Adaptive modelling of offshore wind power fluctuations

Pierre Pinson[1,*], Henrik Madsen[1], Poul E. Sørensen[2], Nicolaos A. Cutululis[2]
[1] Technical University of Denmark, Informatics and Mathematical Modelling, Lyngby, Denmark.
[*] tlf. +45 4525 3428, e-mail pp@imm.dtu.dk
[2] Technical University of Denmark, Risø National Laboratory, Wind Energy department, Roskilde, Denmark.

Abstract — Better modelling of power fluctuations at large offshore wind farms may significantly enhance control and management strategies of their power output. The paper introduces a new methodology for modelling and forecasting the short-term (i.e. few minutes ahead) fluctuations of wind generation. This methodology is based on a Markov-switching autoregressive model with time-varying coefficients. An advantage of the method is that one can easily derive full predictive densities along with the usually generated point forecasts. The quality of this methodology is demonstrated from the test case of two large offshore wind farms in Denmark. The exercise consists in 1-step ahead forecasting exercise on time-series of wind generation with a time resolution of 10 minutes. The interest of such modelling approach for better understanding power fluctuations is finally discussed.

Index Terms — forecasting, modelling, offshore, power fluctuations, wind power.

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 for instance. 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 [1]. 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 [2]. This smoothing effect is hardly present offshore, and thus the magnitude of power fluctuations may reach very significant levels. Characterizing and modelling the power fluctuations for the specific case of offshore wind farms is a current challenge [3, 4], for better forecasting offshore wind generation, developing 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 [5].

When inspecting offshore wind power production data averaged at a few-minute rate, one observes variations that are due to slower local atmospheric changes e.g. frontline passages and rain showers [6]. 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, [7] 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. Though, a drawback of the method described in [7] is that model coefficients are not time-varying, while it is known that wind generation is a process with long-term variations due to e.g. changes in the wind farm environment, seasonality or climate change, see e.g. [8]. The main objective of the present paper is to describe a method for adaptive estimation in Markov-switching autoregression that indeed allows for time-varying coefficients. This method utilizes a parameterization inspired by those proposed in [9, 10] and in [11]. 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 dampen the variability of the model coefficient estimates. A recursive estimation procedure permits to lower computational costs by updating estimates based on newly available observations only. In parallel, advantage is taken of the possibility to express predictive densities from Markov-switching autoregressive models for associating one-step ahead 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. Markov-switching autoregressions with time-varying coefficients, is introduced in Section 1. The specific model parameterization employed is also described. Then, Section 2 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 3-step 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. Then, Markov-switching autoregression with time-varying coefficients is applied for modelling power fluctuations at offshore wind farms in Section 5. Data originates from both Nysted and Horns Rev wind farms, and consists in power averages with a 10-minute temporal resolution. The characteristics of the estimated models are discussed. Concluding remarks in Section 6 end the paper.

2. Markov-switching autoregression with time-varying 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 10 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 process governed by the regime sequence $\{z_t\}$ in the following way

$$y_t = \left(\boldsymbol{\theta}_t^{(z_t)}\right)^\top \mathbf{x}_t + \varepsilon_t^{(z_t)} \tag{1}$$

with

$$\boldsymbol{\theta}_t^{(z)} = [\theta_{t,0}^{(z)} \ \theta_{t,1}^{(z)} \ \ldots \ \theta_{t,p}^{(z)}]^\top \tag{2}$$

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

and where p is the order of the autoregressive 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. The set of parameters for the Markov-switching model introduced above, denoted by $\boldsymbol{\Theta}_t$, is described here. The t-subscript is used for indicating that the autoregressive coefficients are time-dependent, though assumed to be slowly varying. $\{\varepsilon_t^{(z)}\}$ is a white noise process in regime z, i.e. a sequence of independent random variables such that $\mathbb{E}[\varepsilon^{(z)}] = 0$ and $\sigma_t^{(z)} < \infty$. Let us denote by $\eta^{(z)}$ the density function of the innovations in regime z, which we will refer to as a conditional density in the following. For simplicity, it is assumed that innovations in each regime are distributed Gaussian, $\varepsilon_t^{(z)} \sim \mathcal{N}(0, \sigma_t^{(z)})$, $\forall t$, and thus

$$\eta^{(z)}(\varepsilon; \boldsymbol{\Theta}_t) = \frac{1}{\sigma_t^{(z)}\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{\varepsilon}{\sigma_t^{(z)}}\right)^2\right) \tag{4}$$

with the t-subscript indicating that standard deviations of conditional densities are allowed to slowly change over time.

