Modeling and forecasting of wind power generation - Regime-switching approaches

Trombe, Pierre-Julien ; Madsen, Henrik; Pinson, Pierre

Publication date:
2013

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

Modeling and forecasting of wind power generation – Regime-switching approaches

Pierre-Julien Trombe

Kongens Lyngby 2012
IMM-PHD-2012-287

Technical University of Denmark
Informatics and Mathematical Modelling
Building 321, DK-2800 Kongens Lyngby, Denmark
Phone +45 45253351, Fax +45 45882673
reception@imm.dtu.dk
www.imm.dtu.dk

IMM-PHD: ISSN 0909-3192

Preface

This thesis was prepared at the department of Informatics and Mathematical Modeling of the Technical University of Denmark in partial fulfillment of the requirements for acquiring the Ph.D. degree in Engineering.

The thesis deals with different aspects of the modeling and forecasting of offshore wind power generation. The main focus is on the application of regime-switching time series models, but also the exploration of weather radar observations as a new source of information for these models.

The thesis consists of a summary report and a collection of four research papers written during the period 2009–2012.

Lyngby, November 2012

Pierre-Julien Trombe

Acknowledgements

I would like to thank Pierre and Henrik for their patience, support and guidance during this PhD, Janne for technical assistance with life matters, Philip and Peder for never loosing hope of improving my skills in Danish and Emacs editing, Marcito for his caffeinic support and supplying huge amounts of Wasa knækbrød, Ewa and Julija for spreading happiness around them and feeding me with sugar and gossips, Juan Miguel for his liquid life style, Roland for keeping a smile up on his face under all circumstances, and finally all my other colleagues for making this PhD a very interesting, positive and surprising journey.

In addition, I would like to express my gratitude to all my friends. In particular those I met in Copenhagen and was fortunate to spend time with. Special mention goes to Anna Helga for all the great times we had together.

I would also like to acknowledge the "Mesoscale", "SafeWind" and "Radar@Sea" projects for partly funding this PhD.

Summary

The present thesis addresses a number of challenges emerging from the increasing penetration of renewable energy sources into power systems. Focus is placed on wind energy and large-scale offshore wind farms. Indeed, offshore wind power variability is becoming a serious obstacle to the integration of more renewable energy into power systems since these systems are subjected to maintain a strict balance between electricity consumption and production, at any time. For this purpose, wind power forecasts offer an essential support to power system operators. In particular, there is a growing demand for improved forecasts over very short lead times, from a few minutes up to a few hours, because these forecasts, when generated with traditional approaches, are characterized by large uncertainty. In this thesis, this issue is considered from a statistical perspective, with time series models. The primary case study is the Horns Rev wind farm located in the North Sea.

Regime-switching aspects of offshore wind power fluctuations are investigated. Several formulations of Markov-Switching models are proposed in order to better characterize the stochastic behavior of the underlying process and improve its predictability. These models assume the existence of a hidden or unobservable regime sequence. Estimation methods are presented in both Bayesian and Frequentist frameworks. Markov-Switching models enable to highlight structural breaks in the dynamics of offshore wind power generation, with alternating periods of high and low variability. They also yield substantial gains in probabilistic forecast accuracy for lead times of a few minutes. However, these models only integrate historical and local measurements of wind power and thus have a limited ability for notifying regime changes for larger lead times. For that purpose, there is a long tradition in using meteorological forecasts of wind speed and direction that are converted into wind power forecasts. Nevertheless, meteorological forecasts are not informative on the intra-hour

wind variability and thus cannot be used in the present context focusing on temporal resolutions of a few minutes. Instead, this thesis investigates the use of weather radar observations for monitoring weather conditions in the vicinity of offshore wind farms, with the ambition of establishing a link between the passage of precipitation systems and high wind variability. The underlying motivation of this approach is twofold. First, it aims at providing a meteorological interpretation of the hidden regimes as estimated by regime-switching models. Second, it aims at determining an observed sequence of regimes based on the information extracted from the observations supplied weather radar observations. This approach, combining both meteorological and statistical expertise, opens up new possibilities for designing prediction systems in wind energy.

Resumé

Denne afhandling beskriver en række metoder til løsning af række udfordringer, der opstår når en betydelig mængde vedvarende energiproduktion skal integreres i elsystemet. Fokus er på vindenergi, specielt på store havvindmølleparker. Et alvorligt problem for en fortsat vækst af elproduktion med havvindmølleparker er udsving i deres elproduktion, da der stilles store krav til at balancen mellem elproduktionen og elforbruget skal holdes, uden undtagelser. For at opnå denne balance er prognoser af elproduktionen fra vindmøller et essentielt redskab for systemoperatører. Der er i særlig grad et stigende behov for bedre prognoser med kort horisont, fra få minutter og op til et par timer frem, da de nuværende prognoser er behæftet med relativ stor usikkerhed for korte horisonter. I afhandlingen beskrives metoder, til forudsigelse af vindproduktionen på korte horisonter, baseret på statistisk modellering. Det primære casestudie er Horns Rev vindmøllepark i Nordsøen.

Aspekter vedrørende tilstandsskift i variabiliteten af elproduktionen fra havvindmølleparker undersøges. En række formuleringer af Markov-switching modeller foreslås til karakterisering af de underliggende stokastiske processer og danner basis for forbedrede prognosemodeller. Markov-switching modellerne er baseret på en antagelse af, at der i processerne er sekvenser af skjulte tilstande, som kan estimeres, hvilket gøres med både med en frekventistisk og Baysiansk tilgang. Modellerne gør det muligt af estimere strukturelle skift i havvindmølle elproduktionens dynamik og identificere perioder skiftende mellem høj og lav variabilitet. Desuden forbedres kvaliteten af probabilistiske prognoser med horisonter på op til få minutter betydeligt.

Markov-switching modellerne anvendes med historisk data og lokale observationer. For at opnå yderligere forbedringer af prognoserne benyttes traditionelt meteorologiske prognoser af vindhastighed og vindretning. Dette er

dog ikke optimalt i den nuværende kontekst til beskrivelse af variabiliteten i vinden med henblik på prognoser med opløsning i minutter og med horisonter op til få timer. Derfor er der i afhandlingen beskrevet nye metoder til at integrere vejrradar observationer med det formål, at etablere en forbindelse mellem passerende nedbørssystemer og vindvariabilitet ved kraftig vind. Den underliggende motivation har to vigtige aspekter. For det første, at tilvejebringe en meteorologisk fortolkning af de skjulte tilstande, som er estimeret i Markov-switching modellerne. For det andet, at bestemme en sekvens af tilstande frem i tiden på baggrund af informationen i vejrradar observationer. Ved på denne måde at kombinere en meteorologisk og statistisk tilgang åbnes nye muligheder for at designe prognosesystemer.

List of publications

Papers included in the thesis

A Pierre-Julien Trombe, Pierre Pinson, Henrik Madsen (2012). A general probabilistic forecasting framework for offshore wind power fluctuations. *Energies*, 32:621-657.

B Pierre-Julien Trombe, Pierre Pinson (2012). High-resolution forecasting of wind power generation with regime-switching models and off-site observations. IMM Technical Report 2012-15.

C Pierre-Julien Trombe, Pierre Pinson, Thomas Bøvith, Nicolaos Antonio Cutululis, Caroline Draxl, Gregor Giebel, Andrea Hahmann, Niels Einar Jensen, Bo Præstgaard Jensen, Nina F. Le, Henrik Madsen, Lisbeth Birch Pedersen, Anders Sommer, Claire Vincent (2012). Weather radars – The new eyes for offshore wind farms? Submitted to: *Wind Energy (2012)*.

D Pierre-Julien Trombe, Pierre Pinson, Henrik Madsen (2012). Automatic classification of offshore wind regimes with weather radar observations. Submitted to: *IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing (2012)*.

Other Publications

In addition to the papers listed before, the following contributions were also prepared during the project period.

Conference publications (Peer reviewed)

- Braulio Barahona, Nicolaos Antonio Cutululis, Pierre-Julien Trombe, Pierre Pinson (2013) Regime-based control to reduce power fluctuations from offshore wind power plants. PowerTech, Grenoble, France.

- Pierre-Julien Trombe, Pierre Pinson, Henrik Madsen, Niels Einar Jensen, Lisbeth Birch Pedersen, Anders Sommer, Nina F. Le (2011). Can weather radars help monitoring and forecasting wind power fluctuations at large offshore wind farms? 10th International workshop on large-scale integration of wind power into power systems, Aarhus, Denmark.

Other conference publications (no proceedings)

- Pierre-Julien Trombe, Pierre Pinson, Claire Vincent, Henrik Madsen (2012). Weather radars for offshore wind power applications. 9th European Conference on Applied Climatology, Lodz, Poland. **(Oral presentation)**

- Pierre-Julien Trombe, Pierre Pinson, Claire Vincent, Henrik Madsen, Niels Einar Jensen, Thomas Bøvith, Nina F. Le, Anders Sommer (2012). Weather radars — A new pair of eyes for offshore wind farms? EWEA Annual Event 2012, Copenhagen, Denmark. **(Poster - Best poster award)**

- Pierre-Julien Trombe, Pierre Pinson, Sven Creutz Thomsen, Henrik Madsen (2011) Radar@Sea – Towards improving short-term wind power forecasts. European Geosciences Union, Vienna, Austria. **(Poster)**

- Pierre-Julien Trombe, P. Pinson, Sven Creutz Thomsen, Henrik Madsen (2010) A robust optical flow method for deriving 2-D rain fields from weather radar images. DTU Vision Days, Lyngby, Denmark. **(Poster)**

- Pierre-Julien Trombe, Pierre Pinson, Henrik Madsen (2009). Modeling and forecasting of wind power fluctuations at large offshore wind farms. European Offshore Wind Conference, Stockholm, Sweden. **(Oral presentation)**

- Pierre-Julien Trombe, Pierre Pinson, Henrik Madsen (2009). Modeling and Forecasting of wind power fluctuations using Markov-switching AR-GARCH models. International Symposium on Forecasting, Hong Kong, China. **(Oral presentation)**

Technical Reports

- Julija Tastu, Pierre Pinson, Pierre-Julien Trombe, Henrik Madsen (2011). Spatio-temporal correction targeting Nysted offshore – Probabilistic forecasts. Technical report, Technical University of Denmark, Dpt. of Informatics and Mathematical Modeling.

- Søren Larsen, Xiaoli Guo Larsen, Claire Vincent, Poul Sørensen, Pierre Pinson, Pierre-Julien Trombe, Henrik Madsen, Nicolaos Antonio Cutululis (2011). Mesoscale Wind Variability, Final Report (ISBN: 978-87-550-3937-7).

- Søren Ostergaard Jensen, Christian Nielsen, Christian Heerup, Henrik Madsen, Lars Olsen, Jørn Toftum, Philip Andersen Delff, Pierre-Julien Trombe, Søren Laursen (2010). Characterization and optimized control by means of multi-parameter controllers (ISBN: 87-7756-772-2).

Contents

Part I

Summary Report

Introduction

Wind energy in the present energy context

With the emergence of new energy demands linked to population growth or continuous industrial development, the world energy consumption is expected to keep on growing in the coming decades (see Figure 1). As of today, our society is heavily reliant on fossil fuels which account for more than 80% of this consumption (International Energy Agency, 2011). However, coal, oil and gas are available in limited quantities and the processes for converting them into electricity are responsible for large CO_2 emissions in the atmosphere, contributing to global warming. Therefore, the transition towards a more sustainable

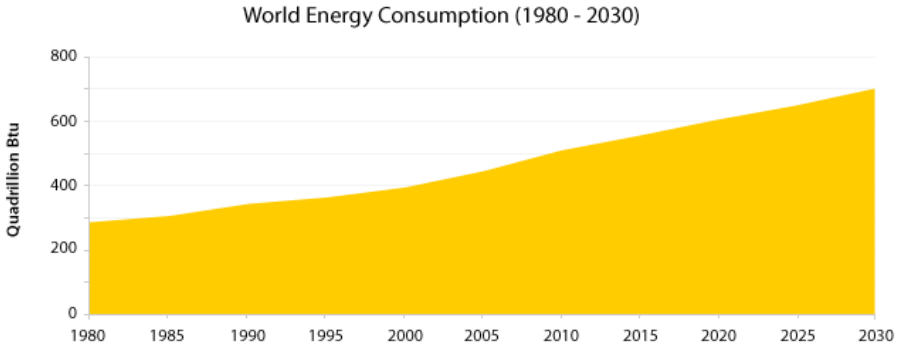


Figure 1: World energy consumption. (1980-2010) Historical data, (2010-2030) Projection. (Source: U.S. Energy Information Administration)

society is conditioned upon changes in energy consumption and production patterns.

In that global energy context, renewable energy sources such as sun or wind are considered credible candidates for meeting new energy demands and partly substituting fossil fuels. First, because these resources are available in large quantities. For instance, there is more potential energy in one hour of solar radiations hitting the atmosphere and the surface of the Earth than the world consumes in one year. Secondly, because their conversion into electricity is almost CO_2 free. Thirdly, because their price as a fuel is null. Finally, because they could enable countries, and particularly, developing countries in becoming more self-sufficient in energy.

Of all alternative and renewable energy sources, wind energy is the one experiencing the fastest growth. Despite the many barriers – social, political, economic, technical (see Beck and Martinot (2004)) –, a number of countries such as Spain and Denmark already experience a large penetration, above 10 and 20% respectively, and ambitious targets have been set for the future. Denmark plans to meet 35% of its final energy consumption with wind power by 2020, for instance (Danish Ministry of Climate, Energy and Building, 2012). However, integrating such large amounts of wind power and, more generally, renewable energy into power systems does not go without problems. In particular, wind energy characteristics represent a challenge to power system design and operation (Ackermann et al., 2005).

Integrating wind power into power systems: what is the issue?

Power systems are managed as dynamical systems containing uncertainty and subjected to constraints. These systems consist of complex networks of interconnected power generators (e.g., wind farms, coal power plants, hydro power plants, etc), transmission and delivery lines for dispatching electrical power to consumers. A first set of constraints is imposed by the layout of these power systems (i.e., the spatial arrangement of lines and generators) and by the respective capacities of these generators and lines. An additional constraint imposes Transmission System Operators (TSOs) to maintain a strict balance between electricity production and consumption, at any time, in order to ensure grid stability. In particular, one of the main issues in operating power systems is to meet peaks in the electricity demand often associated with cold weather in the winter. The uncertainty in managing power systems stems from the spatio-temporal variability of the electricity demand (i.e., the load) and generation,

as well as the availability of power generators. For instance, the behavior of an electricity consumer is not known in advance, neither is the time when a fault causes a power plant to stop generating electricity.

The integration of large amounts of wind power into power systems is currently a challenge because wind differs from conventional fuels (e.g., coal, gas) in many aspects. Firstly, wind is not dispatchable. This means that wind power can only be generated in places where wind is available in sufficient quantities, potentially far from where the power is needed and/or in remote locations (e.g., out at sea or mountainous regions). Secondly, wind cannot be stored, and the generated power is either instantaneously integrated into power systems, potentially forcing conventional power units to decrease or stop their production, or lost. Thirdly, wind is governed by the dynamics of the atmosphere which makes it variable. Consequently, its output power is also variable.

Historically, power systems were designed to handle the variability of the load while that of the generation was considered of lesser importance, in comparison. However, with the growing penetration of wind power, the inherent and potentially extreme variability of wind power substantially increases the uncertainty in managing power systems. In that respect, there is not a unique solution for handling this uncertainty as power systems have different specificities. Rather, a number of changes can be applied to power system design and operation to optimize their safe and economic management: (i) the deployment of energy storage technologies such as pumped storage plants or heat pumps (Hewitt, 2012), (ii) an increase in backup capacity with more fast-acting generators such as coal and heat power (CHP) plants, and (iii) new transmission and interconnection lines, or an increased use of existing ones. Yet, all these changes go towards a same direction, an extension of the power system infrastructure which comes at a high cost. Furthermore, an increased use of CHP plants would not be in line with policies aiming at reducing CO_2 emissions. Another possible change is to apply Demand Side Management (DSM) principles. They consist of introducing more flexibility in the load by inducing changes in electricity consumption patterns, through financial incentives for instance. However, the deployment of DSM technologies will not be immediate, mainly because of social acceptance issues (e.g., , installation of metering devices in individual homes, changes in consumer mentality), as discussed in (Strbac, 2008). Ultimately, the most cost-effective change for reducing that uncertainty is through an increased use of wind power forecasts. The use of forecast information offers the advantage of already being a well spread practise in managing power systems. In particular, TSOs have long relied on the availability of accurate load forecasts for reducing the uncertainty associated with the load variability (Gross and Galiana, 1987). Today, wind power forecasts are already used by many electrical utilities for their operations and their

value is clearly acknowledged as they reduce operating costs (Giebel et al., 2011).

Wind power forecasts

A wealth of approaches and models have been proposed for generating accurate wind power forecasts (Giebel et al., 2011). However, with the growing share of wind power into power systems, improving the accuracy of these forecasts is paramount (Jones and Clark, 2011). Wind power forecasts are required at different spatio-temporal scales and horizons, depending on their intended application. Traditionally, five time scales are considered:

1. *Ultra short-term* (second range): applications include the control of individual wind turbine control (e.g., the pitch angle of blades).

2. *Very short-term* (minute range, up to 1 hour ahead): applications include the management of the immediate regulating and spinning reserves. The former reserve is activated over time intervals of 15–20 minutes, after the system experiences a sudden and large deviation between scheduled and actual power generation. The latter reserve corresponds to the extra capacity available by increasing the power output of generators already connected to the power system.

3. *Short-term* (hour range, from 0 up to 2-3 days ahead): applications include the operation of supplemental reserve (e.g., the extra capacity non connected to the power system that requires a delay to be activated), scheduling unit commitment and economic dispatch, trading of electricity on energy markets,

4. *Medium term* (day range, from 0 up to 7 days ahead): economic dispatch and unit commitment of large power plants.

5. *Long-term (week range)*: applications include planning maintenance operations of wind farms.

In addition, these forecasts can be issued at different spatial scales, from single wind farms, to regions or a whole power system. In that respect, wind power predictability over regions tends to improve with the spatial dispersion of wind farms, owing to the smoothing effect of wind power variability (Focken et al., 2002).

A unique type of model or approach cannot be used for meeting the requirements over all spatio-temporal scales. Rather, the choice of an approach is

modulated by the horizon of interest. The first type of approach is physical. It relies on Numerical Weather Prediction (NWP) models and dynamical equations of atmospheric flows for generating meteorological forecasts. These forecasts can potentially be refined by integrating information on the terrain such as the roughness or orography. Meteorological forecasts are then converted to wind power forecasts through an idealized power curve, as shown in Figure 2. NWP models are usually run from 2 to 4 times a day. The temporal resolution of NWP forecasts is between 1 and 3 hours. Physical approaches are well suited for short and medium term forecasts. A good introduction on NWP models is given in Monteiro et al. (2009). The second type of approach is mathematical and consists of using statistical models (e.g., time series models, artificial neural networks) to find out spatio-temporal dependencies between the wind power production and explanatory variables (e.g., historical observations of wind or wind power). This type of approach usually outperforms physical approaches for very short-term forecasts and up to 6 hours ahead. A third type of approach consists of combining both NWP and statistical models. The NWP models are first used for generating forecasts of meteorological variables (e.g., wind speed and direction, temperature, air density) that can be converted to wind power forecasts with kernel smoothing techniques a posteriori. This hybrid approach is usually used for short and medium term forecasts (Giebel et al., 2011).

Whatever the spatio-temporal scales and horizons of interest, and irrespectively of the approach employed, several types of forecasts can be issued: deterministic forecasts, probabilistic forecasts and scenarios. Scenarios are out of the scope of this thesis and thus are not discussed here. Instead, we refer to Pinson et al. (2009) for a comprehensive introduction on these forecasts. Deterministic or point forecasts are provided as a single value for each look-ahead time. They are informative on the conditional expectation of the wind power generation. This type of forecast remains largely used by TSOs for optimizing the management of power system thanks to their high interpretability (Jones and Clark, 2011). However, wind power generation is not perfectly predictable since our knowledge of the mechanisms governing its variability is incomplete. Consequently, each point forecast contain some uncertainty. This uncertainty can be expressed in the form of probabilistic forecasts (e.g., predictive densities, prediction intervals) around point forecasts. An example of such probabilistic forecasts is given in Figure 3. The additional value of using probabilistic forecasts, compared to the sole point forecasts, has been demonstrated in the case of a wind power producer aiming at trading its production, yielding higher incomes (Pinson et al., 2007). More generally, probabilistic forecasts are a prerequisite for optimal decision-making under uncertainty, as discussed in Gneiting (2008).

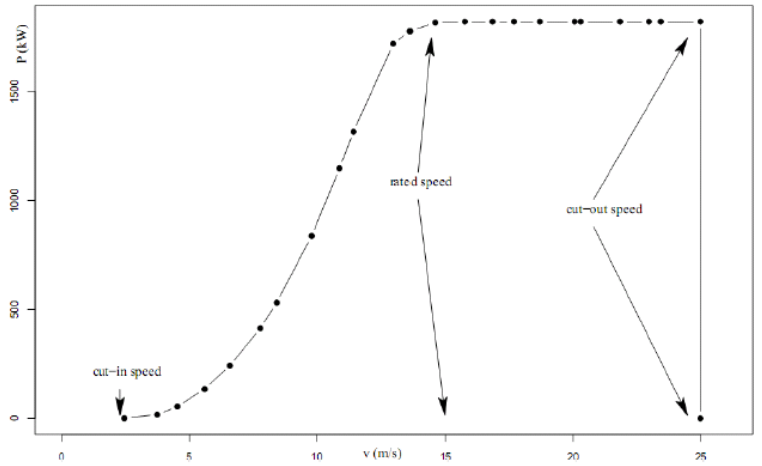


Figure 2: A simple example of an idealized power curve for a single wind turbine where wind power is a functon of wind speed. The cut-in speed corresponds to the minimum wind speed for producing power. Rated speed is the minimum wind speed at which wind turbines produce rated or maximum power. Cut-out speed is speed at which wind turbines stop producing power for safety reasons. (Source: Monteiro et al. (2009))

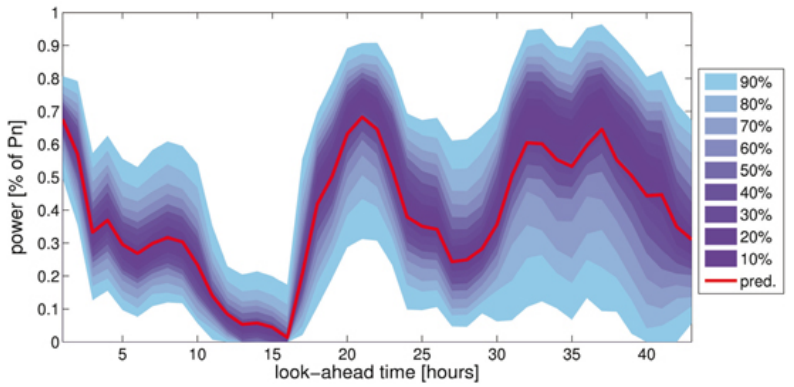


Figure 3: Example of point forecasts and prediction intervals up to 48 hours ahead. (Source: `www.pierrepinson.com`)

Thesis motivation – The Horns Rev experience

Historically, the deployment of wind farms took place onshore, because of lower costs for installing wind turbines and connecting them to power systems,

in comparison to offshore environments. However, large and unexploited wind resources over waters, combined to the limited availability of sites for new wind farm projects onshore and social acceptance issues are pushing the installation of new wind farms offshore. Figure 4 illustrates the planned increase of offshore wind power capacity in Europe. Offshore wind farms will more likely be erected in the form of large and dense clusters of wind turbines such as the Horns Rev 1 (HR1) wind farm.

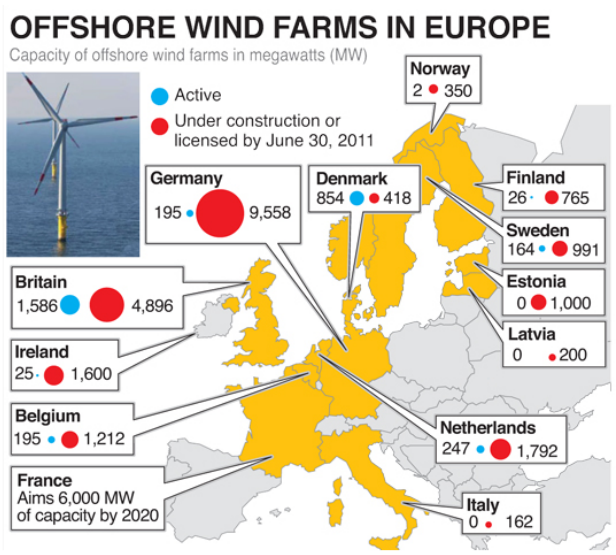


Figure 4: Comparison of available and future offshore wind power capacity in Europe (Sources: Areva, EWEA, media reports).

The HR1 wind farm is the main case study in this thesis. Located in the North Sea, about 15 km off the West coast of Jutland, Denmark (see Figure 5), it is composed of 80 turbines for a rated capacity of 160 MW. Its yearly production corresponds to the consumption of 150000 households using 4000 kWh per year. When it came into operation in 2002, it was the largest offshore wind farms in the world. For that reason, it has attracted a considerable attention in the scientific literature. Research works include modeling of wind turbine wakes for optimizing wind farm layouts (Barthelmie et al., 2009), the observation of wind farm wakes with satellite SAR images for estimating shadowing effects (Christiansen and Hasager, 2005), nacelle wind and yaw angle assimilation for short-term forecasting applications (Draxl et al., 2012), the correction and validation of NWP models with in-situ measurements (Peña and Hahmann, 2012).

The structural particularity of large-scale offshore wind farms like HR1 stems

Figure 5: The Horns Rev 1 wind farm is located in the North Sea. (Source: `www.vattenfall.dk`)

from the high density of large wind turbines, and thereby large wind power capacity, within a small geographical area. Corresponding capacity is spread over a much wider area over land, partly because onshore wind turbines are smaller and partly because of siting constraints. This particularity translates into a reduction of the smoothing effect of wind power variability which occurs with the spatial dispersion of wind turbines (Focken et al., 2002). For instance, wind power generation at HR1 can change by up to 100 MW in 15 to 20 minutes, corresponding to more than 60% of HR1 rated capacity. These large fluctuations have a strong impact on power system and are rarely observed for offshore wind farms (Akhmatov, 2007).

Furthermore, wind flow characteristics change as they move from onshore to offshore environments. In particular, moderate to high wind speeds (i.e., larger than 8 and 15 m s^{-1}, respectively) are more frequent over waters than over land. Diurnal cycles are also much less pronounced. Decoupling of flow is more frequent, translating into stronger vertical shear and different turbulent regimes (Pryor and Barthelmie, 2002). The meeting of wind flow and wind farm over waters results into significant differences in wind power fluctuation patterns in the very short-term when compared to those in onshore environments, as illustrated in Figure 6. Offshore wind power production is higher on average and, more importantly, its variability is magnified.

In order to enhance the integration of its output power, a number of controllers are already implemented at HR1 (Kristoffersen, 2005). Their respective principles are shown in Figure 7. *Absolute power limitation* can be activated for avoiding exceeding the scheduled production, and *ramp rate limitation* can be turned on for dampening fluctuations of large amplitude, for instance. Moreover, despite their specificities, offshore wind farms have to contribute like any other power plant to balancing and backing-up operations. That is the purpose of the

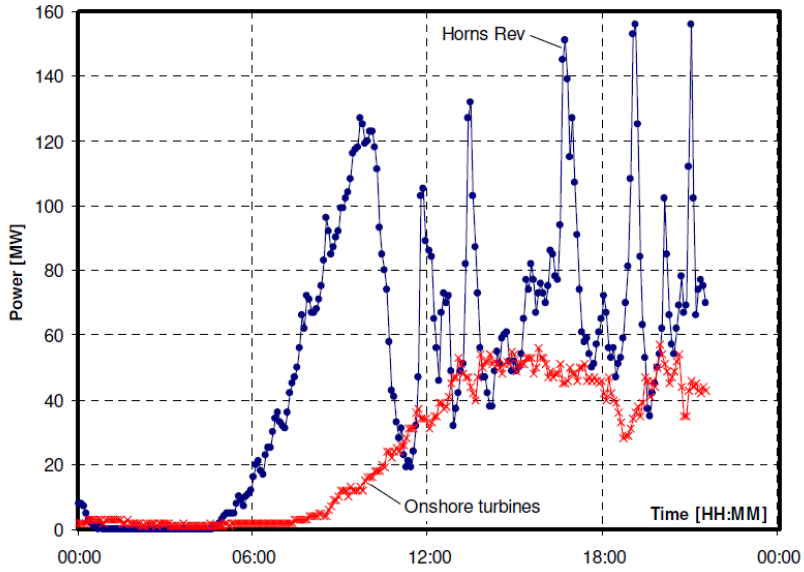


Figure 6: Wind power generation at the Horns Rev 1 wind farm and onshore wind turbines of corresponding capacity. Temporal resolution: 5 minutes. (Source: Kristoffersen (2005))

last two controllers, namely *balance control* and *delta control*, that are activated when requested by the TSO. These controllers and, more specifically, the ramp rate controller requires accurate set points in the form of wind power forecasts in order to be tuned efficiently and to dampen large wind power fluctuations in the very short-term.

Thesis objective & outline

The main objectives of this thesis are to propose new models and explore new methodologies for improving the characterization and predictability of wind power fluctuations from single wind farms in the very short-term, with specific lead times up to 1–2 hours, and temporal resolutions of a few minutes. Wind power forecasts generated with the proposed models could eventually be integrated into the controllers presented in Figure 7.

This thesis addresses a number of important aspects in agreement with the latest recommendations or directions for research in wind power forecasting. First, focus is placed almost exclusively on offshore applications since most of

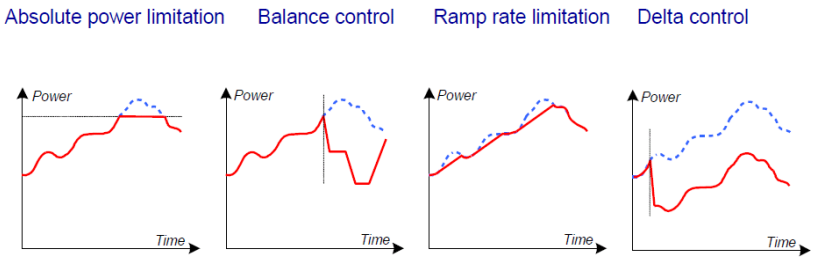


Figure 7: Controllers implemented at the Horns Rev 1 wind farm for dampening wind power fluctuations. (Source: Poul Sørensen, Jesper Kristoffersen (2006) Wind farm control. ECPE Seminar – Renewable Energies, Kassel, Germany)

the new wind power capacity in Europe is expected to be installed over waters (see Figure 4). In that respect, a number of nonlinear models are investigated in order to better account for the specificities of offshore wind power fluctuations. Second, the scale of interest in this work is the very short-term since the availability of accurate wind power forecasts at this scale is likely to become essential for managing wind power systems with large penetrations of wind power, as discussed in Jones and Clark (2011) and GE Energy (2010). However, the recent literature reviews on wind power forecasting by Monteiro et al. (2009) and Giebel et al. (2011) indicate that most of the research efforts have so far concentrated on short term applications. In this context, the work presented in this thesis is a valuable contribution to the field of wind power forecasting in the very short-term. Third, new meteorological observations (i.e., weather radar images) are considered as an alternative to traditional inputs (e.g., meteorological forecasts generated with NWP models). Besides offering the advantage of being available at higher spatio-temporal resolutions than meteorological forecasts, the use of these observations may also enable to avoid one of the main shortcomings of statistical prediction systems based on meteorological forecasts, that is the propagation of the NWP forecast errors. Fourth and last, a strong emphasis is put on methodologies for generating probabilistic forecasts as it is believed to be the way forward for optimizing the management of power systems.

This thesis consists of four papers. Papers A and B present two applications of Markov-Switching Autoregressive (MSAR) model, a class of nonlinear time series models (Zivot and Wang, 2003). The motivation for applying this class of models stems from behavioral changes in the dynamics of wind power fluctuations. Different time series models can hence be used to explain wind power fluctuations at different times. The underlying assumption for using MSAR models is that wind power fluctuations are governed by a regime sequence

that is considered hidden or unobservable. The objectives of these two papers are to improve the characterization of wind power fluctuations and to show that these models can enable a reduction of the uncertainty associated with wind power forecasts.

Paper C and D build on the first two papers and investigate the use of weather radar images as a new source of inputs for statistical models. In particular, the motivation is to characterize the unobserved regime sequence with meteorological observations at high spatio-temporal resolutions. The limitations, potential and perspectives for the integration of weather radar observations into prediction systems are discussed in Paper C. A first classification of wind power regimes with respect to weather radar observations is proposed in D.

Bibliography

Ackermann T. (2005) Wind power in power systems. *Wiley*.

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

Barthelmie RJ, Hansen K, Frandsen ST, Rathmann O, Schepers JG, Schlez W, Phillips J, Rados K, Zervos A, Politis ES, Chaviaropoulos PK. (2009) Modelling and measuring flow and wind turbine wakes in large wind farms offshore. *Wind Energy*, 12:431–444.

Beck F, Martinot E. (2004) Renewable energy policies and barriers. *Encyclopedia of Energy*, 5: 365–83.

Christiansen MB, Hasager CB. (2005) Wake effects of large offshore wind farms identified from satellite SAR. *Remote Sensing of Environment*, 98:251–268.

Danish Ministry of Climate, Energy and Building. (2012) *Energy policy report*. Avilable online: `http://www.ens.dk`

Draxl C, Delle Monache L, Vandenberghe F, Liu Y, Hahmann AN. (2012) Thinning strategies for the assimilation of wind farm observations to improve wind energy predictions. Submitted to *Wind Energy*.

Focken U, Lange M, Mönnich K, Wald HP, Beyer 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.

Focken U, Lange M. (2006) Physical approach to short-term wind power prediction. *Springer*.

GE Energy. (2010). Western wind and solar integration study. Prepared for the *National Renewable Energy Laboratory*.

Giebel G, Brownsword R, Kariniotakis G, Denhard M, Draxl C. (2011) The state-of-the-art in short-term prediction of wind power: A literature overview. Technical Report, ANEMOS.plus.

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

Gross G, Galiana FD. (1987) Short-term load forecasting. *Proceedings of the IEEE*, 75:1558–1573.

Hewitt NJ. (2012) Heat pumps and energy storage – The challenges of implementation. *Applied Energy*, 89:37–44.

International Energy Agency. *World Energy Outlook*, 2011.

Jones L, Clark C. (2011) Wind integration - A survey of global views of grid operators. *In Proceedings of the 10th International Workshop on Large-Scale Integration of Wind Power into Power Systems, Aarhus, Denmark*.

Kristoffersen J. (2005) The Horns Rev wind farm and the operational experience with the wind farm main controller. *In Proceedings of the Offshore Wind International Conference and Exhibition, Copenhagen, Denmark*.

Monteiro C, Bessa R, Miranda V, Botterud A, Wang J, Conzelmann G. (2009) Wind power forecasting: State-of-the-art 2009. Technical Report, Argonne National Laboratory (ANL).

Peña A, and Hahmann AN (2012). Atmospheric stability and turbulence fluxes at Horns Rev – An intercomparison of sonic, bulk and WRF model data. *Wind Energy*, 15:717–731.

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

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

Pinson P, Madsen H, Nielsen HAa, Papaefthymiou, G, Klöckl B. (2009) From probabilistic forecasts to statistical scenarios of short-term wind power production. *Wind Energy*, 12:51–62.

Pinson P. (2012) Very short-term probabilistic forecasting of wind power time-series with generalized Logit-Normal distributions. *Journal of the Royal Statistical Society, Series C*, 61: 555–576.

Pryor SC, Barthelmie, RJ. (2002) Statistical analysis of flow characteristics in the coastal zone. *Journal of Wind Engineering and Industrial Aerodynamics*, 3:201–221.

Strbac G. (2008) Demand side management: Benefits and challenges. *Energy Policy*, 36:4419–4426.

Zivot E, Wang J. (2003) Modeling financial time series with S-PLUS. *Springer Verlag*.

Part II

Papers

PAPER A

A general probabilistic forecasting framework for offshore wind power fluctuations

Authors:
P.-J. Trombe, P. Pinson, H. Madsen

Published in:
Energies (2012), 32(2), pp. 621-657

A general probabilistic forecasting framework for offshore wind power fluctuations

Pierre-Julien Trombe[1], Pierre Pinson[1], Henrik Madsen[1]

Abstract

Accurate wind power forecasts highly contribute to the integration of wind power into power systems. The focus of the present study is on large-scale offshore wind farms and the complexity of generating accurate probabilistic forecasts of wind power fluctuations at time-scales of a few minutes. Such complexity is addressed from three perspectives: (i) the modeling of a nonlinear and non-stationary stochastic process; (ii) the practical implementation of the model we proposed; (iii) the gap between working on synthetic data and real world observations. At time-scales of a few minutes, offshore fluctuations are characterized by highly volatile dynamics which are difficult to capture and predict. Due to the lack of adequate on-site meteorological observations to relate these dynamics to meteorological phenomena, we propose a general model formulation based on a statistical approach and historical wind power measurements only. We introduce an advanced Markov Chain Monte Carlo (MCMC) estimation method to account for the different features observed in an empirical time series of wind power: autocorrelation, heteroscedasticity and regime-switching. The model we propose is an extension of Markov-Switching Autoregressive (MSAR) models with Generalized AutoRegressive Conditional Heteroscedastic (GARCH) errors in each regime to cope with the heteroscedasticity. Then, we analyze the predictive power of our model on a one-step ahead exercise of time series sampled over 10 minute intervals. Its performances are compared to state-of-the-art models and highlight the interest of including a GARCH specification for density forecasts.

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

1 Introduction

Climate change calls for the reduction of greenhouse gas emissions and thus a growing development of renewable energy sources. Benefiting from favorable governmental policies and large wind resources, countries in the north-west of Europe are rapidly increasing their wind power capacities. Historically, onshore installations have prevailed, but offshore wind energy is now growing significantly. In Denmark, the latest figures stated that wind power accounted for about 22% of the domestic electricity supply and, out of 3802 MW wind power capacity, 868 MW were installed offshore (Danish Energy Agency, 2011). The current trend is towards the development of large-scale offshore projects capable of generating several hundreds of MW each. Indeed, sitting wind farms out at sea has substantial advantages of (i) more space available; (ii) a decrease of the frequency and duration of low wind speeds and (iii) an increased persistence for high wind speeds. Offshore wind farms are then expected to have higher capacity factors (i.e., the ratio of the actual power output over a given period of time to the maximum output if the wind farm had been operated at full capacity) (Pryor and Barthelmie, 2002).

However, in practice, integrating significant amounts of wind power into power systems remains a challenge and requires dedicated prediction tools for real-time monitoring, operation scheduling and energy trading. While most of these applications requires wind power forecasts in an hourly resolution, the recent deployment of large-scale offshore wind farms has increased the concern for forecasts with particular lead times of 5–10 minutes (Jones and Clark, 2011). Indeed, power generation at large offshore wind farms turns out to be highly volatile, increasing the risk of imbalance in the power system, in the very short-term. This originates from the specific design of these wind farms which concentrate a large amount of wind power capacity within a relatively small area, increasing the impact of local meteorological phenomena (wind and rain fronts among others) on their short-term power production. For instance, measurements from the offshore site of Horns Rev reveal changes in the output power that may reach an amplitude of 60% the wind farm maximum capacity, within 15–20 minutes (Akhmatov, 2007). Such levels of fluctuations can rarely be observed onshore where similar capacities would be spread over a much wider area, smoothing out the effects of the weather instabilities (Focken et al., 2002). Consequently, maintaining the short-term balance of the transmission system (i.e., matching the power supplied by the wind farm and the electricity demand) and the stability of the power system has become a critical issue and needs to be handled carefully to prevent potential damages (blackouts, *etc*.).

At time-scales of a few minutes, wind power forecasts are preferably generated with statistical models, based on historical data only (Giebel et al., 2011). In

the present paper, our aim is to introduce a case study of statistical modeling and forecasting of offshore wind power fluctuations and its related complexity from three perspectives:

- the modeling of a nonlinear and non-stationary stochastic process for which we propose a model that allows to capture up to three different time series effects: autocorrelation, heteroscedasticity and regime switching (the generic name of our model is MS-AR-GARCH),
- the numerous issues linked to the practical implementation of such model as it requires an advanced estimation method based on a Markov Chain Monte Carlo (MCMC) algorithm,
- the gap between applying such model to synthetic data and real world observations.

