Statistical modelling of space-time processes with application to wind power.

Lenzi, Amanda; Ersbøll, Bjarne Kjær; Clemmensen, Line Katrine Harder; Pinson, Pierre

Publication date:
2017

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

Statistical modelling of space-time processes with application to wind power

Amanda Lenzi

Kongens Lyngby 2017

Technical University of Denmark
Department of Applied Mathematics and Computer Science
Richard Petersens Plads, building 324,
2800 Kongens Lyngby, Denmark
Phone +45 4525 3031
compute@compute.dtu.dk
www.compute.dtu.dk

Summary (English)

Short-term wind power forecasts together with a quantification of uncertainties are required for the reliable operation of power systems with significant wind power penetration. A challenge for utilizing wind power as a source of energy is the intermittent and hardly predictable nature of wind. This thesis aims at contributing to the wind power literature by building and evaluating new statistical techniques for producing forecasts at multiple locations and lead times using spatio-temporal information. By exploring the features of a rich portfolio of wind farms in western Denmark, we investigate different types of models and provide several forms of predictions. Starting with spatial prediction, we then extend the methodology to spatio-temporal prediction of individual wind farms and aggregated wind power at monitored locations as well as at locations where recent observations are not available.

We propose spatial models for predicting wind power generation at two different time scales: for annual average wind power generation and for a high temporal resolution (typically wind power averages over 15-min time steps). In both cases, we use a spatial hierarchical statistical model in which spatial correlation is captured by a latent Gaussian field. We explore how such models can be handled with stochastic partial differential approximations of Matérn Gaussian fields together with integrated nested Laplace approximations. We show that complex hierarchical spatial models are well suited for wind power data and provide results in reasonable computational time. Moreover, the hierarchical

approach for obtaining predictions at a high temporal resolution is found to produce accurate predictions with improved performance compared to a standard geostatistical method at a small additional computational cost. The use of the integrated nested Laplace approximations is motivated by the desire to produce forecasts on large data sets with hundreds of locations, which is critical during periods of high wind penetration.

Subsequently, the extension from spatial to spatio-temporal models is given. Three different hierarchical models are developed for obtaining probabilistic wind power forecasts. First, a time series model consisting of an autoregressive process with a location specific intercept is considered. This approach gives satisfactory results for individual forecasts but fails to generate calibrated aggregated forecasts. The second approach has a common intercept for all farms and a spatio-temporal model that varies in time with first order autoregressive dynamics and has spatially correlated innovations given by a zero mean Gaussian process. The third model, which also has a common intercept as well as an autoregressive process to capture the local variability and the spatio-temporal term from the second approach, is able to produce reliable individual and aggregated forecasts for multiple lead times.

Finally, very-short-term wind power forecasting is considered. Probabilistic forecasts from 15 minutes up to two hours ahead are produced by using anisotropic spatio-temporal correlation models to account for the propagation of weather fronts and a transformed latent Gaussian field is used to accommodate the probability masses that occur in wind power distribution due to chains of zero measurements. Using what is called kriging equations, even the simplest proposed covariance model is able to produce calibrated spatio-temporal predictions of wind power production.

Summary (Danish)

For at opnå en pålidelig drift af strømforsyningssystemer, hvor vindenergi bidrager med en betragtelig andel af energiproduktionen, kræves kortsigtede vindkraftforudsigelser samt en kvantificering af de pågældende usikkerheder. En stor udfordring ved at benytte vindkraft som energikilde er vindens uforudsigelige opførsel. Denne afhandling sigter efter at bidrage til vindenergi-litteraturen ved at opbygge og evaluere nye statistiske metoder, der vha. spatiotemporal information kan producere forudsigelser for adskillelige lokationer og gennemløbstider. Ved at undersøge egenskaberne ved en stor portefølje af vindmølleparker i den vestlige del Danmark, testes forskellige modeller, og flere typer af forudsigelser gives. Vi starter med spatiale forudsigelser og udvider herefter metodologien til spatiotemporale forudsigelser af individuelle vindmølleparker og den samlede vindenergi ved træningssæts-lokationerne samt ved de lokationer, hvor nyere observationer ikke er tilgængelige.

