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

DTU Library

Regime-switching modelling of the fluctuations of offshore wind generation

Pinson, Pierre; Christensen, Lasse; Madsen, Henrik; Sørensen, Poul Ejnar; Donovan, Martin Heyman; Jensen, Leo E.

Published in:
Journal of Wind Engineering and Industrial Aerodynamics

Link to article, DOI:
10.1016/j.jweia.2008.03.010

Publication date:
2008

Document Version
Early version, also known as pre-print

Link back to DTU Orbit

Citation (APA):
Pinson, P., Christensen, L., Madsen, H., Sørensen, P. E., Donovan, M. H., & Jensen, L. E. (2008). Regime-switching modelling of the fluctuations of offshore wind generation. *Journal of Wind Engineering and Industrial Aerodynamics*, *96*(12), 2327-2347. https://doi.org/10.1016/j.jweia.2008.03.010

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.

Regime-switching modelling of the fluctuations of offshore wind generation

P. Pinson [a,*], L.E.A. Christensen [a], H. Madsen [a],
P.E. Sørensen [b], M.H. Donovan [c], L.E. Jensen [c],

[a] *DTU Informatics, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark*

[b] *Risø National Laboratory, Technical University of Denmark, 4000 Roskilde, Denmark*

[c] *DONG Energy, 2400 København & 7000 Fredericia, Denmark*

Abstract

The magnitude of power fluctuations at large offshore wind farms has a significant impact on the control and management strategies of their power output. If focusing on the minute scale, it looks like different regimes yield different behaviours of the wind power output. The use of statistical regime-switching models is thus investigated. Regime-switching approaches relying on observable (i.e. based on recent wind power production) or non-observable (i.e. a hidden Markov chain) regime sequences are considered. The former approach is based on either Self-Exciting Threshold AutoRegressive (SETAR) or Smooth Transition AutoRegressive (STAR) models, while Markov-Switching AutoRegressive (MSAR) models comprise the kernel of the latter one. The particularities of these models are presented, as well as methods for the estimation of their parameters. The competing approaches are evaluated on a one-step ahead forecasting exercise with time-series of power production averaged at a 1, 5, and 10-minute rate, at the Horns Rev and Nysted offshore wind farms in Denmark. For the former wind farm, the one-step ahead Root Mean Square Error (RMSE) is contained between 0.8 and 5% of installed capacity, while it goes from 0.6 to 3.9% of installed capacity for the case of Nysted. It is shown that the regime-switching approach based on MSAR models significantly outperforms those based on observable regime sequences. The reduction in one-step ahead RMSE ranges from 19 to 32% depending on the wind farm and time resolution considered. The presented results clearly demonstrate that the magnitude of fluctuations of offshore wind power cannot be considered as simply influenced by the generation level only.

Key words: control, fluctuations, forecasting, modelling, offshore, regime-switching, 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 (Sweet , 2002).

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, and thus the magnitude of power fluctuations may reach very significant levels. For comparison, while studies of wind power fluctuations by Holttinen (2005) or Siebert (2008) (at the level of the whole Denmark area) mention a standard deviation of wind power production between 20 and 22% of installed capacity, that for large offshore wind farms like Nysted and Horns Rev is in the range of 30-38% of installed capacity. A very comprehensive illustration of the difference between the magnitude of power fluctuations for geographically dispersed wind capacities onshore and for concentrated capacities offshore is given in (Sørensen et al , 2007). Modelling the power fluctuations for the specific case of offshore wind farms is a current challenge (Hendersen et al , 2003), for better forecasting offshore wind generation, developing control strategies, or alternatively for simulating the combination of wind generation with storage. The present paper addresses this issue by exploring the applicability and performance of some statistical models.

Operators of offshore wind farms often observe abrupt changes in power production. The fast variations can be related to the turbulent nature of the wind. They are smoothed out when considering the cumulative production for the wind farm, since turbines are spread over a pretty large area. In parallel, when inspecting power production data averaged at a few-minute rate, one observes succession of periods with power fluctuations of significantly variable magnitude, which may be explained by local atmospheric changes e.g. frontline passages and rain showers (Akhmatov , 2005; Sørensen et al , 2008). This phenomenon can already be clearly seen from time-series of wind speed measurements obtained at the meteorological mast at Horns Rev (freely available

* Corresponding author. Pierre Pinson, DTU Informatics, Technical University of Denmark, bg. 321 (020), 2800 Kgs. Lyngby, Denmark. Tel: +45-45253428; fax: +45-45882673

Email address: pp@imm.dtu.dk (P. Pinson).

at `www.winddata.com`). Offshore wind speed fluctuations are then amplified or dampened by the nonlinear power curve of wind parks, yielding wind power fluctuations exhibiting the same kind of regime-based behaviour. This is illustrated in Fig. 1, which shows a 11-day episode with wind power production at Horns Rev consisting of 10-minute averages, with successive periods with fluctuations of significantly different magnitude (e.g. small on 13.07, medium on 11.07 and very large on 17.07). It seems that local 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 process. The example of the influence of wind direction on wind power fluctuations at Horns Rev is discussed in (Akhmatov , 2007). The succession of periods with power fluctuations of lower and larger magnitude calls for the use of regime-switching models. Here, it is explained how to employ Self-Exciting Threshold AutoRegressive (SETAR) models, Smooth Transition AutoRegressive (STAR) models, as well as Markov-Switching AutoRegressive (MSAR) models for that purpose. Their performance are evaluated on a one-step ahead forecasting exercise, and compared to those of linear models, i.e. AutoRegressive Moving Average (ARMA) models. The available data consist in time-series of power production averaged at a 1, 5, and 10-minute rate, for the Horns Rev and Nysted wind farms. The main objective of the present paper is to show that theunderlying regime is not governed by the level of wind generation (as it is the case when employing STAR or SETAR models), but instead by complex local meteorological phenomena, the influence of which can be captured by MSAR models.

Here is Figure 1

2 From linear to regime-switching models

Generated wind power is considered hereafter as a stochastic process for which statistical models are set up in order to describe its temporal evolution. The notation y_t is used for denoting both the state of the stochastic process at time t and the measured value at that time. All the measured power values over the considered period are gathered in the time-series $\{y_t\}$, $t = 1, \ldots, T$, where T is the total number of successive observations. The set $\Omega_t = (y_1, y_2, \ldots, y_t)$ that contains all the observations up to time t, is referred to as the information set. In the present paper, it is considered that no explanatory variable is used for describing the evolution of $\{y_t\}$, i.e. our framework is that of univariate time-series modelling.

As a starting point, the well-known linear ARMA model is presented. Such model encompasses a Moving Average (MA) part and an AutoRegressive (AR)

part. Such model serves as a basis for constructing the regime-switching models. The term 'regime' originates from the assumption such that the considered stochastic process switches between a finite number of distinct (and most often linear) models. Denote by R the number of these regimes. The SETAR and STAR models are consequently presented, with focus to the estimation of their parameters, and their use for time-series forecasting. For these two families of models, the switches from one regime to the other are governed by an observed process, i.e. by some function of lagged values of $\{y_t\}$ for the specific case of univariate time-series models.

2.1 The baseline ARMA model

The linear ARMA(p, q) model, of order p in its AR part and order q in its MA part, is given by

$$y_t = \theta_0 + \sum_{i=1}^{p} \theta_i y_{t-i} + \sum_{j=1}^{q} \phi_j \varepsilon_{t-j} + \varepsilon_t \tag{1}$$

where $\{\varepsilon_t\}$ is a white noise process, i.e. a purely random process with zero mean and variance ${\sigma_\varepsilon}^2$. The AR(p) part corresponds to the weighted sum of the p lagged values of $\{y_t\}$, and the MA part to the weighted sum of q past values of $\{\varepsilon_t\}$. If q is zero the model simplifies to an AR(p) model, and inversely if p equals zero, the model given by Eq. (1) is simply an MA(q). Let us denote by Θ_a the parameter set, that is,

$$\Theta_a = (\theta_0, \ldots, \theta_p, \phi_1, \ldots, \phi_q, \sigma_\varepsilon)^\top \tag{2}$$

with $.^\top$ the transposition operator.

At time $t-1$, the set of parameters Θ_a can be used for calculating the one-step ahead point forecast $\hat{y}_{t|t-1}$ for the considered ARMA(p, q) process. This prediction corresponds to the conditional expectation of $\hat{y}_t$ given Θ_a and the information set Ω_{t-1}, and is readily given by

$$\hat{y}_{t|t-1} = E(y_t|\Omega_{t-1}, \Theta_a) = \theta_0 + \sum_{i=1}^{p} \theta_i y_{t-i} + \sum_{j=1}^{q} \phi_j \varepsilon_{t-j} \tag{3}$$

For further lead times, the above equation can be used recursively.

ARMA models have already been applied for the modelling of wind power time-series. For instance, Milligan et al (2003) have found them appropriate for producing 10-minute ahead forecasts of wind generation for onshore wind farms. A drawback of these linear models is that the variance of the residuals is the same whatever the level of the predictand or of some external variables that may have an influence on the stochastic process. However, Madsen (1996) has shown that only little could be gained by applying more complex models

i.e. bilinear or STAR models for such short horizons. In the present paper, ARMA models are considered as a benchmark for comparison with the more advanced regime-switching models introduced hereafter. Their parameters are determined using Maximum Likelihood (ML) estimation with a Gaussian assumption on the distribution of residuals (Chatfield , 2004). Note that under the assumption of $\{\varepsilon_t\}$ being Gaussian, ML estimation is equivalent to Minimum Mean Square Error (MMSE) estimation (Madsen , 2007), which will be used instead for some of the models introduced below.

2.2 The SETAR model

The Threshold AutoRegressive model (TAR) is a piecewise linear model, for which an AR model is considered in each of the R regimes. For the specific case of the Self-Exciting TAR (SETAR) model, the current regime is determined by a function of lagged values of the time-series (Tong , 1990), which yields pretty abrupt switches from one regime to the other. These regimes are defined by threshold values r_k, $k = 1, 2, \ldots, R-1$, which correspond to the upper bounds of the intervals where given linear models are active. SETAR models have already been successfully applied for modelling exchange rates (Goering et al , 1998) or real estate time-series (Brooks and Maitland-Smith , 1999) among others.