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 way

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

All the probabilities governing switches from one regime to the others are gathered in the so-called transition matrix $\mathbf{P}(\boldsymbol{\Theta}_t) = \{p_t^{ij}\}_{i,j=1,\ldots,r}$, for which the element p_t^{ij} represents the probability (given the model coefficients at time t, since transition probabilities are also allowed to slowly change over time) 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,

$$\sum_{j=1}^{r} p_t^{ij} = 1 \tag{6}$$

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: $p_t^{ij} \geq 0, \forall i, j, t$, in order to ensure ergodicity, which means that any regime can be reached eventually.

In order for constraint (6) to be met at any time, the transition probabilities are parameterized on a unit sphere, as initially proposed in [9, 10]. Indeed, by having $p_t^{ij} = \left(s_t^{ij}\right)^2$, and for each i, the vector $s_t^{i\cdot} = [s_t^{i1} \ \ldots \ s_t^{ir}]^\top$ describing a location on a r-dimensional sphere, we naturally have

$$\sum_{j=1}^{r} p_t^{ij} = ||s_t^{ij}||_2^2 = 1 \tag{7}$$

For recursive estimation of coefficients in Markov-switching autoregression, [11] argue that a more stable algorithm can be derived by considering the logarithms of the standard deviations of the model innovations, i.e.

$$\tilde{\sigma}_t^{(z)} = \ln\left(\sigma_t^{(z)}\right) \tag{8}$$

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

$$\tilde{s}_t^{ij} = \ln\left(\frac{s_t^{ij}}{1 - s_t^{ij}}\right) \tag{9}$$

Finally, the set of coefficients allowing to fully characterizing the Markov-switching autoregressive model at time t can be summarized as

$$\boldsymbol{\Theta}_t = \left[\boldsymbol{\theta}_t^{(1)\top} \ \ldots \ \boldsymbol{\theta}_t^{(r)\top} \ \tilde{\boldsymbol{\sigma}}_t^\top \ \tilde{\mathbf{s}}^\top\right]^\top \tag{10}$$

where

$$\boldsymbol{\theta}_t^{(j)} = \left[\theta_{t,0}^{(j)} \ \theta_{t,1}^{(j)} \ \ldots \ \theta_{t,p}^{(j)}\right]^\top \tag{11}$$

gives the autoregressive coefficients in regime j and at time t, while

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

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

$$\mathbf{s}_t = \left[\tilde{s}_t^{1\cdot\top} \ \ldots \ \tilde{s}_t^{r\cdot\top}\right]^\top \tag{13}$$

is the vector of logit spherical coefficients summarizing the transition probabilities at that same time.

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, see e.g. [9, 10, 11]. Most of these estimation methods can be extended to the case of Markov-switching autoregressions. However, it is often considered that the underlying model is stationary and that recursive estimation is motivated by online applications 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. This calls for the introduction of an adaptive estimation method permitting to track such long-term changes in the process characteristics.

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 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, given a chosen model structure, this translates to minimizing the negative log-likelihood of the observations given the set of model coefficients $\mathbf{\Theta}$,

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

which can be rewritten as

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

with

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

In contrast, for the case of maximum-likelihood estimation for Markov-switching autoregression with time-varying coefficients, let us introduce the following time-dependent objective function to be minimized at time t

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

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. Note that (17) is a regularized 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 [12]. It may allow to increase the generalization ability of the model when used for prediction. From a numerical point of view, it will permit to derive acceptable estimates even though the condition number of the information matrix used in the recursive estimation procedure is pretty high. Theoretical and numerical properties of Tikhonov regularization are discussed in [13].

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

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

Note that to our knowledge, there is no literature on the properties of model coefficient estimates for Markov-switching autoregressions when the estimation is performed by minimizing (17). We do not aim in the present paper at performing the necessary theoretical developments. A simulation study in [14] shows 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 $\hat{\mathbf{\Theta}}_{t-1}$, 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 $\hat{\mathbf{\Theta}}_t$.

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, $u_t(\hat{\mathbf{\Theta}}_{t-1})$ can be rewritten as

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

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

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

given the observations up to that time, and given the most recent estimate of model coefficients, that is, $\hat{\mathbf{\Theta}}_{t-1}$

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

then making $\mathbf{P}^\top(\hat{\mathbf{\Theta}}_{t-1})\boldsymbol{\xi}_{t-1}(\hat{\mathbf{\Theta}}_{t-1})$ 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 $\mathbf{\Theta}_{t-1}$ 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 [15],

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

where $\otimes$ denotes element-by-element multiplication. $\boldsymbol{\xi}_t$ will hence be referred to as the vector of filtered probabilities in the following.