This paper is organized as follows. Section 2 summarizes the latest achievements in wind power meteorology for very short-term applications and states the motivations for this study. Section 3 introduces the data and shows some of their major features. Then, in Section 4, specifications for the model we propose are discussed throughout a brief overview of the literature on Markov-Switching models which constitute a special class of regime switching models, and on GARCH models which are generalized forms of heteroscedastic models. Section 5 gives a detailed description of the estimation method based on a Markov Chain Monte Carlo algorithm and the reasons for such a choice. Applications to both synthetic and real data are presented and the accuracy and robustness of the estimation method are assessed. A forecast evaluation on real data is performed in Section 6 where the performances of our model are compared with current benchmark models for very short-term wind power fluctuations. Finally, Section 7 delivers concluding remarks.

2 Motivations Based on the State-of-the-Art

First, with the planned deployment of large-scale offshore wind farms, there is an urging need to build up on the existing knowledge on these wind power fluctuations by characterizing the dynamics and identifying the factors which drive the wind power fluctuations in the very short-term. As a first step towards this understanding, Akhmatov et al. (2007) reported that at a temporal resolution of 10 minutes, certain weather conditions at Horns Rev and in particular northwesterly winds very much favored large wind power fluctuations. Then, Sørensen et al. (2008) proposed an aggregated model of individual wind turbines and showed its relative ability to simulate consistent wind power fluc-

tuations at different time scales, from a few minutes up to 2 hours ahead. Very recently, a spectral analysis of wind speed measurements at Horns Rev led to the identification of specific seasonal cycles as key features of wind variability (Vincent et al., 2010).

Second, most of the state-of-the-art statistical methods gives focus to large prediction horizons, from 1 hour to a couple of days, and show limited forecasting skills for very short-term horizons, within tens of a minute, at which large wind power fluctuations must be monitored (Kristoffersen and Christiansen, 2003). This low level of predictability is due to the complex nonlinearities in the output power dynamics which cannot be captured by conventional models. Hence, there is a need for dedicated statistical methods capable of generating accurate forecasts for very short-term horizons. In that regard, our approach on forecasting is probabilistic and the respective performance of the models presented in this paper will be evaluated accordingly (Gneiting, 2008).

As a first attempt to deal with the low predictability of the output power of large-scale offshore wind farm, regime-switching approaches and more specifically Markov-Switching models have received a growing interest within the wind power community. Since their very first introduction in econometrics by (Hamilton, 1989), they have been commonly used in many disciplines such as speech recognition (Rabiner and Juang, 2005) or computational biology (Durbin et al., 1998), for instance. This class of models is prized for its ability to account for structural breaks or sudden changes in the process dynamics. In meteorology, Markov-Switching models are often used to estimate an unobservable climate state which ideally governs other climate variables such as wind speed or wind direction. For the specific case of large-scale offshore wind farms, the inferred states or regimes can be interpreted as changes of the wind farm behavior, in terms of power generation. Besides that, Markov Switching AutoRegressive (MSAR) models are shown to have better point forecast performances than AutoRegressive Moving Average (ARMA), Smooth Transition AutoRegressive (STAR) and Self-Exciting Threshold AutoRegressive (SETAR) in Pinson et al. (2008). Alternatively, a MSAR model is proposed in Pinson and Madsen (2010) with adaptive estimation of the parameters which allows parameter estimates to change over time to better account for the long-term variations of the wind characteristics. Density forecasts generated with that method are shown to be much sharper and have a better calibration than those generated with AR models.

Nevertheless, one can argue that keeping the variance constant over time within each regime stands as a strong limitation for the forecasts sharpness when periods of different volatility levels alternate. This may mistakenly lead to overdetermination of the optimal number of states when fitting the model to the data. One class of models capable of relaxing the constant variance assumption

is the Generalized AutoRegressive Conditional Heteroscedasticity (GARCH) model, allowing the conditional variance in each regime to follow an ARMA process (Bollerslev, 1986). The GARCH class of models is appealing because it can cope with volatility clustering which is a clear issue when studying offshore wind power generation at high frequencies. Therefore, the present study proposes to extend MSAR models with a GARCH specification for the conditional variance dynamic in each regime (hence the resulting model name MS-AR-GARCH). This extension of the original MSAR model is expected to allow for a better identification of the volatility clustering effect and to a more parsimonious parametrization regarding the number of regimes.

3 Data from Large Offshore Wind Farms

The data considered in the present study cover the time period from 16 February 2005 to 25 January 2006 and were recorded at Horns Rev I, the second largest offshore wind farm in operation in the world at that time. Horns Rev I is located 15 km away from the west coast of Jutland (Denmark) and consists of 80 turbines of 2 MW, for a nominal capacity of 160 MW. Original data were provided as individual time series of wind power measurements for each of the 80 turbines at one second time intervals.

The original data are averaged in order to generate an aggregated time series of wind power fluctuations for the entire wind farm. A 10 minute resolution is arbitrarily chosen within the range of values over which significant power fluctuations are observed (Akhmatov, 2007). Another reason to justify this choice is that grid operators monitor offshore wind farms at similar temporal resolutions (Kristoffersen and Christiansen, 2003). The sampling procedure first consists in producing spatio-temporal averages over 10 minute intervals for which a minimum of 75% of the data are of good quality. These averages are then normalized by the nominal capacity of the wind farm, following Madsen et al. (2005). No attempt is made to fill in missing data points and many gaps remain present in the data. A 10 day episode of this time series is depicted on Figure 1. It can be noticed that the power generation is a double-bounded process, below and above. As a matter of fact, the power generation of a wind farm can neither be negative nor exceed its maximum capacity.

Moreover, technical specificities and constraints of wind turbines make that wind power generation is not a linear function of the wind speed. The relationship between wind speed and power generation is described by the so-called power curve. This relationship is often estimated to convert wind speed forecasts into wind power forecasts. For a more detailed description of its use in practice, we refer to Sánchez (2006). More generally, the power curve is con-

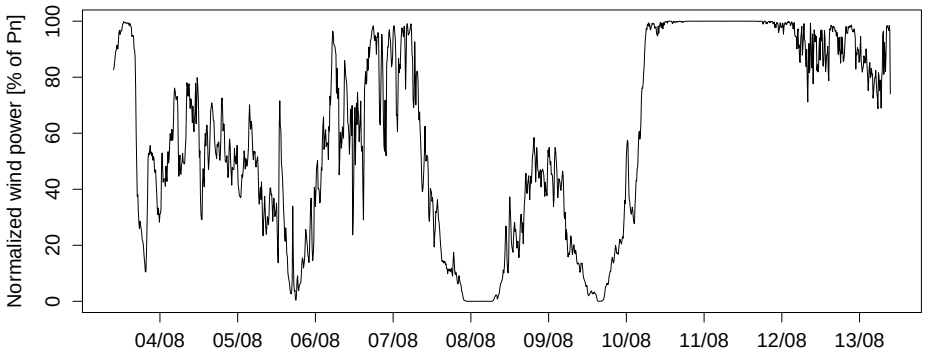


Figure 1: Time series of normalized wind power generation at Horns Rev I over a 10 day episode in August 2005. The time series is sampled with a temporal resolution of 10 minutes.

sidered a function of both the wind speed and the wind direction and must be estimated for every single wind farm. Nevertheless, wind speed and wind direction are not the only two factors that are believed to govern wind farm behavior. In the specific case of large offshore wind farms, it is also commonly assumed that complex local meteorological phenomena have a strong impact on the power generation. Ongoing research works on these phenomena are still in an early stage, and identifying them would require to combine both meteorological and statistical approaches which is not the purpose of this study. As for now, early assumptions based on empirical observations have described these phenomena as combinations of intense precipitations and wind gusts (Vincent et al., 2011).

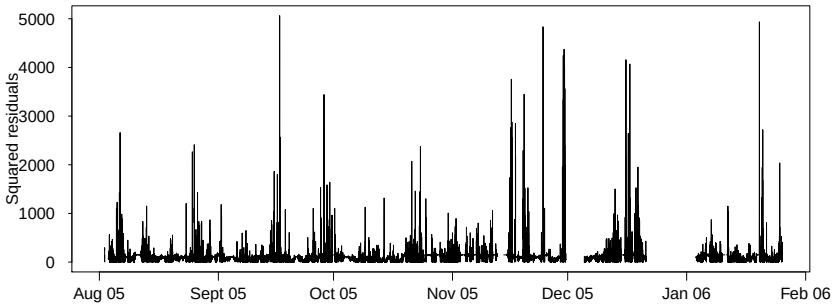
From Figure 1, one can see periods characterized by very different dynamics alternate with various frequencies and durations. This latter observation reveals the non-stationary behavior of this wind power time series, whatever the time scale one considers. This issue is further discussed in Vincent et al. (2010). Non-stationarity is one of the reasons why most linear time series models show limited prediction skills. This feature is further illustrated in Figure 2 which plots the squared residuals of the best autoregressive model (of order 3), the associated autocorrelation function (ACF) and the partial ACF (PACF) for the wind power time series. The model was fitted to the whole time series, but to enhance visualization of the results, the squared residuals are only plotted for the period of time spanning from 1 August 2005 to 26 January 2006. First, a look at the squared residuals highlights the volatility clustering effect, meaning that large errors tend to be followed by large errors and similarly, small errors tend to be followed by small errors. It is a feature often observed for data sampled at a high frequency. Then, the ACF of the squared residuals indicates that the autocorrelation is significant up to very large lags which reveals the het-

eroscedastic behavior of the errors. Finally, the PACF allows one to evaluate the number of significant lags for the time series of squared residuals. It indicates that the conditional variance should be modeled as the weighted sum of approximately the last 20 squared errors. However, for the sake of parsimony, an ARCH process of large order can well be substituted by a GARCH specification (Bollerslev, 1986). This well spread empirical approach offers the double advantage of drastically reducing the number of coefficients to be estimated while conserving the model adequacy. It also introduces a decreasing weight structure, from the most recent to the oldest past squared errors, for the computation of the conditional variance.

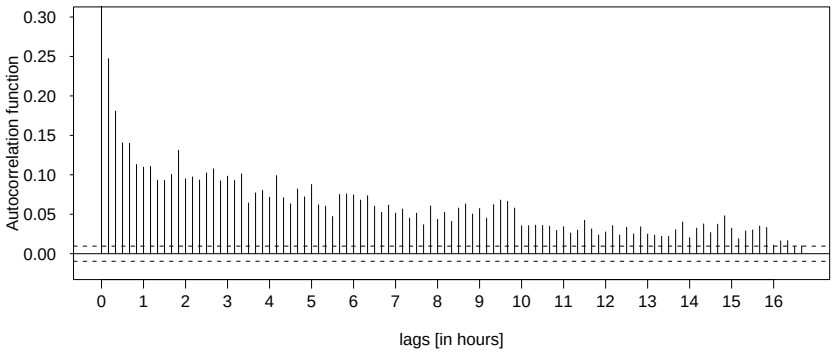
4 Model Specifications

4.1 Wind Power Predictive Density

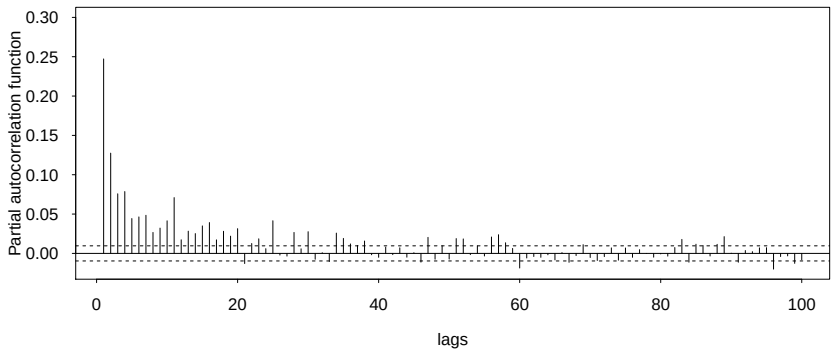
As mentioned in the previous section, the time series of wind power is nonlinear and non-stationary. The smoothing effect outlined when considering a collection of wind turbines scattered over a wide area does not apply in the case of a single large-scale offshore wind farm. Furthermore, wind turbines do not generate electricity for wind speeds below the so called cut-in speed ($\sim$4 m s^{-1}) or above the the cut-off speed ($\sim$25 m s^{-1}). In addition, for wind speeds ranging from 15 m s^{-1} to 25 m s^{-1}, wind turbines operate at full capacity and produce a constant level of power. Consequently, the power generation drops to 0 or reaches its maximum in a significant number of occasions. From a statistical modeling perspective, it means that the process does meet its lower and upper bounds which generates mass points at the extremities of the wind power distribution. This prevents the use of a logistic transformation as adopted in Lau and McSharry (2010) since the mass points would remain, even after transformation. In view of these limitations, truncated and censored normal distributions stand as appealing alternatives to the more classical Normal distribution. Recent developments that use the two former distributions applied to wind data include (Gneiting et al., 2006, Thorarinsdottir and Gneiting, 2010). However, Markov-Switching models imply the computation of distribution mixtures. For the sake of the estimation method simplicity, we choose to consider neither the truncation nor the censoring of the Normal distribution since mixtures of these distributions would be too cumbersome to compute. For similar reasons, the Generalized Logit-Normal distribution as proposed in Pinson (2012) was not considered. Finally, we focused on two symmetric distributions, namely the Student-t and Normal distributions. The Student-t distribution has the advantage of being more heavy-tailed than the Normal distribution, making the regimes more stable (Klaasen, 2002). Its drawback



(a) Squared residuals obtained after fitting an AR(3) model to the wind power time series.



(b) Autocorrelation function of the squared residuals.



(c) Partial autocorrelation function of the squared residuals.

Figure 2: Volatility clustering and heteroscedasticity of the wind power time series.

is that it has one extra parameter (its degree of freedom) which is difficult to

estimate (Gray, 1996). The use of the Normal distribution, though known as not optimal for wind power time series, is therefore considered as a natural starting point for testing the model in this study. We leave questions on more appropriate distributions for further research.

4.2 GARCH Models in Meteorology

An overview of the time series analysis literature shows that GARCH models have been extensively used in econometrics and finance but remains rather unpopular in other fields. In meteorology, GARCH models are often employed in a single regime framework and applied to wind speed or air temperature time series for characterizing their volatility. Tol (1997) first fitted an AR-GARCH model to daily wind speed measurements from Canada and illustrated the better in-sample performance of his heteroscedastic model over homoscedastic ones in presence of high volatility in the data. A bivariate GARCH model was then used in Cripps and Dunsmir (2003) to characterize the wind components (u,v) and their variability at a time scale of 1 min and relate them to local meteorological events in the Sydney harbor. Another meteorological application of GARCH models presented the usefulness of a ARMA-GARCH-in-mean model to estimate the persistence in the volatility of wind speed measurements at different heights (Ewing et al., 2006).

In contrast to these latter studies whose primary focus is in-sample estimation, Taylor and Buizza (2004, 2006) use AR-GARCH models to generate point and density forecasts for temperature and weather derivative pricing, respectively. In addition, the recent work by Taylor et al. (2009) also presents out-of-sample results. It extends the methodology developed in Taylor and Buizza (2004) and used several types of GARCH models to generate daily wind speed density forecasts and converts them into wind power forecasts. This work demonstrates the good ability of GARCH models for generating density forecasts when compared to atmospheric models for early look ahead horizons, from 1 up to 4 days. Another methodology is proposed by Lau and McSharry (2010) in which an ARIMA-GARCH model is used to generate multi-step density forecasts of wind power, outperforming current benchmark models in the short-term, from 15 minutes up to 6–12 hours. Interestingly, all these studies give empirical evidence of the strong potential of using the GARCH class of models for predicting weather related variables in the very short-term when these variables are highly volatile.

4.3 Existing Markov Switching Models with GARCH Errors

Seminal references of combining Markov-Switching and AutoRegressive Conditional Heteroscedasticity (MS-ARCH) include Cai (1994) and Hamilton and Susmel (1994). In practice, capturing time-varying variance with a reasonable number of ARCH terms remains an issue. It often calls for a GARCH specification instead in order to reduce the number of coefficients to be estimated. The difficulty that arises when generalizing MS-ARCH to MS-GARCH relates to the historical path dependency of the conditional variance which is intractable, making that generalization almost computationally infeasible.

Nevertheless, there exist a few approaches to avoid that problem. Regarding maximum likelihood methods, the idea consists in approximating the conditional variance as a sum of past conditional variance expectations as in Gray (1996). This model was later extended by Klaasen (2002) yielding improved volatility forecasts. Alternatively, Haas et al. (2004) suggested a new formulation for MS-GARCH models by disaggregating the overall variance process into separate processes in each regime. Another way of tackling the path dependency problem consists in using Monte Carlo Markov Chain (MCMC) simulations to infer that path by sampling from the conditional distribution of the states of the Markov chain. This can be implemented by data augmentation as described in Fruhwirth-Schnatter (2006). The strength of this approach is that it can be applied for the estimation of many variants of Markov-Switching models. Closer to our problem, Henneke et al. (2011), Chen et al. (2009), Bauwens et al. (2010) proposed three different MCMC algorithms for the Bayesian estimation of MS-ARMA-GARCH, MS-ARX-GARCH and MS-GARCH models, respectively.

Some other difficulties arise when estimating MS-GARCH models. They may be caused by the structural specification of the model or else by the numerical tools used for parameter estimation. For instance, maximum likelihood estimation methods implemented with a numerical optimizer often encounter specific optimization problems due to starting values, inequality constraints or else local minima. Besides, the two formulations of the MS-GARCH model developed in Gray (1996) and Klaasen (2002) are based on an approximation for the recursive update of the conditional variance which leads to further estimation complexity. As for the MS-GARCH model in Haas et al. (2004), it loses its initial appeal of being analytically tractable along with the inclusion of autoregressive terms in the conditional mean equation which does not match with our model specification to combine AR and GARCH effects with Markov-Switching. Along that last comment, it is important to emphasize that most of the studies involving likelihood estimation of MS-GARCH models have as a prime concern the capture of the heteroscedasticity present in the time series and were not designed to cope with data also featuring strong autocorrelation.

In comparison, Bayesian inference offers an alternative framework which allows to overcome most of likelihood estimation problems:

- the robustness of MCMC samplers to starting values can be evaluated by running several Markov chains with different starting values and tested for differences in their outputs,
- inequality constraints can be handled through the definition of prior distributions (Gibbs sampler) or through a rejection step when the constraint is violated (Metropolis–Hastings sampler),
- theoretically, local minima pitfalls are avoided by simulating the Markov chain over a sufficiently large number of iterations (law of large numbers),
- misspecification of the number of states of the Markov chain can be assessed by a visual inspection of the parameter posterior distributions (check for multiple modes).

Moreover, model parametrization limitations linked to the integration of autoregressive terms in the mean equation do not apply in Bayesian estimation and there is no fundamental implementation differences in estimating a MS-GARCH and a MS-ARMA-GARCH model. Of course, the present study would be very partial if the main bottlenecks in using MCMC simulations such as computational greediness or the tuning of the prior distributions were not mentioned. Therefore, we refer to Subsection 4.4 for a detailed description of the main implementation issues of MCMC samplers. In addition, studies on the respective advantages and drawbacks of maximum likelihood and Bayesian estimation methods are available in Rydén (2008). To conclude this discussion, let us say that our goal is not to contribute to the pros and cons debate of maximum likelihood against Bayesian estimation but rather to find the method that is the most suitable for our problem. In this light, our choice to estimate the MS-AR-GARCH model in a Bayesian fashion was motivated by the enhanced flexibility in combining AR and GARCH effects under the assumption of structural breaks in the process.

4.4 The Model Definition

To model the stochastic behavior of a given time series of wind power $\{y_t\}$, a MS(m)-AR(r)-GARCH(p,q) model is proposed as follows:

$$y_t = \theta_0^{(S_t)} + \sum_{i=1}^{r} \theta_i^{(S_t)} y_{t-i} + \sqrt{h_t}\varepsilon_t \tag{1}$$

$$h_t = \alpha_0^{(S_t)} + \sum_{i=1}^{q} \alpha_i^{(S_t)} \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j^{(S_t)} h_{t-j} \tag{2}$$

where $\{h_t\}$ is the conditional variance at time t, $\{\varepsilon_t\}$ is a sequence of independently distributed random variables following a Normal distribution $\mathcal{N}(0,1)$ and $\boldsymbol{S} = (S_1,\ldots,S_T)$ is a first order Markov chain with a discrete and finite number of states m and transition probability matrix $\mathbf{P}$ of elements:

$$Pr(S_t = j|S_{t-1} = i) = p_{ij} \quad \text{for} \quad i,j = \{1,\ldots,m\} \tag{3}$$

For full flexibility, all AR and GARCH coefficients are chosen to be state dependent. In addition, to ensure positivity of the conditional variance, constraints on the model coefficients are imposed as follows:

$$\alpha_0^{(k)} \geq 0,\ \alpha_i^{(k)} > 0,\ \beta_j^{(k)} \geq 0 \quad \text{for} \quad i = \{1,\ldots,p\},\ j = \{1,\ldots,q\},\ k = \{1,\ldots,m\} \tag{4}$$

Finally, the following inequality constraints are applied to ensure covariance stationarity:

$$0 < \sum_{i=1}^{q} \alpha_i^{(k)} + \sum_{j=1}^{p} \beta_j^{(k)} < 1 \quad \text{for} \quad k = \{1,\ldots,m\} \tag{5}$$

From here on, we adopt the following notations:

$$y = (y_1, y_2, \ldots, y_T) \tag{6}$$
$$y_{[1,t]} = (y_1, \ldots, y_t) \tag{7}$$
$$S_{[1,t]} = (S_1, \ldots, S_t) \tag{8}$$
$$S_{\neq t} = (S_1, \ldots, S_{t-1}, S_{t+1}, \ldots, S_T) \tag{9}$$
$$\pi_k = (p_{k1}, \ldots, p_{km})' \quad \text{for} \quad k = \{1,\ldots,m\} \tag{10}$$
$$\theta^{(k)} = [\theta_0^{(k)}, \ldots, \theta_r^{(k)}]' \quad \text{for} \quad k = \{1,\ldots,m\} \tag{11}$$
$$\alpha^{(k)} = [\alpha_0^{(k)}, \ldots, \alpha_q^{(k)}, \beta_1^{(k)}, \ldots, \beta_p^{(k)}]' \quad \text{for} \quad k = \{1,\ldots,m\} \tag{12}$$
$$\Theta = [\theta^{(1)}, \ldots, \theta^{(m)}, \alpha^{(1)}, \ldots, \alpha^{(m)}, \pi_1, \ldots, \pi_m] \tag{13}$$

5 MCMC Implementation

Bayesian inference applied to complex models and large amounts of data has been strongly enhanced by the development of computational methods such as Markov chain simulations. Besides providing a robust and easy-to-implement solution to circumvent the path dependency problem when estimating the MS-GARCH class of models, MCMC techniques offer broader possibilities such as incorporating existing information on the parameter distributions and estimating their full conditional posterior distributions, for instance. Their major interest is the possibility to divide the set of unknown parameters Θ into smaller

blocks to sample from the block conditional posterior distributions instead of sampling from the complex and joint posterior of the full set of parameters. For a practical presentation of MCMC techniques, we refer to Gilks et al. (1996).

Estimating MS-AR-GARCH models in a Bayesian framework is a procedure that implies sampling from the augmented parameter distribution $p(\boldsymbol{S},\Theta|y)$:

$$p(\boldsymbol{S},\Theta|y) \propto p(y|\boldsymbol{S},\Theta)p(\boldsymbol{S}|\Theta)p(\Theta) \tag{14}$$

This can be achieved through a 3 step procedure by implementing a MCMC algorithm that iterates as follows:

- sample the regime sequence by data augmentation,
- sample the transition probabilities from a Dirichlet distribution,
- sample the AR and GARCH coefficients with the Griddy-Gibbs sampler.

5.1 Sampling the Regime Sequence

Generating sample paths of the regime sequence $\boldsymbol{S}$ for Markov-Switching models is facilitated by a class of techniques known as data augmentation. The early idea by Tanner and Wong (1987) is to recursively consider each of the latent state variables S_t of the hidden Markov chain as missing and compute its conditional distribution $p(S_t|S_{\neq t},\Theta)$. It becomes then possible to generate a random draw from that conditional distribution with the Gibbs sampler as in Robert et al. (1993). This procedure is called *single-move sampling* and requires the number of regimes m to be known and finite. Later variants for Hidden Markov Models (HMM) and Markov-Switching models are respectively reviewed in Scott (2002) and Fruhwirth-Schnatter (2006).

At a given time t, the conditional distribution of the latent state variable S_t is obtained as follows:

$$\begin{aligned}\forall\ \ k \in \{1,\dots,m\},\quad P(S_t = k|y,S_{\neq t},\Theta) &= \frac{P(y,S_t = k,S_{\neq t},\Theta)}{P(y,S_{\neq t},\Theta)} \\ &= \frac{P(y|S_t = k,S_{\neq t},\Theta)P(S_t = k,S_{\neq t},\Theta)}{P(y|S_{\neq t},\Theta)P(S_{\neq t},\Theta)} \\ &= \frac{P(y|S_t = k,S_{\neq t},\Theta)P(S_t = k|S_{\neq t},\Theta)}{P(y|S_{\neq t},\Theta)}\end{aligned} \tag{15}$$

And after discarding the scaling factor $P(y|S_{\neq t},\Theta)$, we obtain:

$$P(S_t = k|S_{\neq t},y,\Theta) \propto P(y|S_t = k,S_{\neq t},\Theta)P(S_t = k|S_{\neq t},\Theta) \tag{16}$$

In the equation above, two different quantities have to be computed. First, $P(y|S_t = k, S_{\neq t}, \Theta)$ is the complete data likelihood, conditioned on the chain being in state k at time t and given the full set of parameters Θ and can be calculated as follows:

$$\begin{aligned} P(y|S_t = k, S_{\neq t}, \Theta) &= \prod_{t=max(r,p,q)}^{T} P(y_t|, S_t = k, S_{[1,t-1]}, y_{[1,t-1]}, \Theta) \\ &= \prod_{t=max(r,p,q)}^{T} \frac{1}{\sqrt{2\pi h_t}} \exp\left(-\frac{(y_t - \theta_0^{(S_t)} - \sum_{i=1}^{r} \theta_i^{(S_t)} y_{t-i})^2}{2h_t} \right) \end{aligned} \tag{17}$$

with h_t being defined as in Equation (2).

Second, the Markov property applies on $P(S_t = k|S_{\neq t}, \Theta)$. Given a sample $S_{\neq t}$ of the entire regime sequence but at time t, the state variable S_t only depends on S_{t-1}, and S_{t+1} only depends on S_t:

$$\begin{aligned} P(S_t = k|S_{\neq t}, \Theta) &= P(S_t = k|S_{t-1} = i, S_{t+1} = j) \\ &= \frac{p_{ik} p_{kj}}{\sum_{k=1}^{m} p_{ik} p_{kj}} \end{aligned} \tag{18}$$

Finally, the Gibbs sampler (Geman and Geman, 1984) is used to generate a random sample of the latent state variable S_t from its updated conditional distribution. The state of the Markov chain at time t can then be updated and this sampling procedure is recursively repeated for the remaining state variables of the hidden Markov chain.

Because of the path dependency structure of MS-GARCH models, computing marginal likelihood of the state variables is not feasible as it is for MSAR or MS-ARCH models (Fruhwirth-Schnatter, 2006). Hence, the posterior distributions of the state variables can only be obtained in the form of smoothed probabilities. Let us recall that one can derive different quantities for the optimal inference of the regime sequence:

- *the filtered probabilities* $P(S_t = k|y_{[1,t]}, \Theta)$ which infer the state variable S_t conditioning upon the vector of parameters and all past and present information $y_{[1,t]}$,
- *the smoothed probabilities* $P(S_t = k|y, \Theta)$ which are the outputs of the inference of S_t using the past, present and future information $y = y_{[1,T]}$,
- *the predicted probabilities* $P(S_{t+1} = k|y_{[1,t]}, \Theta)$ which correspond to the one-step ahead inference S_{t+1} at time t and only use past information $y = y_{[1,t]}$.

For a given state variable S_t, its posterior distribution $P(S_t = k|y)$ is computed by averaging the number of occurrences of the Markov chain being in state k

at time t over the N iterations of the algorithm:

$$P(S_t = k|y) = \frac{1}{N}\sum_{n=1}^{N} \mathbb{1}\{S_t^{(n)} = k\} \quad \text{for} \quad k = \{1,\ldots,m\} \tag{19}$$

with $S_t^{(n)}$ being the draw of S_t at the n^{th} iteration of the MCMC algorithm.

5.2 Transition Probability Matrix Sampling

Sampling the transition probability matrix **P** is done by using a Dirichlet distribution (Fruhwirth-Schnatter, 2006). The key assumption is that the rows of **P** are mutually independent since **P** only depends on the regime sequence **S**. Therefore, they can be sampled in a random order. Given an independent prior distribution $p(\pi_k)$ and using Bayes' theorem, we obtain the conditional distribution of the k^{th} row of **P** as follows:

$$\begin{aligned} p(\pi_k|y,\boldsymbol{S},\Theta_{-\pi_k}) &\propto p(\pi_k)p(\pi_k|\boldsymbol{S}) \\ &\propto p(\pi_k)\prod_{r}^{T}(d_{k1})^{\eta_{k1}}\ldots(d_{km})^{\eta_{km}} \end{aligned} \tag{20}$$

where the η_{ki}'s correspond to the numbers of one-step transitions from regime k to regime i in the hidden Markov chain and the d_{ki}'s are the parameters of the multivariate distribution modelling the transition probabilities.

For a 2 state Markov chain, the Beta distribution is traditionally used as prior for binomial proportions, with parameters d_{k1} and d_{k2}, resulting in the conditional distribution of the k^{th} row of P being Beta distributed:

$$p(\pi_k|y,\boldsymbol{S},\Theta_{-\pi_k}) \sim B(\eta_{k1}+d_{k1},\eta_{k2}+d_{k2}) \tag{21}$$

For a m state Markov chain, and $m \geq 2$, the posterior Beta distribution can be generalized to a Dirichlet distribution (Chib, 1996):

$$p(\pi_k|y,\boldsymbol{S},\Theta_{-\pi_k}) \sim D(\eta_{k1}+d_{k1},\eta_{k2}+d_{k2},\ldots,\eta_{km}+d_{km}) \tag{22}$$

with $d_{k1},d_{k2},\ldots,d_{km}$ being the parameters of the Dirichlet distribution used as prior.

The posterior estimates of the transition probabilities are obtained as the empirical means of the posterior densities:

$$\hat{p}_{ij} = \frac{1}{N}\sum_{n=1}^{N} p_{ij}^{(n)} \quad \text{for} \quad i,j = \{1,\ldots,m\} \tag{23}$$

with $p_{ij}^{(n)}$ being the random draw of p_{ij} at the n^{th} iteration of the MCMC algorithm.

5.3 AR and GARCH Coefficient Sampling

Existing MCMC algorithms for the estimation of MS-AR-GARCH models are proposed in Henneke et al. (2011) and Chen et al. (2009). Alternatively, it is possible to apply a MCMC algorithm for MS-GARCH models presented in Bauwens et al. (2010) and include extra autoregressive terms in the mean equation, instead of a single intercept. The difference in those three algorithms lays in the sampler used for the estimation of the autoregressive and heteroscedastic coefficients. The two formers sample the posterior distributions of the model coefficients with the Metropolis–Hastings sampler (MH) whereas the latter uses the Griddy Gibbs sampler (GG). The MH sampler (Hastings, 1970) is based on an acceptance/rejection rule and was designed to generate samples from a target distribution. However, the rate of acceptance can turn out to be very small for complex models and slow down the convergence of the chain. As for the GG sampler (Ritter and Tanner, 1992), it is based on a principle similar to the Gibbs sampler. The key idea is to discretize the support of the parameter to be estimated. At each knot point, the likelihood of the parameter is evaluated and by a numerical integration rule, the conditional distribution of the parameter can then be approximated.

Unlike the MH sampler, the GG sampler does not require to define the analytical form of the posterior distribution a priori. It is notably useful when the conditional posterior to sample from has a complex shape (multimodality, strongly skewed, heavy tails) or when one does not want to impose a shape a priori because of a lack of knowledge. Its implementation fully relies in the informativeness of the data likelihood $p(y|\boldsymbol{S},\Theta)$ and all priors are uniform, even for short time series. Tips for implementing the GG sampler for accurate estimation of posterior distributions are given in Ritter and Tanner (1992). Its main drawback is its high computational cost because of the many likelihood evaluations at each iteration but this can be overcome by parallelization of the code. Empirical results presented in Bauwens and Lubrano (1998) and Asai (2006) for the classical GARCH model are consistent and conclude that estimation methods based on the MH or the GG sampler lead to posterior estimates of similar accuracy. One of the most notable differences is that the MH sampler does not fully explore the distribution tails. This is due to the shape of the target distribution chosen which in some cases may mislead the exploration of the posterior distribution. This type of problems is avoided when estimating posterior distributions with a GG sampler because it does not require the posterior density to be known in closed form. Taking these considerations into account, it was chosen to follow the methodology presented in Bauwens et al. (2010) which uses the GG sampler for estimating MS-GARCH models. Adding extra autoregressive terms for the estimation of MS-AR-GARCH models is then straightforward.

Conditional posterior distributions of our model coefficients are derived from the Bayes' theorem. Let us consider the case of an unknown AR or GARCH coefficients that will be noted γ, and $p(\gamma)$ its prior. Its conditional posterior distribution is defined as follows:

$$p(\gamma|y,\boldsymbol{S},\Theta_{-\gamma}) \propto p(\gamma)p(y|\boldsymbol{S},\Theta) \tag{24}$$

The conditional density and cumulative distribution function (cdf) of γ are noted g_γ and G_γ. Their numerical approximation are noted $f_\gamma = f(\gamma|y,\boldsymbol{S},\Theta_{-\gamma})$ and F_γ, respectively. At each iteration, the GG sampler builds a numerical approximation of the conditional posterior density of each AR and GARCH coefficient. The support of γ is first discretized with n knot points $(x_1,\ldots,x_n)$. Further details on how to set up n are discussed in the next subsection. Then, the complete data likelihood $P(y|\gamma = x_i,\boldsymbol{S},\Theta_{-\gamma})$ is evaluated for each knot point x_i and by a numerical rule of integration, we obtain an approximation $f_\gamma(x_i)$ of the conditional density g_γ. Linear interpolation in between 2 successive knot points was found to be satisfactory in term of accuracy. Therefore, we use the trapezoidal integration method to compute f_γ. From there, approximating the cdf G_γ is direct. Finally, a random number is uniformly generated on $[0,1]$ and by inverse transformation of F_γ, we obtain a random sample of γ. The principle of the GG sampler is graphically summarized in Figure 3. The posterior estimates of the AR and GARCH coefficients are obtained by computing the means of the posterior densities.

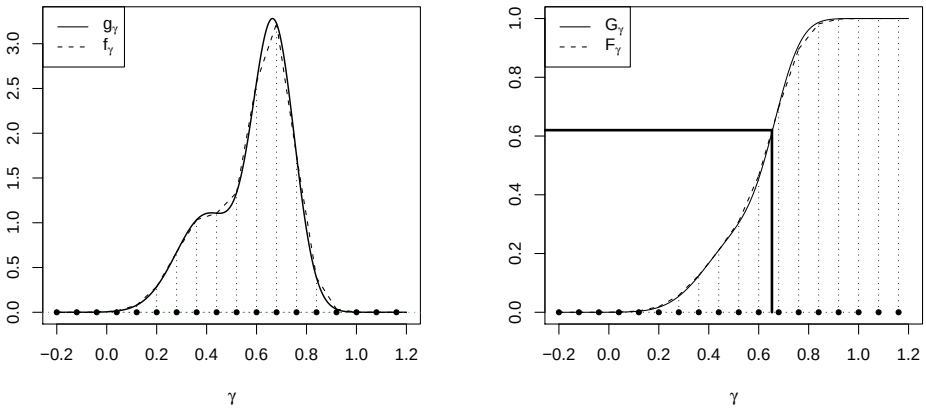


Figure 3: The conditional density g_γ of a given coefficient γ is approximated by numerical integration over a grid of points (left). An approximation F_γ of the cdf G_γ can then be computed. Finally, a random number is uniformly generated on $[0,1]$ and by inverse transformation of F_γ, a random draw of γ is obtained (right).

5.4 Implementation Details

The most simple version of the GG sampler can be coded within a few lines. However, for complex models with many parameters to be estimated, there is a number of issues that have to be handled carefully and makes it implementation less straightforward: choice of prior distributions, label switching, grid shape, mixing efficiency.

Prior Distributions

First, prior distributions have to be defined for sampling the transition probabilities. For a given regime $k \in \{1,\dots,m\}$, setting the parameters $d_{kk} > d_{ki}$ with $i \neq k$ is one way to reflect the prior knowledge that the probability of persistence (staying in the same regime) is larger than the probability of switching from regime k to i. For instance, a $B(8,2)$ distribution is used as prior in Chen et al. (2009) whereas a uniform $B(1,1)$ is preferred in Bauwens et al. (2010). Several simulations with various values for the d_{ij} parameters were run on synthetic time series with more than 1000 data points. The influence of the prior distributions was noticeable for d_{ij} of very high orders of magnitude, due to the length of the time series. For instance, a $B(80,20)$ clearly influences the posterior distribution estimates of the transition probabilities while a $B(8,2)$ almost not, even though these two distributions have equal means. Arguably, we found it relatively risky to favor some regimes over others. Therefore, we favored the approach with uniform priors, meaning that $d_{k1} = d_{k2} = \cdots = d_{km} = 1$.

Secondly, and most importantly, uniform distributions are required for the GG sampler. Defining these priors consists in setting their bounds which is all the more difficult when one has very little prior knowledge of the process being considered. For each AR and GARCH coefficient, one has to make sure that the bounds of the uniform prior encompass the entire support of the true conditional density. Poor settings of the prior bounds may either prevent the convergence of the Markov chain or lead to wrong posterior density and mean estimates. One solution is to use a coarse-to-fine strategy for the MCMC simulation which is divided into three phases:

- a burn-in phase whose draws are discarded until the Markov chain reaches its stationary distribution,
- a second phase at the end of which posterior density estimates are computed and prior bounds are refined (the draws generated during this second phase are also discarded),
- a last phase with adjusted prior bounds at the end of which the final posterior densities are computed.

Refinement of the prior bounds consists in computing the posterior mean and the standard deviation of the densities. The priors are then adjusted and centered around their respective mean with their radius set to 5 standard deviations. That way the uniform priors are shrunk when they were initially too large and enlarged when too small. This approach proved to be robust enough even in case of fat-tailed posterior densities.

Label Switching

Not least, fine settings of the prior bounds can prevent the label switching problem affecting HMM models estimated with Bayesian methods. Since posterior densities are invariant to relabeling the states, that problem can cause erroneous multimodal posterior densities. This can be circumvent by imposing structural constraints on the regimes which can be identified with the permutation sampler presented in Fruhwirth-Schnatter (2006). For the specific case of MS-AR-GARCH models, the most effective constraint against label switching was set on the intercept parameters of the GARCH equation as follows: $\alpha_0^{(1)} < \alpha_0^{(2)} < \cdots < \alpha_0^{(m)}$. At each iteration, the inequality is checked and if not true, regimes are permutated. Another way to make sure that this constraint is true is to define the bounds of the uniform priors of the $\alpha_0^{(k)}$ such that they do not fully overlap.

Grid Shape

Support discretization for the GG sampler implies choosing a suitable structure for the grid along with a fine number of knot points n. As for the structure, Ritter and Tanner (1992) advised to use an evolutive grid with more knot points over areas of high mass and fewer knot points over areas of low mass. Simulations on synthetic data show that this type of grid is difficult to implement in practice and that it yields relatively low gains in accuracy. The use of such a grid is not necessary in this study and instead a grid with equidistant knot points is preferred. A grid made of 42 knot points is generated for each coefficient to be estimated, with the likelihood of the 2 knot points at the extremities of the grid being set to 0, by default. This number was found sufficiently large to accurately approximate conditional densities and is comparable to the 33 knot points used in Bauwens et al. (2010).