The SETAR$(R; p_1, p_2, \ldots, p_R)$ model is given by

$$y_t = \theta_0^{(m_t)} + \sum_{i=1}^{p_{m_t}} \theta_i^{(m_t)} y_{t-i} + \sigma_{m_t} \varepsilon_t \tag{4}$$

where for a given regime k, p_k and ${\sigma_k}^2$ denote the order of the AR model and the related variance of the noise sequence. $\{\varepsilon_t\}$ is a Gaussian white noise process with unit variance, $\varepsilon_t \sim \mathcal{N}(0,1)$, such that ε_t is independent of Ω_{t-1}. $\{m_t\}$ is the sequence of regimes, taking values in $\{1, 2, \ldots, R\}$, for which each m_t is defined by

$$m_t = \begin{cases} 1, & y_{t-d} \in \,]-\infty; r_1] \text{ (regime 1)} \\ 2, & y_{t-d} \in \,]r_1; r_2] \quad \text{ (regime 2)} \\ \vdots & \vdots \\ R, & y_{t-d} \in \,]r_{R-1}; \infty[\text{ (regime } R) \end{cases} \tag{5}$$

with d seen as a lag parameter. By definition, the SETAR model allows one to have different variances of the residuals in each regime. Note that this family of models could be easily extended to SETARMA models, where the stochastic process is modelled with an ARMA model in each regime, or to open-loop

TAR models, for which explanatory variables can be used for both describing the regime-switching and the models in each regime. Here, we only consider the SETAR model in its simplest form.

The parameters of the SETAR model are estimated with the Minimum Mean Square Error (MMSE) estimation method. It is assumed that the number of regimes and the order of each AR part are already known. Let us first write the parameter set Θ_e for the SETAR model as

$$\Theta_e = (\boldsymbol{\theta}, \mathbf{r}, \boldsymbol{\sigma})^\top \tag{6}$$

with

$$\boldsymbol{\theta} = (\theta_0^{(1)}, \ldots, \theta_{p_1}^{(1)}, \ldots, \theta_0^{(R)}, \ldots, \theta_{p_R}^{(R)})^\top \tag{7}$$
$$\mathbf{r} = (r_1, r_2, \ldots, r_{R-1})^\top \tag{8}$$
$$\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \ldots, \sigma_R)^\top \tag{9}$$

that is, as the collection of the AR model coefficients, the vector of threshold values, and the vector of standard deviations of the noise sequence in each regime, respectively.

Then, the objective function to be minimized is

$$S(\Theta_e) = \sum_{t=p_{\max}+1}^{T} (y_t - \hat{y}_{t|t-1})^2 \tag{10}$$

where $p_{\max} = \max(p_1, p_2, \ldots, p_R)$ is the maximum order of all the AR parts. Also, $\hat{y}_{t|t-1}$ denotes the one-step ahead prediction, which can be readily obtained as the conditional expectation of y_t given Ω_{t-1} and the set of parameters Θ_e. This writes

$$\hat{y}_{t|t-1} = E(y_t|\Omega_{t-1}, \Theta_e) = \theta_0^{(m_t)} + \sum_{i=1}^{p_{m_t}} \theta_i^{(m_t)} y_{t-i} \tag{11}$$

with the regime m_t at time t determined according to Eq. (5). The optimal parameter set $\hat{\Theta}_e$ is finally given by

$$\hat{\Theta}_e = \arg\min_{\Theta_e} S(\Theta_e) \tag{12}$$

The above minimization problem can actually reduce to a linear least square problem for the estimation of the parameters of the AR models given the threshold values, by concentrating the sum of squares. In this case, the Weighted

Least-Squares (WLS) estimate $\hat{\boldsymbol{\theta}}$ of the AR parameters can be calculated as

$$\hat{\boldsymbol{\theta}}(\mathbf{r}) = \left(\sum_{t=1+p_{\max}}^{T} \mathbf{x}_t \mathbf{x}_t^\top \right)^{-1} \left(\sum_{t=1+p_{\max}}^{T} \mathbf{x}_t y_t \right) = \left(\tilde{\mathbf{x}}^\top \tilde{\mathbf{x}} \right)^{-1} \tilde{\mathbf{x}} \mathbf{y} \tag{13}$$

where $\tilde{\mathbf{x}}$ is a matrix for which every row contains past values of y_t for each regime multiplied with 0 or 1 depending on the regime sequence. For instance, if the process is in regime R at time t, the t^{th} row of $\tilde{\mathbf{x}}$ is built as

$$\mathbf{x}_t{}^\top = \mathbf{x}_{t|m_t=R}{}^\top = (0, \ldots, 0, 1, y_{t-1}, \ldots, y_{t-p_R}) \tag{14}$$

From this WLS formulation of the AR parameter estimation, the objective function formulated in Eq. (10) can now simplify to a function of the thresholds only

$$S(\mathbf{r}) = \sum_{t=p_{\max}+1}^{T} (y_t - \hat{\boldsymbol{\theta}}(\mathbf{r})^\top \mathbf{x}_t)^2 \tag{15}$$

The optimal threshold values are then found as

$$\hat{\mathbf{r}} = \arg\min_{\mathbf{r}} S(\mathbf{r}) \tag{16}$$

and the corresponding AR parameter estimates $\hat{\boldsymbol{\theta}}(\mathbf{r})$ (in a MMSE sense) are finally computed with Eq. (13).

Since the objective function $S(\mathbf{r})$ might prove to have a lot of local minima, the initialization of the optimization process is crucial. Here, it has been initialized with different starting points spread over the set of possible $\mathbf{r}$, which have been chosen after inspecting the data.

2.3 The STAR model

Often, abrupt changes between regimes are not satisfactory for modelling stochastic processes, even though separate regimes have been clearly identified. For that reason, Smooth Transition AutoRegressive (STAR) models have been introduced in the literature, in order to feature smooth (and controllable) transitions between regimes (Chan and Tong , 1986; Teräsvirta and Anderson , 1992). The Smooth Transition Bilinear (STBL) model, which belongs to the family of STAR models, has already been successfully applied for describing wind speed variations (Madsen , 1996). For one-step ahead forecasting of half-hourly averaged data, Madsen (1996) has described its performance as slightly better than that of a simple AR(1) model. Here, we turn our attention to the multiple-regime STAR (that we will, for convenience, refer to as STAR only), for which the value of the considered stochastic process $\{y_t\}$ at time

t is given as a weighted average of several AR parts. Since we consider an univariate time-series, the weights assigned to the AR parts are a function of lagged values of $\{y_t\}$. For a number of regimes R, with an AR model of order p_k in the k^{th} regime, the STAR$(R; p_1, p_2, \ldots, p_R)$ is given by

$$y_t = \sum_{k=1}^{R-1} \left(\left(\theta_0^{(k)} + \sum_{i=1}^{p_k} \theta_i^{(k)} y_{t-i} \right) \tilde{G}_k(z_t) + \left(\theta_0^{(k+1)} + \sum_{j=1}^{p_{k+1}} \theta_j^{(k+1)} y_{t-j} \right) G_k(z_t) \right) + \varepsilon_t \tag{17}$$

with

$$\tilde{G}_k(z_t) = 1 - G_k(z_t) \tag{18}$$

where $\{\varepsilon_t\}$ is a Gaussian white noise process with variance ${\sigma_\varepsilon}^2$, $\varepsilon_t \sim \mathcal{N}(0, {\sigma_\varepsilon}^2)$, and G_k is a smooth function that controls the transition between the k^{th} and $(k+1)^{\text{th}}$ regimes. $G_k(z)$ takes values in the unit interval. The regime variable z_t can be defined as a lagged value y_{t-d} of the stochastic process (d is then the lag parameter), or alternatively as an average of a set of lagged values.

The choice of the transition function depends on which type of behaviour is to be modelled. The two most popular transition functions are the exponential and logistic ones. The latter is a smooth increasing function, while the former is a smooth symmetric function around c. The latter is chosen here, since it permits to more clearly separate the different regimes. The logistic function is a 2-parameter function defined as

$$G(z) = \left(1 + \exp\left(-\gamma(z - c)\right)\right)^{-1}, \quad \gamma > 0 \tag{19}$$

where γ is the slope parameter, which controls the transition speed between the regimes, and c is the midpoint between the two regimes. Note that a STAR model with a logistic transition is equivalent to the SETAR model introduced above when $\gamma \to \infty$ with $z_t = y_{t-d}$. Inversely, when $\gamma \to 0$ (which is equivalent to $G(z_t) \to 0.5, \forall z_t$), it reduces to a simple linear AR model, since the AR parts of each regime are equally weighted.

In the multiple regime case, one uses a pair of parameters (γ_k, c_k) for defining the transition function G_k between the k^{th} and $(k+1)^{\text{th}}$ regimes, $k = 1, \ldots, R-1$

$$G_k(z) = \left(1 + \exp\left(-\gamma_k(z - c_k)\right)\right)^{-1}, \quad \gamma_k > 0 \tag{20}$$

The estimation method of the AR parameters for the STAR model is pretty similar to that described above for the case of SETAR models. Write

$$\Theta_s = (\boldsymbol{\theta}, \boldsymbol{\Gamma}, \mathbf{c}, \sigma_\varepsilon)^\top \tag{21}$$

with

$$\Gamma = (\gamma_1, \gamma_2, \ldots, \gamma_{R-1})^\top \tag{22}$$
$$\mathbf{c} = (c_1, c_2, \ldots, c_{R-1})^\top \tag{23}$$

which are the parameters of the transition functions, ${\sigma_\varepsilon}^2$ the variance of the white noise process, and with $\boldsymbol{\theta}$ the parameters of the AR models in each regime, as given by Eq. (7) for SETAR models. Indeed, the MMSE estimate of Θ_s is obtained by minimizing an objective function that is equivalent to that of Eq. (10), which has been introduced for the case of SETAR models

$$S(\Theta_s) = \sum_{t=p_{\max}+1}^{T} (y_t - \hat{y}_{t|t-1})^2 \tag{24}$$

where p_{max} is still the maximum order of the AR models, and the one-step ahead prediction $\hat{y}_{t|t-1}$ is this time calculated with

$$\hat{y}_{t|t-1} = \sum_{k=1}^{R-1} \left(\left(\theta_0^{(k)} + \sum_{i=1}^{p_k} \theta_i^{(k)} y_{t-i} \right) \tilde{G}_k(z_t) + \left(\theta_0^{(k+1)} + \sum_{j=1}^{p_{k+1}} \theta_j^{(k+1)} y_{t-j} \right) G_k(z_t) \right) \tag{25}$$