Vi foreslår spatiale modeller til forudsigelse af vindenergiproduktionen på to forskellige tidsskalaer: for den årlige gennemsnitsproduktion og for en høj tidsmæssig opløsning (typisk den gennemsnitlige produktion i 15-min tidsskridt). I begge tilfælde bruger vi en spatial hierarkisk statistisk model, hvori den spatiale korrelation er beskrevet via et latent Gaussisk felt. Vi undersøger, hvordan disse modeller kan håndteres med stokastiske partielle differential-approksimationer af Matern Gaussiske felter sammen med integrerede indlejrede Laplace approksimationer. Vi viser, at komplekse hierarkiske spatiale modeller er velegnede

til vindenergi-data og leverer resultater inden for en fornuftig beregningstid. Den hierarkiske fremgangsmåde til at opnå forudsigelser med høj tidsmæssig opløsning viser sig desuden at give præcise forudsigelser, som, med kun en lille ekstra beregningsmæssig omkostning, præsterer bedre end standard geostatistiske metoder. Brugen af de integrerede, indlejrede Laplace approksimationer er motiveret af ønsket om at lave forudsigelser på en stor spatial skala, hvilket er essentielt under perioder med høj, øjeblikkelig vindpenetration. Efterfølgende præsenteres en udvidelse fra en spatial til en spatiotemporal model. Tre forskellige hierarkiske modeller til probabilistiske vindenergi-forudsigelser udvikles. Først betragter vi en tidrækkemodel, som består af en autoregressiv proces med et stedsspecifikt konstantled. Denne fremgangsmåde giver tilfredsstillende resultater for individuelle forudsigelser, men fejler i at generere kalibrerede samlede forudsigelser. Den anden fremgangsmåde har et fælles konstantled for alle vindmølleparker og en spatiotemporal model, som varierer med tiden med førsteordens autoregressiv dynamik og har spatial korrelerede fejlled, som bliver repræsenteret af en Gaussisk proces med middelværdi lig nul. Det tredje model, som har et fælles konstantled, en autoregressiv proces til at fange den lokale variabilitet samt det spatiotemporale led, er i stand til at producere pålidelige individuelle og samlede forudsigelser for flere forskellige gennemløbstider.

Til sidst betragter vi meget kortsigtede vindenergi forudsigelser. Probabilistiske forudsigelser fra 15 minutter og op til to timer forud bliver genereret ved hjælp af anisotropiske spatiotemporal korrelerede modeller til at tage højde for udbredelsen af vejrfronter og en transformeret latent Gaussisk felt model til at imødekomme sandsynlighedsmassen, der forekommer i vindenergi-fordelinger på grund af kæder af nul-målinger. Ved brug af såkaldte Kriging-ligninger er selv den mest simple foreslåede ko-variansmodel i stand til at give spatiotemporale forudsigelser af vindenergiproduktionen.

Preface

This thesis was prepared at the Section of Statistics and Data Analysis of the Department of Applied Mathematics and Computer Science at the Technical University of Denmark (DTU), in partial fulfilment of the requirements for acquiring the Ph.D. degree in Applied Mathematics.

The thesis deals with spatial and spatio-temporal models for wind power forecasting. Accurate and reliable forecasts of wind power production, together with a quantification of the uncertainty, is essential to optimally integrate wind energy into power systems. The main focus is on developing statistical models that uses spatio-temporal information. Special attention is given to probabilistic wind power forecasting at both individual and aggregated levels. The novelty of the Bayesian hierarchical models for predicting wind power is addressed.

The thesis consists of a summary report and three research papers written during the study period.

Lyngby, 01-May-2017

Amanda Lenzi

List of publications

Papers included in the thesis

[A] Lenzi, A., Pinson, P., Clemmensen, L. H., Guillot, G. (2016). Spatial models for probabilistic prediction of wind power with application to annual-average and high temporal resolution data. *Stoch Environ Res Risk Assess* DOI 10.1007/s00477-016-1329-0, pp. 1-17

[B] Lenzi, A., Steinsland, I., Pinson, P. (2017). Benefits of spatio-temporal modelling for short term wind power forecasting at both individual and aggregated levels. Submitted to *Environmetrics* (under review)

[C] Baxevani, A., Lenzi, A. (2017). Very short-term spatio-temporal wind power prediction using a censored Gaussian field. Submitted to *Stoch Environ Res Risk Assess* (under review)