Mixing of the MCMC Chain

MCMC simulations on synthetic time series reveal that, within a same regime, AR coefficients are strongly correlated with each others, resulting in a poorly mixing chain, slow convergence rate and significant estimation errors. The same observations were made for the GARCH parameters. In order to improve the mixing of the chain, the GG sampler is implemented with random sweeps (Liu et al., 1995). At each iteration of the MCMC algorithm, instead of updating the AR and GARCH coefficients in a deterministic order, we generate a random permutation of the sequence $(1,\ldots,m(2+r+p+q))$ to determine which coefficients to update first, second and so on. For the empirical study on the wind power time series, it was found that the mixing of the chain could be further improved by repeating the sampling of the AR and GARCH coefficients a given number of times for every update of the state sequence. These implementation details positively enhance the well mixing behavior of the chain and lead to much sharper posterior densities (i.e., smaller estimation errors and standard deviations) of the AR and GARCH coefficients, notably.

Implementation Summary

In order to enhance the implementation understanding and to summarize the key steps of our method, we report its structure in Algorithm 1. For the sake of the notation simplicity, let us note γ_i the i^{th} AR or GARCH coefficients of the vector of parameters $(\theta^{(1)},\ldots,\theta^{(m)},\alpha^{(1)},\ldots,\alpha^{(m)})$. The vector of parameters is now noted $(\gamma_1,\ldots,\gamma_{m(2+r+p+q)})$.

5.5 Simulation on Synthetic Time Series

Before moving on to the time series of wind power, the MCMC estimation procedure is tested on a synthetic MS-AR-GARCH process that is plotted in Figure 4 and whose coefficients are reported in Table 1. This process is composed of 2 regimes, each one of them combining an autoregressive structure of order 2 for the conditional mean equation along with a GARCH(1,1) specification for the conditional variance. The values of its coefficients are chosen so as to generate a simplistic series with two well differentiated dynamics for the 2 regimes. The values of the autoregressive coefficients are set so that the autoregressive process in each regime is stationary. The GARCH coefficients in each regime are defined so that the constraint ensuring a finite variance holds. Finally, the errors are normally distributed. The process simulated hereafter neither aims at recreating nor mimicking the wind power fluctuations presented in Section

Algorithm 1 MCMC procedure for the estimation of MS-AR-GARCH models

Initialize prior distribution: $p(\gamma_1),\ldots,p(\gamma_{(m(2+r+p+q))})$
Initialize regime sequence and parameter: $\boldsymbol{S}^{(0)},\Theta^{(0)}$
$n = 0$
while Convergence of the Markov chain is not reached **do**
  $n = n + 1$
  for $t = 1$ to T **do**
    Sample $S_t^{(n)}$ from $p(S_t^{(n)} = k|S_1^{(n)},\ldots,S_{t-1}^{(n)},S_{t+1}^{(n-1)},\ldots,S_T^{(n-1)},\Theta^{(n-1)},y)$ by the single-move procedure
  end for
  Compute the Dirichlet parameters $\eta_{11}^{(n)},\ldots,\eta_{mm}^{(n)}$
  for $k = 1$ to m **do**
    Sample $\pi_k^{(n)}$ from $D(\eta_{k1}^{(n)}+1,\eta_{k2}^{(n)}+1,\ldots,\eta_{km}^{(n)}+1)$
  end for
  Generate a random permutation ρ of $\{1,\ldots,m(2+r+p+q)\}$
  for $i = 1$ to $m(2+r+p+q)$ **do**
    Sample $\gamma_{\rho(i)}^{(n)}$ from $p(\gamma_{\rho(i)}|\boldsymbol{S}^{(n)},\boldsymbol{P}^{(n)},\gamma_{\rho(1)}^{(n)},\ldots,\gamma_{\rho(i-1)}^{(n)},\gamma_{\rho(i+1)}^{(n-1)},\gamma_{\rho(m(2+r+p+q))}^{(n-1)},y)$ with the Griddy-Gibbs sampler
  end for
  if End of the second phase is reached **then**
    Adjust/update the prior distributions
  end if
end while

3. It simply stands for a test case to assess the robustness and the efficiency of our estimation method.

50 series of 1500 data points are generated. Following the coarse-to-fine strategy described in the previous subsection, the bounds of the uniform prior distributions are set coarsely so as not to be too informative on the true coefficient values. The goal is to check whether the MCMC method is robust enough not to get trapped by local minima. The coefficient supports are then discretized with 42 equidistant points. Starting values for the regime sequence and all 16 parameters are randomly initialized within the range of possible values defined by their respective prior support. 50000 iterations of the MCMC algorithm are run, of which the last 30000 iterations are used for posterior inference, the first 10000 being discarded as burn-in and the second 10000 being used to refine the prior supports. For each simulation, convergence of the chain is assessed with the diagnostic proposed in Gelman and Rubin (1992) by running 3 chains in parallel, with different starting values. No evidence of non-convergence was noticed. When considering single sample, large estimation bias can be observed on both AR and GARCH coefficients. More satisfactorily, when considering 50 samples, absolute estimation errors for all parameters are smaller than their corresponding posterior standard deviations. As observed in Chen et al. (2009), the largest estimation errors are found for the posterior distributions of the GARCH coefficients whereas AR coefficients are estimated with a much higher accuracy. In each of the two regimes, β_1 is biased downwards and α_0 is biased upwards, which is a known issue with MS-GARCH models. For a given parameter, the coverage probability (CP) corresponds to the probability of its true value being encompassed within the interval defined by the 2.5% and 97.5% quantiles of its posterior distribution. In other words, these probabilities are the nominal 95% confidence intervals of the posterior estimates. Large deviations could indicate recurrent failure of the estimation method for some parameters. Globally, the estimated CP are all close to 95% and no large deviation is observed which is satisfactory. The grid refinement procedure shows that the supports of the AR coefficients are significantly smaller than the initial supports coarsely set. As for the final supports of GARCH coefficients, they consist of small adjustments of their initial supports. The verification for label switching is performed by analyzing the full posterior densities displayed in Figure 5 where no bimodality is observed. We can also add that the sampler performs quite well in terms of mixing since the densities are rather peaky and have small tails.

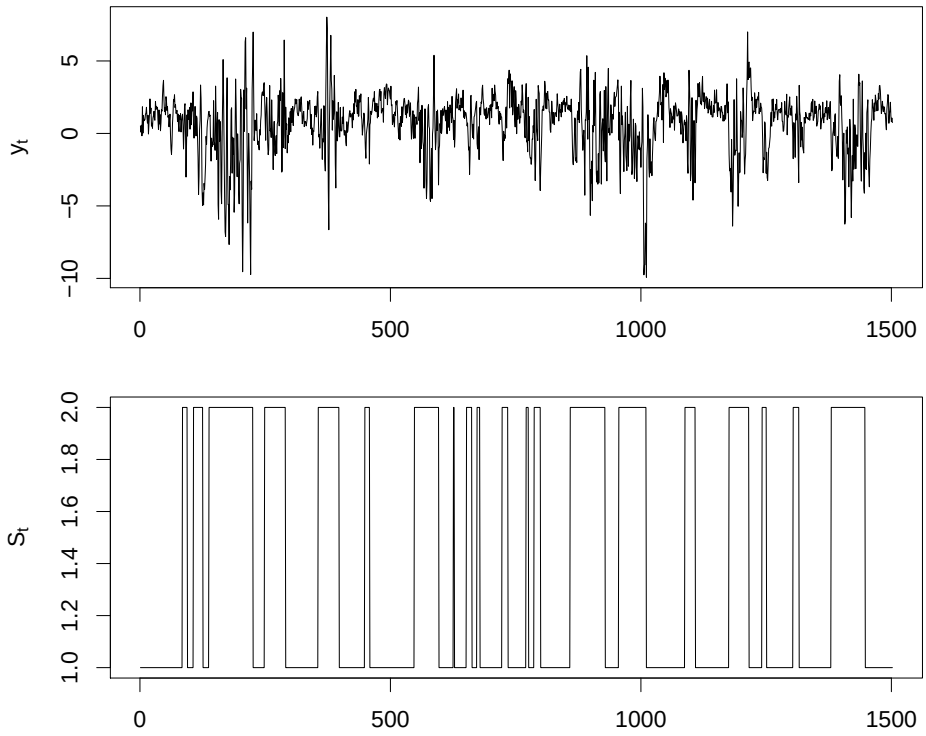


Figure 4: Simulation of a MS(2)-AR(2)-GARCH(1,1) whose coefficients and transition probability values are given in Table 1. Top: simulated process $y = (y_1, \ldots, y_T)$; Bottom: regime sequence $\boldsymbol{S} = (S_1, \ldots, S_T)$.

Table 1: Statistics on the posterior estimates for a synthetic MS(2)-AR(2)-GARCH(1,1) process, for 1 and 50 samples: Posterior means, standard deviations and coverage probabilities (CP).

			50 samples			1 sample		
	True value	Initial prior support	Posterior mean	Posterior std. dev.	CP	Refined prior support	Posterior mean	Posterior std. dev.
$\theta_0^{(1)}$	**0.5**	[-0.2 ; 1.2]	**0.500**	0.072	96%	[0.20 ; 0.78]	0.488	0.050
$\theta_1^{(1)}$	**0.5**	[-0.2 ; 1.2]	**0.502**	0.054	98%	[0.26 ; 0.72]	0.495	0.037
$\theta_2^{(1)}$	**0.2**	[-0.5 ; 0.9]	**0.197**	0.051	98%	[-0.01 ; 0.43]	0.212	0.035
$\alpha_0^{(1)}$	**0.1**	[0 ; 0.5]	**0.109**	0.041	94%	[0 ; 0.17]	0.084	0.020
$\alpha_1^{(1)}$	**0.2**	]0 ; 0.5]	**0.195**	0.068	94%	]0 ; 0.38]	0.175	0.046
$\beta_1^{(1)}$	**0.6**	[0 ; 1]	**0.593**	0.101	94%	[0.36 ; 0.88]	0.621	0.059
$\theta_0^{(2)}$	**0**	[-0.7 ; 0.7]	**-0.015**	0.041	94%	[-0.44 ; 0.36]	-0.038	0.100
$\theta_1^{(2)}$	**0.7**	[0 ; 1.4]	**0.689**	0.081	98%	[0.55 ; 0.99]	0.764	0.051
$\theta_2^{(2)}$	**-0.3**	[-1 ; 0.2]	**-0.308**	0.081	98%	[-0.59 ; -0.17]	-0.381	0.052
$\alpha_0^{(2)}$	**0.4**	[0.1 ; 0.8]	**0.512**	0.189	98%	[0 ; 0.82]	0.373	0.105
$\alpha_1^{(2)}$	**0.1**	]0 ; 0.5]	**0.114**	0.073	92%	]0 ; 0.33]	0.135	0.041
$\beta_1^{(2)}$	**0.85**	[0 ; 1]	**0.813**	0.087	96%	[0.62 ; 1[	0.831	0.044
p_{11}	**0.98**	]0 ; 1[	**0.977**	0.009	90%	]0 ; 1[	0.983	0.005
p_{22}	**0.96**	]0 ; 1[	**0.950**	0.023	92%	]0 ; 1[	0.961	0.012

Inference on the regime sequence can also be performed. However, methods for global decoding such as the Viterbi algorithm (Forney Jr, 1973) are not applicable to MCMC outputs since the sole smoothed probabilities of the regime sequence can be computed. Instead, we use a simple labelling rule to infer the regime sequence: state variables with a smoothed probability of being in regime k larger than 0.5 are classified as being in regime k. Following that rule, we can compute the successful regime inference rate and the probability of regime retrieval (the probability of the true regime being k knowing that the inferred regime is k). Results are reported in Table 2. Ideally, these quantities should be as close to 1 as possible. The rate of successful inference is higher for regime 1 (96%) than for regime 2 (90%). The same result holds for the probability of successful regime retrieval. These results are reasonably good according to the complexity of the model dynamics. Three of the model features may explain these differences: (i) regime 1 is characterized by a higher persistence probability than regime 2 ($p_{11} > p_{22}$); (ii) the unconditional variance ($\sigma^{(k)} = \alpha_0^{(k)}/(1-\alpha_1^{(k)}-\beta_1^{(k)})$ in regime 1 ($\sigma^{(1)} = 0.5$) is lower than in regime 2 ($\sigma^{(2)} = 8$) and (iii) persistence of shocks measured by $\alpha_1^{(k)} + \beta_1^{(k)}$ is also lower in regime 1 than in regime 2. Because of the higher persistence probability, parameters defining the first regime can be estimated over a larger number of data points and over longer time intervals clear off any structural break, on average, which leads to more accurate posterior estimates. The lower unconditional variance combined to the lower persistence to shocks in regime 1 makes the autoregressive and the conditional variance dynamics easier to identify and to separate. These latter comments are confirmed by the estimated posterior standard deviations of the model parameters (see Table 1) which are smaller in regime 1 than in regime 2, for corresponding parameters.

Table 2: Statistics on the inferred regime sequence.

Rate of successful regime inference	**Probability of regime retrieval**
$P(\widehat{S}_t = 1 \mid S_t = 1) = 0.96$	$P(S_t = 1 \mid \widehat{S}_t = 1) = 0.95$
$P(\widehat{S}_t = 2 \mid S_t = 2) = 0.90$	$P(S_t = 2 \mid \widehat{S}_t = 2) = 0.91$

5.6 Study on an Empirical Time Series of Wind Power

One of the main issue that arises when fitting Markov-Switching models to an empirical time series is the determination of the number of states m of the Markov chain. Theoretically, its determination is not to be separated of the autoregressive and conditional variance structure (orders r, p and q in Equations (1) and (2)). Along that idea, Psaradakis and Spagnolo (2006) review dif-

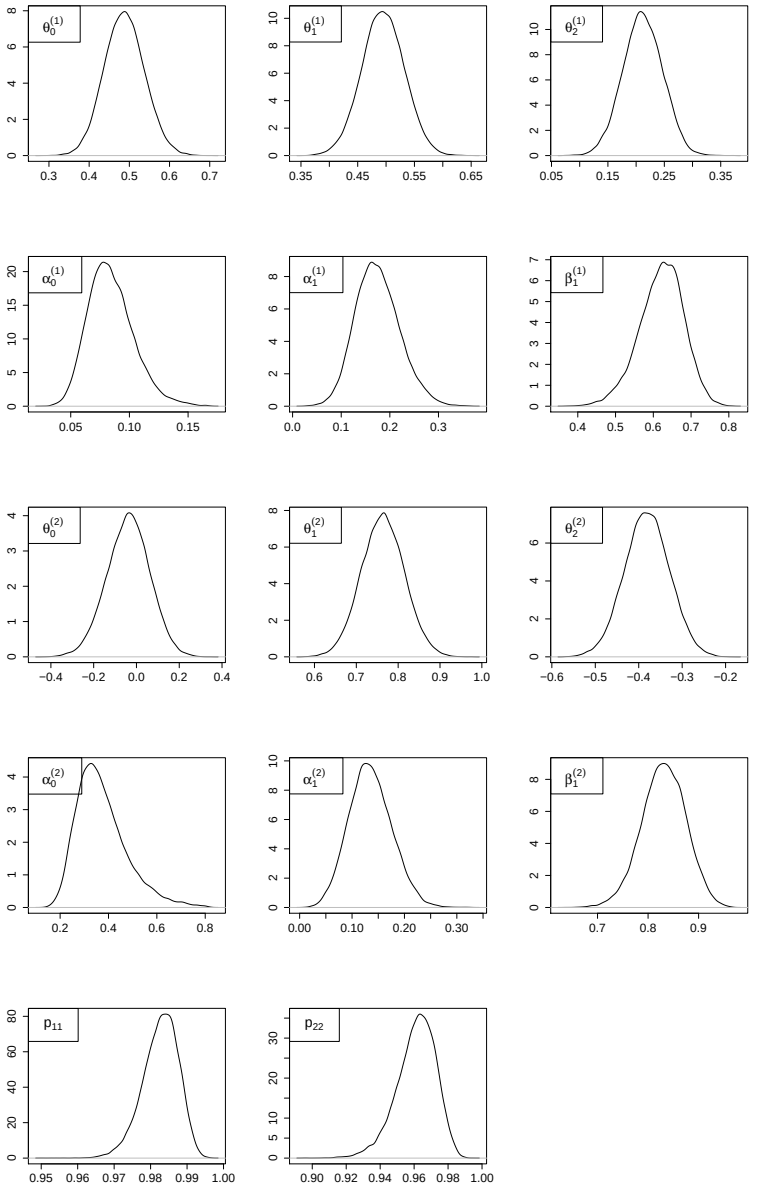


Figure 5: Estimated posterior densities of the simulated MS(2)-AR(2)-GARCH(1,1).

ferent penalized likelihood criteria for the joint determination of the number of hidden states and autoregressive order for MSAR models. However, in practise, misspecification in the parametrization of the model may result in overestimation of the optimal number of regimes. For instance, ignored volatility clustering effects can falsely be reported as regime-switching effects Cheung and Miu (2009).

The model identification approach taken in this study is to define the autoregressive and conditional variance orders a priori and determine the optimal number of regimes accordingly. Most studies involving Markov-Switching test a limited number of regimes, from 1 to 4. The underlying theoretical reason is that regime switchings occur infrequently. The more practical reason is that the number of parameters to be estimated grows quadratically with respect to the number of regimes, and constraints for regime identification become more difficult to define.

One reason to proceed that way and not by computing the Bayesian Information Criterion is that there is no method for computing the marginal likelihood of MS-GARCH models to our knowledge. An empirical cross-validation procedure is used instead. The time series of interest is the one presented in Section 3 for which measurements from the Horns Rev 1 wind farm are averaged over 10 min intervals. All available observations from August 2005 (i.e., 4125 observations) are used for estimating the posterior distributions of the MS-AR-GARCH model. Several parametrizations with respect to m, r, p and q are tested. Then, all available observations from September 2005 (i.e., 4320 observations) are used for cross-validation and the parametrization resulting in the best one-step ahead Continuous Ranked Probability Score (Gneiting, 2008) was chosen. The best performances were obtained for models with 3 autoregressive lags and a GARCH(1,1) structure for the conditional variance in each regime. The autoregressive order is in agreement with previous studies on the same data set (Pinson and Madsen, 2010, Gallego et al., 2011). To keep the computational complexity and burden reasonable, only models defined with 1 and 2 regimes were tested. Furthermore, no constraint for regime identification could be found for a number of regimes larger than 2. Posterior estimates for MS(m)-AR(3)-GARCH(1,1) with $m = 1$ and $m = 2$ are reported in Table 3. Posterior densities for the MS(2)-AR(3)-GARCH(1,1) are shown in Figure 6.

Table 3: Statistics on the posterior estimates of the AR(3)-GARCH(1,1) and MS(2)-AR(3)-GARCH(1,1) model fitted to the time series of wind power.

	1 Regime: AR(3)-GARCH(1,1)				2 Regimes: MS(2)-AR(3)-GARCH(1,1)			
	Initial prior support	**Refined prior support**	**Posterior mean**	**Posterior std. dev.**	**Initial prior support**	**Refined prior support**	**Posterior mean**	**Posterior std. dev.**
$\theta_0^{(1)}$	$[-0.01\ ;\ 0.01]$	$[-0.007\ ;\ 0.006]$	-2.10^{-4}	0.002	$[-0.04\ ;\ 0.04]$	$[-0.004\ ;\ 0.004]$	-3.10^{-5}	6.10^{-4}
$\theta_1^{(1)}$	$[1\ ;\ 1.7]$	$[0.68\ ;\ 2.11]$	1.358	0.232	$[1\ ;\ 1.8]$	$[0.64\ ;\ 2.18]$	1.417	0.273
$\theta_2^{(1)}$	$[-0.85\ ;\ -0.05]$	$[-1.33\ ;\ 0.34]$	-0.460	0.284	$[-0.95\ ;\ -0.15]$	$[-1.36\ ;\ 0.21]$	-0.574	0.304
$\theta_3^{(1)}$	$[-0.15\ ;\ 0.35]$	$[-0.52;\ 0.72]$	0.107	0.206	$[-0.35\ ;\ 0.55]$	$[-0.67\ ;\ 0.99]$	0.156	0.300
$\alpha_0^{(1)}$	$[0\ ;\ 3.10^{-4}]$	$[0\ ;\ 3.10^{-4}]$	7.10^{-5}	6.10^{-5}	$[5.10^{-6}\ ;\ 10^{-4}]$	$[2.10^{-6}\ ;\ 10^{-5}]$	3.10^{-6}	2.10^{-7}
$\alpha_1^{(1)}$	$[0.2\ ;\ 1]$	$[0.03\ ;\ 1]$	0.513	0.161	$[0\ ;\ 1]$	$[0.23\ ;\ 0.74]$	0.499	0.077
$\beta_1^{(1)}$	$[0\ ;\ 0.7]$	$[0\ ;\ 0.95]$	0.467	0.161	$[0\ ;\ 1]$	$[0.25\ ;\ 0.74]$	0.489	0.074
$\theta_0^{(2)}$	-	-	-	-	$[-0.06\ ;\ 0.10]$	$[-0.04\ ;\ 0.09]$	0.011	0.013
$\theta_1^{(2)}$	-	-	-	-	$[0.7\ ;\ 1.7]$	$[0.27\ ;\ 2.02]$	1.178	0.285
$\theta_2^{(2)}$	-	-	-	-	$[-0.7\ ;\ 0.3]$	$[-1.22\ ;\ 0.58]$	-0.323	0.341
$\theta_3^{(2)}$	-	-	-	-	$[-0.4\ ;\ 0.6]$	$[-0.76\ ;\ 1.01]$	0.126	0.284
$\alpha_0^{(2)}$	-	-	-	-	$[1.10^{-3}\ ;\ 8.10^{-3}]$	$[0\ ;\ 4.10^{-3}]$	5.10^{-4}	3.10^{-4}
$\alpha_1^{(2)}$	-	-	-	-	$[0\ ;\ 1]$	$[0\ ;\ 0.54]$	0.079	0.080
$\beta_1^{(2)}$	-	-	-	-	$[0\ ;\ 1]$	$[0\ ;\ 1]$	0.892	0.088
p_{11}	-	-	-	-	$[0\ ;\ 1]$	$[0\ ;\ 1]$	0.913	0.029
p_{22}	-	-	-	-	$[0\ ;\ 1]$	$[0\ ;\ 1]$	0.783	0.114

One of the reason why we prefer the GG over the MH sampler is that it can estimate posterior densities of various shape without prior knowledge of their closed form. From Figure 6, it can be noticed that the posterior densities of the GARCH equation are asymmetric, more notably in regime 2. This is due to the constraints imposed in Equations (4) and (5) and the asymmetry becomes stronger as the posterior mean of a given parameter is close to the bounds of the constraints. $\alpha_0^{(1)}$ is numerically close to 0 and its posterior density has the shape of a mass point. Omitting this parameter for fitting the model makes the regimes less stable and it is decided to keep it in the formulation of the MS(2)-AR(3)-GARCH(1,1) model. The posterior densities of the AR equation have symmetric shapes. However, they are characterized by large posterior standard deviations and rather flat shapes which is the consequence of the strong autocorrelation between coefficients within a same regime, as mentioned earlier in this Section. That problem was neither encountered in our simulations on synthetic data nor in other studies such as Chen et al. (2009), Bauwens et al. (2010), Henneke et al. (2011), since the parametrization of the conditional mean equation is restricted to one lag at most. Since it may affect the final posterior mean estimates used for prediction, further research will be dedicated to investigate potential techniques to overcome it.

In addition, analyzing the posterior estimates of our model may reveal interesting features on the very short-term wind power fluctuations of the Horns Rev 1 wind farm. The low (respectively high) frequency wind power fluctuations are captured by the AR (respectively GARCH) coefficients of the model and different profiles of fluctuations are expected across regimes. In addition, transition probability estimates may indicate whether one regime is more persistent over time than the other.

Regarding the model with one regime, AR(3)-GARCH(1,1), we report its posterior estimates in order to illustrate the transition from a single regime model to a two regime model and appraise how the posterior estimates of the 2 regime model may relate to those of the single regime model. Initial prior bounds were defined based on the estimates obtained by numerical maximization of the likelihood function (NML). The posterior estimates of the AR coefficients are in close agreement with those obtained by NML while the posterior estimates of the GARCH coefficients deviate more. After verification, this can be due to a bimodality on the posterior density of the α_0 coefficient, which makes its estimated posterior mean larger than the one estimated by NML. These results are not presented here in order to save space but are available upon request.

As for the MS(2)-AR(3)-GARCH(1,1), the autoregressive dynamics are rather similar in the two regimes but for the intercept terms $\theta_0^{(1)}$ and $\theta_0^{(2)}$ which confirms the earliest results in Pinson and Madsen (2010). More interestingly, the dynamics of the conditional variance in the two regimes differ in several ways.

First, the intercept terms in regime 1 is significantly lower than in regime 2 ($\alpha_0^{(1)} \ll \alpha_0^{(2)}$), which means that regime 2 can be interpreted as the regime for which the amplitude of the wind power fluctuations are the largest. Then the posterior mean estimates of the GARCH coefficients in regime 1, $\alpha_1^{(1)}$ and $\beta_1^{(1)}$ are approximately equal, which indicates that small prediction errors are followed by fast decreases of the conditional variance value while large errors give rise to sudden explosions. In regime 2, because $\beta_1^{(2)} \gg \alpha_1^{(2)}$, the conditional variance level is more stable between successive observations and has a longer memory of large errors. Finally, one can also notice that $p_{11} > p_{22}$, which translates into regime 1 being more persistent than regime 2 (i.e., periods of low volatility last longer than periods of high volatility).

An illustration of the estimated sequence of smoothed probabilities for the MS-AR-GARCH model is given in Figure 7. In particular, it depicts the smoothed probabilities of being in regime 1. It can be noticed that the two regimes do not seem to be well separated but for periods where the wind power generation is null or close to its nominal capacity P_n, with smoothed probabilities close to 1. Even though a clear separation of the regimes is a very desirable feature, it does not automatically translate into a loss of predictive power of the Markov-Switching model. This aspect will be further addressed in the next section of this study.

First, simulations on synthetic data have allowed us to design and tune our estimation method for MS-AR-GARCH models. Then, its applicability to an empirical time series of wind power is tested and demonstrated a good ability to estimate posterior densities of various shapes despite some limitations regarding the posterior densities of the autoregressive coefficients. Nevertheless, our will is not to identify the best class of models for the modeling of very short-term wind power fluctuations but rather to investigate new alternatives such as the proposed MS-AR-GARCH model for (i) providing additional insights on these wind power fluctuations and (ii) investigating on their potential predictive power.

6 Wind Power Forecast Evaluation

Forecasting wind power fluctuations of large offshore wind farms at a time scale of a few minutes is a relatively new and difficult challenge. The difficulty stems from the lack of meteorological observations in the neighborhood of the wind farm. The consequences are that state-of-the-art models often fail in predicting wind power fluctuations of large amplitude caused by sudden changes in the weather conditions nearby the wind farm. In practise, naive forecasts

are difficult to significantly outperform (Pinson et al., 2008).

The literature on short-term wind power forecasting is abundant and a recent overview is available in Giebel et al. (2011). Originally, the quality and accuracy of statistical forecasts of wind power were evaluated with respect to point prediction scores. From a decision making perspective, the drawback of such an approach is that it clearly neglects the uncertainty associated with the forecast, often leading to sub-optimal control strategies. Therefore, quantifying the probability of all potential outcomes greatly enhances the usefulness of wind power forecasts (Pinson et al., 2007). These probabilistic forecasts can either take the form of density functions or prediction intervals when numerically approximated and should preferably be evaluated with respect to their calibration and sharpness (Gneiting, 2008). Accurate quantification of the uncertainty associated with a point forecast is an information as valuable as the value of the forecast itself. It could first assist wind farm operators in anticipating the risks of unexpected wind power fluctuations when point forecast fails in doing so. And, ultimately, it could help them in determining backup strategies based on available energy reserves.

One of the drawbacks of MS-GARCH models is that the conditional variance becomes intractable with the addition of autoregressive terms in the model formulation. This stands as a clear limitation for the use of such class of models for prediction applications. To bypass that problem, the approach chosen in Chen et al. (2009) is to repeat the estimation of the model over a sliding window and generate one-step ahead forecasts based on the new set of estimates. We think that this approach is too computationally intensive and instead, we prefer to use the recursive update formula of the conditional variance as presented by Gray in Gray (1996).

6.1 Approximating the Conditional Variance for Prediction Applications

The formula developed in Gray (1996) recursively approximates the conditional variance as the weighted average of past conditional variances. One of its advantages is that it is flexible and it can be extended to include autoregressive terms. One may then argue and wonder why we did not use that formula to estimate our MS-AR-GARCH model. We did investigate the possibility of using it with an estimation method based on numerical maximization of the Likelihood function. Nevertheless, due to the complexity of the Likelihood function, parameter either ended up on the bounds of the constraints Equations (4) and (5) or convergence could not be reached, which prevented its use for the estimation step of the study.

For a MS(m)-AR(r)-GARCH(1,1) model, the approximated conditional variance at time t, h_t, is defined as follows:

$$h_t = E[y_t^2|y_{[1,t-1]},\Theta] - E[y_t|y_{[1,t-1]},\Theta]^2 \tag{25}$$

First, the term $E[y_t|y_{[1,t-1]},\Theta]$ is the optimal one-step predictor and, under normality conditions, can be calculated as the weighted sum of the predictions in each regime:

$$E[y_t|y_{[1,t-1]},\Theta] = \hat{y}_{t|t-1} = \sum_{k=1}^{m} \hat{\xi}_{t|t-1}^{(k)}(\theta_0^{(k)} + \sum_{i=1}^{r} \theta_i^{(k)} y_{t-i}) \tag{26}$$

Second, the term $E[y_t^2|y_{[1,t-1]},\Theta]$ can be computed as follows:

$$E[y_t^2|y_{[1,t-1]},\Theta] = \sum_{k=1}^{m} \hat{\xi}_{t|t-1}^{(k)}(h_t^{(k)} + (\theta_0^{(k)} + \sum_{i=1}^{r} \theta_i^{(k)} y_{t-i})^2) \tag{27}$$

with $h_t^{(k)}$ the one-step ahead predicted conditional variance in regime k computed as follows:

$$h_t^{(k)} = \alpha_0^{(k)} + \alpha_1^{(k)} \varepsilon_{t-1}^2 + \beta_1^{(k)} h_{t-1} \tag{28}$$

and $\hat{\xi}_{t|t-1}^{(k)}$ the predictive probability of being in regime k at time t, given all information available at time $t-1$. The vector of predictive probabilities $\hat{\xi}_{t|t-1} = [\hat{\xi}_{t|t-1}^{(1)},\ldots,\hat{\xi}_{t|t-1}^{(m)}]^T$ can be computed in a recursive manner as follows:

$$\hat{\xi}_{t|t-1} = \boldsymbol{P}^T \hat{\xi}_{t-1|t-1} \tag{29}$$

with $\hat{\xi}_{t-1|t-1} = [\hat{\xi}_{t-1|t-1}^{(1)},\ldots,\hat{\xi}_{t-1|t-1}^{(m)}]^T$ the vector of filtered probabilities at time $t-1$ whose elements can be computed as follows:

$$\hat{\xi}_{t-1|t-1}^{(k)} = \frac{\hat{\xi}_{t-1|t-2}^{(k)} \times f(y_{t-1}|S_{t-1}=k, y_{[1,t-2]},\Theta)}{\sum_{k=1}^{m} \hat{\xi}_{t-1|t-2}^{(k)} \times f(y_{t-1}|S_{t-1}=k, y_{[1,t-2]},\Theta)} \tag{30}$$

where $f(y_{t-1}|S_{t-1}=k, y_{[1,t-2]},\Theta)$ is the conditional density of y_{t-1} given the set of information available at time $t-2$.

We are aware that the approximation presented here above is not optimal for prediction applications, since it may introduce a permanent bias in the computation of the conditional variance. It is a choice governed by the necessity to bypass a problem not yet solved and to minimize its computational cost. It could then be expected that the prediction skills of our model would benefit from advances towards a better tracking of the conditional variance for MS-AR-GARCH models. As for now, we can proceed to the evaluation of the prediction skills of our model.

6.2 Evaluation of Point Forecasts

The out-of-sample predictive power of our MS-AR-GARCH model is evaluated based on its performance on one-step ahead forecasts. Point forecast skills are first considered and compared to common benchmark models for very short-term wind power fluctuations as well as state-of-the-art models. Common benchmark models include persistence (i.e., $\hat{y}_t = y_{t-1}$) and the simple but robust AR model. State-of-the-art models include the class of MSAR models as initially applied to wind power time series in Pinson et al. (2008). MSAR models were not estimated with the method presented in the previous section since more robust estimation methods exist for that type of models. Instead, they were estimated by numerical maximization of the Likelihood function. Following the standardized framework for the performance evaluation of wind power forecasts discussed in Madsen et al. (2005), the proposed score functions to be minimized are the Normalized Mean Absolute Error (NMAE) and Root Mean Square Error (NRMSE). A higher importance is given to the NRMSE over the NMAE in the final evaluation of point forecast skills because the RMSE is a quadratic score function and is more likely to highlight the power of a given model to reduce large errors. Reducing these large prediction errors is indeed a very desirable ability of prediction models that we aim at developing. The out-of-sample evaluation is performed over approximately 17,000 data points of which more than 3000 are missing (from October 2005 to January 2006). The optimal parametrization for each of the models cited here above was defined by cross validation in the same way as for the MS-AR-GARCH model. NMAE and NRMSE scores are computed for all models and reported in Tables 4 and 5. For Markov-Switching models, the optimal one-step ahead predictor is given by Equation (26).

As it could have been expected, MSAR models, with 2 or 3 regimes, outperform all other models for both the NMAE and NRMSE. The best improvement in NMAE over persistence is about 5.1% while it is 4.4% for the NRMSE. These levels of improvement agree with earlier results in Pinson et al. (2008) and Gallego et al. (2011). If moving from AR to MSAR models leads to appreciable improvements, moving from AR to AR-GARCH models results in the opposite effect. However, moving from single regime AR-GARCH to regime switching AR-GARCH has a significant positive effect, more notably for the NRMSE. The relatively good performances of the MS-AR-GARCH model are comparable to those of the MSAR model with 2 regimes. All these results tend to indicate that the MSAR class of models, explicitly designed to capture regime switching and autocorrelation effects, has better point prediction skills.

If accounting for heteroscedastic effects in regime switching models makes that part of the dynamics originally captured by the AR component of MSAR models is instead captured by the GARCH component and results in lower perfor-

Table 4: NMAE score given in percentage of the nominal capacity of the Horns Rev 1 wind farm. Results are given for persistence, an AR model with 3 lags AR(3), a MSAR model with 2 regimes and 3 lags in the conditional mean equation MSAR(2,3), a MSAR model with 3 regimes and 3 lags in the conditional mean equation MSAR(3,3), an AR-GARCH model with 3 lags in the conditional mean equation and a GARCH(1,1) specification for the conditional variance, and finally for the MS-AR-GARCH model estimated in Section 5.

Model	**Oct.**	**Nov.**	**Dec.**	**Jan.**	**Total**
Persistence	2.41	2.58	3.01	2.47	2.55
AR(3)	2.36	2.64	2.98	2.46	2.53
AR(3)-GARCH(1,1)	2.29	2.60	2.95	2.41	2.49
MS(2)-AR(3)-GARCH(1,1)	2.27	2.50	2.89	2.38	2.44
MSAR(2,3)	2.28	2.49	2.89	2.37	2.44
MSAR(3,3)	2.26	2.49	2.89	2.36	2.42

Table 5: NRMSE score given in percentage of the nominal capacity of the Horns Rev 1 wind farm. Results are given for the same models as for the NMAE.

Model	**Oct.**	**Nov.**	**Dec.**	**Jan.**	**Total**
Persistence	4.17	6.22	5.76	4.28	5.02
AR(3)-GARCH(1,1)	4.00	6.18	5.72	4.24	4.93
AR(3)	3.98	5.99	5.56	4.17	4.83
MS(2)-AR(3)-GARCH(1,1)	3.96	6.00	5.55	4.15	4.82
MSAR(2,3)	3.98	5.95	5.55	4.17	4.81
MSAR(3,3)	3.96	5.95	5.55	4.17	4.80

mances in point forecasting. It can then be expected that this will translate into better performances for probabilistic forecasts of models explicitly designed to capture the heteroscedastic effects, such as the AR-GARCH and MS-AR-GARCH models.

6.3 Evaluation of Interval and Density Forecasts

Probabilistic forecasts are very useful in the sense that they provide us with a measure of the uncertainty associated with a point forecast. They can either take the form of density or interval forecasts. For their evaluation we follow the framework presented in Gneiting et al. (2007).

First, we consider the overall skill of the probabilistic forecasts generated by the proposed MS-AR-GARCH model. The traditional approach consists in evaluating the calibration and sharpness of the density forecasts. The calibration of a forecast relates to its statistical consistency (i.e., the conditional bias of the observations given the forecasts). As for the sharpness of a forecast, it refers to its concentration or, in other words, to its variance. The smaller the variance, the better, given calibration. One score function known to assess both the calibration and sharpness of density forecasts simultaneously is the Continuous Ranked Probability Score (CRPS), as defined in Gneiting et al. (2007). The exercise consists in generating one-step ahead density forecasts. For the single regime model, these density forecasts take the form of Normal density functions, while for Markov-Switching models they take the form of mixtures of conditional Normal distributions weighted by the predictive probabilities of being in each of the given regime. The CRPS criterion is computed for the same models as for the point prediction exercise and the results are reported in Table 6.

From Table 6, it can noticed that the proposed MS-AR-GARCH model has the best overall skill. Its improvement over AR models is about 12.6%. More generally, GARCH models outperform non-GARCH models even though the improvements are very small in some cases. The relatively good performance of the MSAR model with 3 regimes tend to indicate that the volatility clustering effect captured by GARCH models may partly be captured as a regime switching effect by MSAR models. This may appear as a paradox but it is not, in our opinion. As noticed in Pinson and Madsen (2010), the respective dynamics in the three regimes of the MSAR model can be more easily characterized with respect to the values of their respective variance rather than their respective conditional mean dynamics. While GARCH models are explicitly designed for capturing the heteroscedastic effect, the formulation of MSAR models makes that the same effect can be captured in an implicit manner by the combination of several dynamics with different variances. The consequence of these find-

Table 6: CRPS criterion given in percentage of the nominal capacity of the Horns Rev 1 wind farm. Results are given for the same models as for the point prediction exercise.

Model	Oct.	Nov.	Dec.	Jan.	Total
AR(3)	1.99	2.33	2.48	2.02	2.15
MSAR(2,3)	1.81	2.01	2.26	1.88	1.94
MSAR(3,3)	1.78	1.98	2.24	1.85	1.91
AR(3)-GARCH(1,1)	1.76	1.99	2.24	1.85	1.91
MS(2)-AR(3)-GARCH(1,1)	1.76	1.95	2.20	1.83	1.88

ings is that MS-AR-GARCH models which combine both a Markov-Switching and GARCH formulation are not very powerful for separating the regimes (see Figure 7), since there may be a conflict in their formulation. However, it does not automatically affect their predictive power since a clear separation of the regimes may not automatically translate into better prediction skills. Instead, it is reflected in a more parsimonious parametrization of the MS-AR-GARCH models regarding the optimal number of regimes.

In order to better evaluate the contribution of the calibration to the overall skill of probabilistic forecasts, one can compare the empirical coverage rates of intervals forecasts to the nominal ones. Intervals forecasts can be computed by means of two quantiles which define a lower and an upper bound. They are centered around the median (i.e., the quantile with nominal proportion 0.5). For instance, the interval forecast with a coverage rate of 0.8 is defined by the two quantiles with nominal proportion 0.1 and 0.9. Empirical coverage rates of interval forecasts generated from an AR, MSAR and MS-AR-GARCH are computed and reported in Table 7. A graphical example of the dynamical shape of these interval forecasts is given in Figure 8, for the MS-AR-GARCH model and a coverage rate of 90%. From Table 7, recurrent and large positive deviations are observed for the interval forecasts generated from the AR model, indicating that the intervals are too wide. In contrast, the empirical coverage rate of the interval forecasts generated from the MSAR model exhibits a relatively good match with the nominal coverage rates. The maximum deviation is around 6%. While these intervals seem too wide for small nominal coverage rates (i.e., from 10 up to 50%), they become too narrow for large nominal coverages. As for the intervals generated from the MS-AR-GARCH models, the agreement is excellent for the smallest nominal coverage rates (i.e., from 10 up to 40%) and the largest one (i.e., 90%), whereas it significantly deviates from the nominal coverage of intermediate widths. This latter result may be the consequence of a bias introduced by the approximation of the conditional variance as presented earlier. This also tends to indicate that the relatively good overall skill of prob-

Table 7: Nominal coverage rates and empirical coverage rates of interval forecasts generated by the following three models: AR(3), MSAR(3,3) and MS(2)-AR(3)-GARCH(1,1). The coverage rates are expressed in %.