Then, assuming that R and the order of the AR part in each regime, as well as Γ and $\mathbf{c}$ are known, the MMSE estimate of $\boldsymbol{\theta}$ can be readily obtained from a WLS formulation

$$\hat{\boldsymbol{\theta}}(\Gamma, \mathbf{c}) = \left(\tilde{\mathbf{x}}^\top \tilde{\mathbf{x}} \right)^{-1} \tilde{\mathbf{x}} \mathbf{y} \tag{26}$$

where $\tilde{\mathbf{x}}$ is the weighted regression matrix, for which every row contains for each regime lagged values of $\{y_t\}$, weighted by the value of the transition function at the given time step. For instance, if considering a STAR$(2; 1, 1)$ model, with γ and c the parameters of the logistic transition function G, the t^{th} row of $\tilde{\mathbf{x}}$ is given by

$${\mathbf{x}_t}^\top = \mathbf{x}_t(\gamma, c)^\top = \left(\tilde{G}(z_t), y_{t-1}\tilde{G}(z_t), G(z_t), y_{t-1}G(z_t) \right) \tag{27}$$

From this WLS formulation, the MMSE estimate of Θ_s is obtained by minimizing a reduced form of the objective function defined by Eq. (24)

$$\hat{\Theta}_s = \underset{(\Gamma, \mathbf{c})}{\arg\min}\, S(\Gamma, \mathbf{c}) = \underset{(\Gamma, \mathbf{c})}{\arg\min} \sum_{t=1}^{T} (y_t - \hat{\boldsymbol{\theta}}(\Gamma, \mathbf{c})^\top \mathbf{x}_t)^2 \tag{28}$$

with an appropriate nonlinear optimizer. $\hat{\boldsymbol{\theta}}$ is consequently calculated with Eq. (26). Like for the case of SETAR models, the optimization process may be sensitive to the choice of initial values for Γ and $\mathbf{c}$, and may thus reach local optima. This optimization process is therefore initialized with a set of threshold values spread over the set of possible values, chosen after inspection of the data. In parallel, the initial Γ is chosen to be a unit vector.

The models described in the present Section, i.e. linear ARMA models, SETAR and STAR regime-switching models, as well as related estimation methods, will be employed in Section 4 for the modelling and forecasting of offshore wind power fluctuations.

3 A regime-switching model governed by a hidden Markov chain

The models presented in Section 2 rely on an observable process for determining the actual regime. When considering univariate time-series (like time-series of measured wind power output), the regime is thus compulsorily determined as a function of past values of the process. Markov Switching AutoRegressive (MSAR) models propose an alternative to this observable regime-switching modelling, by allowing the switches to be governed by an unobservable process. This unobservable process is assumed to be a Markov chain. A nice feature of this approach is that it permits to reflect the impact of some external factors on the evolution of the time-series (Hamilton , 1989). Indeed, this class of models has been found particularly suitable for modelling the temporal evolution of weather variables, such as daily rainfall occurrences (Robertson et al , 2003) or wind fields (Ailliot and Monbet , 2006; Ailliot et al , 2006) especially because it manages to capture the influence of some complex meteorological features e.g. related to the motion of large meteorological structures. For the specific case of the fluctuations of offshore wind generation, our aim is to use this hidden Markov chain for describing meteorological features governing the regimes that cannot be determined from past values of measured power production only. MSAR models are described in a first part of the section, the issue of estimation for this class of models is consequently addressed, and a brief last paragraph deals with the use of MSAR models for forecasting.

3.1 Description of MSAR models

MSAR models resemble SETAR models in their formulation. If considering R regimes and AR models of orders $p_1, p_2, \ldots, p_R$ for each of these regimes, the corresponding MSAR$(R; p_1, \ldots, p_R)$ model is indeed given by

$$y_t = \theta_0^{(s_t)} + \sum_{i=1}^{p_{s_t}} \theta_i^{(s_t)} y_{t-i} + \sigma_{s_t} \varepsilon_t \tag{29}$$

where $\{\varepsilon_t\}$ is a Gaussian white noise process with unit variance, $\varepsilon_t \sim \mathcal{N}(0,1)$, ${\sigma_k}^2$ the variance of the noise sequence in the k^{th} regime, and $\{s_t\}$ the regime sequence. But then, MSAR and SETAR models differ in the way the regime sequence is modelled. While this sequence was directly given as a function of

measured power values for SETAR models (cf. Eq. (5)), it is now unobservable. Though, it is assumed that $\{s_t\}$ follows a first order Markov chain on the finite space $\{1, \ldots, R\}$: the regime at time t is determined from the regime at time $t-1$ only, in a probabilistic way

$$P(s_t = j | s_{t-1} = i, s_{t-2}, \ldots, s_0) = P(s_t = j | s_{t-1} = i), \quad \forall i, j, t \tag{30}$$

All the probabilities governing the switches from one regime to the other are gathered in the so-called transition matrix,

$$\mathbf{P} = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1R} \\ p_{21} & p_{22} & \cdots & p_{2R} \\ \vdots & \vdots & \ddots & \vdots \\ p_{R1} & p_{R2} & \cdots & p_{RR} \end{pmatrix}. \tag{31}$$

for which the element p_{ij} represents the probability of being in regime j given that the previous regime was i, as formulated in Eq. (30). 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,

$$\sum_{j=1}^{R} p_{ij} = 1, \quad \forall i \tag{32}$$

since the R regimes represent all the possible regimes that can be reached at any time. And 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.

The set Θ_m of model parameters for MSAR models,

$$\Theta_m = (\boldsymbol{\theta}^{(1)}, \ldots, \boldsymbol{\theta}^{(R)}, \boldsymbol{\sigma}, \mathbf{P})^\top \tag{33}$$

gathers the parameters of the AR parts in each regime,

$$\boldsymbol{\theta}^{(j)} = (\theta_0^{(j)}, \theta_1^{(j)}, \ldots, \theta_{p_j}^{(j)})^\top, \quad k = 1, \ldots, R \tag{34}$$

the standard deviation of the noise sequence in all regimes,

$$\boldsymbol{\sigma}^\top = (\sigma_1, \sigma_2, \ldots, \sigma_R)^\top \tag{35}$$

as well as the transition matrix $\mathbf{P}$.

For illustrating the properties and some characteristics of the model estimation procedure, we simulate in this section a MSAR$(2; 1, 1)$ model. The transition

matrix $\mathbf{P}$ is such that

$$\mathbf{P} = \begin{pmatrix} 0.95 & 0.05 \\ 0.05 & 0.95 \end{pmatrix} , \tag{36}$$

and the other model parameters (i.e. AR coefficients and variance in each regime) are

$$(\boldsymbol{\theta}^{(1)\top}, \boldsymbol{\theta}^{(2)\top}) = (1, .9, 5, .8) \tag{37}$$

$$\boldsymbol{\sigma}^{\top} = (0.8, 1.4) \tag{38}$$

The evolution of this MSAR process over a period of $T = 500$ time-steps is depicted in Fig. 2. The top part of the Figure shows the simulated process $\{y_t\}$, while the bottom part is related to the evolution of the regime sequence $\{s_t\}$. Owing to the choice of transition probabilities the switches between the two regimes are pretty rare.

Here is Figure 2

3.2 Estimation

Estimating the parameters of MSAR models is more complicated than for the case of the regime-switching models introduced in the above sections, due to the unobservable regime sequence $\{s_t\}$. The method described in the following is based on maximum likelihood estimation. More precisely, we apply the Expectation Maximization (EM) algorithm, which consists in an iterative method for maximizing the likelihood (Dempster et al , 1977). This two-step algorithm includes first an expectation step, for which the optimal inference of the regime sequence is determined, and a maximization step, where the parameters of the AR parts are updated by using the likelihood. A first paragraph is devoted to the optimal inference issue, while the EM algorithm is presented in a second one.

3.2.1 Optimal Inference of Regimes

A necessary assumption for determining the optimal inference of the regime sequence is that the number of regimes R, the order of the AR parts, as well as the set of parameters Θ_m are known. Even in this case, it is not possible to readily say in which regime the process belongs to for each observation. The solution to that problem is to consider a filtered probabilistic inference of the

hidden regime sequence given the data (Hamilton , 1989). Define the filtered probability $\xi_{t|t}^{(j)}$ as the conditional probability of s_t being in regime j, given the information set Ω_t at time t and Θ_m, i.e.

$$\hat{\xi}_{t|t}^{(j)} = P(s_t = j|\Omega_t, \Theta_m) \tag{39}$$

Straightforwardly, the filtered probabilities for every regime can be arranged in a vector of filtered probabilities, $\hat{\boldsymbol{\xi}}_{t|t}$, as

$$\hat{\boldsymbol{\xi}}_{t|t} = \left(\hat{\xi}_{t|t}^{(1)}, \hat{\xi}_{t|t}^{(2)}, \ldots, \hat{\xi}_{t|t}^{(R)}\right)^\top \tag{40}$$

The filtered probabilistic inference allows one to iteratively calculate $\hat{\boldsymbol{\xi}}_{t|t}$ starting from $t = 1$, by drawing a simple relation between $\hat{\boldsymbol{\xi}}_{t|t}$ and $\hat{\boldsymbol{\xi}}_{t-1|t-1}$ given the observations up to time $t-1$ and the model parameters

$$\hat{\boldsymbol{\xi}}_{t|t} = g_f(\hat{\boldsymbol{\xi}}_{t-1|t-1}, \Omega_{t-1}, \Theta_m) \tag{41}$$

For deriving the filtered probabilistic inference g_f, we first use the definition of conditional probabilities for reformulating Eq. (39)

$$\hat{\xi}_{t|t}^{(j)} = \frac{f(y_t, s_t = j|\Omega_{t-1}, \Theta_m)}{f(y_t|\Omega_{t-1}, \Theta_m)} \tag{42}$$

with the numerator being the conditional joint density of y_t and s_t being in state j, given Ω_{t-1} and Θ_m, while the denominator is the conditional density of y_t given Ω_{t-1} and Θ_m.