Other publications not included in this thesis

- Lenzi, A., Guillot, G., Pinson, P. (2015). A spatial model for the instantaneous estimation of wind power at a large number of unobserved sites. *Procedia Environmental Sciences*, vol. 26, pp. 131-134

- Lenzi, A., de Souza, C. P., Dias, R., Garcia, N. L., Heckman, N. E. (2017). Analysis of aggregated functional data from mixed populations with application to energy consumption. *Environmetrics*, vol. 28, number 2

Acknowledgements

I would like to thank Professor Pierre Pinson for his continuous support and positive attitude throughout the development of this thesis. At difficult times, he always found time to give me valuable advice and guide me in the field.

An important part of my Ph.D was done at the University of Cyprus and at the Norwegian University of Science and Technology. I would like to thank Professor Anastassia Baxevani for inviting and mentoring me during a 2 month visit at the beginning of 2016 and for encouraging me while, at the same time, being critical with my ideas. I am also grateful for the time you took to show me around Nicosia and for the relaxing invitations for a glass of wine after work.

I also would like to express my deep gratitude to professor Ingelin Steinsland for hosting me in Trondheim for a 3 months visit. I have learned a lot from you and I am thankful for the trust you have placed in me and for pushing me forward with great enthusiasm. At NTNU, I have been fortunate to meet and learn from the INLA team. In particular, I would like to thank Elias Krainski for all the fruitful discussions about Statistics and the helpful tips on how to improve my code. Thank you also for introducing me to the Brazilians in Trondheim and making me feel welcome.

I am grateful for the financial support from the CAPES Foundation in Brazil and for my bachelor's and master's degree supervisor, Professor Nancy Garcia, from the University of Campinas. I am thankful to Nancy for being a great

mentor and for giving me advice on innumerable occasions from 2010 until now. You have always inspired me as a researcher over these years.

At DTU, I would like to thank past and present colleagues in the Statistics and Data Analysis section for creating a friendly and enjoyable atmosphere. I would also like to thank Professor Bjarne Ersbøll and Line Clemmensen for their support throughout the Ph.D. Thanks also to Gilles Guillot for proposing the topic of this thesis and for proofreading it.

I thank my parents for their love and support. Particularly, I thank my dad. He is the one that I can always count on and turn to when things are tough. Thank you for encouraging me to pursue my dreams, this thesis is for you. Last, I thank Daniel, the most important person in my life. His support can be summed up by the fact that at this very moment, even though he is not a statistician, he is once again proofreading my thesis.

Contents

Chapter 1

Introduction

The expected main benefit from using wind power as a source of energy instead of fossil fuels is the reduction of carbon emissions. Indeed, wind power, as an alternative to burning fossil fuels, is widely available, renewable, clean, extremely low in greenhouse gas emissions and consumes no clean water. In a society increasingly concerned with sustainability, the share of wind energy in total installed power capacity has grown rapidly in recent years around the world, and it is expected to keep growing significantly in the years to come. Denmark, for instance, has the largest proportion of wind energy capacity relative to the volume of electricity consumption and the Danish government aims at having more than 50% of the energy demand met by wind power by 2020 (Energinet.dk, 2014). Also, the European Union aims to have 20% of its power coming from renewable sources by 2020, and the United States will likely receive 20% of their electrical energy from wind mills by 2030 (Lindenberg and DeMeo, 2008).

Integrating such large amounts of wind power into the existing power grids brings new challenges to power system design and operation (Ackermann, 2005). For example, thousands of wind turbines have been installed where wind is

available in sufficient speed and frequency, which often happens to be in remote locations. Thus, traditional transmission networks that were built for centralized systems, must move towards systems that connect distant renewable generators to load centers, so that power can be transmitted from new wind farms to where the demand is. Moreover, there is currently no feasible way of storing wind energy directly, so it has to be consumed when it is produced.

Wind energy is different from conventional energy sources, since its production is highly dependent on the occurrence and speed of the wind, there are risks of power shortages during periods of low wind speeds and oversupply when there are strong winds. Electricity markets have to constantly adjust electricity supply to instantaneously meet the changing demand. Ensuring security of supply requires careful planning for high renewable energy penetration, since electricity markets were designed for dealing with mainly dispatchable generation and fairly predictable demand. Energy distribution networks, in turn, are usually projected to handle the maximum demand to prevent overload. A method for modelling the typical energy consumption pattern of different types of consumers, that can be used in the planning of energy distribution, was presented in Lenzi et al. (2016).