Nom. cov.	Emp. cov.		
	AR(3)	**MSAR(3,3)**	**MS(2)-AR(3)-GARCH(1,1)**
10	13.2	7.1	9.4
20	42.6	25.8	20.7
30	55.5	35.2	31.3
40	64.3	43.9	42.3
50	71.4	52.4	63.2
60	77.2	60.3	71.2
70	81.6	68.8	78.1
80	89.9	77.7	84.4
90	90.0	86.9	90.0

abilistic forecasts generated from MS-AR-GARCH models are more likely to be the result of sharp rather than consistent forecasts.

7 Discussion and Concluding Remarks

We presented a general framework for the modeling and forecasting of very-short term wind power fluctuations at large offshore wind farms. The dynamics of these fluctuations are very complex and developing models for prediction applications is an ongoing challenge within the wind power community. The interest of the proposed MS-AR-GARCH model is that it extends the state-of-the-art methodology based on MSAR models and specifies the conditional variance in each regime as a GARCH model in order to better account for heteroscedastic effects. This calls for an advanced estimation method to overcome the problem linked to the historical path dependency of the conditional variance. In that regard, Bayesian methods offer an alternative framework to methods based on Maximum Likelihood Estimation. In particular, they allow to break down the complexity of the global estimation problem into a set of smaller problems for which practical approach exists.

In a first stage, we gave a thorough introduction on the estimation method based on a MCMC algorithm. Then, we identified issues linked to its implementation and presented some solutions to overcome them. In a second stage, the estimation method for the proposed MS-AR-GARCH model was tested on

both synthetic and empirical time series. It was successfully applied to synthetic time series. The results on the empirical time series of wind power are more mixed. In particular, the method encountered clear problems in dealing with the high correlation of the AR coefficients of the model, which resulted in rather flat posterior densities. On the opposite, it seemed to work well for the other model parameters (i.e., GARCH coefficients and transition probabilities). In that respect, directions for future research could include the investigation of more appropriate sampling methods for the AR coefficients.

The predictive ability of the MS-AR-GARCH model was evaluated on a one-step ahead forecasting exercise of wind power time series sampled over 10 min intervals. Empirical comparisons of its performances against common benchmark and state-of-the-art models showed that (i) it is slightly outperformed by MSAR models for point forecasts according to NMAE and NRMSE criteria; (ii) it outperforms all other models in terms of overall skill of probabilistic forecasts evaluated with the CRPS criterion. However, these results need to be put into a broader perspective. First, both point forecast improvements of MSAR and MS-AR-GARCH models over the simple but robust AR model are very small for the NRMSE score function, while they are larger for the NMAE score function. This tends to indicate that Markov-Switching models contribute to reducing point forecast errors over periods where the wind power fluctuations are characterized by small rather than large amplitude. Second, and more interestingly, all three MSAR, AR-GARCH and MS-AR-GARCH models are able to capture periods characterized by different volatility levels of wind power fluctuations at the Horns Rev 1 wind farm. Having said that, the overall merit of the proposed MS-AR-GARCH model is to generate improved probabilistic forecasts with respect to their calibration and sharpness. This is important since only a complete description of all potential outcomes, and hence their probability distribution, may lead to optimal decisions in wind energy, as shown in Pinson et al. (2007).

The concerns raised in Section 4.1 about the sub-optimality of the Normal assumption were recently addressed in Pinson (2012) which proposed the use of a Generalized Logit-Normal distribution instead. One aspect of this distribution is that it is more appropriate for modeling the skewness of the errors and the heteroskedastic effects near the bounds of the process. It led to substantial improvements in terms of calibration, sharpness and overall reliability of density forecasts. For instance, the additional improvement in the CRPS criterion for a simple AR model is about 7%–8%. These results are in line with those reported in Gneiting et al. (2006), Thorarinsdottir and Gneiting (2010), Lau and McSharry (2010) which showed the potential of using a truncated Normal distribution for wind speed and wind power prediction applications. Similarly, the use of the Generalized Logit-Normal distribution for Markov-Switching will be investigated with a particular focus on multi-step ahead forecasts.

For the time being and in the absence of meteorological observations to explain the origin of the volatility observed at Horns Rev, statistical models do not have the ability to anticipate the most abrupt changes in the dynamics of the wind power fluctuations. Future approaches based on the integration of observations of local weather conditions are likely to fill in that gap. A first step was achieved in Gallego et al. (2011) with the integration of on-site wind speed and direction measurements into prediction models, resulting in appreciable improvements of wind power fluctuation predictability. Another lead was given in Vincent et al. (2011) with the observations of convective rain cells during episodes of extreme wind speed variability. Following these observations, a weather radar capable of measuring rain reflectivity at high spatio-temporal resolution is currently operated at the offshore site of Horns Rev in order to provide additional insights on these wind power fluctuations and help improving their predictability.

Acknowledgements

The present study was supported by the "SafeWind" project (ENK7-CT2008-213740) and the Danish Public Service Obligation (PSO) fund under the projects "Mesoscale atmospheric variability and the variation of wind and production for offshore wind farms" (contract PSO-7141) and "Radar@Sea" (contract no. 2009-1-0226) which are hereby acknowledged. Vattenfall is greatly acknowledged for sharing the wind power data for the Horns Rev 1 wind farm. We also acknowledge the two anonymous reviewers for their relevant comments and suggestions to improve the present article.

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, Rasmussen C, Eriksen PB, Pedersen J. (2007) Technical aspects of status and expected future trends for wind power in Denmark. *Wind Energy*, 10:31–49.

Asai M. (2006) Comparison of MCMC methods for estimating GARCH models. *Journal of the Japan Statistical Society*, 36:199–212.

Bauwens L, Lubrano M. (1998) Bayesian inference on GARCH models using the Gibbs sampler. *The Econometrics Journal*, 1:23–46.

Bauwens L, Preminger A, Rombouts V. (2010) Theory and inference for a Markov switching GARCH model. *The Econometrics Journal*, 13:218–244.

Bollerslev T. (1986) Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, 31:307–327.

Cai J. (1994) A Markov model of switching-regime ARCH. *Journal of Business & Economic Statistics*, 12:309–316.

Chen C, So M, Lin E. (2009) Volatility forecasting with Double Markov switching GARCH models. *Journal of Forecasting*, 28:681–697.

Cheung C, Miu P. (2009) Currency instability: Regime switching versus volatility clustering. *Quarterly Journal of Finance and Accounting*, 48: 67–81.

Chib S. (1996) Calculating posterior distributions and modal estimates in Markov mixture models. *Journal of Econometrics*, 75: 79–97.

Cripps E, Dunsmuir W. (2003) Modeling the variability of Sydney Harbor wind measurments. *Journal of Applied Meteorology*, 42:1131–1138.

Danish Energy Agency. (2011) Energy Statistics 2010. Available online: `www.ens.dk`

Durbin R, Eddy S, Krogh A, Mitchison G. (1998) Biological sequence analysis. *Cambridge University Press*.

Ewing B, Kruse J, Schreoder J. (2006) Time series analysis of wind speed with time-varying turbulence. *Environmetrics*, 17:119–127.

Focken U, Lange M, Mönnich K, Wald HP, Beyer G, Luig A. 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.

Forney Jr G. (1973) The Viterbi algorithm. *Proceedings of the IEEE*, 61: 268–278.

Frühwirth-Schnatter S. (2006) Finite Mixture and Markov Switching Models. *Springer*.

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

Geman S, Geman D. (1984) Stochastic relaxation, Gibbs distribution and Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:721–741.

Gelman A, Rubin D. (1992) Inference from iterative simulation using multiple sequences. *Statistical Science*,7: 57–472.

Giebel G, Brownsword R, Kariniotakis G, Denhard M, Draxl C. (2011) The state-of-the-art in short-term prediction of wind power: A literature overview. Technical Report, ANEMOS.plus.

Gilks W, Richardson S, Spiegelhalter D. (1996) Markov Chain Monte Carlo in Practice. *Chapman & Hall*.

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

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

Gneiting T. (2008) Editorial: Probabilistic forecasting. *Journal of the Royal Statistical Society, Series A*, 171:319–321.

Gray S. (1996) Modeling the conditional distribution of interest rates as a regime-switching process. *Journal of Financial Economics*, 42:27–62.

Haas M, Mittnik S, Paolella M. (2004) A new approach to Markov-Switching GARCH models. *Journal of Financial Econometrics*, 2:493–530.

Hamilton J. (1989) A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57:357–384.

Hamilton J, Susmel R. (1994) Autoregressive conditional heteroskedasticity and changes in regime. *Journal of Econometrics*, 64:307–333.

Hastings W. (1970)Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57:97–109.

Henneke J, Rachev S, Fabozzi F, Nikolov M. (2011) MCMC-based estimation of Markov-Switching ARMA-GARCH models. *Applied Economics*, 43:259–271.

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

Klaasen F. (2002) Improving GARCH volatility forecasts with regime-switching GARCH. *Empirical Economics*, 27:363–394.

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

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

Liu J, Wong W, Kong A. (1995) Covariance structure and Convergence rate of the Gibbs sampler with various scans. *Journal of the Royal Statistical Society, Series B*, 57:157–169.

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

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

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

Pinson P, Madsen H. (2010) Adaptative modelling and forecasting of offshore wind power fluctuations with Markov-Switching autoregressive models. *Journal of Forecasting*, 31:281–313.

Pinson P. (2012) Very short-term probabilistic forecasting of wind power time-series with generalized Logit-Normal distributions. *Journal of the Royal Statistical Society, Series C*, Available online.

Pryor S, Barthelmie R. (2002) Comparison of potential power production at on- and offshore sites. *Wind Energy*, 4:173–181.

Psaradakis Z, Spagnolo N. (2006) Joint determination of the state dimension and autoregressive order for Markov regime switching. *Journal of Time Series Analysis*, 27:753–766.

Rabiner L, Juang B. (2005) Fundamentals of Speech Recognition. *Prentice-Hall*.

Ritter C, Tanner M. (1992) Facilitating the Gibbs Sampler: The Gibbs stopper and the Griddy Gibbs sampler. *Journal of the American Statistical Association*, 87:861–868.

Robert C, Celeux G, Diebolt J. (1993) Bayesian estimation of hidden Markov chains: A stochastic implementation. *Statistics & Probability Letters*, 16:77–83.

Rydén, T. (2008) EM versus Markov Chain Monte Carlo for estimation of hidden Markov models: A computational perspective. *Bayesian Analysis*, 3:659–688.

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

Scott S. (2002) Bayesian methods for Hidden Markov Models: Recursive computing in the 21st century. *Journal of the American Statistical Association*, 97:337–351.

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

Tanner M, Wong W. (1987) The calculation of the posterior distributions by data augmentation. *Journal of the American Statistical Association*, 82:528–540.

Taylor J, Buizza R. (2004) A comparison of temperature density forecasts from GARCH and atmospheric models. *Journal of Forecasting*, 23:337–355.

Taylor J, Buizza R. (2006) Density forecasting for weather derivative pricing. *International Journal of Forecasting*, 22:29–42.

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

Thorarinsdottir T, Gneiting T. (2010) Probabilistic forecasts of wind speed: Ensemble model output statistics using heteroskedastic censored regression. *Journal of the Royal Statistical Society, Series A*, 173:371–388.

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

Vincent C, Giebel G, Pinson P, Madsen H. (2010) Resolving nonstationary spectral information in wind speed time series using the Hilbert-Huang transform. *Journal of Applied Meteorology and Climatology*, 49:253–269.

Vincent C. (2011) Mesoscale wind fluctuations over Danish waters. PhD Thesis, Risø - Technical University of Denmark, Roskilde, Denmark (ISBN 978-87-550-3864-6).

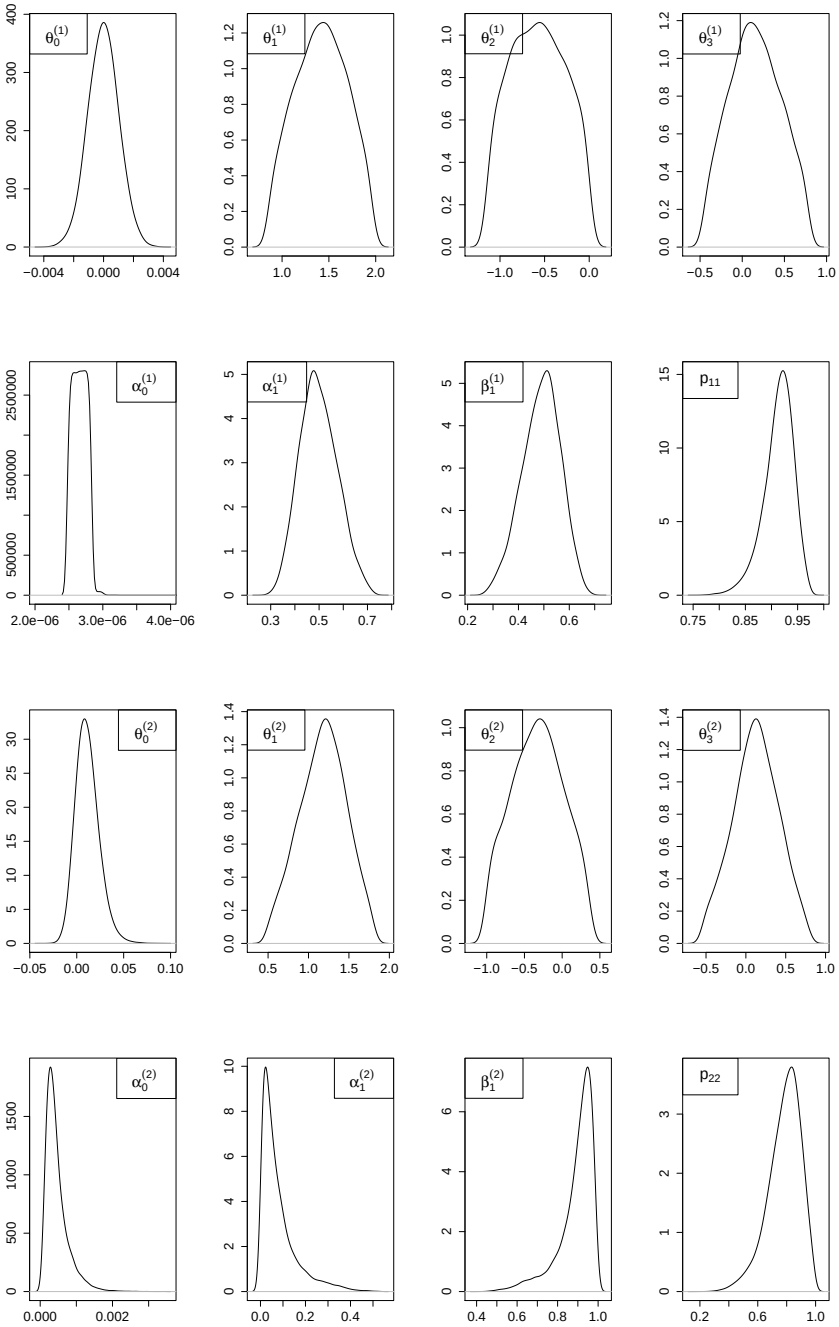


Figure 6: Estimated posterior densities of the MS(2)-AR(3)-GARCH(1,1) model fitted to the time series of wind power.

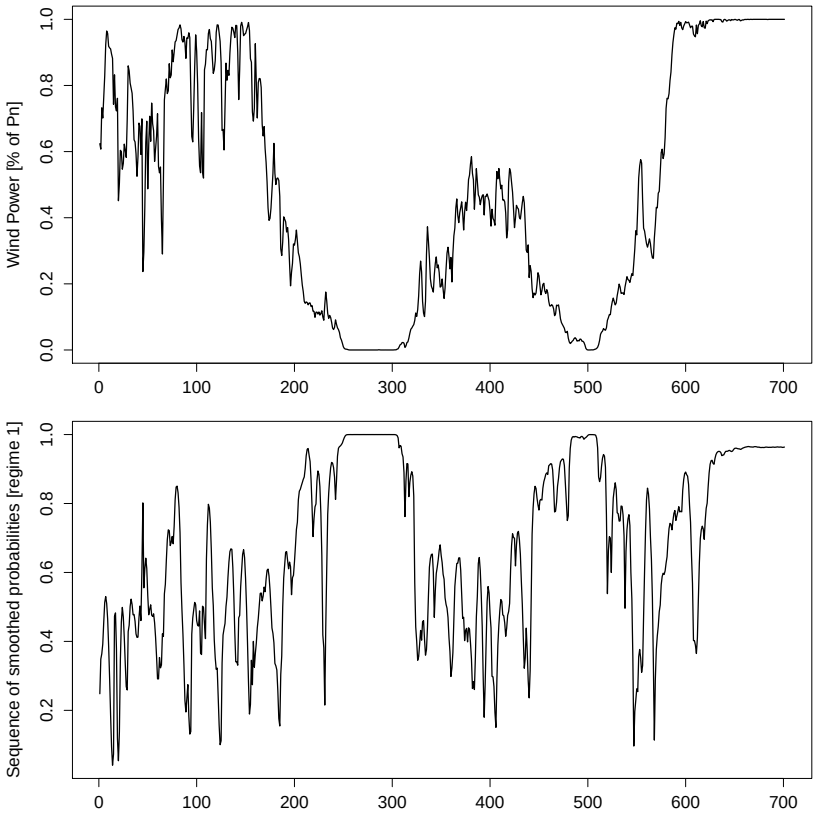


Figure 7: Time series of wind power and estimated sequence of smoothed probabilities of being in regime 1 (i.e., low volatility regime).

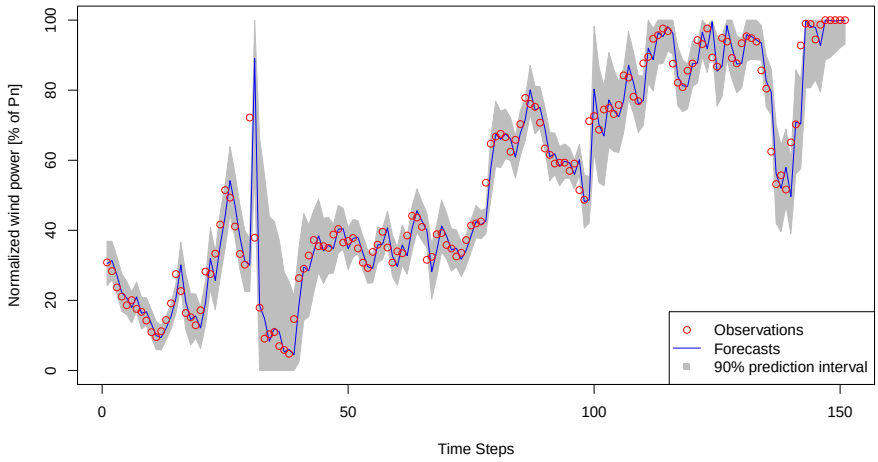


Figure 8: Example of time series of normalized wind power generation (red dots) along with one step-ahead forecasts (blue line) and the prediction interval of 90% coverage rate (shaded area in gray) defined with the two quantiles with nominal proportions 5 and 95%. The forecasts were generated with a MS(2)-AR(3)-GARCH(1,1) model.

PAPER B

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

---

Authors:
P.-J. Trombe, P. Pinson

Published in:
IMM Technical Report 2012-15

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

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

1 Introduction

With the growing penetration of wind power into power systems, electric utilities are called to revise their operational practices. In particular, experts in energy management recommend to increase the scheduling frequency of electricity generation and delivery from hours to minutes, in order to mitigate the impact of wind power variability on power systems (GE Energy, 2010). Transmission System Operators (TSO) expressed concurring views on the integration of large amounts of wind power into power systems (Jones and Clark, 2011). In a few European countries, very short-term wind power forecasts with temporal resolutions from 5 to 15 minutes, and lead times up to 36-48 hours, are already used in a wide range of applications (Holttinen et al., 2011). These include among others optimizing reserve allocation, balancing electricity consumption and production, and controlling wind power fluctuations at large offshore wind farms (Akhmatov et al., 2007, Kristoffersen and Christiansen, 2003). In particular, one application for which forecasts with specific lead times up to 15-20 minutes are needed is the management of the immediate regulating power reserve. This type of reserve is activated over time intervals up to 15-20 minutes, after the system experiences a sudden and large deviation between scheduled and actual wind power generation (Akhmatov, 2007). This issue is paramount in countries or regions with limited interconnections, or with no complementary source of energy (e.g., hydro or pumped hydro) that can be both stored and used for fast-acting generation.

Issuing improved wind power forecasts for supporting decision-making in regulating reserve management has the merit of being more cost-effective when compared to other solutions such as increasing backup capacities. For lead times from a few minutes to a few hours, wind power forecasts are best generated with statistical models using historical data. However, developments in wind power forecasting have long been oriented towards energy market applications, placing focus on forecasts at hourly resolutions, as required by the

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

market structure. These approaches heavily rely on the availability of meteorological forecasts of wind speed and direction owing to the strong relation between wind and wind power, the so-called power curve (Giebel et al., 2011). Employing such a strategy is not realistic when working with lead times of a few minutes. Instead, a number of new modeling and forecasting approaches were recently proposed in view of improving the predictability of wind power fluctuations for very short lead times. These include regime-switching models, off-site predictors and a new type of predictive distribution.

Regime-Switching models – The motivation for applying these models comes from the existence of structural changes in the dynamics of wind power fluctuations at temporal resolutions of a few minutes, hence the term *wind power regime*. Periods of low and high wind power variability alternate, not only modulated by the wind own variability, but also by the power curve that amplifies or dampens wind fluctuations owing to its nonlinear nature. For low or high wind speeds, wind power fluctuations are very small whereas, for moderate wind speeds (i.e., roughly between 7 and 13 $\mathrm{m.s}^{-1}$), wind power fluctuations can become extreme. Originally developed for applications in Econometrics (Tong, 1990), regime-switching models have, since then, also been applied for modeling and forecasting offshore wind power fluctuations in Pinson et al. (2008), Gallego et al. (2011), Trombe et al. (2012), improving the accuracy of wind power forecasts when compared to single regime models. Regime-switching models divide into two categories, those for which regimes are observable and determined by expertise, and those for which they are unobservable and estimated jointly with the model. This translates into two classes of time series models, namely Threshold Autoregressive (TAR) and Markov-Switching Autoregressive (MSAR) models (Tong, 1990, Fruhwirth-Schnatter, 2006).

Off-site predictors – Traditional inputs to statistical prediction models consist of on-site observations (i.e., wind power production, wind speed and direction) and/or meteorological forecasts (wind speed and direction, temperature, atmospheric pressure). However, meteorological forecasts are generated at coarse temporal resolutions, from 1 to 3 hours, and therefore not informative on intra-hour wind fluctuations. Furthermore, wind measurements are rarely available in real-time for applications with lead-times of a few minutes. When wind power data and wind data are not simultaneously available, the difficulty of generating accurate wind power forecasts increases. This is the reason why a number of recent studies explored the potential of off-site observations as new predictors (Alexiadis et al., 1999, Damousis et al., 2004, Gneiting et al., 2006, Larson and Westrick, 2006, Hering and Genton, 2010, Tastu et al., 2010, Lau, 2011). In particular, wind farms and meteorological masts scattered over a region form a net capable of capturing valuable information on the weather conditions. Owing to the synoptic mechanisms in the atmosphere which drive

wind variability in space and time, upwind observations can be informative of upcoming changes in weather conditions and be used as extra predictors (Tastu et al., 2011, Girard and Allard, 2012). Two distinct approaches exist for integrating these off-site predictors into forecasting models, depending on whether (i) the dominant weather conditions are known a priori and the model designed accordingly (Alexiadis et al., 1999, Damousis et al., 2004, Gneiting et al., 2006, Larson and Westrick, 2006), or (ii) there is no a priori information available on weather conditions and it is assumed that the model can capture the associated effects directly from the data (Tastu et al., 2010, Lau, 2011, Hering and Genton, 2010). Despite their high accuracy, models based on the first type of approach have a clear downside, they tend to be very region or site-dependent, lacking of adaptivity when applied to areas with different weather conditions. In contrast, models based on the second type of approach are more data-driven and require less expert knowledge to capture the spatio-temporal dependencies between sites.

The Generalized Logit-Normal distribution – Wind power generation is a double-bounded process since it can neither be negative nor exceed the wind farm rated capacity. In addition, the distribution of wind power forecast errors changes with respect to the conditional expectation of the forecasts (Lange, 2005). In particular, heavy skewness near the bounds and a clear heteroscedastic behavior are generally observed. In a parametric framework, a common approach for dealing with these features consists in combining a statistical model that handles the heteroscedasticity (e.g., Generalized Autoregressive Conditional Heteroscedastic (GARCH) models) with a predictive distribution that deals with the effects of the bounds and, potentially, with skewness (e.g., censored and truncated Normal distributions) as in Lau and McSharry (2010). A generalization of this type of approach was proposed in Pinson (2012) with the Generalized Logit-Normal (GLN) distribution and applied for forecasting wind power fluctuations at large offshore wind farms.

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

evaluate whether additional improvements can be obtained. Focus is placed on wind power fluctuations from a single wind farm.

Wind power forecasts and, more generally, forecasts of any continuous quantity are given in the form of either a single-value (i.e., deterministic forecast) or a full probability distribution or density (i.e., probabilistic forecast). As pointed out in Gneiting (2008), forecasts ought to be probabilistic in order to achieve optimal decision-making under uncertainty. This idea found its echoes with a few TSOs which started using probabilistic information in control rooms (Jones and Clark, 2011). In this work, the accuracy of wind power forecasts is verified with respect to both point and density forecasts even though more importance will be given to the latter ones.

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

2 Data and their characteristics

In this section, we present the data and their characteristics. We also perform a number of analysis to introduce some essential principles that motivate modeling assumptions in section 3. In particular, we give a detailed account on the the GLN predictive distribution as proposed in Pinson (2012), and evaluate spatio-temporal correlations of wind power in view of integrating off-site predictors into time series models.

2.1 Case study

The case study consists of a group of three wind farms located in the South-East of Ireland, the Carnsore wind farm which has a rated capacity (Pn) of 11.9 MW and its two nearest wind farms, Richfield (27 MW) and Ballywater (42 MW), as shown in Figure 1. Ballywater and Richfield are located about 40 km North-East and 17 km West of Carnsore, respectively. The Carnsore wind farm is located at the extreme point of a peninsula, by the sea shore. Richfield and Ballywater are located further away inland but within 5-10 km from the sea, remaining in the zone of influence of the marine weather. In this study, focus is placed on forecasting the wind power generation at the Carnsore wind farm. As aforementioned, no wind measurement is available. Furthermore, available

meteorological forecasts have a too coarse temporal resolution to be informative for lead times of a few minutes and thus cannot be used. Our knowledge of weather conditions in Ireland is restricted to the prevalence of southwesterly winds. In addition, passages of low-pressure systems characterized by large wind variability and developments of storms are more frequent over the period from August to January (Met Eireann, 2012).

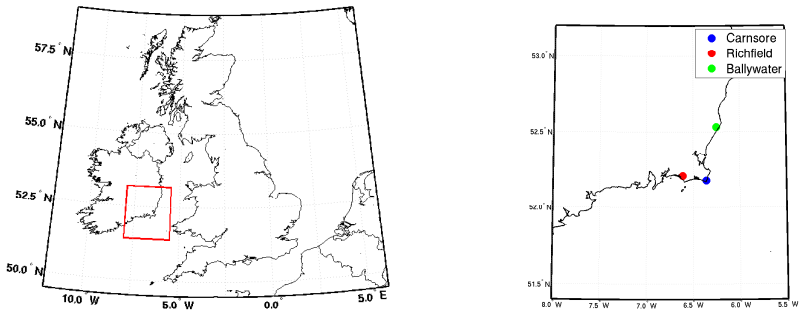


Figure 1: The Carnsore, Richfield and Ballywater wind farms are located in the South-East of Ireland. Carnsore and Richfield are separated by an approximate distance of 17 km, and the distance between Richfield and Ballywater is 40 km.

Ireland and its power system are singular when compared to other countries/regions with high wind power penetrations. Ireland has large wind resource but very limited interconnection capacity with power systems from other countries. More specifically, there exists a single interconnection to Northern Ireland which, in turn, is only connected to the United Kingdom. The target of Ireland is to meet 40% of its energy demand with renewable energy sources by 2020, of which 37% are expected to be covered by the integration of wind power. The small interconnection capacity clearly acts a limiting factor for enabling further wind power into the system since the latter will be unable to spill excess power when needed. Consequently, improved wind power predictability would allow to decrease the frequency of curtailment actions and reduce losses of wind power generation (Holttinen et al., 2011).

2.2 Data quality control

The wind power data used in this study are provided by Eirgrid, the TSO in Ireland. They span the period from December 31, 2006 to June 1, 2009. One time series of wind power production is available for each wind farm, at a

temporal resolution of 15 minutes. Following Madsen et al. (2005), time series are normalized and expressed as a percentage of the wind farm rated capacity. The resulting time series take values on the unit interval $[0,1]$. The raw data records are complete for Carnsore and Richfield but not for Ballywater for which 3071 values (out of 84864) are reported missing. Since the data consist of output power time series, and not available power, a data quality control is performed. We identify several periods where the output power is curtailed, likely indicating that some wind turbines were temporarily out of order or that an absolute power limitation was imposed. An example is given in Figure 2 which shows the time series of wind power for the Carnsore wind farm. The output power never exceeds 92% of the rated power of Carnsore in the second semester of 2007 and the first semester of 2008. Consequently, we only use the period from July 10, 2008 to 27 March, 2009 in this study, corresponding to more than 25000 data points. This period is shaded in grey in Figure 2.

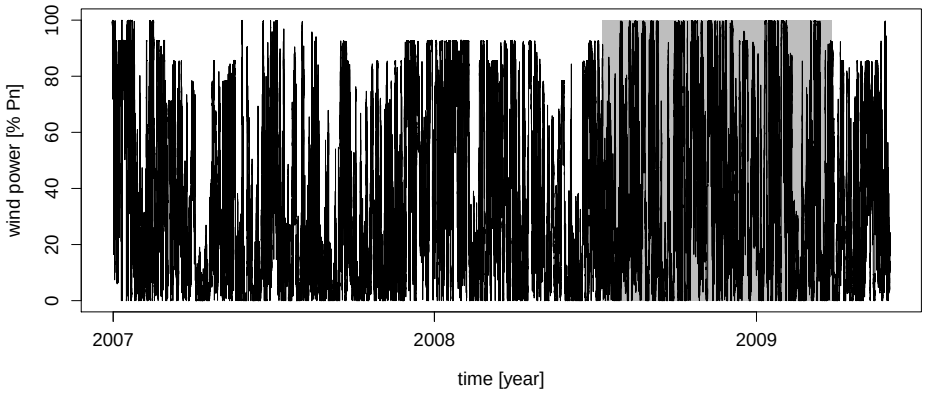


Figure 2: Time series of wind power at Carnsore. The data overlaying the shaded area are considered to be of good quality and used for the experimental part of this study.

2.3 The Generalized Logit-Normal predictive distribution

The conversion from wind to power makes that wind power generation is a double-bounded process, with a potentially high concentration of observations near or at the bounds. This feature is illustrated in Figure 3. In addition, the shape of the distribution of the wind power forecast errors evolves with the conditional expectation of the forecasts. Near the bounds, the conditional distribution of wind power forecast errors tends to have a very small standard deviation and to be heavily skewed. Moving away from these bounds, the standard deviation increases and the skewness decreases (Lange, 2005). When forecasting wind power generation from single wind farms, designing an appro-

priate strategy for taking these features into account is paramount. In Pinson (2012), the author proposed the use of the Generalized Logit-Normal (GLN) distribution. The underlying motivation for using this distribution comes from the work of Box and Cox (1964) where it is shown that appropriate data transformations may enhance characteristics such as linearity, homoscedasticity and additivity.

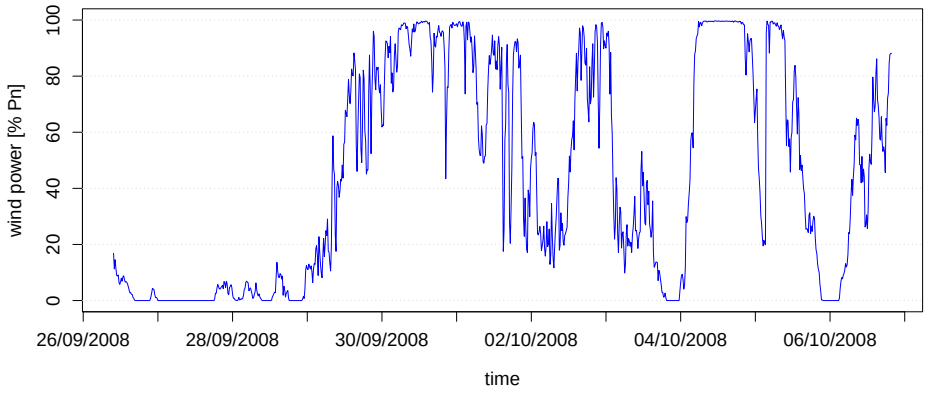


Figure 3: Normalized wind power generation at Carnsore. The temporal resolution of 15 minutes.

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

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

where ν is a shape parameter and the resulting time series $\{\widetilde{y}_t\}$ takes values in $]-\infty, +\infty[$. This transformation, as shown in Figure 4 for a set of different values of ν, aims at outstretching the distribution near the bounds of the interval $[0,1]$. In the original domain $[0,1]$, the assumption of homoscedastic wind power forecast errors does not hold and, even though one may argue that this may still not be the case after transforming the time series, making that such assumption is clearly more appropriate in the transformed domain than in the original one.

However, the concentration of observations at the bounds, in 0 and 1, generates two probability masses that remain in the transformed domain. They are located in $-\infty$ and $+\infty$, respectively. To fix this, the coarsening principle is applied as in Lesaffre et al. (2007). All observations taking values in the open interval $]-\infty, \gamma(\epsilon,\nu)[$ are shifted to $\gamma(\epsilon,\nu)$. Likewise, all observations taking values in $]\gamma(1-\epsilon,\nu), +\infty[$ are shifted to $\gamma(1-\epsilon,\nu)$, with $\epsilon < 0.01$. Two Dirac

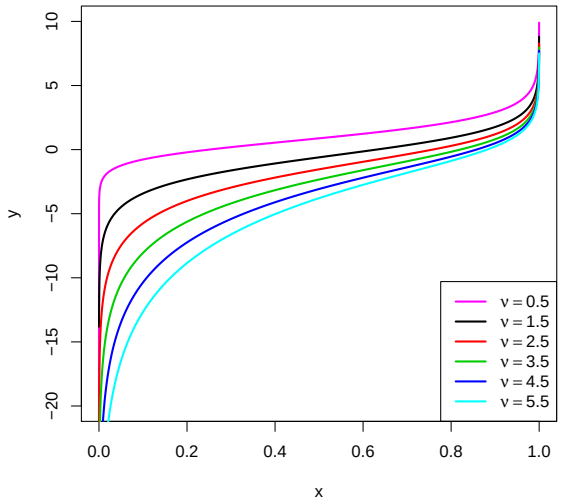


Figure 4: The GLN predictive distribution consists of transforming the original wind power observations in order to deal with the heteroscedasticty near the bounds of the interval $[0,1]$

distributions $\delta_{\gamma(\epsilon,\nu)}$ and $\delta_{\gamma(1-\epsilon,\nu)}$ are introduced so that the one-step ahead predictive distribution in the transformed domain, $Y_{t+1|t}$, is defined as follows:

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

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

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

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

2.4 Spatio-temporal correlations in wind data

Recent studies showed that it was possible to take advantage of spatio-temporal correlations in wind data at an hourly resolution in order to improve the predictability of wind speed or wind power at regional scales (Gneiting et al., 2006, Larson and Westrick, 2006, Hering and Genton, 2010, Tastu et al., 2010). Nevertheless, for higher temporal resolutions, in the order of a few minutes, the wind

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

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

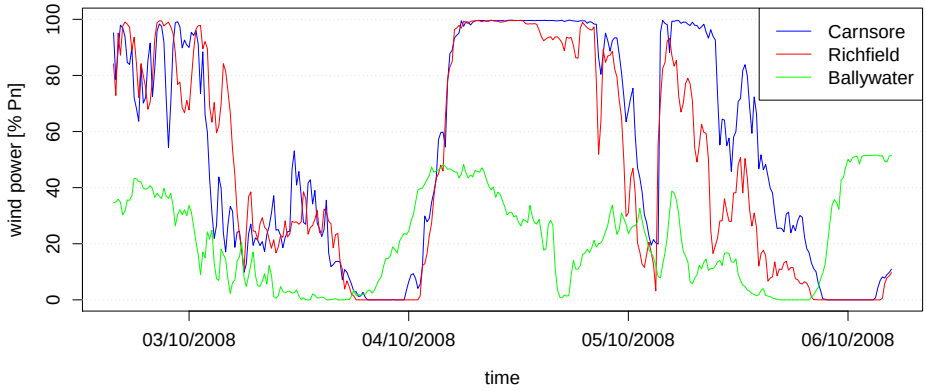


Figure 5: Normalized wind power generation at Carnsore, Richfield and Ballywater with a temporal resolution of 15 minutes.

Before using off-site observations for prediction applications, it is essential to analyze correlations between wind data from distant sites. Following Girard and Allard (2012), we assume that these correlations can appropriately be described and quantified by the traditional linear correlation coefficient. In order to evaluate these correlations, we use the pre-whitening technique presented in Madsen (2008). Let A and B be two wind farms, with their respective time series of wind power generation $\{y_t^{(A)}\}$ and $\{x_t^{(B)}\}$. $\{x_t^{(B)}\}$ is called the input

series and $\{y_t^{(A)}\}$ the output series. The idea is to use the power generation from wind farm B as input for improving the wind power predictability of wind farm A. The procedure is divided into three steps as follows:

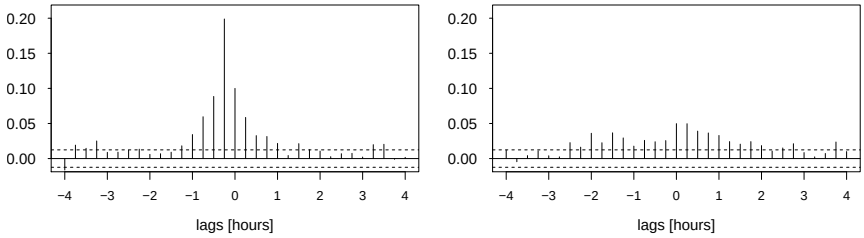
1. An appropriate Autoregressive Moving Average (ARMA) model is fitted to the input series $\{x_t^{(B)}\}$ and a series of residuals $\{e_t^{(B)}\}$ extracted,

2. The output series $\{y_t^{(A)}\}$ is filtered with the same model as in step 1 and a series of residuals $\{e_t^{(A)}\}$ extracted,

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

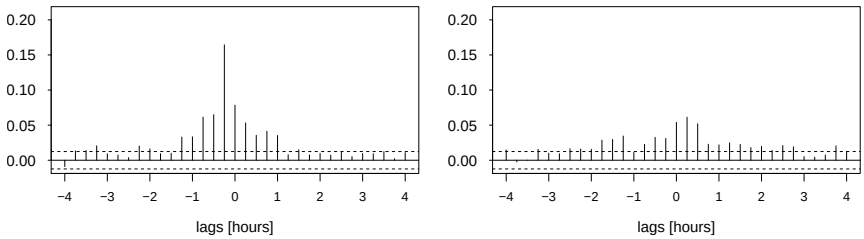
$$\rho_{e^{(A)}e^{(B)}}(\tau) = \frac{\mathrm{cov}(e^{(A)}(t), e^{(B)}(t+\tau))}{\sigma_{e^{(A)}}\sigma_{e^{(B)}}} \tag{5}$$

We repeat the pre-whitening procedure presented hereabove with and without the GLN transformation as given by equation (1) in order to evaluate how this transformation changes the correlation structure between the power generation from two wind farms. The results are reported in Figure 6. Negative lags indicate that wind power fluctuations at Richfield or Ballywater lead those at Carnsore. First, these results reveal larger cross-correlations between Richfield and Carnsore than between Ballywater and Carnsore, thereby confirming the visual observations made from Figure 5. This result is most likely the consequence of the shorter distance separating Carnsore from Richfield than from Ballywater which would be consistent with the empirical analysis in Girard and Allard (2012) where spatio-temporal correlations are shown to quickly decrease within a radius of 50 km. Figure 6 also shows that wind power fluctuations at Richfield tend to lead those at Carnsore up to 30 minutes ahead, on average. In contrast, cross-correlations between Ballywater and Richfield are much lower and it appears more difficult to determine a clear tendency on whether wind power fluctuations propagate preferentially from Richfield to Ballywater, or the opposite. A direct extrapolation from these cross-correlations suggests that off-site observations from Richfield have a higher potential for improving wind power predictability at Carnsore than corresponding observations from Ballywater. Finally, one can see that cross-correlations between Carnsore and Richfield are larger without applying the GLN transformation a priori. Cross-correlations tend to decrease with large values of the shape parameter ν. We can think of two potential causes that explain this feature. First, using the GLN distribution may degrade the linear relationship between the two time series, particularly near the bounds where the respective variances may increase. Secondly, using the GLN distribution may enhance the homoscedasticity of the input time series $\{x_t^{(B)}\}$ so that the residuals series



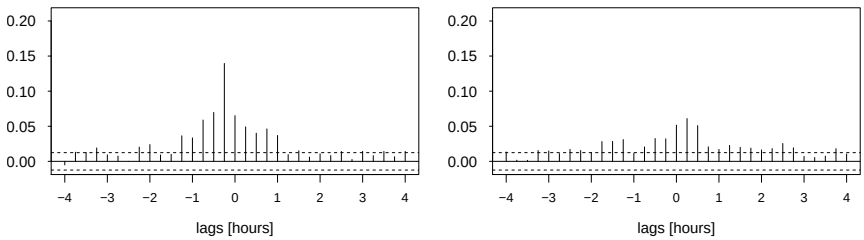
(a) Richfield & Carnsore, No GLN transf.