The numerator in the above equation can be computed by

$$f(y_t, s_t = j|\Omega_{t-1}, \Theta_m) = \hat{\xi}_{t|t-1}^{(j)} f(y_t|s_t = j, \Omega_{t-1}, \Theta_m) \tag{43}$$

where $\hat{\xi}_{t|t-1}^{(j)} = P(s_t = j|\Omega_{t-1}, \Theta_m)$ is the forecast of the probability of s_t being in the regime j, given the information set Ω_{t-1} at time $t-1$ and Θ_m. In parallel, the conditional density of y_t, given that s_t is in regime j at time t, and given Ω_{t-1} and Θ_m, is readily given by

$$f(y_t|s_t = j, \Omega_{t-1}, \Theta_m) = \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left(-\frac{(y_t - \mathbf{x}_{t,j}{}^\top \boldsymbol{\theta}^{(j)})^2}{2{\sigma_j}^2}\right) \tag{44}$$

with $\mathbf{x}_{t,j}{}^\top = (1, y_{t-1}, \ldots, y_{t-p_j})$, and by using a Gaussian assumption on the distribution of the noise sequence in each regime, $\varepsilon_t|s_t \sim \mathcal{N}(0, {\sigma_{s_t}}^2)$. Also, it is assumed that the conditional density of y_t only depends on the current regime. It could be possible to consider a potential dependency with previous regimes, but this issue is not addressed here.

The denominator in Eq. (42) is the conditional density of the observation y_t, given the past observations up to time $t-1$. This function can be calculated as the sum of the conditional joint density functions for all regimes

$$f(y_t|\Omega_{t-1},\Theta_m) = \sum_{j=1}^{R} f(y_t, s_t = j|\Omega_{t-1},\Theta_m) \tag{45}$$

where the conditional joint density function $f(y_t, s_t = j|\Omega_{t-1},\Theta)$ is given by Eq. (43).

In order to use a matrix notation for describing the function g_f, let us define $\boldsymbol{\eta}_t$ as the vector which gathers at time t the conditional densities of y_t, given that the regime sequence is in such or such regime

$$\boldsymbol{\eta}_t = \Big(f(y_t|s_t = 1,\Omega_{t-1},\Theta_m),\ldots,f(y_t|s_t = R,\Omega_{t-1},\Theta_m)\Big)^\top \tag{46}$$

and for which each element is simply given by (44). Then for every time-step t, following the previous developments, the vector of filtered probabilities $\hat{\boldsymbol{\xi}}_{t|t}$ can be computed by

$$\hat{\boldsymbol{\xi}}_{t|t} = \frac{\hat{\boldsymbol{\xi}}_{t|t-1} \odot \boldsymbol{\eta}_t}{\mathbf{1}_R{}^\top(\hat{\boldsymbol{\xi}}_{t|t-1} \odot \boldsymbol{\eta}_t)} \tag{47}$$

where $\odot$ denotes the element-wise multiplication, $\mathbf{1}_R$ is a vector of ones of dimension R, and where $\hat{\boldsymbol{\xi}}_{t|t-1}$ is given by

$$\hat{\boldsymbol{\xi}}_{t|t-1} = \mathbf{P}^\top\hat{\boldsymbol{\xi}}_{t-1|t-1} \tag{48}$$

Finally, the filtered probabilistic inference g_f can be defined as

$$g_f \;:\; (\hat{\boldsymbol{\xi}}_{t-1|t-1},\Omega_{t-1},\Theta_m) \rightarrow \hat{\boldsymbol{\xi}}_{t|t} = \frac{(\mathbf{P}^\top\hat{\boldsymbol{\xi}}_{t-1|t-1}) \odot \boldsymbol{\eta}_t}{\mathbf{1}_R{}^\top((\mathbf{P}^\top\hat{\boldsymbol{\xi}}_{t-1|t-1}) \odot \boldsymbol{\eta}_t)} \tag{49}$$

Eqs. (47) and (48) are often referred to as the Hamilton filter, and can be used in an iterative manner in order the calculate the vector of filtered probabilities for all observations. Note that an interesting feature of this filtered probabilistic inference is that the conditional log-likelihood of Θ_m given the data can be obtained as a by-product of the method. $\ell(\Theta_m|\Omega_T)$ is indeed given by

$$\ell(\Theta_m|\Omega_T) = \sum_{t=1}^{T} \log\left(f(y_t|\Omega_{t-1},\Theta_m)\right) = \sum_{t=1}^{T} \log\Big(\mathbf{1}_R{}^\top(\hat{\boldsymbol{\xi}}_{t|t-1} \odot \boldsymbol{\eta}_t)\Big) \tag{50}$$

As an illustration of the method, the sequence of filtered probabilities $\{\hat{\boldsymbol{\xi}}_{t|t}\}$ from the data shown in Fig. 2, and thus related to the MSAR$(2;1,1)$ introduced in Section 3.1, is depicted in Fig. 3 (top part). The bottom part shows

the sequence having the highest probability, which is very similar to the true regime sequence shown in the bottom part of Fig. 2.

Here is Figure 3

3.2.2 Estimation of the model parameters with the EM algorithm

As briefly explained in the introductive part of the present Section, the EM algorithm is an iterative method that permits for MSAR models to estimate the transition matrix, the parameters of the AR parts and the variance in each regime, with the aim of maximizing the conditional log-likelihood $\ell(\Theta_m|\Omega_T)$ of the model parameters. The first step of the algorithm i.e. the expectation step, relates to the inference of the regime. For that purpose, we use the Hamilton filter introduced in the above paragraph. The second part, i.e. the maximization step, consists in the application of a set of update equations for the subsets of models parameters $\boldsymbol{\theta}^{(j)}$, $\boldsymbol{\sigma}$, and $\mathbf{P}$. The maximization of the likelihood is carried by successively applying expectation and maximization steps. Even though one notes that the conditional log-likelihood is not directly used in the method, it can be shown that the EM algorithm asymptotically maximizes that likelihood (Bishop , 1995). In addition, the evolution of its value can be monitored since it can be computed at each expectation step.

As a first stage of the maximization step, one needs to calculate the sequence of smoothed probabilities related to the regime sequence. The smoothed probability $\hat{\xi}_{t|T}^{(j)}$ is defined as the conditional probability of being in regime j at time t given the whole dataset Ω_T and the model parameters Θ_m

$$\hat{\xi}_{t|T}^{(j)} = P(s_t = j|\Omega_T, \Theta_m) \tag{51}$$

For computing $\hat{\xi}_{t|T}^{(j)}$, one uses the smoothed probabilistic inference g_s, which is defined as

$$g_s \ : \ (\hat{\boldsymbol{\xi}}_{t|t}, \hat{\boldsymbol{\xi}}_{t+1|T}, \Omega_T, \Theta_m) \rightarrow \hat{\boldsymbol{\xi}}_{t|T} = \hat{\boldsymbol{\xi}}_{t|t} \odot \left(\mathbf{P}\left(\hat{\boldsymbol{\xi}}_{t+1|T} \oslash \mathbf{P}^\top \hat{\boldsymbol{\xi}}_{t|t}\right)\right) \tag{52}$$

with $\odot$ and $\oslash$ the element-wise multiplication and division, respectively. The smoothed probabilistic inference is derived in a similar fashion than the filtered probabilistic inference introduced in the above paragraph (Hamilton , 1994). Eq. (52) is used recursively, in order to obtain the smoothed probability at time t.

In a second stage, each element p_{ij} of the transition matrix $\mathbf{P}$ can be estimated by visiting the sequence of smoothed probabilities and by applying the following equation

$$\hat{p}_{ij} = \frac{\sum_{t=p_{\max}+1}^{T} P(s_t = j, s_{t-1} = i|\Omega_T, \Theta_m)}{\sum_{t=p_{\max}+1}^{T} P(s_{t-1} = i|\Omega_T, \Theta_m)} \tag{53}$$

$$= \left(\sum_{t=p_{\max}+1}^{T} \hat{\xi}_{t-1|T}^{(i)} \right)^{-1} \sum_{t=p_{\max}+1}^{T} \hat{\xi}_{t|T}^{(j)} \hat{\xi}_{t-1|T}^{(i)} \tag{54}$$

where $p_{\max}$ is the maximum order of the AR models.

Consequently, the parameters $\boldsymbol{\theta}^{(j)}$ of the AR part related to regime j can be re-estimated by using a WLS formulation. It is indeed possible since the probabilities of being in such or such regime at any time t are known. They are given by the sequence of smoothed probabilities $\{\hat{\boldsymbol{\xi}}_{t|T}\}$. In the WLS formulation, each observation is weighted by the smoothed probability of being in that regime.

For that purpose, arrange the smoothed probabilities related to regime j in a weight matrix Σ_j (of dimension $T \times T$), for which the t^{th} element on the diagonal corresponds to the smoothed probability of regime j at time t, that is,

$$\Sigma_j = \begin{pmatrix} \hat{\xi}_{1|T}^{(j)} & & 0 \\ & \ddots & \\ 0 & & \hat{\xi}_{T|T}^{(j)} \end{pmatrix} \tag{55}$$

Then, for each regime j, an estimate of the AR parameters $\boldsymbol{\theta}^{(j)}$ can be computed by solving the usual equation for WLS estimation

$$\hat{\boldsymbol{\theta}}^{(j)} = (\tilde{\mathbf{x}}_j^\top \Sigma_j \tilde{\mathbf{x}}_j)^{-1} \tilde{\mathbf{x}}_j^\top \Sigma_j \mathbf{y}_j \tag{56}$$

where

$$\tilde{\mathbf{x}}_j = \begin{pmatrix} {\mathbf{x}_{1,j}}^\top \\ \vdots \\ {\mathbf{x}_{T,j}}^\top \end{pmatrix} = \begin{pmatrix} 1 & y_{p_j} & \cdots & y_1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & y_{T-1} & \cdots & y_{T-p_j} \end{pmatrix}, \quad \text{and} \quad \mathbf{y}_j = \begin{pmatrix} y_{p_j+1} \\ \vdots \\ y_T \end{pmatrix} \tag{57}$$

Finally, updating the variance ${\sigma_j}^2$ in regime j can be readily done with

$${\hat{\sigma}_j}^2 = \frac{1}{T - (2 + p_{\max})} \left(\sum_{t=1+p_{\max}}^{T} \hat{\xi}_{t|T}^{(j)} \right)^{-1} \sum_{t=1+p_{\max}}^{T} \left(y_t - {\mathbf{x}_{t,j}}^\top \hat{\boldsymbol{\theta}}^{(j)} \right)^2 \hat{\xi}_{t|T}^{(j)} \tag{58}$$

that is, by summing over t the squared residuals of the AR model related to regime j, weighted by the smoothed probability of being in regime j at time

t, and normalized by the sum over t of the smoothed probabilities for this regime.

