted.*

nom. cov. [%]	obs. cov. Horns Rev [%]	obs. cov. Nysted [%]
10	11.32	12.09
20	19.72	23.56
30	28.69	35.98
40	37.75	47.66
50	48.36	57.58
60	57.91	65.61
70	67.46	74.09
80	76.29	81.80
90	84.95	89.66