From the point of view of a producer participating in the day-ahead electricity market, lack of predictability at a wind power production site results in imbalance costs, given the financial penalties that are incurred for deviating from the power levels that are declared in the contract settled by the Transmission System Operator (TSO) and the market operator (Girard et al., 2013). In this sense, accurate forecasting of wind power generation helps to reduce imbalance charges to producers and makes wind power more competitive in the energy market.

A number of approaches can be considered to mitigate the uncertainty in power systems and electricity markets that arises from the large scale integration of wind energy into the power grids. Several ideas, such as energy storage technologies like pumped storage plants or heat pumps, are based on the extension of the power system infrastructure. However, even while research in storage technologies is still ongoing, one has to bear in mind that increasing the reserves and extending the power system infrastructure might reduce the environmental and economic benefits brought by wind power. Another important component

of the integration of variable power sources into the grids is accurate forecasting (Morales et al., 2013). Forecasts of wind power production are, in fact, widely used by electrical utilities, as they are not only important in order to efficiently handle the energy demand (Katzenstein et al., 2010), but accurate forecasts also increase the revenues from the electricity market with the optimization of bidding strategies (Pinson et al., 2007). In particular, specific attention has been given to short-term wind power forecasting, i.e., within a time frame of a few hours ahead, which is the most important horizon for optimal planning of power dispatch (Landberg, 1999).

1.1 Motivation and objectives

Ultimately, with the increasing abundance of wind farms on many power systems today, it is desirable to build predictor systems that are cost efficient, easy to implement and can be applied on a regional or national scale, with hundreds of wind farms installed. This calls for a new framework that mimics the spatio-temporal dependence structure of the wind power process, and that is able to give an overview of the power generation at all wind power generation sites over a region. However, understanding a spatio-temporal process throughout a country, such as wind power production in Denmark, is a difficult problem due to the complex temporal and spatial structures. In this direction, Girard and Allard (2013) provided a detailed analysis on the propagation of wind power forecast errors in space and time. In view of this evolving context and of the limitations with existing approaches, the methodologies described in this thesis can be seen as part of this ongoing effort towards making better use of the available data by utilizing the spatial and spatio-temporal characteristics present in wind power generation.

The motivation for this thesis is to contribute to the wind power forecasting state-of-the-art methodologies by utilizing the spatial correlation among wind farms. The initial idea stems from the fact that most of the wind power prediction systems provide methods that are optimized for each and every location individually - be it a turbine, a wind farm or an aggregated portfolio of wind farms. In those cases, traditional inputs to forecasting models consist of wind power measurements, wind speed, wind direction and meteorological forecasts

at a single site. However, these techniques do not consider information coming from neighbouring territories, whose inclusion offers many attractive benefits and has been currently overlooked. Firstly, borrowing information by utilizing the spatial correlation among individual wind farms has been shown to reduce the errors in forecasts significantly. Secondly, it provides the option of generating forecasts for locations that are not within the observation samples, which is essential for operational problems where online data is not available for all wind farms in a region, while decision-making problems may require information at all sites over a region.

The purpose of this research is to develop new prediction techniques for application to short-term and very-short-term wind power forecasting. Density forecast at multiple locations is obtained by considering recent wind power measurements at the neighbouring areas, but not necessarily at the site where prediction is desired. This is achieved by taking into account the spatial and spatio-temporal dependence structures at all wind power generation sites in a region. Keeping in mind that decision making problems require a quantification of the uncertainty of the forecast, all the techniques studied in this thesis provide predictive probability density functions of wind power generation.