In order to initiate the EM algorithm, one has to choose an initial value of $\hat{\boldsymbol{\xi}}_{1|0}$. Two main approaches can be used: *(i)* one can set each element of $\hat{\boldsymbol{\xi}}_{1|0}$ to equal probability, i.e. $\hat{\xi}_{1|0}^{(j)} = R^{-1}, \forall j$; *(ii)* the initial value estimation can be included in the likelihood maximization problem solved with the EM algorithm (Hamilton , 1994). For the sake of simplicity, the first approach is chosen. In parallel, the initial transition probabilities gathered in $\mathbf{P}$, as well as the initial variances and AR parameters, are derived after inspection of the dataset.

3.3 Forecasting with MSAR

At time $t-1$, producing a one-step ahead forecast with MSAR models consists in determining the conditional expectation of y_t given the information set Ω_{t-1} and the model parameters Θ_m. For that purpose, one needs first to predict the probabilities to be in such or such regime at time t. This is done with

$$\hat{\boldsymbol{\xi}}_{t|t-1} = \hat{\mathbf{P}}^{\top} \hat{\boldsymbol{\xi}}_{t-1|t-1} \tag{59}$$

where $\hat{\mathbf{P}}$ is the transition matrix estimated over the training set with the EM algorithm described above, and $\hat{\boldsymbol{\xi}}_{t-1|t-1}$ gathers the probabilities of being in such or such regime at time $t-1$ (cf. definition (47)). Then, the one-step ahead prediction for the stochastic process itself is calculated as

$$\hat{y}_{t|t-1} = E(y_t|\Omega_{t-1}, \hat{\Theta}_m) = \hat{\mathcal{A}}_t^{\top} \hat{\boldsymbol{\xi}}_{t|t-1} = \left({\mathbf{x}_{t,1}}^{\top} \hat{\boldsymbol{\theta}}^{(1)}, \ldots, {\mathbf{x}_{t,R}}^{\top} \hat{\boldsymbol{\theta}}^{(R)} \right)^{\top} \hat{\boldsymbol{\xi}}_{t|t-1} \tag{60}$$

where $\hat{\mathcal{A}}_t$ is the vector of the forecasts of each of the AR parts related to the R regimes. The one-step ahead forecast thus consists in a weighted sum of the AR forecasts for every regime, the weights being given by the probability of being in these regimes.

4 Results from offshore case studies

In order to analyse and compare the performance of the various linear and regime-switching models presented above, we use them for describing the fluctuations of offshore wind generation on two real-world case studies. The exercise consists in one-step ahead forecasting of time-series of wind power production. The data for these two offshore wind farms are described in a first paragraph. 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.

4.1 Case studies

The two offshore wind farms are Horns Rev and Nysted, located in Denmark, off the west coast of Jutland and off the south cost of Zealand, 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 power data consist in one-second measurements for each wind turbine. Focus is given to the total power output at Horns Rev and Nysted. Following Sørensen et al (2007, 2008), it has been chosen to model each wind farm as a single wind turbine, the production of which consists in the average of the power generated by all the available wind turbines. These turbines are of nominal capacity 2000 kW and 2300 kW for Horns Rev and Nysted, respectively. Then, a sampling procedure has been developed in order to obtain time-series of 1, 5, and 10-minute power averages. These sampling rates are 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. Whatever the sampling rate, the chosen threshold 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.

From the averaged data, i.e. the three time-series of wind power production averaged at different rates and for the two offshore wind farms, 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 Horns Rev, the training set relates to September 2005. The testing set is composed by 19 periods whose lengths are between 2 and 16 days, identified in the remaining of the whole dataset. Regarding Nysted, the training set corresponds to the period from the 15$^{\text{th}}$ February 2005 to the 9$^{\text{th}}$ March 2005, while the test set gathers 14 periods of length 6-27 days from the rest of the available data.

4.2 Models, estimation setup and evaluation criteria

The various time-series of wind power production are modelled with the linear ARMA and regime-switching SETAR, STAR and MSAR models presented in Sections 2 and 3. The order of the AR and MA parts of ARMA models are chosen to vary between 1 and 5, thus leading to 25 competing ARMA models.

For the case of SETAR and STAR models, expert knowledge indicates that a number of 3 regimes should be defined. Indeed, this choice for 3 regimes is motivated by the influence of the turbine characteristic curve on the variance of wind generation: this variance is smaller in the low and high power range, while it is much larger in the steep slope part of the power curve (Lange , 2005; Pinson , 2006). Consequently, regimes are defined by 2 thresholds. Their optimal values are determined from the nonlinear optimization procedures described in paragraphs 2.2 and 2.3. The lag parameter d is chosen to be 1. Thresholds are initialized by considering various combinations of lower and higher threshold values. The lower ones are picked in the set $\{200, 500, 800\}$ for both wind farms, while the higher ones are picked in the sets $\{1300, 1600, 1900\}$ and $\{1500, 1800, 2100\}$ for Horns Rev and Nysted, respectively. This yields 9 combinations of initial threshold values for each wind farm. For the particular case of STAR models, the shape of the logistic functions is fixed by setting the slope parameter γ to 1. In parallel, the order of the AR parts in each regime is chosen to vary between 1 and 5. This finally results in 1125 competing models in each of the SETAR and STAR model families.

The AR parts in each regime of MSAR models also have orders varying from 1 to 5. In order to compare with SETAR and STAR models, the same number of 3 regimes is defined. Note then that the forecast accuracy of MSAR models presented and discussed below may not be optimal, as it may improve by defining more (or maybe less) regimes. The initialization of the EM algorithm used for parameter estimation consists of picking an initial transition matrix $\mathbf{P}$, as well as initial AR parts, by specifying their parameters and their variances, such that the resulting MSAR model is stationary. A stationary MSAR model is defined as a MSAR model whose AR part in each regime is stationary (cf. the definition of a stationary AR model given in (Chatfield , 2004)). The approach chosen here is to impose the transition matrix and the variances of the AR parts, while having the set of AR parameters varying. The initial $\mathbf{P}$ and $\boldsymbol{\sigma}$ are

$$\mathbf{P} = \begin{pmatrix} 0.8 & 0.1 & 0.1 \\ 0.1 & 0.8 & 0.1 \\ 0.1 & 0.1 & 0.8 \end{pmatrix} \quad , \quad \text{and} \quad \boldsymbol{\sigma} = \begin{pmatrix} 40 \\ 40 \\ 40 \end{pmatrix}$$

The initial parameters of the AR parts take point of departure in the simple

three regime MSAR$(3; 1, 1, 1)$ model with

$$\left(\boldsymbol{\theta}^{(1)\top}, \boldsymbol{\theta}^{(2)\top}, \boldsymbol{\theta}^{(3)\top}\right) = (1, 0.7, 50, 0.9, 100, 0.9)$$

Then, when increasing the order of one of the AR part, the new AR parameter to be initialized is given by a randomly chosen real number. The stationarity of the resulting MSAR model is verified. If this model is not stationary, another random number is drawn. This procedure is repeated until a stationary MSAR model is obtained. For each order of the MSAR model, we consider 10 different initial parameter sets yielding a stationary model. This raises the number of competing MSAR models to 1250.

Either with the ML estimation method and a Gaussian assumption on the residual distributions, or with the MMSE estimation method, the parameters of the models are determined with the aim of minimizing a quadratic error criterion. Therefore, in order to be consistent with the way parameters are estimated, models are also evaluated with a quadratic criterion on the testing set. More precisely, from the large panel of error measures available for evaluating wind power predictions (see (Madsen et al , 2005) for discussion on these error measures), the Root Mean Square Error (RMSE) criterion is chosen.

4.3 Results and discussion

In a first stage, Table 1 lists the best models of each class — best in terms of a minimum RMSE on the testing set — for the time-series related to the Nysted wind farm. For instance for the 1-minute averaged data, the best of the 25 competing ARMA models has been found to be the ARMA$(5, 4)$. In addition, models are ranked from minimum to maximum RMSE.

This Table also gives the characteristics of the optimal SETAR and STAR models, that is, the thresholds that were determined from the optimization procedure. Note that for the 5 and 10-minute averaged data, the thresholds related to the lower regime for the SETAR models are very low (equal to 2.2 and 6 kW, respectively), thus isolating the no-production cases as a regime itself.

Here is Table I

Whatever the sampling rate, the ARMA, SETAR and STAR models have a similar level of performance, while the RMSE for the MSAR models is much lower. The improvement obtained with the Markov-switching models with

respect to the three other types of models ranges from 19% to 32% with respect to the other models, depending on the sampling rate. Then, it seems that STAR models have an advantage versus the two others since it is ranked second in all cases, even though the reduction in RMSE may not be highly significant. The performance of the SETAR and ARMA models are a lot alike. It appears that considering separate regimes does not give any improvement against the classical linear models unless the switches between regimes are smoothed and controlled by some transition function. And, the hypothesis of some succession of regimes that could be captured with a first order Markov chain is validated by these results.

The testing set for Nysted is composed by 14 periods of different lengths and with different characteristics. These characteristics, i.e. various mean production levels and standard deviation of the wind power output are summarized in Fig. 4, for the time-series with a temporal resolution of 5 minutes. Mean production and standard deviation values are very similar for the other 2 averaging rates. In parallel, the detail of the performance of the various models is given in Fig. 5, which depicts the RMSE of the models listed in the above Table for each period. In this Figure, there are only few periods for which the level of performance of the MSAR is worse than that of the other models. In general, the performance of all models is stable from one period to the other, and it does not seem that certain type of conditions would advantage such or such type of model. Also, by noticing that the curves for ARMA, SETAR and STAR models lie on top of each other whatever the period, one understands that modelling the regime-switching with a lagged value of measured wind power output does not yield a more dynamic modelling of the power fluctuations.

Here is Figure 4

Here is Figure 5

In a second stage, the same type of exercise is carried out for the Horns Rev case study. Table 2 gives the sorted list of the best models of each category for the three sampling rates, as well as their characteristics. One sees that the thresholds of the SETAR model for the 5-minute averaged data, and those of the STAR model for the 10-minute sample data, are very close, showing that we almost converged towards two-regime models. The number of regimes has been imposed here, based on the knowledge of the effects of the non-linear and bounded conversion process on power fluctuations. However, the number of regimes could also be considered as a model parameter to be optimized in the future, in order to see its influence on the resulting model performance.