In order to derive the recursive estimation procedure, the method employed is based on using a Newton-Raphson step for obtaining the estimate $\hat{\mathbf{\Theta}}_t$ as a function of the previous estimate $\hat{\mathbf{\Theta}}_{t-1}$, see e.g. [16],

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

After some mathematical developments, which are described in [14], one obtains a 2-step scheme for updating the set of model coefficients at every time step. If denoting by $\mathbf{h}_t$ the information vetor at time t, i.e.

$$\mathbf{h}_t = \nabla_{\mathbf{\Theta}} \ln(u_t(\mathbf{\Theta}_{t-1})) = \frac{\nabla_{\mathbf{\Theta}} u_t(\mathbf{\Theta}_{t-1})}{u_t(\mathbf{\Theta}_{t-1})} \tag{25}$$

and by $\mathbf{R}_t$ the related inverse covariance matrix, the 2-step updating scheme can be summarized as

$$\mathbf{R}_t = \lambda \mathbf{R}_{t-1} + (1-\lambda)\left(\mathbf{h}_t\mathbf{h}_t^\top + \nu \mathbf{I}\right) \tag{26}$$

$$\begin{aligned}\hat{\mathbf{\Theta}}_t = \pi_s\Big\{ &\left(\mathbf{I} + \nu \mathbf{R}_t^{-1}\right)^{-1} \left[\left(\mathbf{I} + \lambda\nu \mathbf{R}_t^{-1}\right) \hat{\mathbf{\Theta}}_{t-1} \right. \\ &\left. + (1-\lambda)\mathbf{R}_t^{-1}\mathbf{h}_t \right] \Big\}\end{aligned} \tag{27}$$

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. Note that this procedure is applied after having calculated the vector of filtered probabilities $\boldsymbol{\xi}_t$. For that reason, the overall updating procedure is referred to as a 3-step procedure.

One clearly sees in (26)-(27) the effects of regularization. It consists of a constant loading on the diagonal of the inverse covariance matrix, thus permitting to control the condition number of $\mathbf{R}_t$ to be inverted in (27). Then, the second equation for model coefficients includes a dampening of previous estimates before and after updating with new information. Note that when $\nu = 0$ one retrieves a somehow classical updating formula for model coefficients tracked with Recursive Least Squares (RLS) of Recursive Maximum Likelihood (RML) methods. For more details, see e.g. [16].

For initializing the recursive procedure without any information on the process considered, one may use equal probabilities of being in the various states, set the autoregressive coefficients to zero, put a large load on the diagonal elements of the transition matrix, 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. In parallel, the inverse covariance matrix $\mathbf{R}_0$ can be initialized with a matrix of zeros. Then, for the first few steps of the recursive estimation procedure, only (26) is used for gaining information as long as $\mathbf{R}_t$ is not invertible. After that, (27) 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 easily be expressed as

$$\begin{aligned}\hat{f}_{t|t-1}(y) = \sum_{j=1}^{r} \hat{\xi}^{(j)}_{t|t-1} \Big[&{\boldsymbol{\theta}^{(j)}_{t-1}}^\top \mathbf{x_t} \\ &+ \eta^{(j)}_{t-1}\left(y - {\boldsymbol{\theta}^{(j)}_{t-1}}^\top \mathbf{x_t}; \mathbf{\Theta}_{t-1}\right)\Big]\end{aligned} \tag{28}$$

where $\hat{\xi}^{(j)}_{t|t-1}$ 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

$$\hat{\boldsymbol{\xi}}_{t|t-1} = \mathbf{P}^\top(\hat{\mathbf{\Theta}}_{t-1})\boldsymbol{\xi}_{t-1}(\mathbf{\Theta}_{t-1}) \tag{29}$$

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 (28) as

$$\hat{y}_{t|t-1} = \sum_{j=1}^{r} \hat{\xi}^{(j)}_{t|t-1} {\hat{\boldsymbol{\theta}}^{(j)}_{t-1}}{}^\top \mathbf{x_t} \tag{30}$$

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

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 $\hat{q}^{(\alpha)}_{t|t-1}$ for a given proportion α is

$$\hat{q}^{(\alpha)}_{t|t-1} = {\hat{F}_{t|t-1}}^{-1}(\alpha) \tag{31}$$

The calculation of quantiles for finite mixtures of Normal densities is discussed in [17].

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 parameter of the Markov-switching autoregressive 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 recursive complicates this question even more. However, the use of bootstrap methods may be envisaged, as initially proposed in [18]. And, concerning the recursive estimation issue, one may consider adapting the nonparametric block bootstrap method introduced in [19] to the case of the models considered here.

5. Results

In order to analyze the performance of the proposed Markov-switching autoregressions and related adaptive estimation method for the modelling of offshore wind power fluctuations, they are used on a real-world case study. The exercise consists in 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.