(b) Ballywater & Carnsore, No GLN transf.



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

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



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

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

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

$\{e_t^{(B)}\}$ is closer to being a white noise process, and thereby is less informative.

3 Time series modeling

The stochastic nature of wind power generation is described hereafter with time series models. We start by considering linear models (i.e., ARX and ARX-

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

3.1 ARX models

While it is generally acknowledged that wind power generation is a nonlinear process, operational wind power forecasting systems usually rely on linearity assumptions (Nielsen et al., 2007). ARX models are some of the most widely used in practice. There are several reasons for this. First, their formulation is very intuitive and simply consists of a linear combination of lagged variables which leads to fast estimation procedures. Secondly, they stand as very competitive models for generating point forecasts owing to their parsimony (i.e., low number of parameters to be estimated). Thirdly, there exists closed-form formula for generating multi-step ahead forecasts (Madsen, 2008).

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

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

where

$$\boldsymbol{\theta} = [\theta_0, \theta_1, \ldots, \theta_p] \tag{7}$$

$$\boldsymbol{\psi} = [\psi_{r_B}^{(B)}, \ldots, \psi_{s_B}^{(B)}, \psi_{r_C}^{(C)}, \ldots, \psi_{s_C}^{(C)}, \ldots] \tag{8}$$

$$\boldsymbol{Y_t} = [1, y_{t-1}^{(A)}, \ldots, y_{t-p}^{(A)}, x_{r_B}^{(B)}, \ldots, x_{s_B}^{(B)}, x_{r_C}^{(C)}, \ldots, x_{s_C}^{(C)}, \ldots]^T \tag{9}$$

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

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

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

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

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

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

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

In order to obtain the one-step ahead GLN density, additional steps are needed. First, the transformation given in (1) must be applied for estimating the vector of parameters $\widehat{\Theta}_{MLE}$ in the transformed domain. Second, the one-step ahead predictive density in the transformed domain is obtained by following the formula (2-4). Last, the inverse GLN transformation presented in Pinson (2012) is applied on a quantile per quantile basis for generating the GLN density in the original domain.

3.2 ARX-GARCH models

ARX-GARCH models are a popular extension of ARX models as they can relax the assumption of constant variance without data transformation. GARCH models were first introduced in Econometrics by Bollerslev (1986). A short review of meteorological applications of GARCH models is available in Trombe et al. (2012). This class of model proposes to capture the dynamical structure of the conditional variance, jointly to that of the process conditional expectation. The conditional variance h_t^2 is modeled as an ARMA process for the squared errors ε_t^2. It was shown in a number of studies that a GARCH(1,1) structure is in most cases appropriate to capture the temporal dynamics of h_t^2. The linear AR(p)-X(q)-GARCH(1,1) model with p autoregressive and q exogenous predic-

tors is given by:

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

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

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

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

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

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

where ε_t is given by (12) and h_t^2 is given by (14). For the implementation of the model, analytical formula for the first and second order derivatives of the negative log-likelihood function are given in (Fiorentini et al., 1998).

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

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

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

3.3 TARX models

TARX models are the first regime-switching models considered in this study. They are piecewise linear, and the transitions between regimes are governed in a deterministic way by a lagged variable, and are hence observable. See (Tong, 1990) for a more detailed introduction to these models. The TAR($p_1, \ldots, p_R$)-X($q_1, \ldots, q_R$) model with R regimes, p_j autoregressive and q_j exogenous predictors in regime j, with $j = 1, \ldots, R$, is given by:

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

where

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

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

and $\{\varepsilon_t\}$ is an i.i.d sequence of random variables with 0 mean and unit variance, $\sigma^{(j)}$ the standard deviation in the regime j, z_{t-d} the lagged variable; $d \in N^+$ the delay parameter with usually $d \leq \max(p_1, \ldots, p_R)$, and r_j the threshold values separating the regimes. The regime-switching effect translates into the autoregressive and exogenous coefficients as well as the standard deviation of the error term being state-dependent. Applications of TAR models for forecasting wind power fluctuations can be found in Pinson et al. (2008), Gallego et al. (2011) which alternatively use lagged observations of wind speed, wind direction or wind power for controlling transitions between regimes. A special class of TAR model is the Self-Exciting TAR (SETAR) model which corresponds to the case where the dependent variable is chosen as the lagged variable.

The major issue with TAR models is the joint determination of the delay d and thresholds $r_j, j = 1, \ldots, R$. In particular, the most spread technique for the determination of the r_j is based on the visual assessment of scatter plots of t-ratios (see Tong (1990)). In order to fill in the lack of consistency of such approach, an automated procedure for determining the number of regimes and threshold values of TAR models was recently proposed in Bermejo et al. (2011). It consists of detecting jumps in the values of the estimates of an arranged autoregression by using a recursive least squares (RLS) estimation method. This method can be extended to deal with exogenous predictors without complicating its procedure. Once the threshold values known, the parameters for a given regime can be estimated independently of the parameters of the other regimes by applying the formula given in formula (10-12) for each regime, and predictive densities can be generated as with ARX models.

3.4 MSARX models

MSARX models are the second type of regime-switching models in this study. Structurally, the major difference between MSARX and TARX models lays in the way the sequence of regimes is determined. With TAR models, this sequence is determined explicitly by a lagged variable, and the transitions between regimes are therefore discontinuous. With MSARX models, the sequence is assumed hidden and estimated directly from the data. More specifically, MSARX models assume that an unobservable Markov process governs the distribution of the observations (Fruhwirth-Schnatter, 2006). This enables smooth transition between regimes.

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

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

where

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

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

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

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

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

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

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

where

$$\mathcal{L}(\Theta | \Omega_{\boldsymbol{T}}) = \boldsymbol{\delta} \left(\prod_{t=1}^{n} \boldsymbol{P} \boldsymbol{D}_t \right) \mathbf{1}^T \tag{28}$$

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

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

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

where $\boldsymbol{\delta}$ is the stationary distribution of the Markov chain, $\mathbf{1}$ is a unit vector of size R, $\boldsymbol{I}_R$ and $\boldsymbol{U}_R$ Identity and Unity matrices of size $R \times R$, $\boldsymbol{D}_t$ a diagonal matrix and ϕ the probability density function of the Normal distribution. Practical solutions for the implementation of MSARX models are given in Zucchini and MacDonald (2009).

With MSARX models, predictive densities take the form of mixture of densities (Fruhwirth-Schnatter, 2006, Zucchini and MacDonald, 2009). For the case where the errors are Normally distributed in each regime, the resulting predictive density is a mixture of R Normal densities that is censored in 0 and 1 later on. At time t, given the vector of estimated parameters $\widehat{\Theta}_{MLE}$ and the set of

observations $\Omega_{\boldsymbol{t}}$, the one-step ahead density can be obtained as follows:

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

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

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

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

3.5 Estimation procedure

As mentioned in section 2, the data we selected cover the period from July 10, 2008 to 27 March, 2009. This corresponds to about 25000 observations, for each of the three time series (i.e., Carnsore, Ballywater, Richfield). Focus is placed on predicting the wind power generation at the Carnsore wind farm. The first 15000 observations are used for fitting the models. The following 5000 observations are used for performing a one-fold cross-validation and determining the optimal parametrisation of each model. The last 5000 observations, corresponding to about 63 days, are kept for forecast evaluation.

Cross-validation is jointly performed on the structure of the model (i.e., selection of the optimal AR lags from 1 up to 8, and X lags from 1 to 5, number of regimes R) and a set of values for the shape parameter ν of the GLN distribution (from 0.1 to 3.1 with steps of 0.1). Because of that, and because the likelihood function is unbounded, neither the respective goodness-of-fit nor the predictive power of the models can be compared with respect to likelihood based scores. Instead, the cross-validation procedure is performed by minimizing the Continuous Ranked Probability Score (CRPS) for one-step ahead density forecasts. The CRPS quantifies the accuracy of conditional density forecasts based on two principles: calibration (i.e., the relative position of a forecast with respect to the observed value) and sharpness (i.e., the concentration of the predictive distribution around the observed value) (Gneiting et al., 2007).

For each class of models presented in this section, we estimated four different models with: (N) a censored Normal distribution, (X-N) a censored Normal

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

4 Experimental results and forecast evaluation

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

4.1 Point forecasts

Electric utilities have a long tradition of using point or deterministic forecasts of wind power (Jones and Clark, 2011, Giebel et al., 2011). In this study, point forecast accuracy is evaluated with respect to the Normalized Mean Absolute Error (NMAE). There is an inverse relationship between point forecast accuracy and the NMAE score: the lower the NMAE, the better. Following Gneiting (2011), we use the median of the predictive densities as the optimal point forecast, due to the nature of the NMAE which is based on a symmetric piecewise linear scoring rule. All models are benchmarked against Persistence since it is one of the most competitive benchmarks for such short lead times. Persistence

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

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

usually outperforms other common benchmarks such as Climatology, Moving average or Constant forecast (see for instance Pinson (2012), Lau (2011)) which are not included here. It is an Autoregressive model of order 1 with no intercept term and its coefficient value equal to 1. Point forecast results are given in Table 2. It is interesting to note that not all models outperform Persistence and that even the largest improvement does not exceed 3%. Overall, MSARX and ARX-GARCH with a GLN distribution give the best results. When considering each class of models independently of the others, we observe two trends. The first one concerns AR and TAR models for which the use of either off-site information or the GLN distribution yields substantial gains in wind power predictability. These gains are further improved by using both. The second trend regards AR-GARCH and MSAR models for which the use of the GLN distribution alone, without off-site information, leads to negligible gains whereas the opposite (i.e., no GLN distribution and off-site information) leads to appreciable gains.

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

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

4.2 Density Forecasts

Forecasts of any quantity contain an inherent part of uncertainty. Supplying information on this uncertainty is paramount for developing efficient decision-making strategies, as shown in the context of wind power trading by Pinson et al. (2007). Here, information on this uncertainty is provided in the form of full predictive densities of wind power, for all four classes of models. The accuracy of these densities is assessed with respect to the Normalized CRPS (NCRPS). This score is a generalization of the NMAE score for probabilistic forecasts and measures the difference between the observed cumulative distribution functions and those predicted (Gneiting et al., 2007). It can be interpreted in a similar way as the NMAE, meaning the lower the NCRPS the better. All models are benchmarked against an AR model with a censored Normal distribution (AR-N). Results for one-step ahead densities are reported in Table 2. The best result is given by the MSAR model with off-site information and the use of the GLN distribution (MSAR-X-GLN), with a relative improvement of

almost 20% when compared to an AR-N model. In addition, we observe a common trend across all four classes of models when considered independently of the others. Their ranking is dominated by models including both off-site observations and the GLN distribution (X-GLN), then come models specified with the GLN distribution and no off-site predictors (GLN), then models with off-site predictors but without GLN distribution (X), and finally models with neither the GLN distribution nor off-site predictor (N).

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

4.3 Forecast reliability

The CRPS is a global score that averages the predictive accuracy of conditional densities based on their calibration and associated sharpness. However, it is not informative on the behavior of these densities in terms of probabilistic reliability. Reliability measures how well the predicted probabilities of an event correspond to their observed frequencies. For instance, one may want to measure the proportion of observations actually lower than the 5^{th} percent quantile or larger then the 95^{th} percent quantile for evaluating the ability of the predictive density tails in predicting extreme or rare events. In this study, the reliability of the predictive densities of wind power is evaluated with four reliability diagrams as shown in Figure 9. These diagrams are generated for each of the four classes of models by comparing the nominal (i.e., theoretical) proportions of a set of quantiles with the observed proportions of the same set. Here, we used 19 quantiles, from the 5^{th} percent quantile to the 95^{th} percent quantile with a step of 5^{th} percent. The best reliability is given by the model whose diagram is closer to the ideal case in Figure 9, that is the MSAR-X-GLN model.

4.4 Discussion

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

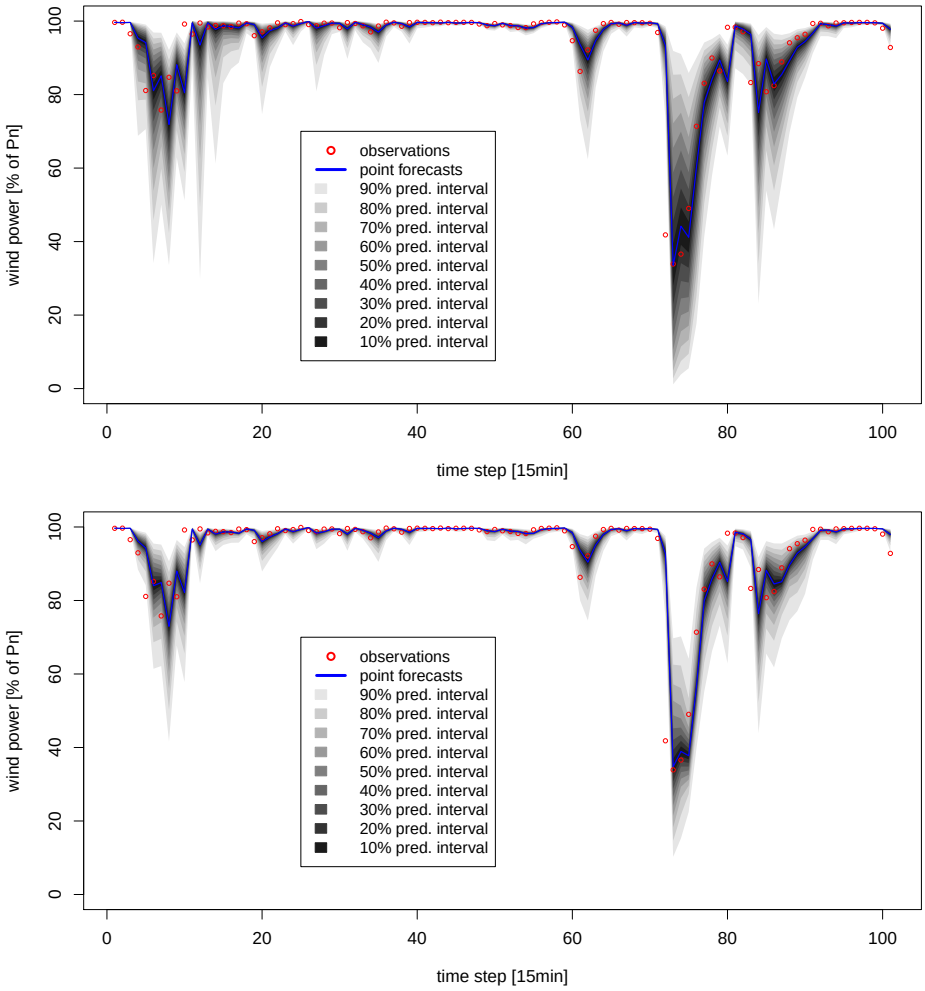


Figure 7: Example 1. Time series of normalized wind power generation at Carnsore and one-step ahead point forecasts and prediction intervals with nominal coverage from 10 to 90%. ARX-GARCH-GLN model (Top panel), MSAR-X-GLN model (Bottom panel).

1. In the Irish case study chosen for this work, the variability of wind power fluctuations can be considered as extreme. For instance, the NMAE value of the Persistence is about 50% larger than that at the Horns Rev 1 wind farm where wind power fluctuations are known to be characterized by a high variability (Pinson et al., 2008, Trombe et al., 2012). In that sense, this case study offered a difficult test to all models, enhancing the impact

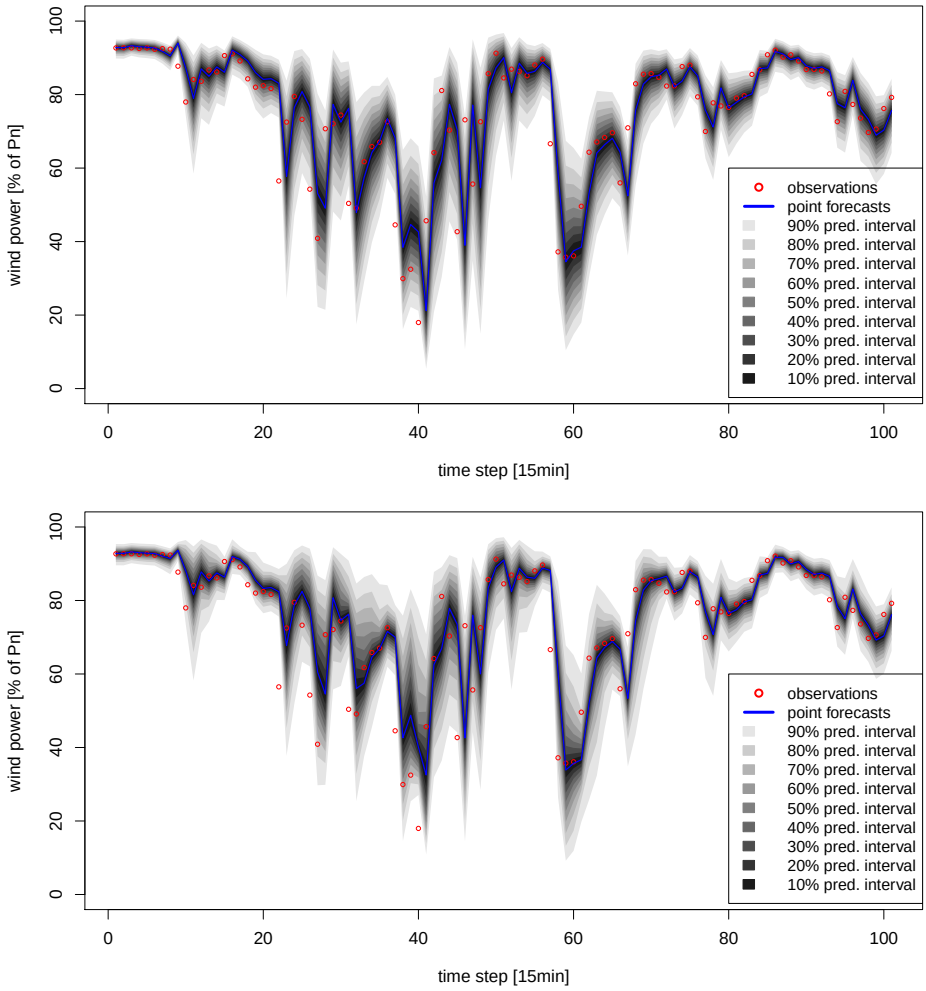


Figure 8: Example 2. Time series of normalized wind power generation at Carnsore and one-step ahead point forecasts and prediction intervals with nominal coverage from 10 to 90%. ARX-GARCH-GLN (Top panel) model, MSARX-GLN model (Bottom panel).

of the results obtained.

2. Irrespectively of the availability of off-site measurements, the use of the GLN distribution is recommended for very short-term forecasts. In particular, it enables an improved modeling of the heteroscedastic behavior of wind power time series, which translates to substantial gains in pre-

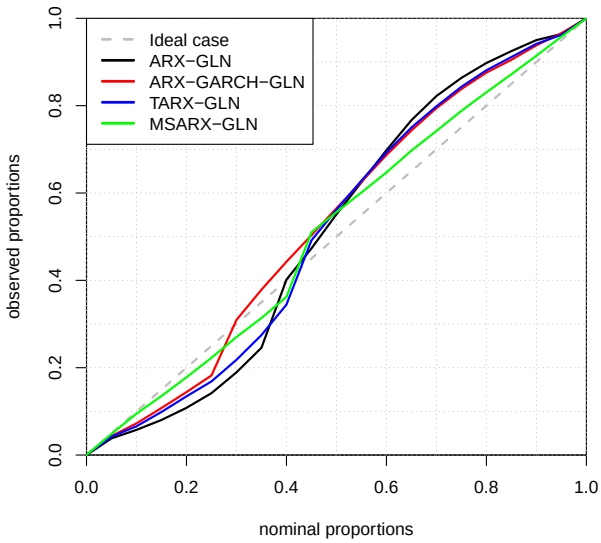


Figure 9: Reliability diagram of predictive densities of wind power.

dictability even for models already explicitly accounting for heteroscedasticity in their formulation (i.e., MSARX and ARX-GARCH) . However, it calls for further research on its potential for multi-step ahead forecasts. This issue was not addressed here but will be investigated in the future. In addition, focus should be placed on developing a more consistent framework than cross-validation for estimating the optimal value of the shape parameter ν of the GLN distribution. For instance, the estimation of ν could be performed jointly with the estimation of the model via the Expectation-Maximization (EM) algorithm (see Dempster et al. (1977)).

3. The results obtained with TARX models are relatively disappointing, particularly, when analyzed from a perspective including the cost complexity of these models and the level of expertise required to tune them. It is also worth noting that TAR models are outperformed by linear in mean ARX-GARCH models. It could be expected that TARX models perform much better for point forecasting especially in combination with the GLN distribution since the introduction of regimes via the thresholds could reduce the strong influence of the probability masses in $\gamma(\epsilon,\nu)$ and $\gamma(1-\epsilon,\nu)$ on the autoregressive coefficient estimates.

4. Density forecasts of wind power generated with Markov-Switching models have superior calibration and sharpness when compared to those generated with other models in this study. Beyond this result, it is important to stress the underlying assumption in MSAR models which leads to such

result, that is the existence of an unobservable regime sequence which governs the wind power generation. As of today, our knowledge is limited and we can only assume that the estimated regime sequence is linked to some weather regime. Therefore, it would be useful to investigate the use of data (e.g., quick scan satellite images, weather radar images) that can describe weather conditions over large spatial areas and high temporal resolutions for improving the characterization of this regime sequence.

5 Conclusion

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

Through an application with three wind farms in Ireland, we show that regime-switching models for which the sequence of regime is unobservable (i.e., Markov-Switching) generate more accurate point forecasts, better calibrated and sharper conditional densities, than single regime or other regime-switching models for which the regimes are observable. Furthermore, gains in wind power predictability can be increased by taking advantage of off-site information when available or using a more appropriate predictive distribution such as the GLN distribution, as introduced in Pinson (2012). The highest gains were obtained by using simultaneously off-site observation and the GLN distribution.

The superior predictive power of Markov-Switching models is interesting in two aspects. First, because this type of models is rather generic and thus non site-dependent, requiring very little expert knowledge to be tuned. It confirms the potential shown for offshore applications (Pinson et al., 2008, Trombe et al., 2012). Second, because Markov-Switching models assume the existence of an unobservable regime sequence that can be interpreted as a hid-

den weather regime. This indicates that substantial gains in wind power predictability could be obtained by integrating more meteorological data at high spatio-temporal resolution such as satellite images, weather radar images, or meteorological forecasts. In particular, this a prerequisite for extending regime-switching approaches to multi-step ahead wind power forecasts.

Acknowledgments

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

References

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, Rasmussen C, Eriksen PB, Pedersen J. (2007) Technical aspects of status and expected future trends for wind power in Denmark. *Wind Energy*, 10:31–49.

Alexiadis MC, Dokopoulos PS, Sahsamanoglou HS. (1999) Wind speed and power forecasting based on spatial correlation models. *IEEE Transactions on Energy Conversion*, 14:836–842.

Bermejo MA, Peña D, Sánchez I. (2011) Identification of TAR models using recursive estimation. *Journal of Forecasting*, 30:31–50.

Bollerslev T. (1986) Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31:307–327.

Box GEP, Cox DR. (1964) An analysis of transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*, 26:211–252.

Damousis IG, Alexiadis MC, Theocharis JB, Dokopoulos JS. A fuzzy model for wind speed prediction and power generation in wind parks using spatial correlation. *IEEE Transactions on Energy Conversion*, 19:352–361.

Dempster AP, Laird NM, Rubin DB. (1977) Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 1–38.

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

Fruhwirth-Schnatter S. (2006) Finite mixture and Markov-Switching models. *Springer*.

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

GE Energy (2010). Western wind and solar integration study. Prepared for the *National Renewable Energy Laboratory*.

Giebel G, Brownsword R, Kariniotakis G, Denhard M, Draxl C. (2011) The state-of-the-art in short-term prediction of wind power: A literature overview. Technical Report, ANEMOS.plus.

Girard R, Allard D. (2012) Spatio-temporal propagation of wind power prediction errors. *Wind Energy* (Available online).

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

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

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

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

Hering AS, Genton MG. (2010) Powering up with space-time wind forecasting. *Journal of the American Statistical Association*, 105:92–104.

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

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

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

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

Larson KA, Westrick K. (2006) Short-term wind forecasting using off-site observations. *Wind Energy*, 9:55–62.

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

Lau A. (2011) Probabilistic wind power forecasts: from aggregated approach to spatio-temporal models. PhD Thesis, University of Oxford.

Lesaffre E, Rizopoulos D, Tsonaka R. (2007) The logistic transform for bounded outcome scores. *Biostatistics*, 8:72–85.

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

Madsen H. (2008) Time series analysis. *Chapman & Hall/CRC*.

Met Eireann, the Irish National Meteorological Service. `http://www.met.ie`.

Nielsen HA, Pinson P, Nielsen TS, Christiansen LE, Madsen H, Giebel G, Badger J, Larsén XG, Ravn HV, Tøfting J. (2007) Intelligent wind power prediction systems: Final report. Technical report, Informatics and Mathematical Modelling, Technical University of Denmark, DTU.

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

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

Pinson P. (2012) Very short-term probabilistic forecasting of wind power time-series with generalized Logit-Normal distributions. *Journal of the Royal Statistical Society, Series C*, 61: 555–576.

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

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

Tong H. (1990) Non-linear time series: a dynamical system approach. *Oxford University Press*.

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

Zucchini W, MacDonald IL. (2009) Hidden Markov models for time series: An introduction using R. *Chapman & Hall/CRC*.

PAPER C

Weather Radars - The new eyes for offshore wind farms?

Authors:
P.-J. Trombe, P. Pinson, T. Bøvith, N.A. Cutululis, C. Draxl, G. Giebel, A.N. Hahmann, N.E. Jensen, B.P. Jensen, N.F. Le, H. Madsen, L.B. Pedersen, A. Sommer and C. Vincent

Submitted to:
Wind Energy (2012)

Weather radars - The new eyes for offshore wind farms?

Pierre-Julien Trombe[1], Pierre Pinson[1], Thomas Bøvith[2], Nicolaos A. Cutululis[3], Caroline Draxl[3], Gregor Giebel[3], Andrea N. Hahmann[3], Niels E. Jensen[4], Bo P. Jensen[4], Nina F. Le[5], Henrik Madsen[1], Lisbeth B. Pedersen[4], Anders Sommer[6], Claire Vincent[3]

Abstract

Offshore wind fluctuations are such that dedicated prediction and control systems are needed for optimizing the management of wind farms in real-time. In this paper, we present a pioneer experiment – Radar@Sea – in which weather radars are used for monitoring the weather at the Horns Rev offshore wind farm, in the North Sea. First, they enable the collection of meteorological observations at high spatio-temporal resolutions for enhancing the understanding of meteorological phenomena that drive wind fluctuations. And second, with the extended visibility they offer, they can provide relevant inputs to prediction systems for anticipating changes in the wind fluctuation dynamics, generating improved wind power forecasts and developing specific control strategies. However, integrating weather radar observations into automated decision support systems is not a plug-and-play task and it is important to develop a multi-disciplinary approach linking meteorology and statistics. Here, (i) we describe the settings of the Radar@Sea experiment, (ii) we report the experience gained with these new remote sensing tools, (iii) we illustrate their capabilities with some concrete meteorological events observed at Horns Rev, (iv) we discuss the future perspectives for weather radars in wind energy.

[1]DTU Informatics, Technical University of Denmark, Kgs. Lyngby, Denmark
[2]Danish Meteorological Institute, Copenhagen, Denmark
[3]DTU Wind Energy, Technical University of Denmark, Roskilde, Denmark
[4]Danish Hydrological Institute (DHI), Aarhus, Denmark
[5]DONG Energy A/S, Gentofte, Denmark
[6]Vattenfall Denmark A/S, Fredericia, Denmark

1 Introduction

A substantial number of large-scale offshore wind farms have been deployed in Northern Europe over the last few years, and the plan is to keep on expanding offshore wind power in the near future (Danish Energy Agency, 2007). Along that expansion, the development of specific methodologies for wind resource assessment in offshore environments has received much attention. In particular, the use of remote sensing techniques has led to significant advances in that domain (Sempreviva et al., 2008). In comparison, much less attention has been given to operational issues linked to the predictability and controllability of these large offshore wind farms Sørensen et al. (2007). And yet, the potential magnitude of wind fluctuations is such that advanced control strategies are indispensable and have to be performed in real-time (Kristoffersen, 2005), even more when weather conditions become extreme (Cutululis et al., 2011). Offshore wind power fluctuations also induce additional challenges for Transmission Systems Operators (TSO) in maintaining the balance between electricity production and demand (Akhmatov et al., 2007). For these applications, the availability of accurate wind power forecasts is a prerequisite. In particular, there is a large consensus on the growing importance of such forecasts at specific temporal resolutions of 5-10 minutes, and look-ahead times of a few hours (Jones and Clark, 2011).

Short-term wind power forecasts, from a few minutes up to a few hours, are preferably generated with statistical models using historical data. However, today, operational prediction systems for offshore wind farms are not fundamentally different than for onshore wind farms (Giebel et al., 2011). They traditionally rely on meteorological forecasts (e.g., wind speed and direction) whose temporal resolution is usually between 1 and 3 hours, and up to a forecast length of 48-72 hours. This acts as a limitation when it comes to capturing the intra-hour volatility of offshore wind power fluctuations induced by meteorological phenomena in the boundary layer, even more when meteorological forecasts are misleading (e.g., phase errors). Furthermore, it is a well-known issue that the layout of offshore wind farms, concentrating a high density of wind turbines within a small geographical area, makes the impact of local meteorological phenomena on their power production stronger than at onshore sites where smoothing effects occur. These issues were addressed in several recent studies which alternatively proposed the use of regime-switching models (Pinson et al., 2008, Trombe et al., 2012), a new type of predictive density (**?**), or local wind speed and direction measurements as new inputs (Gallego et al., 2011). However, even though these models give evidence of their interesting predictive power, their ability to accurately predict the most severe fluctuations remain very limited and offshore wind power forecasts are characterized by large uncertainties. This also highlights the limitations of local wind measure-

ments (e.g., from nacelle anemometry and SCADA systems) when it comes to upcoming changes in weather conditions on spatial scales of kilometers. Meteorological observations that cover a broader spatial area are thus required, not only to improve our understanding of the phenomena driving mesoscale wind fluctuations, but also to provide more informative inputs to prediction models.

In wind power forecasting, there is a need for new and multi-disciplinary approaches combining the expertise of meteorologists, forecasters, control engineers and wind farm operators. This is the idea developed in an ongoing experiment – Radar@Sea – which proposes the use of weather radars, novel remote sensing tools in wind energy, for the online observation of the atmosphere at offshore sites. This experiment is motivated by recent advances in the modeling of wind fluctuations at Horns Rev, Denmark, and the identification of several climatological patterns correlated with periods of increased wind speed variability, for time scales from 10 minutes up to 1 hour (Vincent et al., 2011). In particular, precipitation and large wind speed fluctuations are often observed simultaneously. Weather radars are the ideal tools to detect, locate and quantify precipitation. They have become essential tools in real-time decision support systems for tracking and predicting natural hazards. More generally, owing to their techniques, they offer an extended visibility of the weather conditions over substantially large areas. Therefore, they have the potential for anticipating the arrival of weather fronts and other meteorological phenomena which intensify offshore wind fluctuations. It is even more important for some offshore wind farms that cannot benefit from upwind information, being the first hit by the onset of particular weather regimes.

The experiment we present in this paper is the first of this type for wind energy applications worldwide, to our knowledge. Yet, lessons learnt from the use of weather radars in hydrological and meteorological sciences show that integrating weather radar observations into automated decision support systems is not a plug-and-play task. The volume and complexity of weather radar observations are such that specific diagnosis tools have to be developed for data quality control, data visualization and feature extraction (see, for instance, Lakshmanan et al. (2007) for a detailed description of the WDSS-II system for severe weather nowcasting). Therefore, a thorough understanding of the weather radar techniques, capabilities and limitations, as well as the field of application are expected to influence the design of the final decision support system. For those reasons, we think that the experience gained through the Radar@Sea experiment could be a valuable source of information to other researchers following a similar approach.

The structure of this paper is as follows. In section 2, we give an introduction to the meteorological conditions (precipitation and wind fluctuations patterns) over Denmark and the North Sea. In section 3, weather radars principles, capa-

bilities and limitations are presented. In section 4, we describe the Radar@Sea experiment along with the two weather radar systems used for the experiment. In section 5, we show four precipitation events and analyze how they relate to wind speed and wind power fluctuations observed at Horns Rev. In section 6, we discussed the future perspectives for weather radars in wind energy applications. Finally, section 7 delivers concluding remarks.

2 Meteorological context

Automating the integration of complex and large meteorological observation sets into prediction systems requires a preliminary understanding of the meteorological phenomena over the region of interest, both at the synoptic scale and the mesoscale. More specifically, we are interested in using precipitation observations as indicators for weather conditions featuring high wind variability. Therefore, a clear view on the relationship between meteorological variables and the development of precipitation is likely to help interpreting weather radar observations. In this section, the focus is placed on the coastal area of Denmark and, in particular, the North Sea.

2.1 Synoptic scale

Denmark is located at the border between the North Sea and the European continent. The atmospheric circulation patterns are dominated by westerly flows coming from the Atlantic Ocean and the North Sea. The average wind direction can often be associated with particular weather conditions, and each weather phenomenon has a unique signature in terms of the local wind variability, precipitation and small scale weather.

For example, cold fronts, which are the boundary between cold and warm air masses, approach the North Sea from the west and are usually associated with a wind direction change from southwesterly to northwesterly. In the winter months, anticyclones over the region often bring cold, clear conditions and light easterly winds, while in the summer months, anticyclones tend to be positioned further to the south and bring warm, sunny weather and still wind conditions. West and South-West are the prevailing wind directions while North and North-East directions are the least frequent (Cappelen and Jørgensen, 1999). A brief summary of the most frequent weather types and their associated precipitation patterns is provided in Table 1, conditioned upon wind direction and season. For the purposes of this article, we consider that there are only two

seasons in Denmark, a winter season from October to March, and a summer season from April to September.

Table 1: Weather types and their associated precipitation patterns in Denmark, as described in Cappelen and Jørgensen (1999).

	Winter	Summer
Northerly winds	• Northerly winds are the least frequent in Denmark and are mostly observed during the winter, • They bring dry and very cold air which result in very occasional showers in the East of Denmark,	
Easterly winds	• Easterly winds are most frequent during the late winter and early summer and are the consequence of the continental influence, • Rare and light precipitation (mostly during the winter).	
Southerly winds	• They bring cold continental air, • Occasional light precipitation.	• They bring warm continental air, • Moist air transforms into heavy rain showers (and strong wind downdrafts), • Occasional thunderstorms.
Westerly winds	• Westerly winds bring depressions associated with frontal systems and trailing precipitation (occasional snow in the winter) or heavy rain showers, • Successive arrival of depressions may repeat over weeks, being separated by one or two days.	
	• Cold air carried out by fronts passing over warm sea often results in strong convection and rain showers.	• Under westerly winds, precipitation are usually more abundant in the summer than in the winter.

Severe phenomena and large wind fluctuations are mainly associated with two types of synoptic scale systems. First, low pressure systems and their associated cold fronts, coming from the Atlantic Ocean, are very dynamic and favor the development of squall lines and thunderstorms accompanied by heavy rain

showers. These low pressure systems may contain more than one cold front. Hence, their effects may persist over several days. The level of severity associated with these low pressure systems is generally higher in the winter than in the summer. Second, the continental influence may be more pronounced during the summer than the winter and result in warm and moist air being driven from the South over Denmark. This initiates a favorable context for the development of thunderstorms. In Van Delden (2001), a 4-year climatological study of these thunderstorm events showed that their frequency was relatively low in Northern Europe, when compared to Western Europe. In Denmark, that study also showed that thunderstorms tended to occur at a higher frequency over the coastal area and the North Sea than over land.

2.2 Mesoscale

Mesoscale phenomena have length scales between a few kilometers and several hundred kilometers, and it follows that they are associated with wind fluctuations with periods between a few minutes and a few hours. Therefore, the wind fluctuations of interest in this paper are driven by mesoscale phenomena, which are driven by the prevailing synoptic conditions.

In Vincent et al. (2011), mesoscale wind fluctuations observed at the Horns Rev 1 (HR1) wind farm were analyzed and it was shown that the largest amplitude fluctuations tended to occur when the wind direction was from the westerly sector, a result that was consistent with Gallego et al. (2011) and Akhmatov (2007), who reported large power fluctuations and large forecast uncertainty in the same sector. Furthermore, large wind fluctuations were found in the presence of precipitation, when the mean sea level pressure was dropping rapidly (indicating post-frontal conditions) and during the late summer and early winter months when the North Sea is often warmer than the near-surface air. In Vincent et al. (2012), the authors examined a case of large wind fluctuations at HR1, and used mesoscale modelling to demonstrate the potential for open cellular convection over the North Sea, which forms in maritime flow under unstable, post-frontal conditions to cause high wind variability. The lattice of hexagonal shaped cells that are characteristic of open cellular convection can often be clearly identified in satellite pictures over the North Sea during post-frontal conditions (see Figure 1). This phenomenon is of particular interest here, because it may be identified in radar pictures in cases where there is precipitation associated with the cloudy cell walls. Further characteristics of open cellular convection phenomena are described in Atkinson and Zhang. (1996).

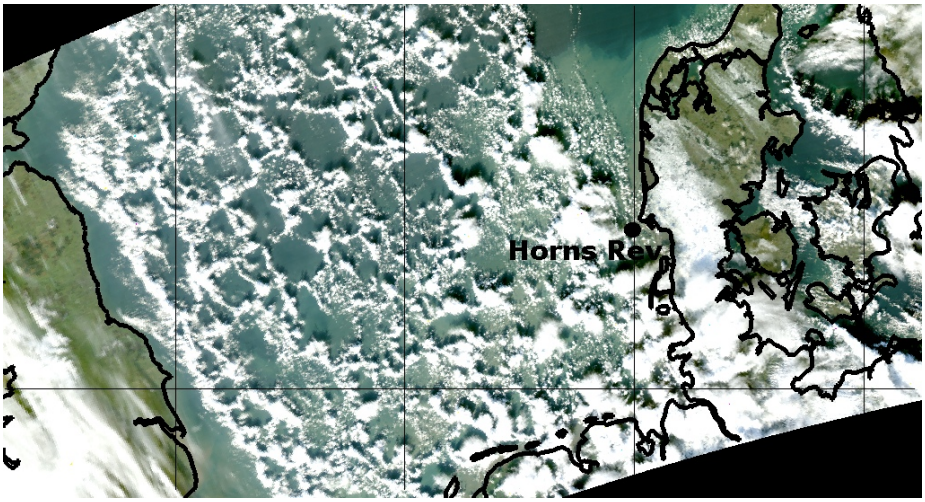


Figure 1: Satellite image of a case of open cellular convection over the North Sea. The cloud tops are shaped like a honeycomb, with cloud rings on the edge and cloud-free centers. The image is from the MODIS TERRA satellite: `http:/ladsweb.nascom.nasa.gov`

3 Weather radars

Remote sensing tools have enabled the collection of large amounts of meteorological data and their importance for the development of wind energy projects is constantly growing (International Energy Agency, 2007). For instance, ground-based tools such as LiDAR and SoDAR are used for estimating wind profiles at high heights. Alternatively, LiDAR can be mounted on a wind turbine hub or rotating spinner to measure the approaching wind flow in view of optimizing wind turbine control (Harris et al., 2006, Mikkelsen et al., 2010). Airborne radars can contribute to the observation of wake effects at large offshore wind farms, and offshore wind maps can be generated from satellite observations (Hasager et al., 2008). However, applications of remote sensing tools in wind energy often converge towards a common goal, which is an improved assessment of the wind resource. In addition, their outputs tend to be either spatially limited (e.g., LiDAR and SoDAR) or temporally sparse (e.g., satellite observations). In contrast, one of the clear strengths of weather radar systems is their superior capacity to generate observations at high resolutions, both in time and space, which is a very desirable capability for the short-term forecasting of wind power fluctuations. In this section, we provide some insights on weather radar principles, capabilities and limitations which are further illustrated by concrete examples taken from Radar@Sea in the subsequent sections.