Here is Table II

Again, the STAR models have a slight advantage against the SETAR models, and these latter ones are also slightly better than linear ARMA models. But, they are significantly outperformed by the MSAR models, whatever the sampling rate. Indeed, the improvement proposed by this class of models with respect to the others ranges between 20 and 32%. This confirms once again the interest of considering a hidden Markov chain for modelling the regime-switching. In parallel, note that both for the Nysted and Horns Rev test cases, the average level of RMSE increases as the temporal resolution of time-series increases from 1 to 10 minutes. The persistent nature of wind generation makes that variations in actual wind power output can be more easily modelled and predicted from recent power measures when the lead time is shorter. In addition, the average level of RMSE is significantly larger for Horns Rev than for Nysted, and this whatever the sampling rate. Since the estimated models are globally unbiased, this reveals that the variance of the model residuals is higher for the former wind farm, and hence that the random part of the fluctuations have a larger magnitude. This is certainly due to a more turbulent wind at Horns Rev.

While the characteristics of the 19 periods composing the evaluation set at Horns Rev are gathered in Fig. 6 (in terms of mean and standard deviation of the wind power production for time-series with a 5-minute temporal resolution), the performance of the various models listed in Table 2 are detailed in Fig. 7. As it was the case above for Nysted, mean production and standard deviation values over the various evaluation periods are very similar for the other 2 averaging rates. From Fig. 7 it appears that forecast performance is more variable at Horns Rev. However, MSAR models are still significantly better than the other models for almost all periods, except for periods number 15 and 17. A particularity of these two periods is that they consist in fast successions of drops and increases of wind power output. One may think that in these specific periods the SETAR and STAR models may be more appropriate since they have different AR parts depending on the level of power output, while it is not the case for MSAR and ARMA models. Though, since SETAR and STAR models do not exhibit a more significant improvement with respect to ARMA models for these two periods, this reveals that the regime-switching based on lagged values of power output does not have a higher value in these situations. Therefore, the poorer performance of MSAR models over periods 15 and 17 may simply be explained by the fact that the probabilistic inference of the regime-sequence was not very representative over these periods, owing to some more seldom meteorological phenomena.

Here is Figure 6

Here is Figure 7

5 Concluding remarks

Particular attention has to be given to the modelling of the fluctuations of offshore wind generation, since dedicated models are needed for enhancing the existing control and energy management strategies at offshore wind parks. This issue has been addressed in the present paper, by applying some statistical regime-switching models. The choice for regime-switching approaches has been motivated by the succession of periods with fluctuations of lower and larger magnitudes that can be easily noticed when inspecting time-series of offshore wind power production averaged at a minute rate.

Two different types of regime-switching models have been applied. On the one hand, SETAR and STAR models rely on explicit rules for determining what the current regime is. It is in practice given by some function of past values of measured power. On the other hand, MSAR models are based on the idea that the regime-switching is governed by a hidden Markov process. This, from a theoretical point of view, may allow one to capture some complex influence of meteorological conditions on the wind power fluctuations. For verifying this a priori nice feature of MSAR models, they have been compared to SETAR, STAR and ARMA models on a one-step ahead forecasting exercise, with the aim of minimizing a quadratic error criterion. The time-series considered have been time-series of wind power averaged at a 1, 5, and 10-minute rate for the Nysted and Horns Rev offshore wind farm in Denmark. In all cases, it has been found that MSAR models significantly outperform the other ones: the error reduction ranges between 19 and 32% depending on the test case and the sampling rate. The gain of applying SETAR or STAR models instead of simple linear ARMA models does exist, but is not really significant. In contrast, MSAR models indeed manage to capture the influence of some complex meteorological features on the power fluctuations. This hence demonstrates the complexity of the regime sequence governing successive periods with different behaviours of wind power fluctutions, which cannot be considered as a simple function of the level of wind generation only. It will be of particular interest to study the relation between the temporal evolution of some meteorological variables and the regime sequences of MSAR models in order to determine which of these variables have a direct impact on the magnitude of power fluctuations. Integrating this knowledge in existing forecasting methods will permit to significantly increase their skill for the specific case of very short-term prediction (from some minutes to few hours) at offshore sites.

The results for the two offshore wind farms encourage further investigation

on the use of Markov-switching approaches to the modelling of time-series of offshore wind generation at the minute scale. First, we have used an AR model in each regime, for which the variance of the residuals is fixed and the parameters unconditional. Though, we could allow this variance to vary by using some kind of Generalized AutoRegressive with Conditional Heteroskedasticity (GARCH) models in each regime. Alternatively, we may propose to use AR models whose parameters are conditional to the level of the predictand. In such a case, the Gaussian assumption must be rethought, as it is known that conditional distributions of wind generation given the level of power output are not Gaussian (Pinson , 2006). If a parametric assumption is to be made, a β-distribution assumption is much more suitable. The development and application of conditional β-MSAR models will be the focus of further research works. Finally, wind generation is a non-stationary process, and it would hence be appropriate to enhance the proposed models in order for them to have time-varying parameters.

Broader perspectives regarding follow-up studies include the development of stochastic models for simulating the interaction of offshore wind generation with conventional generation or storage, used as a backup for smoothing the fast power fluctuations at offshore wind farms. Better control strategies will result from the application of these models, which will significantly reduce the potential large costs induced by unwanted large power fluctuations.

Acknowledgements

The results presented have been generated as a part of the project 'Power Fluctuations in Large Offshore Wind Farms' sponsored by the Danish PSO fund (PSO 105622 / FU 4104), which is hereby greatly acknowledged. Energi E2 and Elsam (now part of DONG Energy) are acknowledged for providing the wind power data for the Nysted and Horns Rev offshore wind farms. The authors are grateful for the helpful comments provided by two anonymous reviewers.

References

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

Akhmatov, V., Abildgaard, H., Pedersen, J., Eriksen, P.B., 2005. Integration of offshore wind power into Western Danish power system, in: Proceedings of the 2005 Copenhagen Offshore Wind International Conference and Exhibition, Copenhagen, Denmark.

Ailliot, P., Monbet, V., 2006. Markov switching autoregressive models for wind time series. submitted to Journal of Statistical Planning and Infererence.

Ailliot, P., Monbet, V., Prevosto, M., 2006. An autoregressive model with time-varying coefficients for wind fields. Environmetrics 19, 107-117.

Bishop, C.M., 1995. Neural Networks for Pattern Recognition (first edition). Oxford University Press, Oxford.

Brooks, C., Maitland-Smith, J.K., 1999. Threshold autoregressive and Markov-switching models: an application to commercial real estate. Journal of Property Research 19, 1-19.

Chan, K.S., Tong, H., 1986. On estimating thresholds in autoregressive models. Journal of Times Series Analysis 7, 178-190.

Chatfield, C., 2004. The Analysis of Time Series: An Introduction (sixth edition). Chapman & Hall/CRC, New York.

Dempster, A.P., Laird, N.M., Rubin, D., 1977. Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society B 39, 1-38.

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.

Goering, G.E., Pippenger, M.K., Brooks, C., Maitland-Smith, J.K., 1998. Exchange rate forecasting: results from a threshold autoregressive model. Open Economies Review 9, 157-170.

Hamilton, J.D., 1989. A new approach to the economic analysis of nonstationary time-series and business cycles. Econometrica 57, 357-384.

Hamilton, J.D., 1994. Time Series Analysis. Princeton University Press, Princeton.

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

Holttinen, H., 2005. Hourly wind power variations in Nordic countries. Wind Energy 8, 173-195.

Lange, M., 2005. On the uncertainty in wind power predictions - Analysis of the forecast accuracy and statistical distribution of errors. Journal of Solar Energy Engineering 27, 177-184.

Madsen, H., 1996. Models and Methods for Wind Power Forecasting. Eltra, Skærbæk, Denmark. ISBN 87-87090-29-5.

Madsen, H., 2007. Time Series Analysis. Chapman & Hall/CRC, London.

Madsen, H., Pinson, P., Nielsen, T.S., Nielsen, H.Aa., Kariniotakis, G., 2005. Standardizing the performance evaluation of short-term wind power prediction models. Wind Engineering 25, 475-489.

Milligan, M., Swartz, M., Wan, Y., 2003. Statistical wind power forecasting models. Results for U.S. wind farms. National Renewable Energy Laboratory, Golden (Colorado).

Siebert, N., 2008. Development of Methods for Regional Wind Power forecasting. Ph.D. Dissertation, Ecole des Mines de Paris, Paris (France).

Pinson, P., 2006. Estimation of the Uncertainty in Wind Power Forecasting. Ph.D. Dissertation, Ecole des Mines de Paris, Paris (France).

Robertson, A.W., Kirshner, S., Smyth, P., 2003. Hidden Markov models for modeling daily rainfall occurence over Brazil. Report UCI-ICS-03-27, Information and Computer Sciences, University of California, Irvine (California).

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

Sørensen, P., Cutululis, N.A., Vigueras-Rodríguez A., Madsen, H., Pinson, P., Jensen, L.E., Hjerrild, J., Donovan, M.H., 2008. Modelling of power fluctuations from large offshore wind farms. Wind Energy 11, 29-43.

Sweet, W., 2002. Reap the wild wind. IEEE Spectrum 39, 34-39.

Teräsvirta, T., Anderson, H.M., 1992. Characterising nonlinearities in business cycles using smooth transition autoregressive models. Journal of Applied Econometrics 7, 119-139.

Tong, H., 1990. Non-Linear Time Series - A Dynamical Approach (first edition). Oxford University Press, Oxford.

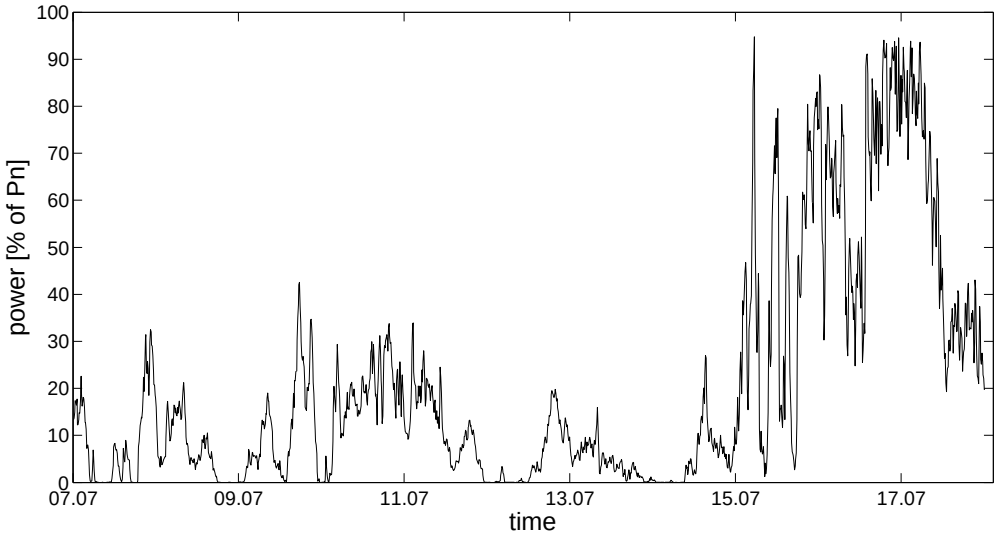

Figure 1. Wind power generated at Horns Rev over a 11-day episode in July 2005. Power values consist in 10-minute averages, normalized by the nominal power P_n of the wind farm (160 MW).