5.1. Case studies

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

For both wind farms, the original power measurement data consist of one-second measurements for each wind turbine. Focus is given to the total power output at Horns Rev and Nysted. Following [6], it has been chosen to model each wind farm as a single representative wind turbine, the production of which consists of the average of the power generated by all the available wind turbines. These turbines have a nominal capacity P_n of 2000 kW and 2300 kW for Horns Rev and Nysted, respectively. Time series of power production are then normalized by these rated capacities. 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 sampling rate is selected so 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 in 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 are from 16$^{\text{th}}$ February 2005 to 25$^{\text{th}}$ January 2006. And, for Nysted, these data have been gathered for the period ranging from 1$^{\text{st}}$ January to 30$^{\text{th}}$ September 2005.

5.2. Model configuration and estimation setup

From the averaged data, it is necessary to define periods that are used for training the statistical models and periods that are used for evaluating what the performance of these models may be in operational conditions. These two types of datasets are referred to as learning and testing sets. We do not want these datasets to have any data considered as not valid. Sufficiently long periods without any invalid data are then identified and permit to define the necessary datasets. 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. The autoregressive order of the Markov-switching models is arbitrarily set to $p = 3$, and the number of regimes to $r = 3$. For more information on cross validation, we refer to [20]. The error measure that is to be minimized over the cross validation set is the Normalized Root Mean Square Error (NRMSE), since it is aimed at having 1-step ahead forecast that would minimize such criterion over the evaluation set.

For all simulations, the autoregressive coefficients and standard deviations of conditional densities in each regime are initialized as

$$\begin{aligned}
\boldsymbol{\theta}_0^{(1)} &= [0.2\ \ 0\ \ 0\ \ 0]^\top, & \sigma_0^{(1)} &= 0.15\\
\boldsymbol{\theta}_0^{(2)} &= [0.5\ \ 0\ \ 0\ \ 0]^\top, & \sigma_0^{(2)} &= 0.15\\
\boldsymbol{\theta}_0^{(3)} &= [0.8\ \ 0\ \ 0\ \ 0]^\top, & \sigma_0^{(3)} &= 0.15
\end{aligned}$$

while the initial matrix of transition probabilities is set to

$$\mathbf{P}_0 = \begin{bmatrix} 0.8 & 0.2 & 0 \\ 0.1 & 0.8 & 0.1 \\ 0 & 0.2 & 0.8 \end{bmatrix}$$

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. 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 for compensating for very bad model specification. No restriction is imposed on the potential range of values for the regularization parameter ν.

5.3. Point forecasting results

The results from the cross-validation procedure, i.e. the values of the forgetting factor λ and regularization parameter ν that minimize the 1-step ahead NRMSE over the validation set, are gathered in Table 1. In both cases, the forgetting factor takes value very close to 1, indicating that changes in process characteristics are indeed slow. The values in the Table correspond to number of effective observations of 500 and 250 for Nysted and Horns Rev, respectively, or seen differently to periods covering the last 3.5 and 1.75 days. Fast and abrupt changes are dealt with thanks to the Markov-switching mechanism. In addition, regularization parameter values are not equal to zero, showing the benefits of the proposal. Note that one could actually increase this value even more if interested in dampening variations in model estimates, though this would affect forecasting performance.

Table 1. Forgetting factor λ and regularization parameter ν obtained from the cross validation procedure for the Nysted and Horns Rev wind farms.

	λ	ν
Nysted	0.998	0.005
Horns Rev	0.996	0.007

For evaluation of out-of-sample forecast accuracy, we follow the approach presented in [21] for the evaluation of short-term wind power forecasts. Focus is given to the use of error measures such as NRMSE and Normalized Mean Absolute Error (NMAE). In addition, forecasts from the proposed Markov-switching autoregressive models are benchmarked against those obtained from persistence. Persistence is the most simple way of producing a forecast and is based on a random walk model. A 1-step ahead persistence forecast is equal the last power measure. Despite its apparent simplicity, this benchmark method is difficult to beat for short-term look-ahead time such as that considered in the present paper.

The forecast performance assessment over the evaluation set is summarized in Table 2. 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 is higher for the Nysted wind farm than for the Horns Rev wind farm. In addition, the level of error is in general higher for the latter wind farm. This confirms the findings in [7], where it is shown that the level of forecast performance, whatever the chosen approach, 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.

Table 2. One-step ahead forecast performance over the evaluation set for Nysted and Horns Rev. Results are both for persistence and Markov-switching models. Performance is summarized with NMAE and NRMSE criteria, given in percentage of the nominal capacity P_n of the representative single turbine.

	persistence		Markov-switching model	
	NMAE	NRMSE	NMAE	NRMSE
Nysted	2.37	4.11	2.20	3.79
Horns Rev	2.71	5.06	2.70	4.96