3.1 Principles & Capabilities

Weather radars are airborne or ground-based remote sensing tools. In this paper, we only deal with ground-based weather radars. The data acquisition process consists of a circular and volumetric scanning of the atmosphere. Microwave radiation is emitted and reflected by precipitation particles. Data collected by weather radars correspond to quantitative estimations of precipitation reflectivity. Precipitation intensity estimation can be obtained through the so-called Z-R relationship (Marschall, 1948). The volumes scanned are traditionally summarized to deliver standardized output displays such as images of precipitation reflectivity at different altitudes. For a technical introduction on weather radars, we refer to Meischner (2004).

There exist a wide variety of weather radars and their specificities depend on their wavelength: X-Band, C-Band or S-Band for the most common ones (listed here from the shortest to the longest wavelength; from 3.2 cm, to 5.4 and 10 cm). Typically, the longer the wavelength, the further away the radar waves can travel in the atmosphere and detect precipitation. S-Band radars have an operational range beyond 450 km and are preferably used for severe weather monitoring (e.g., forecasting of environmental hazards such as flash floods and tornadoes; tracking of severe meteorological events such as thunderstorms and lightnings (Serafin and Wilson, 2000)), C-Band radars operate up to 200-300 km and are often used for quantitative precipitation estimation for monitoring river catchment or urban drainage systems, whereas X-Band radars have a range within 100 km and are useful for local applications. The reason for the difference in the applicable range is that at lower wavelengths the attenuation of the electromagnetic signal is higher. However, shorter wavelengths are more sensitive to small precipitation particles and more suitable for the observation of drizzle or even fog. S and C-band radars are usually used for medium to long range applications for which reason data are typically available at medium spatial resolutions of 500 m to 2000 m and temporal resolutions from 5 to 15 minutes. X-Band radars often implement a faster temporal update cycle down to 1 minute and spatial resolutions at or below 500 m. These characteristics depend on the specifications of the radar system such as the scanning strategy (e.g., antenna rotation speed, pulse repetition frequency, sampling frequency, number of elevations) and the antenna design (e.g., beam width). Other important differences between the three types of weather radars relate to their cost effectiveness and the size of their installation. X-Band radars are the most cost-effective and their small size makes them well suited for mobile installations. In contrast, the size of the antenna of C and S-Band radars reduces the range of possibilities for siting them.

Weather radar capabilities are also modulated by their techniques: Doppler and/or Polarimetric, or neither. In particular, the range of capabilities of weather

radar with Doppler technique is not limited to the detection and quantitative estimation of precipitation. They can also estimate the radial velocity of precipitation particles, revealing very useful insights on the spatio-temporal structure of complex meteorological phenomena. Polarimetric weather radars are, on the other hand, favored for their improved ability to characterize precipitation type (rain, snow, hail, etc.) as well as better capabilities for distinguishing between meteorological and non-meteorological targets. Contemporary weather radar networks operated in Europe (Holleman et al., 2008) or the United States (Crum and Alberty, 1993) mostly consist of Doppler radars. These networks are traditionally operated by national meteorological institutes and observations are available in real-time over large areas. Furthermore, overlapping observations of several weather radars can be merged to create composite images which can cover the whole Western Europe or the United States and their respective coastal areas.

3.2 Limitations

Weather radars have some shortcomings as there is an inherent uncertainty associated with their measurements. It is acknowledged that the measurement uncertainty increases with the intensity of precipitation. In Radar@Sea, we prefer working directly on the reflectivity values to avoid approximating precipitation intensity through the Z-R relationship (Marschall, 1948). In addition, various problems may arise during the data acquisition process and applying mitigation techniques is a prerequisite before integrating weather radar observations into automated systems. These problems are addressed in detail in Bøvith (2008) and we report here some examples:

- Radar waves can be intercepted, reflected or even completely blocked by non-meteorological targets such as ground, sea, buildings, mountains, etc. This problem is referred to as clutter. In this regard, the choice of an appropriate site for installing a weather radar is crucial as it reduces the risk of clutter;

- Short wavelength radars (e.g., X-Band) can be affected by beam attenuation problems in case of intense precipitation, resulting in the quality of the measurements altered at far ranges and, more specifically, large underestimation of precipitation reflectivity;

- Specific atmospheric conditions (e.g., inversion of the vertical temperature or moisture gradient in the atmosphere) may cause anomalous propagation of the radar waves which are super-refracted and bent towards the ground or the sea instead of propagating in the atmosphere;

- During convective events, the scale of precipitation cells may be relatively small compared to the volume scanned by weather radars, resulting in underestimating precipitation reflectivity, this problem is known as beam filling and become more serious at far ranges;
- Due to the curvature of the Earth, the height at which radar waves propagate increases with the range, leading to potential underestimation of near surface precipitation at far ranges, this problem is known as overshooting.

Furthermore, a growing source of concerns regarding measurement accuracy is linked to the deployment of wind farms nearby weather radar installations, generating large clutter (Isom et al., 2009). In particular, wind farms echoes are comparable to those of small storm cells. The larger the wind farm, the larger the area and the strength of the clutter are. The closer the weather radar and wind farm are, the further away the problems propagate. Impacts of wind turbines on weather radar observations can even be identified at far ranges, up to 100 km (Chèze and Haase, 2010).

4 The Radar@Sea experiment

Radar@Sea, the first experiment involving weather radars for offshore wind energy applications, started in 2009 and is expected to run until the end of the year 2012. It consisted of the installation, operation, and maintenance of a Local Area Weather Radar (LAWR) based on X-Band technology, at the offshore site of Horns Rev, Denmark. Observations from a nearby Doppler C-Band weather radar were used to complement the initial data set. Finally, wind speed, wind direction and wind power measurements from the HR1 wind farm came to complete what is by now a unique data set in the wind energy community. The respective geographical locations and spatial coverage of the two radars and the HR1 wind farm are shown in Figure 2.

4.1 Local Area Weather Radar

The LAWR is installed on the roof of the accommodation platform of the Horns Rev 2 (HR2) wind farm (see Figure 3), in the North Sea, about 20 km off the West coast of Jutland, Denmark. The LAWR is a light configuration weather radar system, ideal for remote locations (see Pedersen et al. (2010) for a complete presentation of the system). The data collection campaign with the LAWR started in 2010. The LAWR is located 19 km away from HR1 and is run with a coverage range of 60 km. In order to produce one image, 24 continuous scans

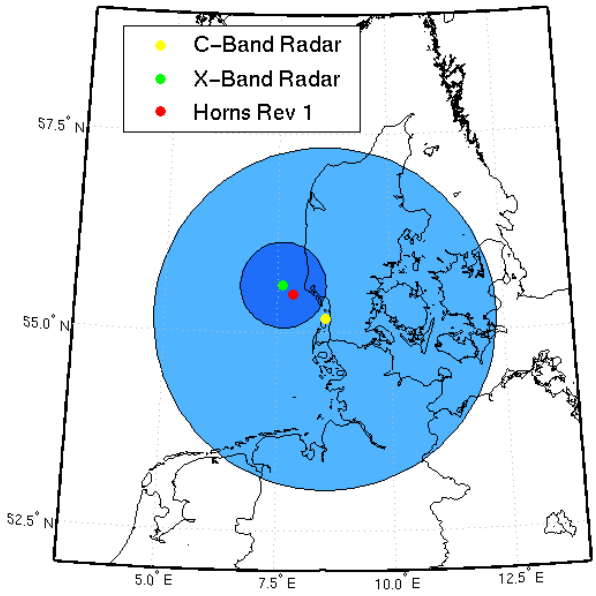


Figure 2: Geographical location of the X-Band radar (LAWR), the C-Band radar and the Horns Rev 1 wind farm, 20 km off the west coast of Jutland, Denmark. The area shaded in dark blue is the area covered by the X-band radar (range of 60 km) whereas the area shaded in light blue is the area covered by the C-Band radar (range of 240 km).

are performed every minute with a large vertical opening angle of $\pm 10^{\circ}$ and a horizontal opening of 1°. One specificity of the LAWR is that is does not generate direct observations of precipitation reflectivity but, instead, dimensionless count observations (Integer values of range 0-255) that can be converted to precipitation intensity through rain gauge calibration. A sample image generated by the LAWR can be seen in Figure C.4(b). For a summary on the operational settings of the LAWR, see Table 2.

In the course of the Radar@Sea experiment, the observational capabilities of the LAWR have been challenged by several problems. First, it is important to mention that the accommodation platform of the HR2 wind farm, where the LAWR is currently installed, performs many functions other than the LAWR. The result is that, even though the best possible spot on the platform was chosen, there is a large blocking of the beam and observations are not available for

(a) Accommodation platform of the Horns Rev 2 wind farm. A LAWR can be seen on the top left corner of the platform.

(b) Another LAWR is installed at the Horns Rev 1 wind farm for the needs of a separate experiment.

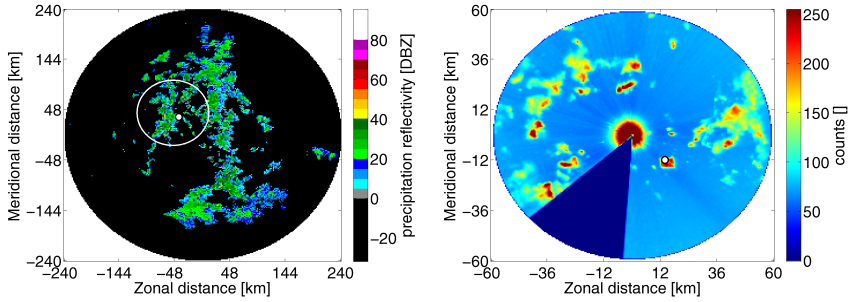
Figure 3: The first experiment of weather radars for offshore wind energy takes place at Horns Rev, Denmark.

southwesterly azimuths (see Figure C.4(b)). Second, the very close proximity of the wind turbines of HR2 contributed to large uncertainties in the measurements at close ranges. Third, due to the shared utilization of the LAWR with another experiment for wave monitoring, its mechanical clutter fence was re-

Table 2: Geographic information, technical specifications and operational settings of the two weather radars used in the Radar@Sea experiment.

		LAWR (X-Band)	**Rømø (Doppler C-Band)**
Geographic information	Location	Offshore (HR2)	Onshore (Rømø)
	Coordinates	55.600°N, 7.623°E	55.173°N, 8.552°E
	Distance to HR1	19 km	57 km
	Height (above sea level)	30 m	15 m
Technical specifications	Frequency	9.41 GHz	5.62 GHz
	Wavelength	3.2 cm	5.4 cm
	Vertical opening angle	±10°	±1°
	Horizontal opening angle	±1°	±1°
	Scanning strategy	Continuous	9 elevations
	Scanning elevation	-	0.5°, 0.7°, 1°, 1.5°, 2.4°, 4.5°, 8.5°, 13°, 15°
Operational settings	Image frequency	1 min.	10 min.
	Range	60 km	240 km
	Grid size	500×500 m	2×2 km
	Data	Dimensionless count values (Integer 0-255)	Decibel of reflectivity (dBZ)

moved. This important component usually ensures that only the reflected energy corresponding to the upper 10° of its vertical opening angle is kept for precipitation sampling. The modification resulted in the measurements being contaminated by sea clutter. On the images, this translates into "dry" pixels having values between 70 and 100, instead of values closer to 0. These problems could easily be avoided if, as part of the design of the platform in the future, a specific spot was allocated for installing measuring instruments. Last but not least, the extreme weather conditions experienced at Horns Rev presented a difficult test for the robustness of the LAWR. Passages of many storms over Denmark were recorded in the winter 2011, with mean wind speeds approaching 30 m s^{-1}, coupled with strong gusts. Running the LAWR during these storms increased the number of rotations of its antenna from 24 to 33-39 rotations per minute, thereby increasing the risk of damaging its structural components. To circumvent that problem, an electronic breaking system was added and has, since then, proved its efficiency, enabling data collection during the subsequent storms.



(a) Sample image generated by the Doppler C-Band weather radar.

(b) Sample image generated by the LAWR.

Figure 4: Sample images generated by the two weather radars on August 29, 2010 at 3.30am. The white circle on Figure (a) indicates the area covered by the LAWR. The position of the Horns Rev 1 wind farm is depicted by a white dot on both images. As a result of their different scanning strategies, the 2 weather radars reveal different features of precipitation fields.

4.2 Rømø weather radar

The Doppler C-Band weather radar used in the Radar@Sea experiment is located in Rømø, Denmark, and operated by DMI, the Danish Meteorological Institute (see Gill et al. (2006) for an introduction on the Danish weather radar network). It is located 57 km away from the HR1 wind farm and has a coverage range of 240 km. Observations were collected using a 9 elevation scan strategy (0.5°, 0.7°, 1°, 1.5°, 2.4°, 4.5°, 8.5°, 13°,15°) every 10 minutes (see Table 2). Raw reflectivity measurements were converted into decibel of reflectivity (dBZ) since it is a more appropriate unit for processing reflectivity images, as demonstrated in Lakshmanan (2012). A sample image generated by the Doppler C-Band weather radar can be seen in Figure C.4(a). The observations DMI provided us with consist of a 1-km height pseudo-CAPPI (Constant Altitude Plan Position Indicator) image product. The images which in our case have a grid spacing of 2 km display the radar reflectivity at an altitude of 1 km by selecting reflectivity from the most appropriate elevation. At ranges further than approximately 80 km where the beam of the lowest elevation exceeds 1 km altitude, data from the lowest elevation are used (hence the "pseudo"-CAPPI). A general pixel-wise interpretation of reflectivity values considers background noise echoes (birds, insects, etc.) to be between 0 and 10 dBZ, light precipitation systems (e.g., stratiform rainfall) between 10 and 30 dBZ and the threshold for convective precipitation systems is often set to between 30 and 40 dBZ. This pixel-wise interpretation is only to be used as a simple heuristic and the charac-

terization of radar reflectivity echoes in terms of precipitation types is a much more complex task that requires the use of advanced algorithms (Biggerstaff and Listemaa, 2000).

In its weather monitoring and forecasting activities, weather radar data are used by DMI and its partners for an increasing number of applications. This implies an increased work on data quality control procedures to improve the observation of precipitation and to mitigate the influence of radar clutter.

4.3 Towards validating the experiment

The experimental part of the project is not limited to the data collection. There are also a substantial number of necessary steps for validating these data, transforming them into ready-to-use products and, more generally, automating their integration into a decision support system. A preliminary step consists of performing a quality control of the data. This operation is necessary for evaluating the level of uncertainty associated with the data and defining appropriate strategies to process them. As explained in section 3, the uncertainty comes from two different sources. One is inherent to weather radar techniques (e.g., limitation for observing near surface precipitation) and the other may be caused by non-meteorological factors (e.g., clutter). In practise, the effects of the latter problems are easier to detect since measurement artifacts are not random and exhibit well-determined statistical signatures. Ideally, artifact detection methods should be robust, in the statistical sense, as they have to accommodate for levels of uncertainty that are changing over time. In Radar@Sea, clutter removal filters were applied routinely on both weather radars. In addition, volume correction and beam attenuation procedures were applied as part of the data acquisition process of the LAWR (Pedersen et al., 2010). However, a posteriori data quality controls revealed recurrent clutter and some consistency issues on measurements from both radars. These results as well as mitigation techniques are presented in Appendix A.

5 Illustrative meteorological events from Horns Rev

In this section, we analyze four meteorological events which show the development and passage of precipitation systems in relation to wind fluctuations at the HR1 wind farm. These events were selected to illustrate the variety of situations that weather radar can help observing. We do not attempt to make any projection related to forecasting issues. Normalized wind power fluctuations at HR1 are also included in order to show their corresponding amplitude

during these events. Wind speed, direction and power measurements were collected from the nacelle anemometry and SCADA systems (Kristoffersen, 2005). To be consistent with section 2, we consider that there are only two seasons in Denmark, a summer or warm season from April to September, and a winter season from October to March. The prevailing synoptic conditions for each of these two seasons are given Table 1.

Note that non-meteorological information has not been perfectly cleaned from the displayed images. Let us acknowledge that removing measurement artifacts with automated algorithms is a highly complex task. In particular, there is always a risk of also removing valuable meteorological information by being too aggressive on the detection criteria. Our approach is to reduce the amount of non-meteorological information down to an acceptable level and adapt the robustness of image analysis methods accordingly.

5.1 Summer storms

The first meteorological event as seen by the Rømø weather radar and wind observations is shown in Figure 5. It is from July 2010 and depicts how the development of typical summer storms driving warm and moist continental air coming from the South relates to wind speed and wind power fluctuations at the HR1 wind farm. The arrows show the wind direction recorded at HR1. (1) It begins with a case of anomalous propagation falsely suggesting the presence of precipitation. This problem is likely to be caused by a temperature or moisture gradient inversion in the vertical stratification of the atmosphere (see Appendix A). (2) The problem is persistent for several hours and also visible on the right part of the second image which shows the development of strong convection. One can notice a storm in the proximity of the HR1 wind farm. It is delimited by a cluster of pixels with high reflectivity values exceeding 40 dBZ. That storm is embedded into a larger precipitation system. The birth and growth of that storm precede the occurrence of a strong wind gust at HR1 quickly followed by a large drop of wind speed. After that, precipitation dissipates until the development of a larger storm, one day later. (3-5) The passage of that second storm across HR1 is coupled with very large wind fluctuations. These fluctuations are likely to be caused by the strong updrafts and downdrafts associated with this type of storms (Houze, 2004). Over the 5 days of this events, the wind exhibits a very chaotic behavior, with sudden and frequent changes of direction. Another interesting aspect of this event is that it illustrates a case of high wind variability coupled with medium mean wind speeds. In terms of wind power fluctuations, the passage of the first storm translates into a sudden drop of power from the rated power of HR1 to 0 within 2-3 hours. The passage of the second cluster of storms generates

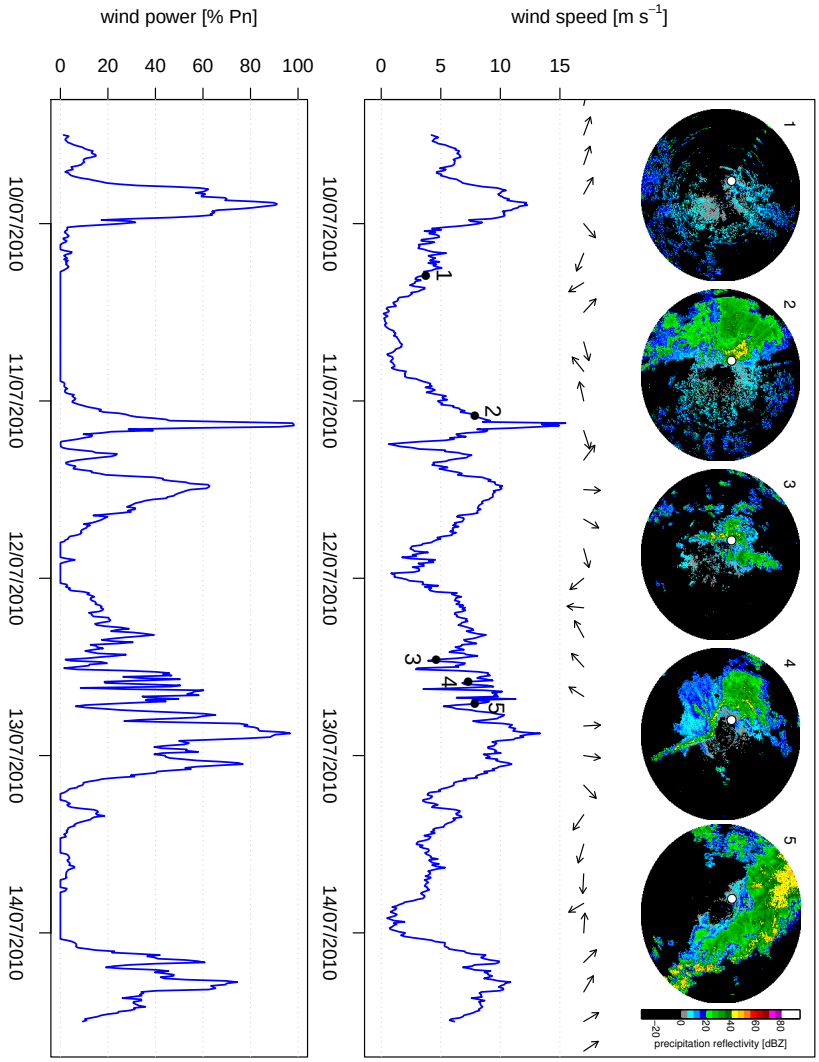


Figure 5: July 2010 - Typical summer storms bringing moist and warm air from the South, resulting in heavy rain showers. The images were generated by the C-Band radar in Rømø. The arrows indicate the wind direction recorded at HR1. (1) No precipitation but a case of anomalous propagation. (2) Development of a large convective rainfall system with an embedded storm just before a strong wind gust is sensed at the HR1 wind farm. (3-5) Development of another large storm associated with increased wind speed and wind power volatility at HR1.

fluctuations of an amplitude equivalent to 50% the rated power of HR1, over a period of 8 hours.

5.2 A cold front in the winter

The second event is shown in Figure 6 and is from December 2010. It illustrates the passage of a cold front over the North Sea and across the HR1 wind farm during the winter. Let us recall that the North Sea surface is warmer than the lower part of the atmosphere at that time of the year, enhancing the development of strong convection (Vincent et al., 2011). (1) It starts with a shift in wind direction at HR1, from the North-East to the South-West, and smoothly increasing wind speed as the front approaches. Meanwhile, light and widespread precipitation is moving from the North-West. (2) Wind fluctuations intensify as the cold front passes across HR1 until a large negative gradient of wind speed is sensed in the transition zone of the front, behind its leading edge. The front leading edge is marked by an area of high reflectivity, between 30 and 40 dBZ, indicating the development of convection. This area of convection is embedded into a larger area of precipitation, characterized by intermediate mean reflectivity. (3-5) In the wake of the front, the wind direction shifts from the South-West to the North-West. In addition, large wind fluctuations are sensed at HR1 simultaneously with the passage of many scattered precipitation cells. These cells are small and are characterized by very short lifetime, growing and decaying within a few hours. Inspecting satellite pictures corresponding to this events reveals well developed open cellular convection covering part of the North Sea. Wind fluctuations have an average period of around 1-2 hours, which is consistent with the spatial scale of the open cellular convection, as discussed in Vincent et al. (2012). Resulting wind power fluctuations reach an amplitude of almost 80% the rated power of HR1, within one hour.

5.3 Precipitation without severe wind fluctuations

The third event is shown in Figure 7 and is from May 2010. It illustrates the development of a relatively large precipitation system which is not associated with severe wind fluctuations at HR1. Precipitation is moving from the North-East whereas the mean wind recorded at Horns Rev is northwesterly. (1-3) The mean wind speed increases steadily as the precipitation system is moving towards HR1. When compared to the previous event showing a cold front passage in the winter, the spatial structure of the leading edge of the present precipitation system is quite similar. It consists of a convective area embedded into a larger area of less intense precipitation. (4-5) Precipitation dissipates and the mean wind speed decreases without noticeable change in its variability.

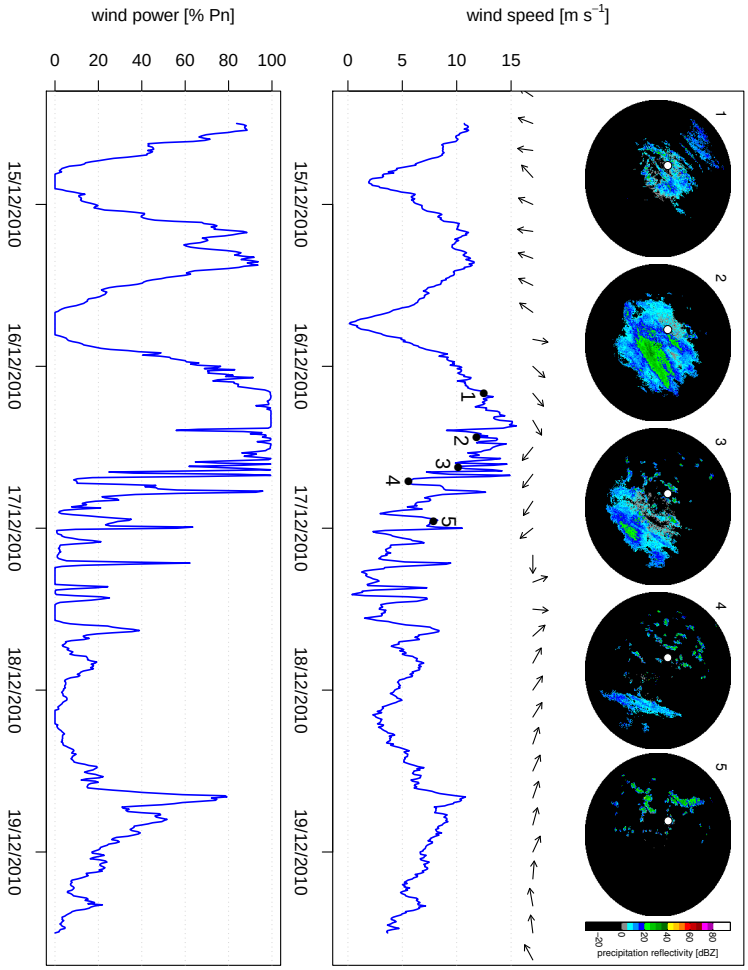


Figure 6: December 2010 - An example of cold front passage over the North Sea and the HR1 wind farm during the winter. The images were generated by the C-Band radar in Rømø. The arrows show the wind direction recorded at HR1. (1) The wind speed is peaking up with the arrival of the cold front. It leading edge is characterized by widespread stratiform precipitation with embedded convection. (2) A first large negative gradient of wind speed is sensed at HR1 while the leading edge of the cold front is passing across the wind farm. (3-5) Very large wind speed fluctuations coupled with a steady decrease of the mean wind speed. Small precipitating cells can be observed in the wake of that cold front. These cells correspond to well developed open cellular convection.

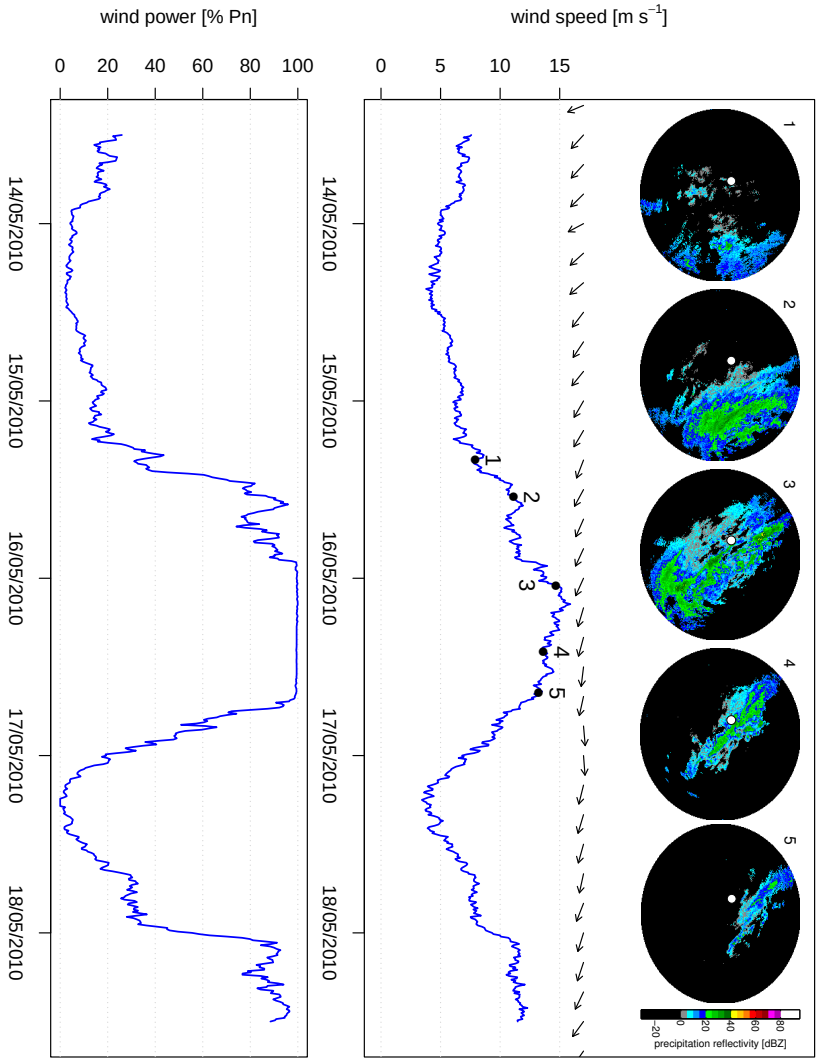


Figure 7: May 2010 - An example of precipitation system which is not associated with severe wind fluctuations at the HR1 wind farm. The images were generated by the C-Band radar in Rømø. The arrows show the wind direction recorded at HR1. (1-3) The wind speed increases steadily with the arrival of the precipitation system from the North-East. (4-5) The wind speed decreases steadily as the precipitation dissipates.

Unlike the previous episode, the leading part of the precipitation system is not followed by any trailing cell. It can also be noted that the resulting wind power

fluctuations are relatively small.

This event shows that the presence of precipitation in the vicinity of the HR1 wind farm is not always associated with severe wind fluctuations. There may be several reasons for this. Firstly, the strength and severity of phenomena producing precipitation usually decreases after they reach their mature stage. In particular, in this event, it can be seen that precipitation dissipates as the convective area reaches the HR1 wind farm. Secondly, the synoptic conditions associated with the development of precipitation may not favor severe weather. Here, precipitation is being driven from the North-East. This setting rarely produces severe phenomena (see Table 1). Finally, what may be the most likely reason is that the precipitation observed by the Rømø radar may be produced high up in the atmosphere where the weather conditions are different than those observed at the nacelle height where the wind speed and direction are recorded.

5.4 Small precipitation cells passing across HR1

The fourth event is shown in Figure 8 and is from August 2010. It illustrates how small precipitation cells can generate relatively large wind power fluctuations. The mean wind is westerly. The visualization of that episode is made more difficult by the removal of clutter pixels during the data quality control stage (see Appendix A). In particular, there is no information available in the center of the images and for southwesterly azimuths. However, it can be seen that the passage of relatively small precipitation cells of high reflectivity across HR1 has a strong impact on the short-term dynamics of the wind power fluctuations. Short wavelength weather radars such as the LAWR are particularly well suited for tracking these cells as they can provide one image per minute and, thus, enable a timely tracking of these cells with an accurate synchronization of when they are going to hit the wind farm.

6 Discussion on future perspectives for weather radars in wind energy

The most common fields of application of weather radar data include hydrology and weather surveillance. Consequently, most of the methodologies for analyzing weather radar data are centered on issues such as the conversion from precipitation reflectivity to intensity, or natural hazard nowcasting. In Radar@Sea, the approach we aim at developing is inspired by existing approaches for storm tracking. However, Radar@Sea is just one among other

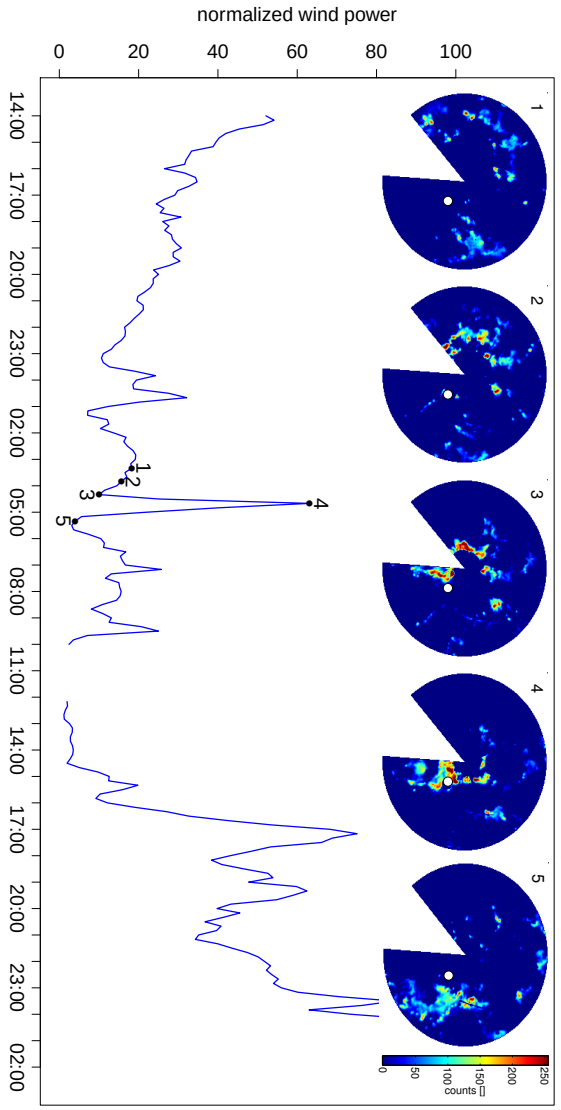


Figure 8: August 2010 - The passage of small precipitation cells through the Horns Rev 1 wind farm causes a large fluctuation of wind power. The mean wind is westerly. The images were generated by the LAWR installed at HR2.

potential wind energy applications of weather radar data. In this section, we describe the future lines of work in Radar@Sea and also discuss the future perspectives for weather radars in wind energy.

6.1 Automating the integration of weather radar observations into a real-time wind power prediction system

Raw weather radar data are useful to meteorologists for diagnosing precipitation systems and their respective severity by visual assessment. However, as the amount of data increases, making consistent decisions becomes more lengthy and difficult. Hence, the real value of weather radar observations can only be obtained through their integration into automated decision support systems (see Lakshmanan et al. (2007) and references therein). Automating a decision support system requires that one or several experts determine a series of rules or criteria to be fulfilled in order to make consistent decisions. Furthermore, the system should also have the capability to learn by itself, in a closed-loop, through the acquisition of new data and experience with potential new events never observed before. For these purposes, it is important to understand the weaknesses and strengths of the weather radar system providing the data.

In Radar@Sea, a clear weakness of the two weather radars is their limited range visibility which is inherent to single weather radar systems, as opposed to networks of radars which cover much larger areas. Note that small range visibility does not mean small temporal visibility. A small range visibility translates into potential difficulties for observing the full extent of precipitation systems in real-time, since weather radars may only observe them partially. For instance, an illustrative example is to compare the second and third events in section 5. At the beginning of both events, convection develops within a relatively large precipitation field. Before and until the time the convective part of the precipitation system reaches the HR1 wind farm, it is not possible to observe what type of weather (i.e., precipitation or not) is developing in its wake, out of the range of the weather radar. In the second event, small precipitation cells corresponding to well developed open cellular convection follow whereas, in the third event, precipitation dissipates. With information on upcoming precipitation available at longer range, severe phenomena could likely be anticipated with a higher accuracy. Comparing events 2 and 3 also shows the difficulty for estimating the stage of development of precipitation (e.g., growing, mature, decaying) which is crucial for predicting the occurrence of severe meteorological phenomena in real-time (Jirak et al., 2003).

As for the strengths, let us mention the high flexibility offered by the two weather radars which have different scanning strategies, spatio-temporal resolutions (see section 4) and thus different capabilities. In our view, the potential of these 2 weather radars could be optimized through a hierarchical approach. Owing to its longer range, the Rømø radar could first be used for characterizing and classifying precipitation regimes with respect to the magnitude of wind fluctuations at Horns Rev, by extracting features linked to the spatial variabil-

ity, the reflectivity distribution or even the motion of precipitation fields. An example of such expert-based classification is given in Baldwin et al. (2005). Tracking specific phenomena such as storm cells or squall lines is also a possibility but is made cumbersome by the high sampling variability between two consecutive images and, in some cases, the very short lifetime of these cells. In a second stage, the high spatio-temporal resolution of the LAWR is expected to enable a timely tracking of the boundary of weather fronts and small precipitation cells before they hit the wind farm.

6.2 Getting the most out of weather radar capabilities

As illustrated in the previous section, 2 dimensional reflectivity images can already be very informative on changes in the local wind conditions. Yet, we are far from tapping the full potential of weather radars. For instance, raw weather radar data comprise a third dimension which can bring valuable information on the vertical variability of precipitation fields and contribute to a better classification of precipitation regimes (e.g., convective precipitation are expected to have a higher vertical extent than stratiform precipitation) and their respective severity, also potentially leading to improved identification of near sea-surface convective phenomena. In addition, the Doppler technique also enables the retrieval of horizontal wind fields as demonstrated in Tuttle and Foote (1990), Laroche and Zawadzki (1995). These data could either be used to complement precipitation reflectivity data or, depending on their accuracy, substitute them since it is more direct to interpret and process wind rather than precipitation data for wind energy applications. In the Radar@Sea experiment, it was decided to first investigate the potential of 2 dimensional reflectivity data before, possibly, extending our investigation to 3 dimensional reflectivity data and horizontal wind fields.

6.3 Future perspectives for wind power meteorology

One of the main objectives of the Radar@Sea experiment is to collect observations of atmospheric variables in view of extending our understanding of the climatology over the North Sea. In particular, these observations are expected to enable the validation of the work on mesoscale wind fluctuations presented in Vincent et al. (2011, 2012).

Furthermore, in meteorology, there is a long tradition in assimilating data into NWP models for generating improved meteorological forecasts (Ghil and Malanotte-Rizzoli, 1991). A reason for assimilating weather radar data into NWP models

is that a fully statistical approach (i.e., weather radar data exclusively and directly used as inputs to statistical models) would likely bound its forecast skill to lead times within 3 hours whereas the requirements for integrating wind power and, more generally, renewables into power systems are such that accurate forecasts are needed, not only for the next 3 hours, but for much longer horizons. In that respect, the forecast improvement resulting from data assimilation into mesoscale NWP models could be substantial up to 12-24 hours ahead. Even though there are many issues to overcome for assimilating weather radar data into high resolution NWP models (Dance, 2004), encouraging results were already obtained in some particular case studies where Doppler observations were used for initializing these models (Zhao et al., 2006).

6.4 Future perspectives on improving offshore wind farm predictability and controllability

A wealth of statistical models have been proposed for the very short-term forecasting of wind power fluctuations but, in practise, simple and parsimonious models remain difficult to outperform (Giebel et al., 2011). For the specific case of offshore wind fluctuations, most research studies have focused on the development of regime-switching models and their application for generating one step-ahead forecasts, with lead times between 1 and 10 minutes (Pinson et al., 2008, Trombe et al., 2012, Pinson, 2012, Gallego et al., 2011). So far, these models rely on local and historical measurements which loose their informative value as the forecast lead time increases. In view of that limitation, a promising line of work is to explicitly determine and predict the sequence of regimes based on the information extracted from the weather radar observations, instead of assuming it hidden and estimating it from the wind time series itself. That way, combining weather radar observations and and statistical models is expected to fill in the gap between 2 consecutive meteorological forecasts and improve wind power predictability up to 2-3 hours, with the interesting potential of correcting for phase errors of NWP models when they occur. This approach meets many recent works in the sense that it focuses on a better exploitation of available observations rather than the development of more complex and overparametrized models. From the controller perspective, the issue is to adapt the wind farm control strategy with respect to the predicted wind power fluctuations (Kristoffersen, 2005). There has been a recent increase of the attention for developing flexible controllers during extreme events, in order to find solutions for better planning of sudden wind farms shut downs (Cutululis et al., 2011).