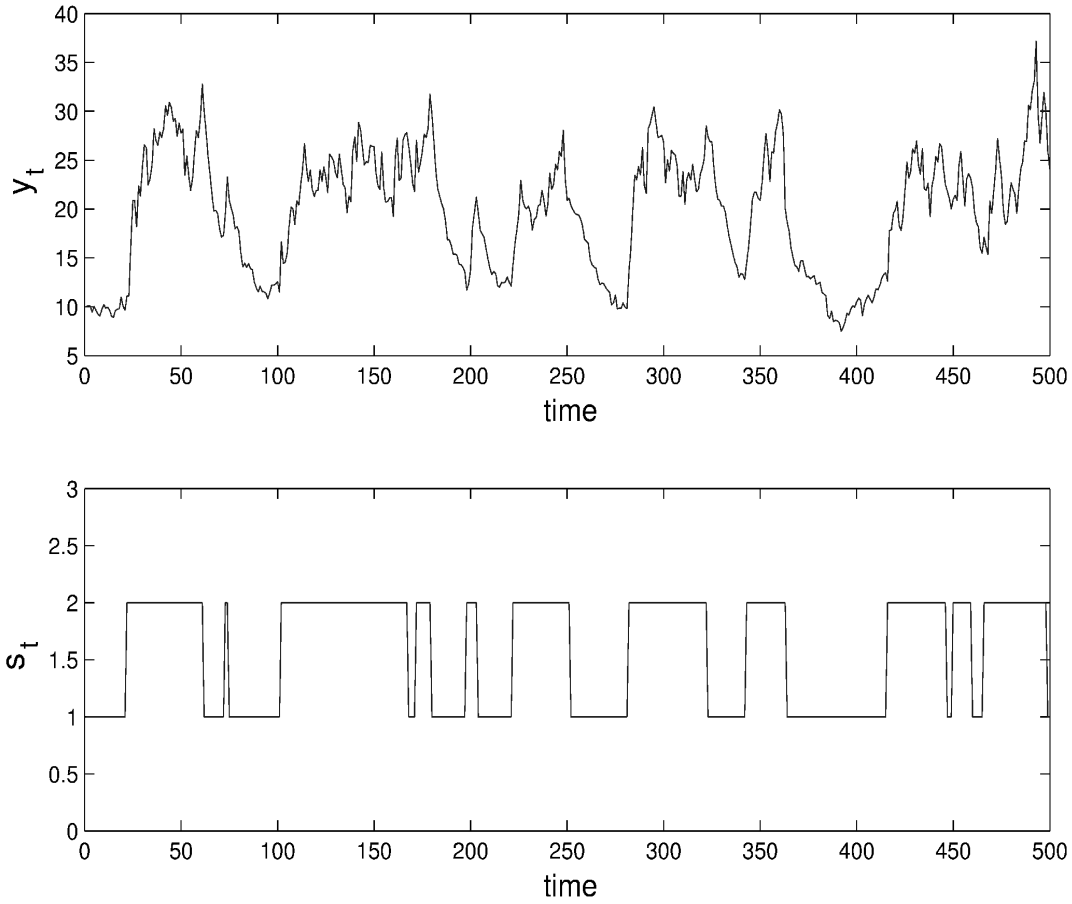

Figure 2. Simulation of a MSAR(2;1,1) process over a period of 500 time-steps. The transition matrix is given by Eq. (36), and the AR coefficients and variance in each regime by equations (37) and (38), respectively. Top: simulated process $\{y_t\}$. Bottom: regime sequence $\{s_t\}$.

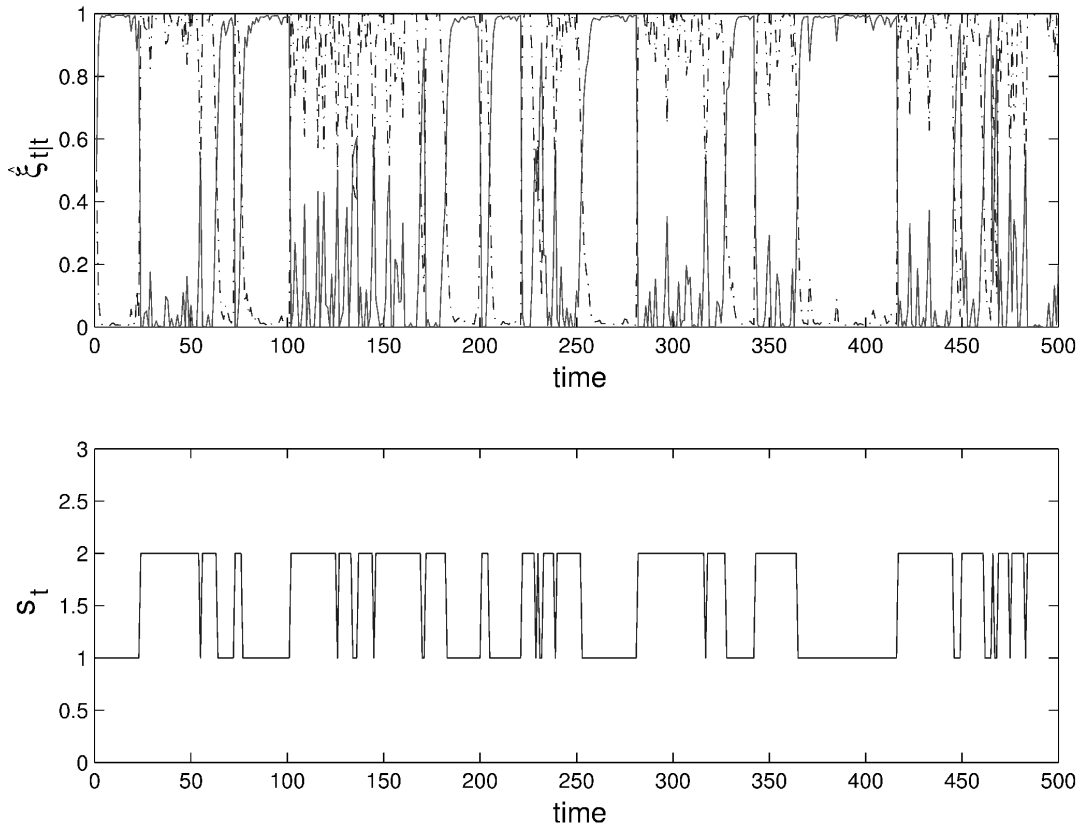

Figure 3. Inference of the regime sequence. The model considered is the MSAR(2; 1, 1) introduced in the above paragraph. Top: The estimated regime sequence $\{\hat{\boldsymbol{\xi}}_{t|t}\}$ (solid line: filtered probability of being in regime 1; dash-dot line: filtered probability of being in regime 2). Bottom: the regime sequence of highest probability.

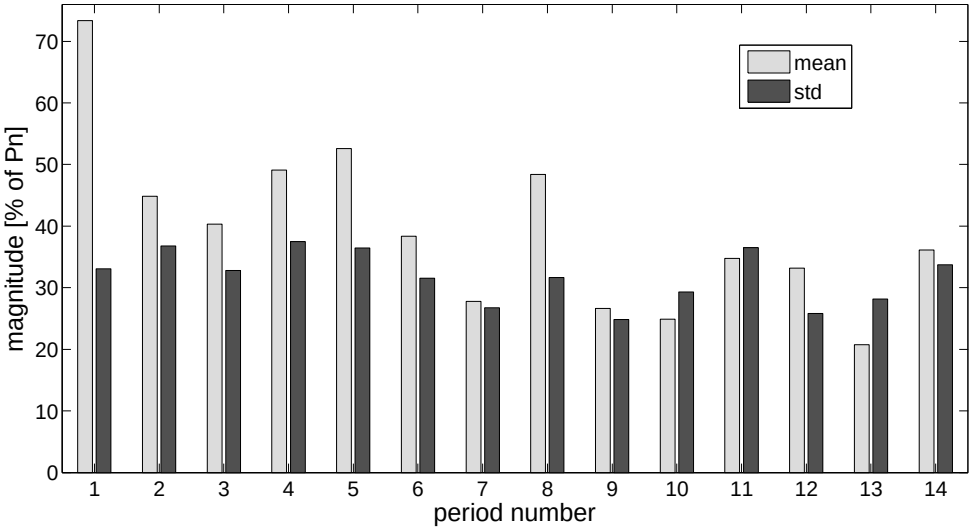

Figure 4. Characteristics of the various evaluation periods at Nysted, in terms of mean production level and standard deviation of the wind power output (both normalized by the nominal power P_n of the wind farm (165 MW)).