An expected 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 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 can be summarized by the model autoregressive coefficients and related standard deviations of related conditional densities,

$$\boldsymbol{\theta}_{N_n}^{(1)} = [0.0 \;\; 1.361 \;\; -0.351 \;\; -0.019]^\top, \; \sigma_{N_n}^{(1)} = 0.0007$$
$$\boldsymbol{\theta}_{N_n}^{(2)} = [0.013 \;\; 1.508 \;\; -0.778 \;\; 0.244]^\top, \; \sigma_{N_n}^{(2)} = 0.041$$
$$\boldsymbol{\theta}_{N_n}^{(3)} = [-0.001 \;\; 1.435 \;\; -0.491 \;\; 0.056]^\top, \sigma_{N_n}^{(3)} = 0.011$$

while the final matrix of transition probabilities is

$$\mathbf{P}_{N_n} = \begin{bmatrix} 0.888 & 0.036 & 0.076 \\ 0.027 & 0.842 & 0.131 \\ 0.051 & 0.075 & 0.874 \end{bmatrix}$$

In parallel for Horns Rev, the autoregressive coefficients and related standard deviations are

$$\boldsymbol{\theta}_{N_h}^{(1)} = [0.002 \;\; 1.253 \;\; -0.248 \;\; -0.008]^\top, \sigma_{N_h}^{(1)} = 0.023$$
$$\boldsymbol{\theta}_{N_h}^{(2)} = [0.022 \;\; 1.178 \;\; -0.3358 \;\; 0.123]^\top, \sigma_{N_h}^{(2)} = 0.066$$
$$\boldsymbol{\theta}_{N_h}^{(3)} = [0.069 \;\; 0.91 \;\; 0.042 \;\; -0.022]^\top, \sigma_{N_h}^{(3)} = 0.005$$

while the final matrix of transition probabilities is

$$\mathbf{P}_{N_h} = \begin{bmatrix} 0.887 & 0.069 & 0.044 \\ 0.222 & 0.710 & 0.068 \\ 0.173 & 0.138 & 0.689 \end{bmatrix}$$

For both wind farms, the first regime is dominant in the sense that it has the highest probability of keeping on with the same regime when it is reached. However, one could argue that the first regime is more dominant for Horns Rev, as the probabilities of staying in second and third regimes are lower, and as the probabilities of going back to first regime are higher. The dominant regimes have different characteristics for the two wind farms. At Nysted, it is the regime with the lower standard deviation of the conditional density, and thus the regime where fluctuations of smaller magnitude are to be expected. It is not the case at Horns Rev, as the dominant regime is that with the medium value of standard deviations of conditional densities. Such finding confirms the fact that power fluctuations seem to be of larger magnitude at Horns Rev than at Nysted.

Let us study an arbitrarily chosen episode of power generation at the Horns Rev wind farm. For confidentiality reason, the dates defining beginning and end of this period cannot be given. The episode consists of 250 successive time-steps with power measurements and corresponding one-step ahead forecasts as obtained by the fitted Markov-switching autoregressive model. These 250 time steps represent a period of around 42 hours. The time-series of power production over this period is shown in Figure 1, along with corresponding one-step ahead forecasts. In parallel, Figure 2 depicts the evolution of the filtered probabilities, i.e. the probabilities given by the model of being in such or such regime at each time step. Finally, the evolution of the standard deviation of conditional densities in each regime is shown in Figure 3.

First of all, it is important to notice that there is a clear difference between the three regimes in terms of magnitude of potential power fluctuations. There is a ratio 10 between the standard deviations of conditional densities between regime 2 and 3. In addition, these regimes are clearly separated, as there is a smooth evolution of the standard deviation parameters over the episode. If focusing on the power time-series of Figure 1, one observes successive periods with fluctuations of lower and larger magnitude. Then, by comparison with the evolution of filtered probabilities in Figure 2, one sees that periods with highly persistent behaviour of power generation are all associated with very high probability of being in the first regime. This is valid for time steps between 20 and 80 for instance. This also shows that regimes are not bviously related to a certain level of power generation, as it would be the case f using e.g. Self-Exciting Tansition Auto-Regressive (SETAR) or Smooth Transition Auto-Regressive (STAR) models [7]. If looking again at the autogressive coefficients in each regime given above for Horns Rev at the end of the evaluation period, one clearly sees that intercept coefficients are almost zero. While regime 1 appears to be the regime with low magnitude fluctuations, both regime 2 and 3 contribute to periods with larger ones. Studying obtained series of filtered probabilities along with the evolution of some meteorological variables is expected to give useful information for better understanding meteorological phenomena that govern such behaviour. This would then permit to develop prediction methods taking advantage of additional explanatory variables.