6.5 Limitations of weather radar data for wind power predictability

In section 3, we reported a number of technical limitations that could reduce the informative power of weather radar data. These limitations are illustrated with examples from Radar@Sea in Appendix A. In particular, we mentioned the importance of mitigating the effects of measurement artifacts for avoiding the generation of false alarms due to clutter or anomalous propagation of the radar beam. Much attention is being given to these problems in view of improving operational weather radar products, and it is expected that data accuracy and overall quality will be taken a step further in the future. Such advances would likely facilitate the integration of weather radar data into wind power prediction systems. However, in our view, the main limitation of using weather radar data for improving wind power predictability is that these data are only informative on meteorological phenomena associated with precipitation. Yet, phenomena generating intense wind fluctuations can also develop without producing precipitation and be invisible to weather radars. A typical example is open cellular convection which do not always produce precipitation.

7 Conclusion

This paper presented the first dedicated experiment of weather radars for offshore wind energy applications. It was shown that weather radar were promising candidates for providing the high-resolution spatio-temporal information required in view of improving offshore wind power predictability. In particular, weather radar images have the capability of observing upcoming precipitation fields associated severe wind speed and wind power fluctuations at offshore sites. However, a number of issues have to be addressed before radar-based wind power prediction systems can become a reality.

Firstly, wind turbine clutter which, until recently, had received very little attention, cannot be efficiently removed by traditional clutter filtering techniques due to its characteristics (Isom et al., 2009). This problem is paramount when operating a weather radar in close proximity to a large offshore wind farm since the small distance between the wind turbines and the radar strongly magnifies the clutter impact. In that respect, the data collected by the LAWR at Horns Rev provide a unique base for investigating new wind turbine clutter detection and mitigation techniques.

Secondly, pattern recognition techniques are needed for identifying precipita-

tion features associated with periods of intense wind fluctuations and, conversely, with small wind fluctuations at offshore sites. Reflectivity patterns can refer to the scale, shape, motion, texture or cell arrangement of precipitation fields. In that respect, patterns should be considered at different spatial scales to distinguish between the information associated with synoptic conditions and that associated with mesoscale phenomena. In particular, a widespread approach in storm nowcasting consists of identifying specific cells or objects (i.e., contiguous pixels having reflectivity values above a given threshold) and tracking their trajectory over a sequence of weather radar images in order to predict their development and motion in the very short-term (Dixon and Wiener, 1993, Lakshmanan et al., 2007).

Thirdly, experiments such as Radar@Sea could contribute to make the wind energy and radar communities work closer. Today, wind turbine acceptance remains a major source of concern for radar operators and users since wind turbines severely degrade the accuracy of weather radar observations and, therefore, their usefulness in other applications (Isom et al., 2009). This has led to a unilateral recommendation from the radar community for excluding wind farm sites in close proximity to radar installations (Chèze and Haase, 2010). In our view, this rather reflects the lack of coordination between the two communities. Eventually, benefits could be mutual and, not only could weather radars bring benefits to the wind energy community, their application in wind energy would also create new business opportunities and attract more attention for research and development on their techniques. For instance, light configuration weather radars, such as the LAWR used in Radar@Sea, are being tested as observational tools of the sea state, for measuring wave heights, in view of improving the planning of maintenance operations at offshore wind farms. Alternatively, weather radars are being used for monitoring bird migration and could provide important information in view of assessing the potential impact of wind farms on bird populations.

Finally, Radar@Sea places focus on the application of weather radars in offshore environments because it is where the largest potential is foreseen, especially, for wind farms for which no upwind information is available. However, weather radar could also be very useful for onshore applications and, particularly, for the detection and correction of phase errors. For instance, mid-latitude squall lines often develop ahead of cold fronts and propagate both over water and land. Tracking squall lines could therefore be useful for assessing the good phasing of meteorological forecasts generated with NWP models.

Acknowledgment

This work was fully supported by the Danish Public Service Obligation (PSO) fund under the project "Radar@Sea" (contract PSO 2009-1-0226) which is gratefully acknowledged. DONG Energy and Vattenfall are acknowledged for sharing the images generated by the LAWR and the wind data for the Horns Rev 1 wind farm, respectively. DHI is thanked for providing assistance with the images. The authors also express their gratitude to the radar meteorologists from the Danish Meteorological Institute (DMI) for providing data from the Rømø radar and sharing their expertise. Finally, we thank Roland Löwe for his constructive comments for improving the present manuscript.

Appendix A – Data Quality Control

A.1 Sea clutter

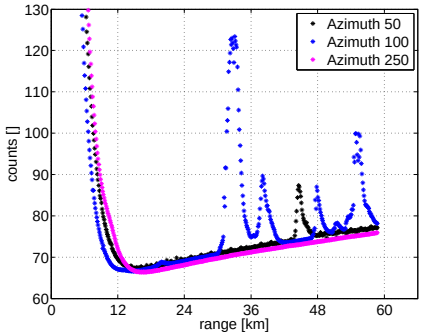
We start by analyzing the effects of the removal of the mechanical clutter fence on the LAWR images. It resulted in a recurrent and widespread sea clutter during the first six months of the data collection campaign, from April to September 2010. For this analysis, we use the original images in polar coordinates because sea clutter is usually azimuth dependent. The polar images are 360×500 and each pixel takes an Integer value between 0 and 1023. Images displaying no precipitation echoes were collected and averaged over time in order to produce a clutter map. For each of the 360 sampled azimuths, there is a systematic bias in the form of a positive and linear relationship between the count values generated by the LAWR and their range. This problem is illustrated in Figure C.9(a) where that relationship is shown for observations sampled in 3 different azimuths. One can notice that many data points lay apart from the lower trend, for all azimuths. They correspond to pixels that are recurrently affected by ground clutter and are identified in a subsequent step, after correcting for the trend. Correcting for systematic and non random artifacts is very important as many weather radar imagery techniques make use of heuristics which are not robust to such artifacts (e.g., thresholding operations to define "wet" and "dry" pixels). In addition, the level of uncertainty introduced by ground clutter contamination varies from one azimuth to another. To estimate the relationship between the count values and its range, we propose a linear regression model for each of the 360 azimuths as follows:

$$Y^{(i)} = \theta_0^{(i)} + \theta_1^{(i)} X + \varepsilon^{(i)}, \quad i = 1, \ldots, 360 \tag{1}$$

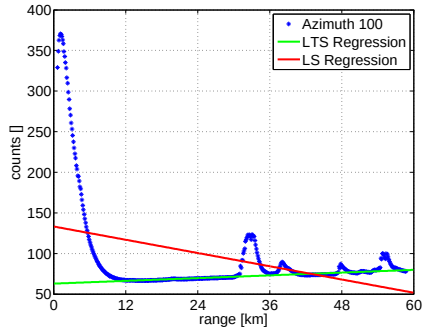
where $Y^{(i)} = (Y_1^{(i)}, \ldots, Y_n^{(i)})^T$ is a vector of n counts values extracted from the i^{th} azimuth of the clutter map, X is the range, $\varepsilon^{(i)}$ is a random variable which is assumed normally distributed with zero mean and standard deviation $\sigma^{(i)}$, and $\Theta^{(i)} = (\theta_0^{(i)}, \theta_1^{(i)})^T$ the vector of unknown parameters to be estimated for each azimuth i. For this model, a widely used estimator is the Least Squares (LS) estimator which is obtained by minimizing the sum of squared residuals, as follows:

$$\widehat{\Theta} = \underset{\Theta}{\operatorname{argmin}} \; S(\Theta) \tag{2}$$

$$\text{with} \quad S(\Theta) = \sum_{j=1}^{n} (Y_j^{(i)} - \theta_0^{(i)} - \theta_1^{(i)} X_j)^2 = \sum_{j=1}^{n} (\varepsilon_j^{(i)})^2 \tag{3}$$



(a) Clutter map for azimuth 50, 100 and 250. For each azimuth, and for range values between 12 and 60 km, count values increase with respect to their range. Observations laying apart from the trend correspond to ground clutter.



(b) Correcting for non random effects should preferably be performed with robust statistics. Here, we used the Least Trimmed Squares (LTS) regression because the estimator used in the Least Squares (LS) regression is not robust to outliers.

Figure 9: A clutter map is computed based on the original measurements in polar coordinates in order to remove the recurrent sea clutter contamination. A robust linear regression method is used for estimating the effects of the sea clutter on the images generated by the LAWR.

However, it is a well-known issue that the LS estimator is not robust to extreme values or outliers, often resulting in a poor fit of the data. Here, to overcome that problem, we use a robust technique based on the Least Trimmed Squares (LTS) (Rousseeuw, 1984). The advantage of using such technique is that it can resist up to 50% of data points laying apart of the main trend. So, instead of minimizing the sum of squared residuals as in the LS technique, we minimize

the sum of the k smallest squared residuals, as follows:

$$S(\Theta) = \sum_{j=1}^{k} (\varepsilon^{(i)^2})_{j:n} \tag{4}$$

$$\text{with} \quad k = \alpha n + 1 \quad \text{and} \quad 0.5 < \alpha < 1 \tag{5}$$

where $(\varepsilon^{(i)^2})_{1:n} < \ldots < (\varepsilon^{(i)^2})_{n:n}$ are the ordered squared residuals, sorted in ascending order. $(1-\alpha)$ corresponds to the percentage of outliers that the method is assumed to resist and it cannot exceed 50%. $(1-\alpha)$ is directly related to the notion of breakdown point which is the smallest percentage of outliers than can cause large deviations of the estimates. An example of the respective performances of the LS and LTS regressions is given in Figure C.9(b). It can be observed that the LS regression is clearly not suitable for such problem. In contrast, the LTS regression performs equally well for all azimuths. In this application of the LTS regeression, we set $\alpha = 0.4$. We assumed the sea clutter to be additive and, for each image and azimuth, we subtracted the fitted trend from the original measurements.

A.2 Ground clutter

Mitigating ground clutter on weather radar images remains a complex process and is best to be performed on the original measurements at different elevations since clutter echoes are usually limited to the lower elevations (Steiner and Smith, 2002). In addition, Doppler radars can take advantage of the reflected Doppler speed to discriminate between clutter which is usually caused by non-moving targets (buildings, mountains, etc) and precipitation which is driven by the wind. In practise, ground clutter translates into non-precipitation or non-meteorological targets having high reflectivity values which may be mistaken for small storm cells. The difficulty in identifying and correcting clutter echoes arises when ground clutter is embbeded or contiguous to precipitation fields. Ground clutter has a specific statistical signature, it is stationary in space. However, it may not be stationary over time and the values of pixels affected by clutter may vary with the weather conditions.

Here, we focus on recurrent ground clutter problems which were not detected by clutter removal filters applied on the original measurements before producing the final images (Pedersen et al., 2010, Gill et al., 2006). We follow the method proposed in Lakshmanan (2012) which is well suited for that problem since it is based on the assumption that clutter is spatially stationary. It formulates the identifaction of clutter as an image thresholding problem in order to separate clutter pixels from clutter-free pixels (Otsu, 1975). This method has several advantages and is:

- automatic and unsupervised, leading to a data-driven determination of the threshold, depending on the level of clutter contamination;
- computationally cheap;
- robust since based on count statistics.

The outline of the method is as follows:

- for each of the N pixels (x, y) of the image, compute the frequency $f_{(x,y)}(\tau)$ of its value exceeding a given threshold τ over a period of time T. In particular, a frequency value close to 1 likely indicates a clutter.
- compute a histogram by binning the N frequency $f_{(x,y)}(\tau)$ values into L levels. Let p_i be the proportion of pixels at level i, for $i = 1, \ldots, L$.
- use the segmentation method proposed in Otsu (1975) for determining a consistent threshold value k^* which separates the pixel population into 2 groups, with the first group G_1 likely being clutter free and the second group G_2 likely bieng affected by clutter. The method consists in an iterative search for the optimal threshold k^* by maximizing the inter-group variance $\sigma_B^2(k)$:

$$k^* = \underset{1<k<L}{\operatorname{argmax}} \ \sigma_B^2(k) \tag{6}$$

$$\text{with} \quad \sigma_B^2(k) = w_1 w_2 (\mu_2 - \mu_1)^2 \tag{7}$$

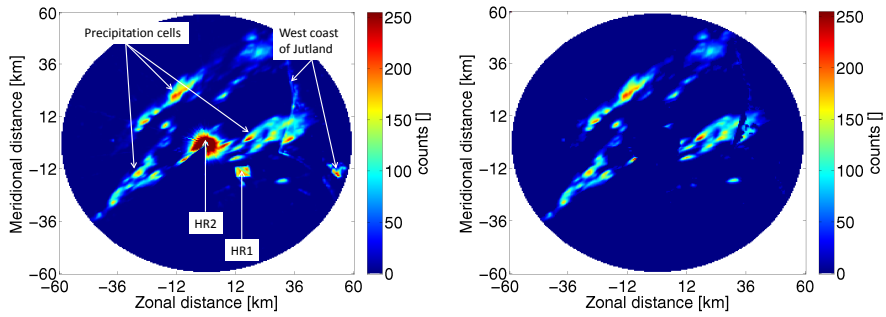
$$w_1 = \sum_{i=1}^{k} p_i \quad \text{and} \quad w_2 = \sum_{i=k+1}^{L} p_i = 1 - w_1 \tag{8}$$

$$\mu_1 = \frac{1}{w_1} \sum_{i=1}^{k} i p_i \quad \text{and} \quad \mu_2 = \frac{1}{w_2} \sum_{i=k+1}^{L} i p_i \tag{9}$$

  where w_1 and w_2 are the respective probability of occurence of G_1 and $G2$, while μ_1 and μ_2 are their respective mean level.

Note that one of the inherent hypothesis of the method described hereabove, is that the histogram to be thresholded is bimodal, implying thus that there is a significant fraction of pixels affected by clutter, at any time. This idea matches with the recurrent clutter we aim at identifying. However, clutter is non stationary over time and some pixels may be clutter over some periods of time and clutter free over some others. In order to account for that feature, the procedure is applied on a rolling window of 24 hours (i.e., 1440 images for the LAWR) and moved forward along the acquisition of new images. All images are used, both those with and without precipitation echoes. An example is

given in Figure C.10(a) which shows an image generated by the LAWR before and after the removal of recurrent clutter. The original threshold τ was determined by experience to reflect the limit between precipitation and non-precipitation targets, Here, we used $\tau = 5$. However, this method does not account for contiguity features, disregarding the potential cluster effect of clutter. A potential line of work for further improving that method could be to refine the segmentation process with a Markov Random Field (Li, 2009) step for taking into account potential spatial correlation between neighboring pixels. As for now, the method was implemented as presented in Lakshmanan (2012) and satisfactory results were obtained. The values of single clutter pixels (i.e., a pixel is clutter whereas its 8 neighbours are clutter free) were interpolated with the median of its 8 neighbours. For large clusters of clutter pixels, no attempt was made to interpolate them and their values set to 0.



(a) Original image with precipitation and clutter.

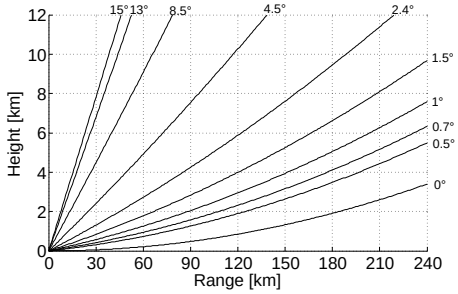
(b) Image after removing ground clutter.

Figure 10: Ground clutter is caused by the wind turbines of HR1 and HR2 and the West coast of Jutland, Denmark. Clutter translates into non-meteorological targets having high reflectivity values which may be mistaken for small storm cells and needs to be removed.

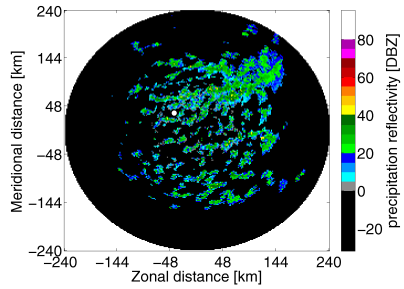
A.3 Potential underestimation of near surface precipitation at far ranges

Among the inherent limitations of long range weather radars listed in section 3, we mentioned the potential underestimation of near surface precipitation. An illustrative example of this problem is given in Figure 11. Figure C.11(a) depicts the theoretical relationship between the ground height of the 9 elevation scans of the Rømø radar with respect to the radar range, under normal wave refractivity conditions in the atmosphere. While the 9 elevation scan strategy enables an efficient sampling of the atmosphere, one can notice that

near surface precipitation, within a 2km height above ground level, cannot be detected at ranges farther than 150-180km, due to the curvature of the Earth (elevation 0°). This limitation is further demonstrated in Figure C.11(b) which shows how it translates on a precipitation reflectivity image from September 17, 2010. One can see that the weather radar detects precipitation in the close ranges (<150km) but looses its observational power at farther ranges because of precipitation fields developing in low altitudes. That limitation can be seen as the consequence of the physical settings of some meteorological phenomena described in section 2. For instance, open cellular convection phenomena develop within 1-2 km of the planetary boundary layer and are capable of producing precipitation. In such case, the visibility the Rømø radar offers is likely to be reduced as illustrated in Figure 11.



(a) Due to the curvature of the Earth, the height at which radar waves propagate increase with the range (here the 9 scan elevations of the C-band weather radar in Rømø).



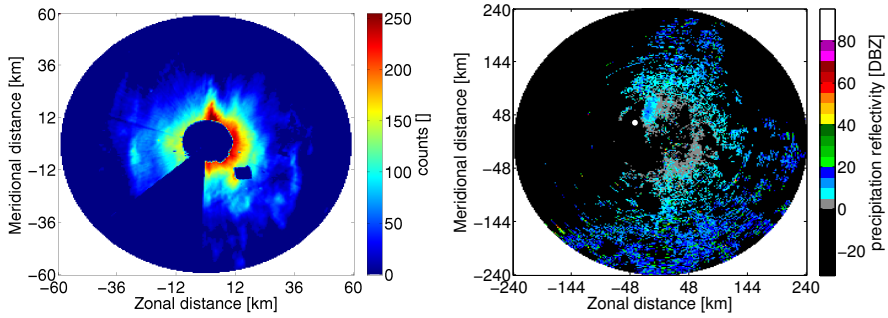
(b) Image on September 17, 2010 generated by the Rømø radar.

Figure 11: A typical example of the inherent limitation of long range weather radars for observing near surface precipitation. In particular, precipitation located within a 2km height above ground level cannot be detected at ranges farther than 150-180km.

A.4 Other limitations

Other limitations or problems enountered by weather radars include anomalous propagation and partial beam-filling. They can well be detected but are more difficult to correct. Partial beam filling is a limitation that occur when the vertical distribution of precipitation fields is not uniform over the volume scanned by the radar. This problem is illustrated in Figure C.12(a). It can be seen that precipitation close to the radar (i.e., in the center of the image) have higher count values than precipitation sampled far from it. In that example, precipitation is low-lying, widespread and relatively uniform along the radar

range. However, as the radar beam propagates downstream, through the atmosphere, the volume scanned increases. The amount of precipitation detected at close ranges, relatively to the volume scanned, is therefore larger than at far ranges. Ongoing research is carried out in order to retrieve the altitude of precipitation and improve beam-filling correction procedures (Pedersen et al., 2010).



(a) Partial beam-filling on the LAWR on May 12, 2010 at 13:00. The volume scanned increases with the range. It results in close range precipitation being better sampled than precipitation at far ranges.

(b) Example of anamalous propagation on the Rømø radar, on July 7, 2010 at 05:40. The beam of the radar is bent towards the sea and the ground, and falsely indicates the presence of precipitation.

Figure 12: Examples of other limitations and problems encountered by weather radars.

Another problem that affects radar observational accuracy is known as anomalous propagation. An example of how it appears on weather radar images is given in Figure C.12(b). It shows echoes likely caused by super-refraction of the radar beam. These echoes falsely indicate the detection of precipitation. It often occurs when there is a temperature inversion in the atmosphere, warm and moist air overlaying cool air. In Denmark, these situations are typical during the summer with southerly winds which bring moist continental air. In contrast to ground clutter, anomalous propagation is not characterized by a strong spatial stationarity. Furthermore, it may even grow and decay in the same way as light precipitation systems which makes it difficult to detect in an automated fashion.

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, Rasmussen C, Eriksen PB, Pedersen J. (2007) Technical aspects of

status and expected future trends for wind power in Denmark. *Wind Energy*, 10:31–49.

Atkinson B, Zhang J. (1996) Mesoscale shallow convection in the atmosphere. *Reviews of Geophysics*, 34:403–431.

Baldwin M, Kain J, Lakshmivarahan S. (2005) Development of an automated classification procedure for rainfall systems. *Monthly Weather Review*, 133:844–862.

Biggerstaff M, Listemaa S. (2000) An improved scheme for convective/stratiform echo classification using radar reflectivity. *Journal of Applied Meteorology*, 39:2129–2150.

Bøvith T. (2008) Detection of weather radar clutter. PhD Thesis, Department of Informatics and Mathematical Modelling, Technical University of Denmark, Kgs. Lyngby. (ISBN: 87-643-0436-1).

Cappelen J, Jørgensen B. (1999) Observed wind speed and direction in Denmark - with climatoligical standards normals, 1961-90. Technical Report, DMI - Danmarks Meteorologiske Institut.

Chèze J, Haase G. (2010) Impact study on radar observations by wind turbines - OPERA deliverable: OPERA-2010-05. Technical Report, EUMETNET.

Crum T, Alberty R. (1993) The WSR-88D and the WSR-88D operational support facility. *Bulletin of the American Meteorological Society*, 74:1669–1688.

Cutululis AN, Detlefsen N, Sørensen P. (2011) Offshore wind power prediction in critical weather conditions. *In Proceedings of the 10^{th} International Workshop on Large-Scale Integration of Wind Power into Power Systems, Aarhus, Denmark, 2011*.

Danish Energy Agency. (2007) Future offshore wind power sites - 2025. *Danish Energy Agency, Ministry of Climate and Energy* (ISBN: 978-87-7844-677-0).

Dixon M, Wiener G. (1993) TITAN: Thunderstorm Identification, Tracking, Analysis, and Nowcasting – A radar-based methodology. *Journal of Atmospheric and Oceanic Technology*, 10:785–797.

Dance S. (2004) Issues in high resolution limited area data assimilation for quantitative precipitation forecasting. *Physica D: Nonlinear Phenomena*, 196:1–27.

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

Giebel G, Brownsword R, Kariniotakis G, Denhard M, Draxl C. (2011) The state-of-the-art in short-term prediction of wind power: A literature overview. Technical Report, ANEMOS.plus.

Gill R, Overgaard S, Bøvith T. (2006) The Danish weather radar network. *In Proceedings of the 4^{t}h European Conference on Radar in Meteorology and Hydrology, Barcelona, Spain, 2006*.

Ghil M, Malanotte-Rizzoli P. (1991) Data assimilation in meteorology and oceanography. *Advances in Geophysics*, 33:141–266.

Hasager C, Peña A, Christiansen M, Astrup P, Nielsen M, Monaldo F, Thompson D, Nielsen P. (2008) Remote sensing observation used in offshore wind energy. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 1:67–79.

Harris M, Hand M, Wright A. (2006) Lidar for turbine control. National Renewable Energy Laboratory, NREL/TP-500-39154, Golden, CO, Tech. Rep.

Holleman I, Delobbe L, Zgonc A. (2008) Update on the European weather radar network (OPERA). *Proceedings of the 5th European Conference on Radar in Meteorology and Hydrology, Helsinki, Finland, 2008*.

Houze Jr R. (2004) Mesoscale convective systems. *Reviews of Geophysics*, 42.

Isom B, Palmer R, Secrest G, Rhoton R, Saxion D, Allmon T, Reed J, Crum T, Vogt R. (2009) Detailed observations of wind turbine clutter with scanning weather radars. *Journal of Atmospheric and Oceanic Technology*, 26:894–910.

International Energy Agency. (2007) State of the art of remote wind speed sensing techniques using Sodar, Lidar and satellites. *51^{s}t IEA Topical Expert Meeting*. Available online: `www.ieawind.org`.

Jirak I, Cotton W, McAnelly R. (2003) Satellite and radar survey of mesoscale convective system development. *Monthly Weather Review*, 131:2428–2449.

Jones L, Clark C. (2011) Wind integration - A survey of global views of grid operators. *In Proceedings of the 10th International Workshop on Large-Scale Integration of Wind Power into Power Systems, Aarhus, Denmark, 2011*.

Kristoffersen J. (2005) The Horns Rev wind farm and the operational experience with the wind farm main controller. *In Proceedings of the Offshore Wind International Conference and Exhibition, Copenhagen, Denmark,2005*.

Lakshmanan V, Smith T, Stumpf G, Hondl K. (2007) The Warning Decision Support System-Integrated Information. *Weather and Forecasting*, 22:596–612.

Lakshmanan V. (2012) Image processing of weather radar reflectivity data: Should it be done in Z or dBZ? *Electronic Journal of Severe Storms Meteorology*, 7:1–8.

Lakshmanan V, Zhang J, Hondl K, Langston C. (2012) A statistical approach to mitigating persistent clutter in radar reflectivity data. *IEEE Journal on Selected Topics in Applied Earth Observations and Remote Sensing*, Available online.

Laroche S, Zawadzki I. (1995) Retrievals of horizontal winds from single-Doppler clear-air data by methods of cross correlation and variational analysis. *Journal of Atmospheric and Oceanic Technology*, 12:721–738.

Li S. (2009) Markov Random Field modeling in image analysis. *Springer*.

Marshall J, Palmer W. (1948) The distribution of raindrops with size. *Journal of Atmospheric Sciences*, 5:165–166.

Meischner P. (2004) Weather radar: Principles and advanced applications. *Springer*.

Mikkelsen T, Hansen K, Angelou N, Sjöholm M, Harris M, Hadley P, Scullion R, Ellis G, Vives G. (2010) Lidar wind speed measurements from a rotating spinner. *In Proceedings of the European Wind Energy Conference and Exhibition, Warsaw, Poland, 2010*.

Otsu N. (1975) A threshold selection method from gray-level histograms. *Automatica*, 11:285–296.

Pedersen L, Jensen N, Madsen H. (2010) Calibration of Local Area Weather Radar - Identifying significant factors affecting the calibration. *Atmospheric Research*, 97:129–143.

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

Pinson P. (2012) Very short-term probabilistic forecasting of wind power time-series with generalized Logit-Normal distributions. *Journal of the Royal Statistical Society, Series C*, 61:555–576.

Rousseeuw P. (1984) Least median of squares regression. *Journal of the American Statistical Association*, 79:871–880.

Sempreviva A, Barthelmie R, Pryor S. (2008) Review of methodologies for offshore wind resource assessment in European seas. *Surveys in Geophysics*, 29:471–497.

Serafin R, Wilson J. (2000) Operational weather radar in the United States: Progress and opportunity. *Bulletin of the American Meteorological Society*, 81:501–518.

Sørensen P, Cutululis AN, Vigueras-Rodriguez A, Jensen L, Hjerrild J, Donovan M, Madsen H. (2007) Power fluctuations from large wind farms. *IEEE Transactions on Power Systems*, 22:958–965.

Steiner M, Smith J. (2002) Use of three-dimensional reflectivity structure for automated detection and removal of nonprecipitating echoes in radar data. *Journal of Atmospheric and Oceanic Technology*, 19:673–686.

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

Tuttle J, Foote G. (1990) Determination of the boundary layer airflow from a single Doppler radar. *Journal of Atmospheric and Oceanic Technology*, 7:218–232.

Van Delden A. (2001) The synoptic setting of thunderstorms in Western Europe. *Atmospheric Research*, 56:89–110.

Vincent C, Pinson P, Giebel G. (2011) Wind fluctuations over the North Sea. *International Journal of Climatology*, 31:1584–1595.

Vincent C, Hahmann A, Kelly M. (2012) Idealized mesoscale model simulations of open cellular convection over the sea. *Boundary-Layer Meteorology*, 142:103–121.

Zhao Q, Cook J, Xu Q, Harasti P. (2006) Using radar wind observations to improve mesoscale numerical weather prediction. *Weather and Forecasting*, 21:502–522.

PAPER D

Automatic classification of offshore wind regimes with weather radar observations

Authors:
P.-J. Trombe, P. Pinson and H. Madsen

Submitted to:
IEEE Journal Journal of Selected Topics in Applied Earth Observations & Remote Sensing (2012)

Automatic classification of offshore wind regimes with weather radar observations

Pierre-Julien Trombe[1], Pierre Pinson[1], Henrik Madsen[1],

Abstract

Weather radar observations are called to play an important role in offshore wind energy. In particular, they can enable the monitoring of weather conditions in the vicinity of large-scale offshore wind farms and thereby notify the arrival of precipitation systems associated with severe wind fluctuations. The information they provide could then be integrated into an advanced prediction system for improving offshore wind power predictability and controllability.

In this paper, we address the automatic classification of offshore wind regimes (i.e., wind fluctuations with specific frequency and amplitude) using reflectivity observations from a single weather radar system. A categorical sequence of most likely wind regimes is estimated from a wind speed time series by combining a Markov-Switching model and a global decoding technique, the Viterbi algorithm. In parallel, attributes of precipitation systems are extracted from weather radar images. These attributes describe the global intensity, spatial continuity and motion of precipitation echoes on the images. Finally, a CART classification tree is used to find the broad relationships between precipitation attributes and wind regimes.

1 Introduction

Unlike fossil fuels or nuclear energy, the availability of renewable sources of energy (e.g., solar, hydro, wind power) is directly governed by the dynamics of the atmosphere. It is therefore important to monitor weather conditions for assessing, forecasting and integrating these resources into power systems. In that respect, remote sensing observations of the atmosphere have become essential for the management of energy systems and, in offshore wind energy, they have already led to significant advances in a wide range of applications. These applications include the use of satellite SAR images for improving the accuracy of

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

wind maps over coastal areas, airborne SAR measurements for studying wake effects at large offshore wind farms, and LiDAR and SoDAR measurements for sampling vertical wind profiles (see Hasager et al. (2008) and references therein).

A new application of remote sensing tools in wind energy is now under experimentation at Horns Rev, in the North Sea. It consists of using weather radar observations for monitoring weather conditions in the vicinity of large-scale offshore wind farms (Trombe et al., 2012). This application is motivated by the need to improve offshore wind power predictability at high temporal resolutions (Jones and Clark, 2011). In particular, the high variability of offshore wind fluctuations is a serious problem for wind farm and transmission system operators because it increases the uncertainty associated with the short-term prediction of wind power (Akhmatov et al., 2007). Statistical analysis of wind data from Horns Rev showed that this variability was actually the result of frequent and sudden changes of wind regimes (i.e., wind fluctuations with specific frequency and amplitude) over waters (Pinson et al., 2008, Vincent et al., 2010). Subsequent analysis showed that large wind fluctuations tended to be coupled with specific climatological patterns and, particularly, the occurrence of precipitation (Vincent et al., 2011). This suggests that precipitation could be used as an early indicator for high wind variability. Our idea is thus to take advantage of the extended visibility provided by weather radars for notifying the arrival of precipitation systems in the vicinity of offshore wind farms, and adapting the forecasting strategy accordingly.

In view of integrating weather radar observations into wind power prediction systems, it is necessary to understand the precipitation settings associated with high wind variability at offshore sites. In some other meteorological contexts, the settings favoring the development of severe weather with the formation of precipitation are well documented (Bluestein and Jain, 1985, Bluestein et al., 1987). However, no detailed precipitation climatology over the North Sea exists to our knowledge. As a first step towards this understanding, we start by analyzing precipitation over the largest spatial scale enabled by the weather radar system used for monitoring the weather at Horns Rev, that is within a window of radius 240 km. Weather radar observations show that the passage of some meteorological phenomena producing precipitation was coupled with severe wind fluctuations while that of some other phenomena, also producing precipitation, was not (Trombe et al., 2012). Capturing the differences between precipitation systems by "eye" becomes increasingly difficult with the volume of data. This difficulty may further be increased by other factors such as (i) the relatively small range of single weather radar systems which only enables a partial observation of precipitation systems; (ii) seasonal variations of precipitation which implies that two similar events on weather radar images at two different times of the year may have different levels of severity. This calls for

the use of statistical classifiers for generating a consistent catalogue of situations where the variability of wind fluctuations is explained by attributes (i.e., characteristics) of precipitation systems.

Traditionally, classification applications using precipitation attributes aim at improving the understanding of precipitation itself. For instance, an automated classification procedure for rainfall systems is proposed in Baldwin et al. (2005). Alternatively, Lakshmanan and Smith (2009), Gagne et al. (2009) address the classification of precipitation objects (i.e., storms) that require to be defined and identified a priori. Yet, a major drawback of these approaches is that they rely on an expert training performed manually with its inherent shortcomings: (i) the potential lack of consistency since two experts may disagree on how to classify an event, or a same expert may classify two similar events differently; (ii) it is limited in the volume of data that can be treated. Our study differs in two aspects. First, the target variable is not precipitation but wind. And second, it does not require any expert training for the classification and therefore avoid the aforementioned shortcomings. Instead, a categorical sequence of wind regimes is automatically estimated from a wind speed time series by combining a global decoding algorithm, the Viterbi algorithm (Forney, 1973), with the Markov-Switching model proposed in Pinson et al. (2008). In parallel, a number of precipitation attributes are computed from weather radar images. These attributes describe the global intensity, spatial continuity and motion of precipitation echoes on the images. Finally, a CART classification tree, is used for finding relationships between precipitation attributes and wind regimes observed at Horns Rev. The motivation for using such a classification technique is that it can explore large amounts of data and, yet, produce a simple partition with interpretable rules (Hastie et al., 2001).

The rest of the paper is organized as follows. In Section 2, we describe the data. In Section 3, we give an overview of the procedure for extracting the most likely sequence of regimes from wind speed time series. In Section 4, we compute a number of precipitation attributes from weather radar images. In Section 5, we present the classification tree technique and apply it to the problem of the automatic classification of offshore wind regimes. Finally, Section 6 delivers concluding remarks.

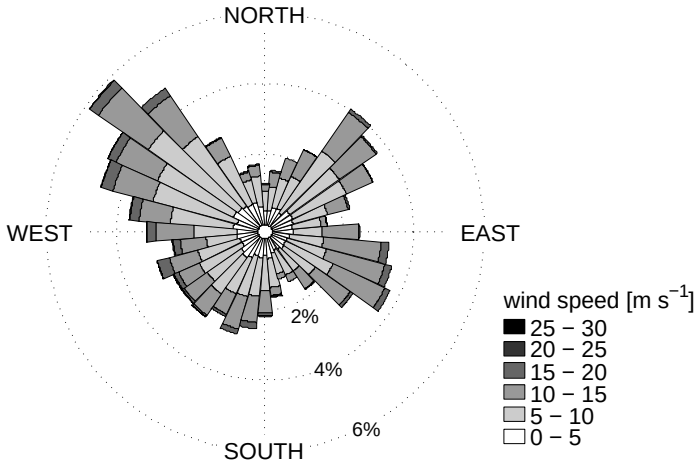
2 Data

2.1 Wind data

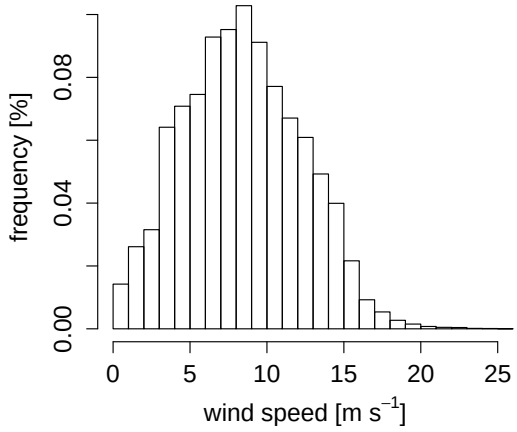
Wind data were collected from the nacelle anemometry and SCADA systems of the Horns Rev (HR1) wind farm (Kristoffersen, 2005). The original measurements consisted of individual time series of wind speed and wind direction, for each of the 80 wind turbines of HR1. Two aggregated time series of wind speed and wind direction were obtained by jointly averaging these individual time series over 10 minute intervals. The time series span the year 2010. Due to some technical problems, measurements are missing over 2 periods of about 5 and 12 days, respectively. There are 2664 missing values out of 52560 (i.e., 94.9% of data availability). No attempt was made to fill in those gaps. The wind distribution is shown in Figure 1. The wind rose shows 3 preferred wind directions. While the prevalence of northwesterly directions is consistent with other wind data analysis at Horns Rev (see Vincent et al. (2010)), the frequent occurrences of northeasterly winds are more exceptional since it is usually the direction where the wind is suppressed in Denmark. This phenomenon can be explained by a strong annual wind variability in 2010. Note also that strong winds, above 15 m s^{-1}, are more frequent for westerly than easterly directions.

2.2 Weather radar data

Weather radar data consist of 2D images of precipitation reflectivity. More specifically, they correspond to 1 km height pseudo-CAPPI (Constant Altitude Plan Position Indicator) image products, with a 2×2 km grid resolution. They were produced by a C-Band Doppler radar located in Rømø, approximately 57 km to the East of the HR1 wind farm. The radar is operated by the Danish Meteorological Institute (DMI), using a 9 elevation scan strategy and an operational range of 240 km (Gill et al., 2006). One image is generated every 10 minutes. Clutter removal filters are applied during the data acquisition process. Data quality control is also performed a posteriori and persistent clutter is removed following the automatic method introduced in Lakshmanan (2012). For a complete description of the radar settings and images, we refer to Trombe et al. (2012). About 2000 images are missing over the year 2010 (i.e., 96.1% of data availability).



(a) Wind rose. Angles indicate the direction from which the wind blows (meteorological conventions).



(b) Frequency histogram of wind speed.

Figure 1: Wind distribution at the Horns Rev 1 wind farm, in 2010. Data were collected from the nacelle anemometry and SCADA systems (Kristoffersen, 2005).

3 Estimation of wind regimes

In this section, we estimate a categorical sequence of wind regimes from the time series of wind speed presented in Section 2. Such a procedure can also be viewed as a segmentation of the time series where the latter is partitioned into

homogeneous sections. Our plan is to use this sequence of wind regimes as the dependent variable (i.e., the variable to predict) for growing a classification tree in Section 5.

Numerous studies have pointed out the nonstationary behavior of offshore or near-offshore wind fluctuations at the minute scale (Cripps and Dunsmir, 2003, Vincent et al., 2010, Davy et al., 2010). Numerically, this nonstationarity translates into sudden shifts in the amplitude and/or frequency of wind fluctuations. Such patterns of fluctuations can be analyzed either in the frequency domain, with an empirical spectral decomposition technique as in Vincent et al. (2010), or in the time domain with Generalized AutoRegressive Conditional Heteroskedasticity (GARCH) models (Cripps and Dunsmir, 2003), or Markov-Switching AutoRegressive (MSAR) models (Pinson et al., 2008). The advantage of MSAR models over other techniques is that they are clearly tailored to address the extraction of a hidden sequence of regimes, as discussed in Zucchini and MacDonald (2009).

3.1 Regime-switching modeling with MSAR models

MSAR models are an extension of Hidden Markov Models (HMM). They are widely used for the modeling of time series characterized by structural breaks in their dynamics. The underlying assumption of these models, both HMM and MSAR, is that there is an unobservable Markov process which governs the distribution of the observations (Zucchini and MacDonald, 2009). Compared to HMM, MSAR models have an additional capability, they can accommodate autocorrelated data and include autoregressors in the model formulation. Applications of MSAR models to wind data include Ailliot and Montbet (2012), Pinson et al. (2008).