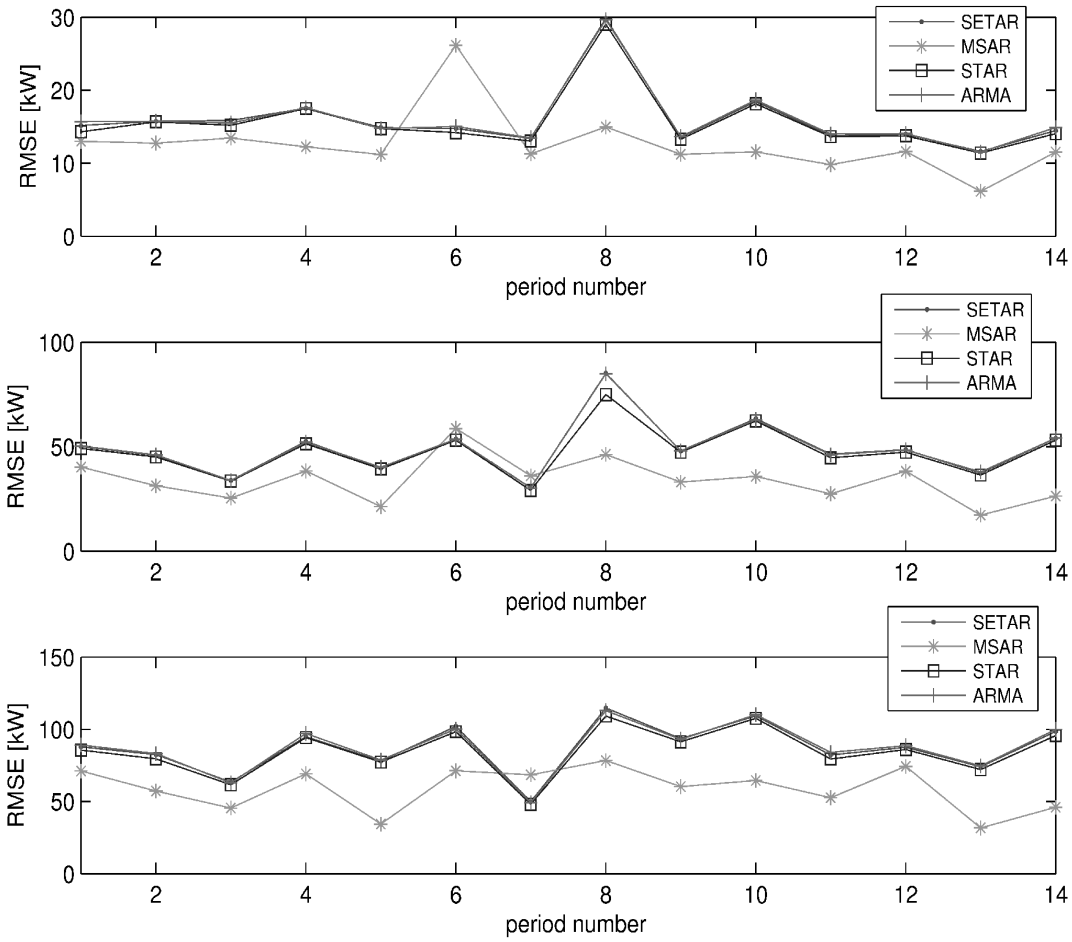

Figure 5. The RMSE on all test data sets from Nysted for each model. Top: 1 minute. Middle: 5 minute. Bottom: 10 minute.

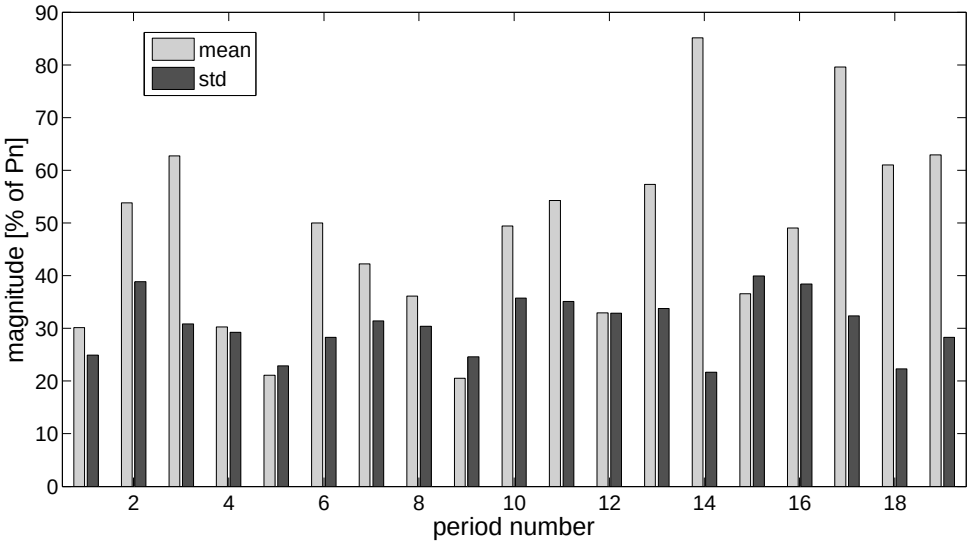

Figure 6. Characteristics of the various evaluation periods at Horns Rev, in terms of mean production level and standard deviation of the wind power output (both normalized by the nominal power P_n of the wind farm (160 MW)).

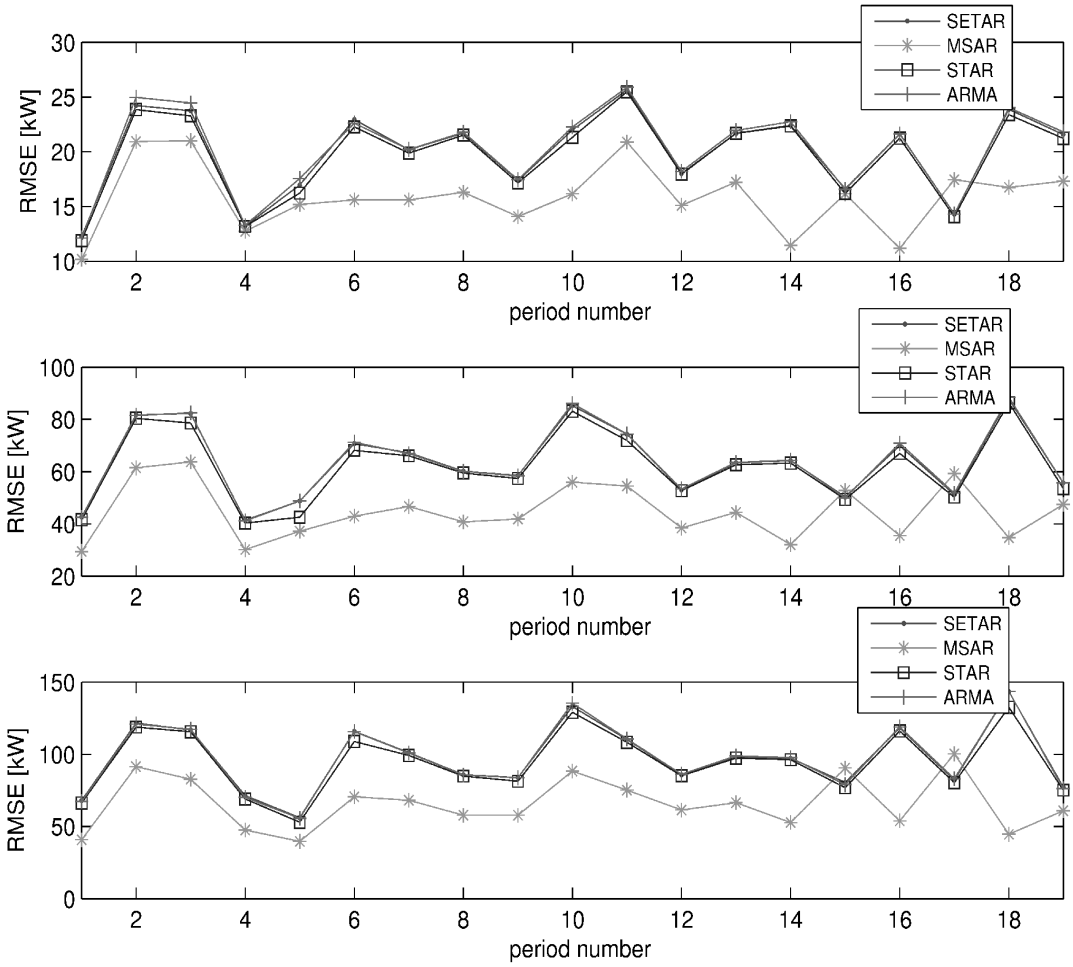

Figure 7. The RMSE on all test data sets from Horns Rev for each model. Top: 1 minute. Middle: 5 minute. Bottom: 10 minute.

Table 1
Performance evaluation of the various linear and regime-switching models for Nysted, with the RMSE criterion, on the times-series averaged at different rates. The left column gives the optimal model orders for each type of model. The optimal threshold values for SETAR models (**r**) and STAR models (**c**) are also given. The models are ranked as a function of their RMSE on the testing set.

(a) 1-minute averaged data

Model	RMSE [kW]	**r** [kW]	**c** [kW]
MSAR(3; 4, 4, 3)	13.3	–	–
STAR(3; 5, 5, 5)	16.1	–	(920.9, 2096.6)
SETAR(3; 4, 4, 4)	16.5	(203.6, 2006.3)	–
ARMA(5, 4)	16.5	–	–

(b) 5-minute averaged data

Model	RMSE [kW]	**r** [kW]	**c** [kW]
MSAR(3; 4, 5, 5)	35.5	–	–
STAR(3; 5, 5, 5)	48.9	–	(827.3, 1638.7)
ARMA(4, 5)	50.8	–	–
SETAR(3; 1, 3, 3)	50.9	(2.2, 2149.8)	–

(c) 10-minute averaged data

Model	RMSE [kW]	**r** [kW]	**c** [kW]
MSAR(3; 2, 4, 4)	60.6	–	–
STAR(3; 5, 5, 5)	86.2	–	(579.4, 1545.4)
SETAR(3; 3, 5, 5)	88.6	(6.0, 1595.4)	–
ARMA(5, 1)	88.9	–	–

Table 2
Performance evaluation of the various linear and regime-switching models for Horns Rev, with the RMSE criterion, on the times-series averaged at different rates. The left column gives the optimal model orders for each type of model. The optimal threshold values for SETAR models (**r**) and STAR models (**c**) are also given. The models are ranked as a function of their RMSE on the testing set.

(a) 1-minute averaged data

Model	RMSE [kW]	**r** [kW]	**c** [kW]
MSAR$(3;3,2,5)$	16.1	–	–
STAR$(3;5,5,5)$	20.1	–	(505.5, 1824.2)
SETAR$(3;4,4,4)$	20.4	(432.2, 1824.3)	–
ARMA$(2,1)$	20.6	–	–

(b) 5-minute averaged data

Model	RMSE [kW]	**r** [kW]	**c** [kW]
MSAR$(3;3,1,5)$	45.0	–	–
STAR$(3;5,5,5)$	63.3	–	(892.9, 1673.0)
SETAR$(3;3,2,3)$	65.1	(744.4, 760.6)	–
ARMA$(2,2)$	65.3	–	–

(c) 10-minute averaged data

Model	RMSE [kW]	**r** [kW]	**c** [kW]
MSAR$(3;3,2,4)$	68.1	–	–
STAR$(3;5,4,5)$	96.9	–	(705.0, 779.9)
SETAR$(3;3,3,1)$	99.8	(240.2, 2300) [a]	–
ARMA$(5,2)$	99.9	–	–

[a] Here, the upper threshold for the optimal SETAR model has converged to the nominal power value, indicating that this optimal model is indeed a SETAR(2;3,3) models.