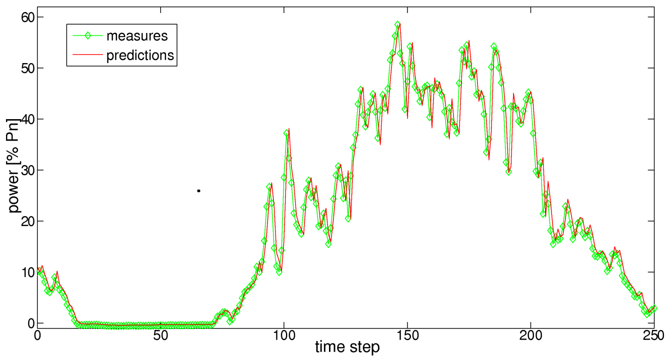

Figure 1. Time-series of normalized power generation at Horns Rev (both measures and one-step ahead predictions) over an arbitrarily chosen episode.

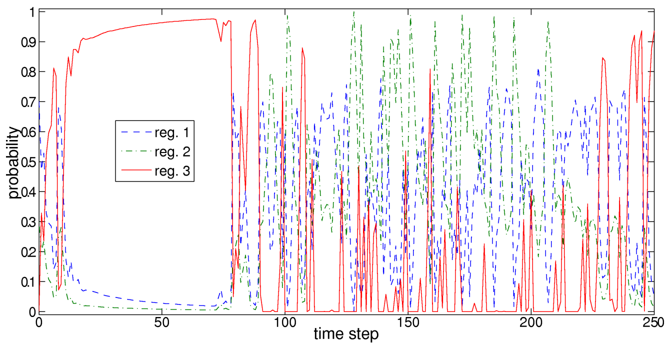

Figure 2. Evolution of filtered probabilities given by the Markov-switching model over the same period.

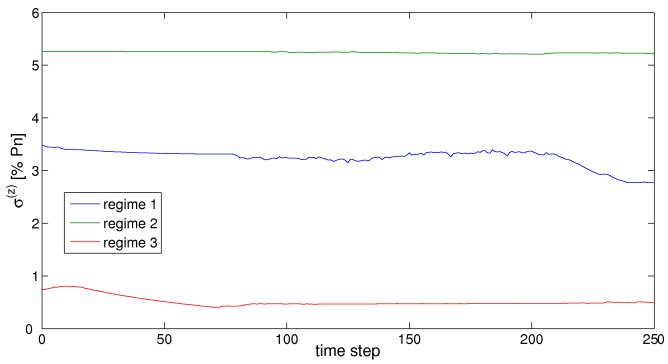

Figure 3. Evolution of the standard deviation of conditional densities in the various regimes for the same episode.

5.4. Interval forecasting results

In a second stage, focus is given to the uncertainty information provided by the Markov-switching autoregressive 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. their nominal coverage rate. Therefore, the possibility of associating point predictions with central prediction intervals is 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 80% has its bounds consisting of the quantile forecasts with nominal proportions 0.1 and 0.9. Therefore, for evaluating the reliability of generated interval forecasts, i.e. their probabilistic correctness, one has to verify the observed proportions of quantiles composing the bounds of intervals. For more information on the evaluation of probabilistic forecasts, and more particularly for the wind power application, we refer to [22, 23].

Table 3. Empirical coverage of the interval forecasts produced from the Markov-switching autoregressive models for Horns Rev and Nysted.

nominal [%]	emp. cov. Horns Rev [%]	emp. cov. Nysted [%]
10	10.09	10.38
20	21.23	19.55
30	31.48	28.69
40	41.67	38.16
50	51.36	48.59
60	61.39	59.18
70	70.45	69.59
80	79.84	79.92
90	89.59	90.92

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 (28). The observed coverage for these various prediction intervals are gathered in Table 3. The agreement between nominal coverage rates and observed one is good, with deviations from perfect reliability overally less than 2%. However as explained above, this valuation has to be carried further by looking at the observed proportions of related quantile forecasts, in order to verify that intervals are indeed properly centred. Such evaluation is performed in Figure 4 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. Note that for operational applications one would be mainly interested in using prediction intervals with high nominal coverage rates, say larger than 80%, thus corresponding to quantile forecasts that are more reliable in the present evaluation. It seems that the Gaussian assumption for conditional densities allows to have predictive densities (in the form of Normal mixtures) that appropriately capture the shape of the tails of predictive distributions, but not their central parts. Using nonparametric density estimation in each regime may allow to correct for that.