The wind speed time series we use for this study does not show any well pronounced diurnal cycle. In addition, we disregard the potential long-term drift and seasonal variations of wind speed since the available time series only spans a one year period. For the sake of simplicity, we do not specifically deal with the wind speed truncation in 0. We only assume that wind speed has an autoregressive behavior in each regime. Let $\{y_t\}, t = 1, \ldots, n$, be the time series of measured wind speed at the HR1 wind farm. The MSAR model with m regimes and autoregressive orders $(p_1, \ldots, p_m)$ is defined as follows:

$$Y_t = \boldsymbol{\theta}^{(Z_t)T} \boldsymbol{X}_t + \sigma^{(Z_t)} \varepsilon_t \tag{1}$$

with

$$\boldsymbol{\theta}^{(Z_t)} = [\theta_1^{(Z_t)} \quad \dots \quad \theta_{p_{Z_t}}^{(Z_t)}]^T \tag{2}$$

$$\boldsymbol{X}_t = [Y_{t-1} \quad \dots \quad Y_{t-p_{Z_t}}]^T \tag{3}$$

where $\{\varepsilon_t\}$ is a sequence of independently distributed random variables following a Normal distribution $\mathcal{N}(0,1)$; and $\boldsymbol{Z} = (Z_1,\dots,Z_n)$ is a first order Markov chain with a discrete and finite number of states (i.e., regimes) m and transition probability matrix $\boldsymbol{P}$ of elements $(p_{ij})_{i,j=1,\dots,m}$ with:

$$p_{ij} = Pr(Z_t = j | Z_{t-1} = i), \quad i,j = 1,\dots,m \tag{4}$$

$$\sum_{j=1}^{m} p_{ij} = 1, \quad i = 1,\dots,m \tag{5}$$

There exist two distinct methods for estimating the parameters of a MSAR model with given number of regimes m and autoregressive orders $(p_1,\dots,p_m)$, the Expectation-Maximization (EM) algorithm and direct numerical maximization of the Likelihood. The respective merits of these 2 methods are discussed in Zucchini and MacDonald (2009), along with practical solutions for their implementation. As for this study, we estimate MSAR models by direct numerical maximization of the Likelihood owing to its lower sensitivity to starting values. Let $\Theta = (\boldsymbol{\theta}^{(1)},\dots,\boldsymbol{\theta}^{(m)},\boldsymbol{P},\boldsymbol{\sigma})$ be the set of parameters to estimate. The Maximum Likelihood Estimator (MLE), $\widehat{\Theta}_{MLE}$, is obtained by maximizing the Likelihood function $L(\Theta)$:

$$\widehat{\Theta}_{MLE} = \arg\max_{\Theta} \mathcal{L}(\Theta|\boldsymbol{Y}) \tag{6}$$

$$= \arg\max_{\Theta} \boldsymbol{\delta}(\prod_{t=1}^{n} \boldsymbol{P}\boldsymbol{D}_t)\mathbf{1}^T \tag{7}$$

where

$$\boldsymbol{\delta} = \mathbf{1}(\boldsymbol{I}_m - \boldsymbol{P} + \boldsymbol{U}_m)^{-1} \tag{8}$$

$$\boldsymbol{D}_t = \mathrm{diag}(\eta(t,1),\dots,\eta(t,m)) \tag{9}$$

$$\eta(t,i) = \frac{1}{\sigma^{(i)}}\phi\left(\frac{Y_t - \boldsymbol{\theta}^{(i)T}\boldsymbol{X}_t}{\sigma^{(i)}}\right), \quad i = 1,\dots,m \tag{10}$$

$\boldsymbol{\delta}$ is the stationary distribution of the Markov chain; $\mathbf{1}$ is a unit vector of size m; $\boldsymbol{I}_m$ and $\boldsymbol{U}_m$ the Identity and Unity matrices of size $m \times m$; $\boldsymbol{D}_t$ a diagonal matrix; and ϕ the probability density function of the Normal distribution.

We estimate four MSAR models, from one up to four regimes. For each of these MSAR models, the optimal autoregressive orders in each regime are de-

termined by following a forward selection procedure based on Likelihood Ratio (LR) tests, as described in Bacher and Madsen (2011). Then, all four models are compared with one another by performing LR tests, leading to the rejection of the MSAR model with four regimes. For MSAR models from one to three regimes, Table 1 summarizes some of the important parameter estimates that help interpreting the regimes. In particular, the elements of the diagonal of the transition probability matrix, $\text{diag}(\boldsymbol{P})$, give an estimation of the mean persistence of the regimes over time. As for the vector of standard deviations $\boldsymbol{\sigma}$, it expresses the relative variability of wind speed fluctuations in each regime. The estimates of the autoregressive coefficients are of lesser importance and, instead, we just report the optimal autoregressive order in each regime. Regimes are ranked by ascending values of standard deviation. Both with 2 and 3 regimes, there is an inverse relationship between wind fluctuation variability and persistence (i.e., the more variable, the less persistent).

Table 1: Summary statistics on msar models fitted to the time series of wind speed.

m	$(p_1,\ldots,p_m)$	$\text{diag}(\boldsymbol{P})$	$\boldsymbol{\sigma}$
1	5	-	0.51
2	(5,5)	(0.98, 0.92)	(0.31, 0.96)
3	(4,3,6)	(0.98, 0.95, 0.89)	(0.25, 0.47, 1.28)

3.2 Global decoding

Global decoding consists of estimating the most likely sequence of regimes $\widehat{\boldsymbol{z}} = (\hat{z}_1,\ldots,\hat{z}_n)$ under a fitted model, as opposed to *local decoding* which consists of estimating the most likely regime at time t, $\hat{z}_t$, independently of the regime values at other times. The most likely sequence of regimes $\widehat{\boldsymbol{z}}$ is found by maximizing the joint probability of the observations and states of the Markov chain:

$$\widehat{\boldsymbol{z}} = \arg\max_{\boldsymbol{z}} \Pr(\boldsymbol{Z} = \boldsymbol{z}, \boldsymbol{Y} = \boldsymbol{y}) \tag{11}$$

where $\boldsymbol{Y} = (Y_1,\ldots,Y_n)$. For estimating $\widehat{\boldsymbol{z}}$, we use the Viterbi algorithm Forney (1973). For that purpose, let us introduce the following notations:

$$\boldsymbol{Y}^{(i)} = (Y_1,\ldots,Y_i) \quad \text{and} \quad \boldsymbol{Z}^{(i)} = (Z_1,\ldots,Z_i) \tag{12}$$

$$\xi_{1i} = \Pr(Z_1 = z_1, Y_1 = y_1) = \delta_i \eta(1,i) \tag{13}$$

$$\xi_{ti} = \max_{\boldsymbol{z}^{(t-1)}} \Pr(\boldsymbol{Z}^{(t-1)} = \boldsymbol{z}^{(t-1)}, Z_t = i, \boldsymbol{Y}^{(t-1)} = \boldsymbol{y}^{(t-1)}) \tag{14}$$

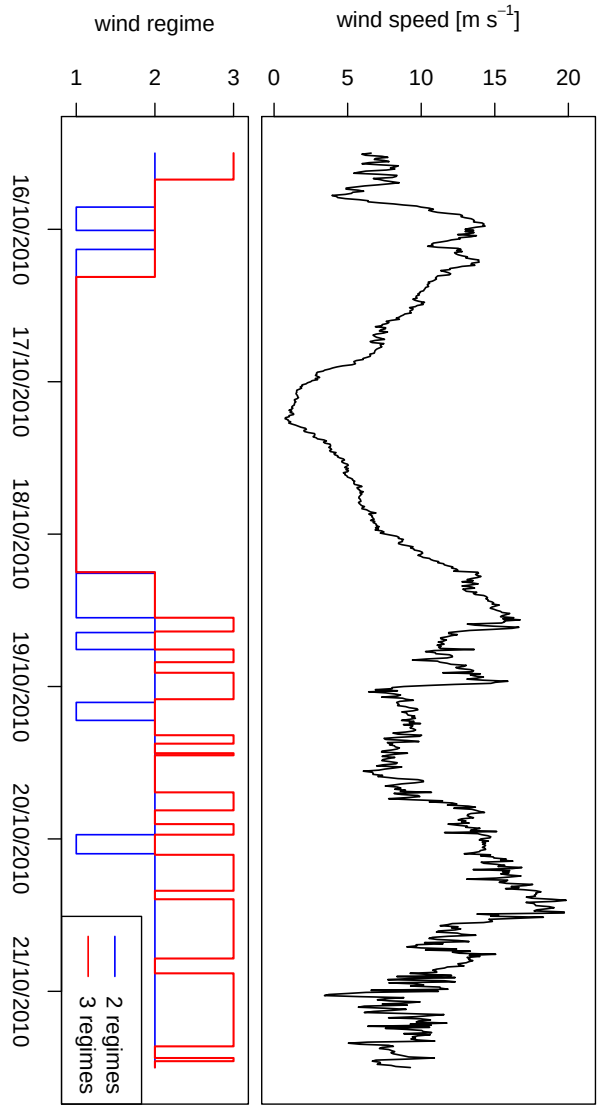


Figure 2: (Upper panel) Time series of wind speed recorded at the Horns Rev 1 wind farm. The temporal resolution is 10 minutes. (Lower panel) Estimated sequence of regimes, for 2 and 3 regimes. Regimes can be interpreted in terms of wind variability, from low in Regime 1 to high variability in Regime 3.

for $t = 2, \ldots, n$. The quantities ξ_{ti} can be seen as the most probable sequence leading to regime i at time t, among all possible sequences $\boldsymbol{Z}^{(t-1)}$. Finally, $\widehat{\boldsymbol{z}}$ is

found by the solving the following backward recursion, starting from n:

$$\hat{z}_n = \arg \max_{i=1,\ldots,m} \xi_{ni} \tag{15}$$

$$\hat{z}_t = \arg \max_{i=1,\ldots,m} \xi_{ti} p_{i,\hat{z}_{t+1}} \quad \text{for} \quad t = n-1,\ldots,1 \tag{16}$$

The most likely sequence of wind regimes was computed under the fitted MSAR models, with both 2 and 3 regimes. The result is illustrated in Figure 2 over a 6 day episode where a clear change of wind speed variability, from low to high, can be observed on October 18, 2010. Note that the regimes are more stable (i.e., there are fewer switchings) for the sequence with 2 regimes than with 3.

4 Precipitation identification and attributes

4.1 Precipitation identification

Weather radar images can contain 2 sources of information which fall either into the meteorological sources (e.g., rain, hail, snow) or into non-meteorological sources (e.g., clutter due to buildings, wind farms, ground, sea). Echoes caused by non-meteorological targets can usually be identified and filtered out during the data acquisition process or a posteriori data quality control when they have non-random patterns (see Bøvith (2008) for illustrative examples on the Danish weather radar networks). However, not all non-meteorological echoes can be removed and, in some cases, significant portions of weather radar images remain contaminated by non-meteorological artifacts (Trombe et al., 2012). Regarding the images used in this study, the most serious problems are due to anomalous propagation (anaprop) of the radar beam. We observe these problems more frequently during the summer season, from April to September in Denmark. In some extreme cases, the contamination can extend up to 20% of the image pixels over several hours. Image pre-processing operations such as median filtering are inefficient for removing anaprop echoes.

In this subsection, our goal is to develop a method for assigning a binary label to each image indicating the detection of precipitation (potentially mixed with noisy echoes) or not. In Wheater et al. (2000), rainfall is identified by computing the proportion of wet pixels (i.e., pixels recording positive rainfall) over the entire image. A rainfall event is then defined as a continuous period of time where the coverage proportion of wet pixels over the whole image is above a threshold of 25%. This approach is clearly an over-simplified view of the problem and could not apply to our images, even by optimizing the

threshold level. In other applications and, particularly, severe weather nowcasting, storm identification is addressed by defining thresholding and contiguity heuristics (Johnson et al., 1998). These later methods are tailored for very specific types of precipitation being depicted by high reflectivity echoes on weather radar images.

We propose an alternative method for identifying precipitation, irrespectively of the mean reflectivity. It is based on the assumption that contiguous pixels recording precipitation have a higher correlation than contiguous pixels contaminated by noise. This assumption is supported by Zawadzki (1973) which shows that the shape of precipitation echoes tends to be elliptical. We use a geostatistical tool, the correlogram, as a measure of spatial correlation of precipitation echoes for each image (Isaaks and Srivastava, 1889). In order to capture the potential anisotropy of precipitation echoes, these correlograms are produced in 2 dimensions, based on the estimation of directional correlograms $\rho(\boldsymbol{h})$ of vector $\boldsymbol{h}$ as follows:

$$\rho(\boldsymbol{h}) = \frac{\gamma(\boldsymbol{h})}{\gamma(0)} \tag{17}$$

$$\gamma(\boldsymbol{h}) = \frac{1}{N(\boldsymbol{h})} \sum_{(p_i,p_j)|h_{p_ip_j}=\boldsymbol{h}} (I_{p_i} - I_{p_j})^2 \tag{18}$$

where $\gamma(\boldsymbol{h})$ is a directional variogram computed by summing over all paired pixels (p_i, p_j) with intensities (I_{p_i}, I_{p_j}) and separated by a vector $\boldsymbol{h}$. $N(\boldsymbol{h})$ is the number of paired pixels (p_i, p_j) matching this latter criterion. These 2-dimensional correlograms are computed with the *gstat* package of the R programming environment (Pebesma, 2004).

Figure 3 shows 4 sample images and their associated correlograms. A zoom in the central part of the correlogram is also provided for illustrating the local continuity of reflectivity values. The images were chosen to reflect various types of precipitation systems (e.g., small and scattered precipitation cells, banded or widespread precipitation system) and a case of anaprop. In particular, the small spatial correlation of anaprop echoes can well be observed, it drops below 0.4 for all 1-lagged (i.e., adjacent) pixels, whatever the direction. Note also the quick decorrelation in space for small scattered cells but, unlike for anaprop, the spatial correlation is larger than 0.4 up to 3-4 lagged pixels. The anisotropy of banded systems can also be well be captured by these correlograms.

For a given image, we consider that precipitation is detected if the correlation is larger than 0.6 for all 1 and 2-lagged pixels (i.e., the central 5x5 neighborhood of the correlogram). Then, we define a precipitation event as a period with a minimum duration of 1 hour (i.e., 6 consecutive images) over which precipi-

tation is detected. If the time between the end of a precipitation event and the beginning of a new one is less than one hour, we consider it to be the same event. Table 2 summarizes the number of events identified and their mean lifetime in 2010.

Table 2: Monthly statistics: Number of precipitation events in 2010 and their mean lifetime

	Jan.	Feb.	March	Apr.	May	June	July
Number of events	23	19	40	20	32	18	24
Mean lifetime [hours]	22.7	24.2	08.6	14.9	11.1	12.3	20.4
	Aug.	Sept.	Oct.	Nov.	Dec.	Total	
Number of events	25	20	24	36	21	302	
Mean lifetime [hours]	21.9	20.5	21.2	16.2	20.4	17.1	

4.2 Precipitation types

Precipitation is commonly described as either stratiform, convective or a mix of these two. In the mid-latitudes, stratiform precipitation develops in a variety of situations where the atmosphere is stably stratified. Typical examples of these situations are warm fronts where masses of warm air gradually lift over cold masses of air. These fronts have the particularity of propagating relatively slowly and spreading over large horizontal scales up to and beyond 100 km. On weather radar images in 2D, stratiform precipitation is thus generally identified as a widespread region of moderate, homogeneous and continuous intensity with a slow dynamics. Winds associated with pure stratiform precipitation usually have a small vertical velocity and low turbulency. In comparison, convective precipitation develops in unstable atmosphere and have a much higher spatial variability, with many scattered and heavy precipitation showers occurring locally, over horizontal scales from a few kilometers up a few tens of kilometers, potentially forming complex convective systems over several hundreds kilometers. In addition, the updraft associated with this type of precipitation is stronger, resulting in highly turbulent winds. In the mid-latitudes, convective precipitation prevails during the summer and over warm oceans. On weather radar images in 2D, convective precipitation is depicted by small clusters of high reflectivity propagating relatively quickly. However, in many cases, convective precipitation can be embedded into stratiform regions and forms more complex precipitation structures.

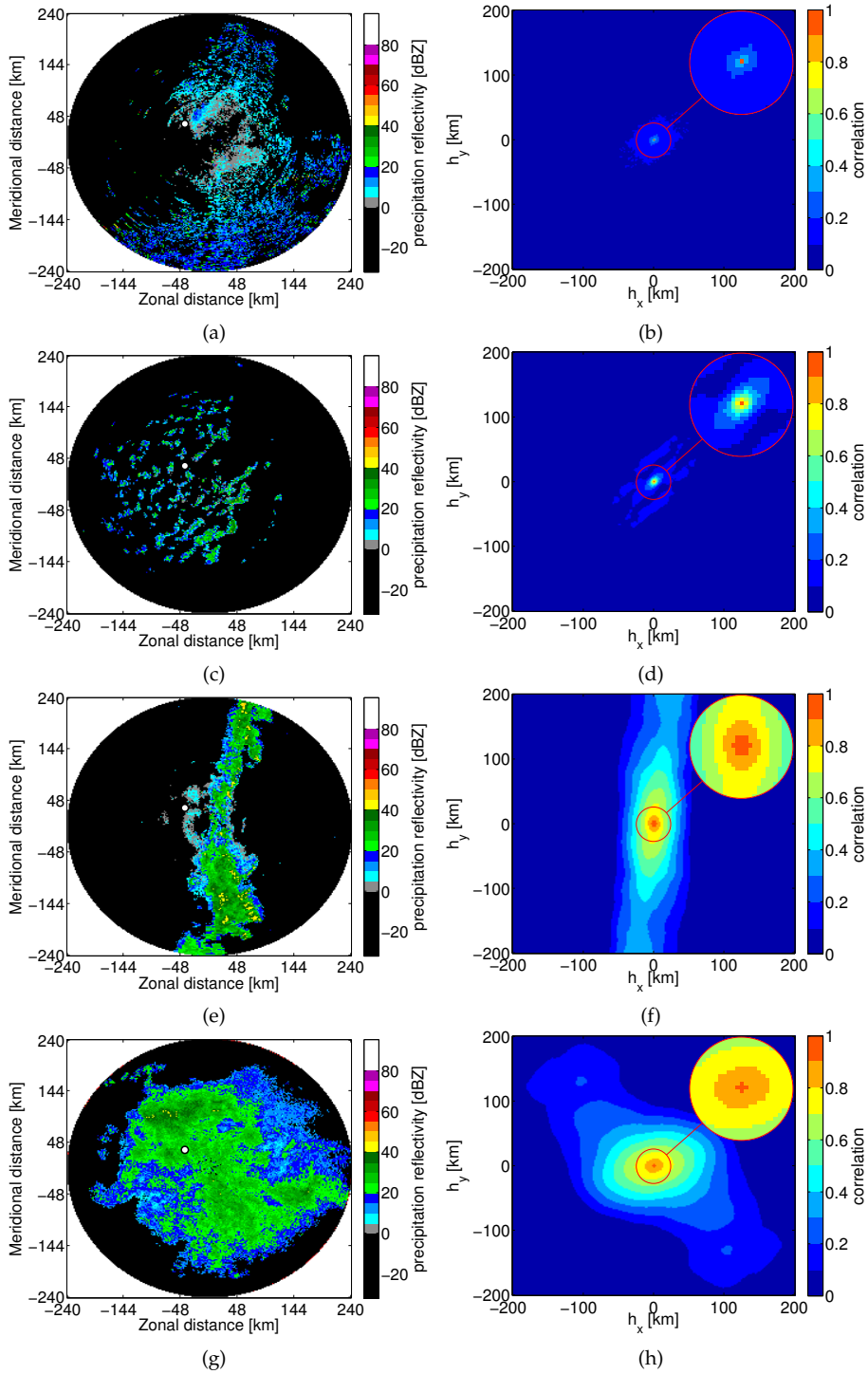


Figure 3: Image samples (left column) and their associated correlograms in 2 dimensions (right column). (a-b) A case of anomalous propagation without precipitation. (c-d) Small scattered convective precipitation cells. (e-f) Banded precipitation system. (g-h) Widespread precipitation system.

4.3 Precipitation attributes

For each image where precipitation is identified, we compute a number of attributes linked the global intensity, spatial continuity and motion of precipitation. These attributes are meant to describe the main characteristics of the different types of precipitation discussed hereabove. They are summarized in Table 3.

On weather radar images, the intensity of precipitation is measured in decibel of reflectivity (dBZ). Within a same precipitation system, the distribution of intensity may not be homogeneous and, with the occurrence of severe weather traditionally associated with high values of reflectivity, it tends to be positively skewed. So, in order to describe the distribution of precipitation intensity, we propose a set of non parametric statistics composed of (i) location measures with the median (i.e., the 50^{th} quantile), the 75^{th}, 90^{th}, 95^{th} and 99^{th} quantiles; (ii) dispersion measures with the interquartile range (i.e., the range between the 25^{th} and 75^{th} quantiles); (iii) shape measures with the skewness to inform on the asymmetry of the distribution, and the kurtosis to inform on its sharpness. Only pixels with strictly positive reflectivity values are considered. Note that we choose to use robust statistics with, for instance, the median in place of the mean and the 99^{th} quantile in place of the maximum in order to filter out the potential effects of residual noise.

For measuring the spatial continuity of precipitation, we again use the correlogram introduced in this Section and follow the procedure presented in Baldwin et al. (2005). It assumes that each correlogram contains an elliptical object that can be described by its eccentricity and area. The procedure is as follows: (1) the correlogram is transformed into a binary image by means of a thresholding operation, with the threshold value arbitrarily chosen between 0 and 1; (2) a connected-component labelling algorithm is used to identify all connected regions on the binary image (Suzuki et al., 2003) and only the region intersecting with the center of the image is kept; (3) the edge of that region is identified with the Canny edge detector (Canny, 1986); (4) an ellipse is fitted on the detected edge by minimizing the least square criterion (Fitzgibbon et al. , 1999). In this study, this procedure is performed twice, for threshold values of 0.4 and 0.7, and the eccentricity (i.e., the ratio of the major axis over the minor axis) and the area of the elliptical object are computed for both values. For the threshold value of 0.4, these attributes are likely to reflect the large-scale continuity of precipitation whereas, for the value of 0.7, they will capture the more local continuity.

The horizontal motion of precipitation is computed with an optical flow method. This type of method is very useful for estimating the visible flow field (u, v) between 2 consecutive images. The two underlying assumptions that define the

optical flow formulation are *brightness constancy* and *spatial smoothness*. Brightness constancy means that the intensity of an object is conserved over time, despite its likely change of position between 2 consecutive images. Spatial smoothness refers to the coherence between neighboring pixels which should ideally have a similar motion (Sun et al., 2010). The formulation we use is the one proposed in Black and Anandan (1996) owing to its robustness to outliers. It is expressed as an Energy minimization problem with the objective function $E(u,v)$ defined as follows:

$$E(u,v) = E_{BC}(u,v) + \lambda E_{SS}(u,v) \tag{19}$$

where λ a regularization parameter (i.e., the larger λ, the denser the flow field); E_{BC} and E_{SS} are the functions resulting from the brightness constancy and spatial smoothness constraints:

$$E_{BC}(u,v) = \sum_{i,j} f(I_1(i,j) - I_2(i+u_{i,j}, j+v_{i,j})) \tag{20}$$

$$\begin{aligned} E_{SS}(u,v) = \sum_{i,j} [&g(u_{i,j} - u_{i+1,j}) + g(u_{i,j} - u_{i,j+1}) \\ &+ g(v_{i,j} - v_{i+1,j}) + g(v_{i,j} - v_{i,j+1})] \end{aligned} \tag{21}$$

where I_1 and I_2 are 2 consecutive images, f and g are 2 penalty functions. Following the implementation of Black and Anandan, we set $f = g = log(1 + \frac{1}{2}(\frac{x}{\sigma})^2)$, the Laurentzian function with scale parameter σ. The expression of E_{SS} is formulated with a pairwise Markov Random Field (MRF) discretization, based on a 4-neighborhood (Li, 2009). Since our goal is to estimate a unique speed and direction for each pair of consecutive images, we extract the median Cartesian flow from the flow field and convert it into its Polar components (i.e., speed and direction). Flow direction is then transformed into a categorical variable by binning its values into 8 sectors (North (N), North-East (NE), East (E),...).

Finally, we also add a seasonal attribute in the form of a categorical variable to allow for potential seasonal patterns of precipitation. We consider that there are only two seasons in Denmark so that the variable takes value *Summer* from March to August, and *Winter* from September to February. In summer, the North Sea is on average colder than the air whereas, in winter, the opposite holds true and favors thermal instabilities in the atmosphere (Vincent et al., 2011).

Table 3: Description of precipitation attributes used for growing the classification tree.

Attribute acronyms	Type (source)	Unit	Description
skew & kurt	Intensity (reflectivity images)	-	Skewness and Kurtosis of reflectivity distribution
q50, q75, q90, q95 & q99	Intensity (reflectivity images)	dBZ	50^{th}, 75^{th}, 90^{th}, 95^{th} & 99^{th} reflectivity quantiles
iqr	Intensity (reflectivity images)	dBZ	Interquartile range (range defined by the 25^{th} and 75^{th} reflectivity quantiles)
speedOF	Motion (optical flow)	m s^{-1}	Median speed of the flow field
dirOF	Motion (optical flow)	N, NE, E, SE, S, SW, W, NW	Median direction (8 sectors) of the flow field. Direction are in meteorological conventions, they indicate the direction of origin.
spaArea04, spaArea07	Spatial continuity (correlogram)	km^2	Area of the ellipse fitted on correlograms for threshold values 0.4 and 0.7
spaEcc04, spaEcc07	Spatial continuity (correlogram)	-	Eccentricity of the ellipse fitted on correlograms for threshold values 0.4 and 0.7
season	Temporal	Sm./Wt.	Summer (from April to September), Winter (from October to March)

5 Automatic Classification

For the automatic classification of precipitation systems, we use a tree-based classification technique called CART, in a supervised learning framework (i.e., the classification is governed by the categorical sequence of wind regimes computed in Section 3). These trees, also known as decision trees, are attractive in many aspects. First, for the relative simplicity of their principles based on a recursive partitioning of the data set. Second, they provide a powerful alternative to more traditional classification techniques (e.g., discriminant analysis and logistic regression) which generate a global model for the entire data set while variables may interact in a highly complex and nonlinear way and require to be fitted locally. Finally, because their interpretation is mainly visual and can lead to a straightforward understanding of the relationships between variables (Hastie et al., 2001). Applications of classification trees to precipitation data extracted from weather radar images can be found in Gagne et al. (2009) and Lakshmanan and Smith (2009).

5.1 CART classification trees

Let Y be the dependent categorical variable taking values $1, 2, \ldots, K$, and $(X_1, \ldots, X_p)$ the set of p predictors (i.e., the independent variables) that can either be continuous or categorical. Growing a classification tree consists of a recursive partitioning of the feature space (i.e., the space composed of the p predictors each with n observations) into rectangular areas. Each split consists of a dichotomy applied on a single predictor (e.g., $X_2 < 3$ if X_2 is continuous or $X_2 = "a"$ if it is categorical). The feature space is first split into 2 groups so that the response of Y is maximized in each of the 2 groups. This procedure is recursively repeated and each of the 2 groups is partitioned into 2 new sub-groups, and so on. Splits are more commonly called *nodes*. A terminal node (i.e., node that cannot be further split) is called a *leaf*.

For each node, the splitting predictor and rule are determined so as to minimize the impurity level in the resulting two nodes. For a given node, let $p = (p_1, \ldots, p_K)$ be the vector of proportions of elements in class $1, \ldots, K$. There exist several impurity measures and the one we use in this study is known as the Gini index. It measures how often a randomly chosen element from the node would be incorrectly labeled if it were labeled according to the frequency distribution of labels in the node. The Gini index $i_G(p)$ is computed as follows:

$$i_G(p) = 1 - \sum_{j=1}^{K} p_j^2 \tag{22}$$

When growing a tree, the tradition is to build a complex tree and simplify it by pruning (i.e., removing the nodes that over-fit the feature space). This is done by minimizing the misclassification rate within leaves over a 10-fold cross-validation procedure.

5.2 Experimental results

The classification is performed using the sequence of wind regimes computed in Section 3 as the dependent variable, and the precipitation attributes extracted from the weather radar images and listed in Table 3 as predictors. Observations where no precipitation is detected are filtered out. After that, more than 29000 observations remain for the classification. We choose to grow the tree for the sequence of wind regimes with 2 regimes. There are 76% of observations in regime 1 and 24% in regime 2. The final tree is shown in Figure 4. Branches going downwards to the left indicate that the splitting rule is satisfied.

The classification tree we grew is interesting in two aspects. First, it reveals the broad patterns of precipitation systems associated with the different wind regimes. For instance, the leftmost leaf which contains 35% of the total number of observations, shows that 93% of the observations for which the speed of precipitation echoes is smaller than 12 m s^{-1} (i.e., speedOF$<$12) and the maximum reflectivity is smaller than 29 dBZ (i.e., reflQ99$<$29) are in Regime 1. On the opposite side of the tree, the rightmost leaf which contains 14% of the total number of observations, indicates that 59% observations for which the speed is larger than 12 m s^{-1}, the maximum reflectivity larger than 30 dBZ and the precipitation comes from North-West, West or South are in Regime 2. One recurrent pattern in this tree is that when precipitation systems comes from North-East, East or South-East, wind fluctuations tend be classified in Regime 1, the regime with the lowest variability. This is consistent with the results in Vincent et al. (2011) that show that wind fluctuations are more variable for westerly flows than for easterly flows.

Secondly, the tree highlights the predictive power of each of the variables used in the classification. Some variables may repeatedly be used for generating new nodes whereas some other variables may not be used at all. This contrasts with the hierarchical clustering technique proposed in Baldwin et al. (2005) where all variables equally contribute to classify observations, with the risk of including non informative variables and degrading the accuracy of the classification. In the present experiment, one can notice that only 4 predictors are used in the final tree, the motion speed and direction of precipitation echoes (i.e., speedOF and dirOF), the season and the maximum reflectivity (i.e., reflQ99). Note that the maximum reflectivity value (i.e., reflQ99) is the only intensity related at-

tribute used in the final tree. This attribute characterizes the most extreme, yet marginal, intensity recorded on the images, highlighting the necessity to consider precipitation information at smaller scales in the future. Moreover, none of the 4 variables derived from the correlogram (i.e., spaArea04, spaArea07, spaEcc04 and spaEcc07) is used. The most likely reason for the small predictive power of correlograms is the too complex organizational structure of precipitation systems. In particular, when there are spatial discontinuities between precipitation echoes (i.e., precipitation echoes are separated by regions recording no precipitation), correlograms are only informative locally and cannot capture the full extent of the precipitation system. Inversely, when small clusters of high intensity are embedded into a large and continuous region of moderate intensity, correlograms tend to only capture the large-scale feature. This suggests the development of hierarchical techniques where precipitation would be analyzed at multi-scale, as a potential line of work in the future.

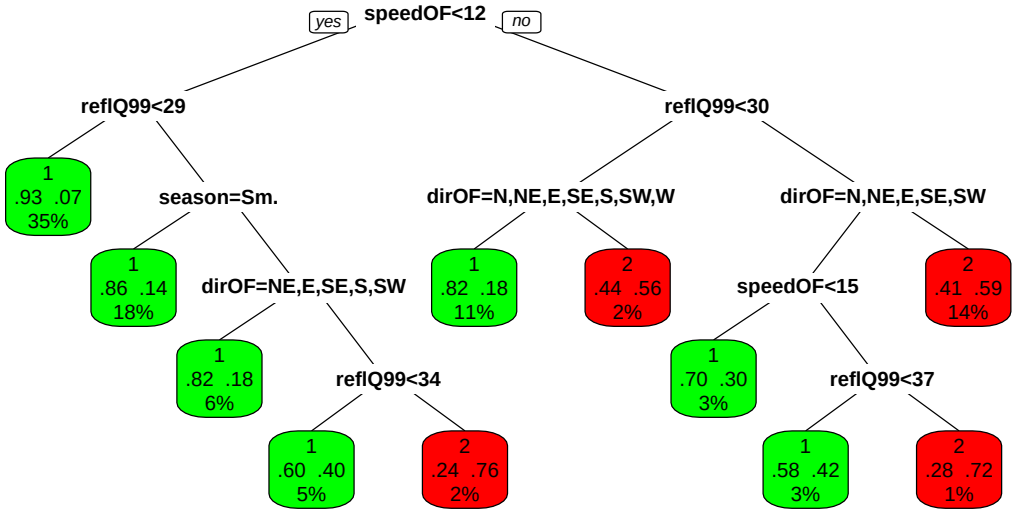


Figure 4: Classification tree explaining wind regimes at the Horns Rev 1 wind farm with precipitation system attribute extracted from weather radar images. Wind regimes and precipitation system attributes are computed in Section IV and V, respectively.

6 Conclusion

In this work, we proposed an automatic procedure for classifying offshore wind regimes based on precipitation attributes extracted from weather radar images. We found that winds with a high variability are more likely to be ob-

served with the passage of precipitations systems being advected at relatively high speeds, preferably from West and North-West, and having large maximum reflectivity values. This result is consistent with earlier data analysis Vincent et al. (2011) and confirms the potential of weather radar observations for providing appropriate information to future wind power prediction systems. However, the insights we gained on the relationship between precipitation and wind are not readily integrable into prediction systems.

We studied wind fluctuations in a univariate framework, only considering wind speed. It has the merit of keeping the complexity of the procedure reasonable. Yet, wind should ideally be considered and treated as a bivariate process of speed and direction because patterns of wind speed fluctuations may either be direction-dependent or coupled with specific patterns of wind direction fluctuations. For instance, larger wind speed fluctuations are observed for westerly flows at Horns Rev (Vincent et al., 2011). However, the statistical modeling of circular time series (e.g., wind direction) that feature autocorrelation is quite cumbersome and it is preferable to transform wind speed and direction into their associated (u,v) components, as in Cripps and Dunsmir (2003), for instance. That way, both variables of the bivariate process are non-circular and unbounded, and traditional methodologies can be applied. In that view, an interesting generalization of our work could consist of applying MSAR models in a vectorial form as introduced in Krolzig (1997), on the transformed (u,v) components of the wind.

As for precipitation, we considered it over a unique and large spatial scale which is suitable for a preliminary investigation aiming at defining a rough climatology of precipitation and wind. However, our approach clearly overlooks the important organizational structure of precipitation systems. This acts as a limiting factor for improving the accuracy of the classification of offshore wind regimes. A potential line of work to overcome that limitation consists of identifying precipitation entities at more appropriate spatial scales, potentially at multi-scales. These entities could then substitute precipitation system as the experimental units for extracting attributes. In our view, there exist two potential techniques to address this problem. First, the extended watershed technique presented in Lakshmanan et al. (2009) which provides a consistent and flexible framework for detecting convective storms over small spatial scales. Second, the multi-scale segmentation technique introduced in Lakshmanan et al. (2003) which enables to split precipitation systems into sub-regions with specific textural properties.

Finally, there are a number of issues that we did not address in this study and that are left for future work. Firstly, the sensitivity of the results to the data length will be analyzed with the acquisition of new data or, if new data were not to become available, the application of resampling techniques such as

bootstrap will be investigated. Secondly, this work aimed at classifying wind regimes at time t based on the weather conditions as seen by a weather radar at the same time t. It is planned to repeat the same study with lagged weather radar images, at time $t - k$, in order to examine the detection of early precipitation patterns. Thirdly, the temporal dimension of the sequence of images was not considered while each time series of precipitation attributes is characterized by a relatively strong autocorrelation. Further research will therefore be encouraged in this direction and data mining techniques dealing with autocorrelated data will receive specific attention.

Acknowledgment

This work was fully supported by the Danish Public Service Obligation (PSO) fund under the project "Radar@Sea" (contract PSO 2009-1-0226) which is gratefully acknowledged. Vattenfall is acknowledged for sharing the wind data from the Horns Rev 1 wind farm. The authors express their gratitude to the radar meteorologists from the Danish Meteorological Institute (DMI) for providing data from the Rømø radar and for their help.

References

Ailliot P, and Monbet V. (2012), Markov-Switching autoregressive models for wind time series. *Environmental Modelling & Software*, 30:92–101.

Akhmatov V, Rasmussen C, Eriksen PB, Pedersen J. (2007) Technical aspects of status and expected future trends for wind power in Denmark. *Wind Energy*, 10:31–49.

Bacher, P. and Madsen, H. (2011) Identifying suitable models for the heat dynamics of buildings. *Energy and Buildings*, 43:1511–1522.

Baldwin M, Kain J, Lakshmivarahan S. (2005) Development of an automated classification procedure for rainfall systems. *Monthly Weather Review*, 133:844–862.

Black M, and Anandan P, The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields. *Computer Vision and Image Understanding*, 63:75–104.

Bluestein H, Jain M. (1985) Formation of mesoscale lines of precipitation: Severe squall lines in Oklahoma during the spring. *Journal of Atmospheric Sciences*, 42:1711–1732.

Bluestein H, Marx G, Jain M. (1987) Formation of mesoscale lines of precipitation: Nonsevere squall lines in Oklahoma during the spring. *Monthly Weather Review*, 115:2719–2727.

Bøvith T. (2008) Detection of weather radar clutter. PhD Thesis, Department of Informatics and Mathematical Modelling, Technical University of Denmark, Kgs. Lyngby. (ISBN: 87-643-0436-1).

Canny J. (1986) A computational approach to edge detection. *IEEE Transactions Pattern Analysis and Machine Intelligence*, 8:679–698.

Cripps E, Dunsmuir W. (2003) Modeling the variability of Sydney harbor wind measurements. *Journal of Applied Meteorology*, 42:1131–1138.

Davy R, Woods M, Russell C, Coppin P. (2010) Statistical downscaling of wind variability from meteorological fields. *Boundary-Layer Meteorology*, 135:161–175.

Fitzgibbon A, Pilu M, Fisher R. (1999) Direct least square fitting of ellipses. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21:476–480.

Forney Jr G. (1973) The Viterbi algorithm. *Proceedings of the IEEE*, 61:268–278.

Gagne D, McGovern A, Brotzge J. (2009) Classification of convective areas using decision trees. *Journal of Atmospheric and Oceanic Technology*, 26:1341–1353.

Gill R, Overgaard S, Bøvith T. (2006) The Danish weather radar network. *In Proceedings of the 4^th European Conference on Radar in Meteorology and Hydrology, Barcelona, Spain, 2006*.

Hasager C, Peña A, Christiansen M, Astrup P, Nielsen M, Monaldo F, Thompson D, Nielsen P. (2008) Remote sensing observation used in offshore wind energy. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 1:67–79.

Hastie T, Tibshirani R, Friedman J. (2001) The Elements of Statistical Learning. *Springer Series in Statistics*.

Isaaks E, Srivastava R. (1989) An Introduction to Applied Geostatistics. *Oxford University Press*.

Johnson J, MacKeen P, Witt A, Mitchell E, Stumpf G, Eilts M, Thomas K, The storm cell identification and tracking algorithm: An enhanced WSR-88D algorithm. *Weather Forecasting*, 13:263–276.

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

Kristoffersen J. (2005) The Horns Rev wind farm and the operational experience with the wind farm main controller. *In Proceedings of the Offshore Wind International Conference and Exhibition, Copenhagen, Denmark,2005*.

Krolzig H. (1997), Markov-Switching Vector Autoregressions: modelling, statistical inference, and application to business cycle analysis. *Springer*.

Lakshmanan V, Rabin R, DeBrunner V. (2003) Multiscale storm identification and forecast. *Atmospheric Research*, 67: 367–380.

Lakshmanan V, Hondl K, Rabin R. (2009) An efficient, general-purpose technique for identifying storm cells in geospatial images. *Journal of Atmospheric and Oceanic Technology*, 26: 523–537.

Lakshmanan V, Smith T. (2009) Data mining storm attributes from spatial grids. *Journal of Atmospheric and Oceanic Technology*, 26: 2353–2365.

Lakshmanan V, Zhang J, Hondl K, Langston C. (2012) A statistical approach to mitigating persistent clutter in radar reflectivity data. *IEEE Journal on Selected Topics in Applied Earth Observations and Remote Sensing*, Available online.

Li S. (2009) Markov Random Field modeling in image analysis. *Springer*.

Pebesma E. (2004) Multivariable geostatistics in S: the gstat package. *Computers & Geosciences*, 30:683–691.

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

Sun D, Roth S, Black M. (2010) Secrets of optical flow estimation and their principles. *IEEE Conference on Computer Vision and Pattern Recognition*, pp 2432–2439.

Suzuki K, Horiba I, Sugie N. (2003) Linear-time connected-component labeling based on sequential local operations. *Computer Vision and Image Understanding*, 89:1–23.

Trombe P-J, Pinson p, Bøvith T, Cutululis N, Draxl C, Giebel G, Hahmann A, Jensen NE, Jensen BP, Le NF, Madsen H, Pedersen L, Sommer A, Vincent C. (2012) Weather radars – The new eyes for offshore wind farms? Working paper under review.

Vincent C, Giebel G, Pinson P, Madsen H. (2010) Resolving nonstationary spectral information in wind speed time series using the Hilbert-Huang transform. *Journal of Applied Meteorology and Climatology*, 49:253–267.

Vincent C, Pinson P, Giebel G. (2011) Wind fluctuations over the North Sea. *International Journal of Climatology*, 31:1584–1595.

Wheater H, Isham V, Onof C, Chandler R, Northrop P, Guiblin P, Bate S, Cox D, Koutsoyiannis D. (2000) Generation of spatially consistent rainfall data. Department of Statistical Science, University College London, Tech. Rep.

Zawadzki I. (1973) Statistical properties of precipitation patterns. *Journal of Applied Meteorology*, 12:459–472.

Zucchini W, MacDonald I. (2009) Hidden Markov Models for time series: An introduction using R. *Chapman & Hall/CRC*.