Finally, Figure 5 depicts the same episode with power measures and corresponding one-step ahead point prediction that than shown in Figure 1 for the Horns Rev wind farm, except

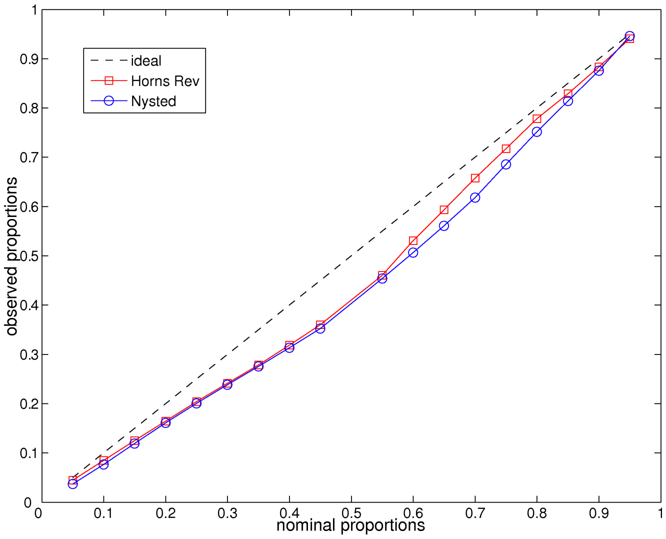

Figure 4. Reliability evaluation of quantile forecasts obtained from the Markov-switching autoregressive models for both Horns Rev and Nysted. Such reliability diagram compare nominal and observed quantile proportions.

that here point predictions are associated with prediction intervals with a nominal coverage rate of 90%. Prediction intervals with such nominal coverage rate are the most relevant for operation applications, and they have been found to be the most reliable in practice. The size of the prediction intervals obviously varies during this 250 time-step period, with their size directly influenced by forecasts of filtered probabilities and standard deviations of conditional densities in each regime (cf. (28)). 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 clearly not. In this episode, prediction intervals are wider during periods with power fluctuations of larger magnitude. Even though point predictions may be less accurate (in a mean square sense) during these periods of larger fluctuations, Markov-switching autoregressive models can provide this valuable information about their potential magnitude.

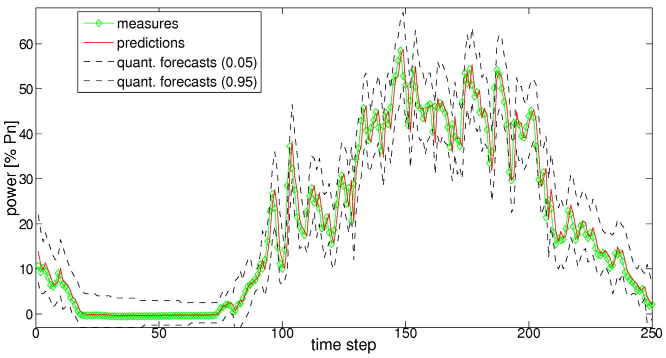

Figure 5. Time-series of normalized power generation at Horns Rev (both measures and one-step ahead predictions) over an arbitrarily chosen episode, accompanied with prediction intervals with a nominal coverage of 90%.

6. Conclusions

Markov-switching autoregressive models are an appealing approach to the modelling of short-term wind power fluctuations at large offshore wind farms. Such models can be used for simulation or forecasting purposes. This class of models has been generalized so that they are 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 estimation method using recursive maximum likelihood and Tikhonov regularization has been introduced. The proposal for including 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 convergence and tracking abilities of the method have been shown from simulations. Then, Markov-switching autoregressive models have been employed for characterizing the 10-minute power fluctuations at Horns Rev and Nysted, the two largest offshore wind farms worldwide. The models and related estimation method have been evaluated on a one-step ahead forecasting exercise, with persistence as a benchmark. For both wind farms, the forecast accuracy of the proposed approach is higher than that of persistence, with the additional benefit of informing on the characteristics of such fluctuations. Indeed, it has been possible to identify regimes with different autoregressive behaviours, and more importantly with different variances in conditional densities. This shows the ability of the proposed approach to characterize periods with lower or larger magnitudes of power fluctuations. In the future, the series of state 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 these meteorological variables.

In addition to generating point predictions of wind generation, it has been shown that the interest of the approach proposed also lied in the possibility of associating prediction intervals or full predictive densities to point predictions. Indeed when focusing on power fluctuations, even if point predictions give useful information, one is mainly interested in the magnitude of potential deviations from these point predictions. It has been shown that for large nominal coverage rates (which are the most appropriate for operational applications) the reliability of prediction intervals was more acceptable than for low nominal coverage rates. 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 [24]. Therefore, in order to better shape predictive densities, the Gaussian assumption should be relaxed in the future. Nonparametric density estimation may be achieved with kernel density estimators, as in e.g. [25], though this may introduce some problems in a recursive maximum likelihood estimation framework e.g. multimodality of conditional densities.

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 conditional densities, 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 procedures similar to that proposed in [19]. This will be the focus of further research.

ACKNOWLEDGMENTS

The results presented have been generated as a part of the project 'Power Fluctuations in Large Offshore Wind Farms' sponsored by the Danish Public Service Obligation (PSO) fund (PSO 105622 / FU 4104), which is hereby greatly acknowledged. DONG Energy Generation A/S and Vattenfall Danmark A/S are also acknowledged for providing the wind power data for the Nysted and Horns Rev offshore wind farms. Finally, the authors would like to acknowledge Tobias Rydén for fruitful discussions.

References

[1] W. Sweet. Reap the wild wind. IEEE Spectrum, vol. 39, pp. 34-39, 2002.

[2] U. Focken, M. Lange,M. Monnich, H.-P. Waldl, H.-G. Beyer, A. Luig. 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, vol. 90, pp. 231-246, 2002.

[3] V. Akhmatov. Influence of wind direction on intense power fluctuations in large offshore windfarms in the North Sea. Wind Engineering, vol. 31, pp. 59-64, 2007.

[4] A.R. Hendersen, C. Morgan, B. Smith B., H.C. Sørensen, R.J. Barthelmie, B. Boesmans. Offshore wind energy in Europe - A review of the state-of-the-art. Wind Energy, vol. 6, pp. 35-52, 2003.

[5] P.E. Sørensen, N.A. Cutululis, A. Vigueras-Rodriguez, L.E. Jensen, J. Hjerrild, M.H. Donovan, H. Madsen. Power fluctuations from large wind farms. IEEE Transactions on Power Systems, vol. 22, pp. 958-965, 2007.

[6] P.E. Sørensen, N.A. Cutululis, A. Vigueras-Rodriguez, H. Madsen, P. Pinson, L.E. Jensen, J. Hjerrild, M.H. Donovan. Modelling of power fluctuations from large offshore wind farms. Wind Energy, to be published, 2007.

[7] P. Pinson, L.E.A. Christensen, H. Madsen, P.E. Sørensen, M.H. Donovan, L.E. Jensen. Regime-switching modelling of the fluctuations of offshore wind generation. Journal of Wind Engineering and Industrial Aerodynamics, submitted for publication, 2006.

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

[9] I.B. Collings, V. Krishnamurthy, J.B. Moore. On-line identification of hidden Markov models via prediction error techniques. IEEE Transactions on Signal Processing, vol. 42, pp. 3535-3539, 1994.

[10] I.B. Collings, T. Rydén. A new maximum likelihood gradient technique algorithm for on-line hidden Markov model identification. in: Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing, Seattle, 1998.

[11] U. Holst, G. Lindgren, J. Holst, M. Thuvesholmen. Recursive estimation in switching autoregressions with a Markov regime. Journal of Time Series Analysis, vol. 15, pp. 489-506, 1994.

[12] A.N. Tikhonov, V.Y. Arsenin. Solutions of Ill-posed Problems. Wiscon: Washington, DC, 1977.

[13] T.A. Johansen. On Tikhonov regularization, bias and variance in nonlinear system identification. Automatica, vol. 33, pp. 441-446, 1997.

[14] P. Pinson, H. Madsen. Markov-switching autoregression with time-varying coefficients for offshore wind power fluctuations. Technical report, Technical University of Denmark, Informatics and Mathematical Modelling, Denmark, 2007.

[15] J.D. Hamilton. A new approach to the economic analysis of nonstationay time-series and business cycles. Econometrica, vol. 57, pp. 357-384, 1989.

[16] H. Madsen. Time-Series Analysis, Second Edition. Technical University of Denmark: Kgs. Lyngby. ISBN: 87-643-0098-6, 2006.

[17] M. Rahman, R. Rahman, L.R. Pearson. Quantiles for finite mixtures of Normal distributions. International Journal of Mathematical Education in Science and Technology, vol. 37, pp. 352-357, 2006.

[18] B.A. Craig, P.P. Sendi. Estimation of the transition matrix of a discrete-time Markov chain. Health Economics, vol. 11,pp. 33-42, 2002.

[19] V. Corradi, N.R. Swanson. Nonparametric bootstrap procedures for predictive inference based on recursive estimation schemes. International Economic Review, vol. 43, pp. 67-109, 2007.

[20] M. Stone. Cross-validation and assessment of statistical predictions (with discussion). Journal of the Royal Statistical Society B, vol. 36, pp. 111-147, 1974.

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

[22] G. Gneiting, F. Balabdaoui, A.E. Raftery. Probabilistic forecasts, calibration and sharpness. Journal of the Royal Statistical Society B, vol. 69, pp. 243-268, 2007.

[23] P. Pinson, H.Aa. Nielsen, J.K. Møller, H. Madsen, G. Kariniotakis. Nonparametric probabilistic forecasts of wind power: required properties and evaluation. Wind Energy, to be published, 2007.

[24] P. Pinson. Estimation of the Uncertainty in Wind Power Forecasting. PhD Thesis, Ecole des Mines de Paris, France, available online: www.pastel.paristech.org, 2006.

[25] C.C.Y. Dorea, L.C. Zhao. Nonparametric density estimation in hidden Markov models. Statistical Inference for Stochastic Processes, vol. 5, pp. 55-64, 2002.