Several types of wind power forecasts are of interest. We start with models for spatial probabilistic prediction. This type of scenario is of relevance for a number of operational problems where wind power generation is only observed at a limited number of wind farms, while decision-making problems may require an overview of power generation at all sites over a region. Next, we move to spatio-temporal probabilistic forecasts of wind power production. We obtain forecasts at wind farms from which data is available, but also at a larger portfolio including wind farms at possible unmonitored locations. We furthermore study the performance of the different approaches in forecasting from individual and aggregated wind farms. Probabilistic forecasts of individual farms are essential for a wind farm operator, since they improve decision making regarding the management of the spinning reserves, which is the extra capacity that is available by increasing the power output of generators already connected to the power system. On the other hand, aggregated wind power generation over pre-defined areas is of particular importance from the point of view of a system operator.

Through the extensive analysis of a Danish wind power data set, we get differ-

ent insights into the modelling and forecasting schemes. First, we investigate suitable transformations for wind power data, that allow us to work within a Gaussian framework. Then, motivated by the need to handle probability masses in the individual wind power distribution, we turn our attention to approaches for modelling nonnegative data with clumping at zero. Further, the critical component of finding correlation models that are able to describe the spatio-temporal correlations between wind power at different locations and times are studied. Moreover, with large data sets, sparsity is essential for numerical robustness and computational efficiency. We explore how models for wind power prediction can be handled with stochastic partial differential approximations (SPDE) of Matérn Gaussian fields, together with integrated nested Laplace approximations (INLA), which offer at the same time flexibility on modelling and speed in computations.

1.2 Contributions

The work presented in this thesis has led to several new statistical methodologies that contribute to the wind power forecasting community and literature. Before moving to the complex spatio-temporal dependence structures in wind power data, the first step in this work consists of building a better understanding of the underlying spatial aspects. Toward this aim, in Paper A, we propose purely spatial statistical models for the prediction of wind power generation. Initially, we focus on a model describing spatial variation of annual average wind power generation. The annual average of all the energy demand and supply is of relevance for both system operation and for electricity companies. We use a hierarchical spatial model based on a beta distribution with a stochastic, spatially structured mean, that depends on a covariate. We proceed to consider wind power generation data with high temporal resolution (typically wind power averages over 15-min time steps). Predicting wind power at a high resolution is particularly important for the management of the immediate regulating power reserve. Having a method that provides an overview of the power generation at all sites over a region, even when observations are provided only at a limited number of wind farms, helps to ensure the continuous balance of the power system. In this case, we use a hurdle model, which consists of a mixture of a degenerated distribution at zero and a skewed continuous distribution for the

non-zero values. We assume that there is a common Gaussian random field with a Matérn covariance function governing those two variables, the occurrence and the amount of wind power generation. To meet the computational requirements, the SPDE approach to spatial modelling is taken (Lindgren et al., 2011) for which fast Bayesian inference can be done using INLA. We apply the proposed methods on a data set from 349 wind farms in western Denmark. We show that predictions, especially at low quantiles (i.e., when wind power is close to zero) for high temporal resolution, are significantly improved when we apply the hierarchical hurdle model instead of the simpler but less realistic kriging method.

After analysing the spatial dependences of wind power data in Paper A, Paper B and C aim at deriving models which capture the spatio-temporal dependency structure in order to improve the quality of typical wind power forecasts based only on measurements from a single site as input.

Another contribution in this thesis is our work from Paper B on providing short-term probabilistic forecasts for individual wind farm and aggregated wind power production. In this work, we analyse wind data from 349 Danish wind farms, which is a much larger data set than those generally being studied in the literature, see, e.g., Tastu et al. (2014), Focken et al. (2002), Wytock and Kolter (2013a) and Tastu et al. (2011). To deal with the non-Gaussianity of wind power series, we use a parametric framework for distributional forecasts based on the logit-normal transformation. Motivated by the need to produce such forecasts on a very large spatial scale, we introduce a scalable inference methodology using the INLA-SPDE approach and illustrate this methodology using three different hierarchical models. We start with a model consisting of a location-specific intercept and an autoregressive component that captures the local variability without considering the dependency between the farms. This model is well suited for individual forecasts, but it is not calibrated for aggregated forecasts. To obtain reliable aggregated forecasts, we introduce two different models that capture the spatio-temporal features present in the data. The first has a common intercept and a spatio-temporal process, in which spatial and temporal dependency is modelled by a latent Gaussian field. The second is a combination of the previous two models, with a common intercept, an autoregressive process and a spatio-temporal term that varies in time with first order autoregressive dynamics. We apply the models to generate probability forecasts

of wind power generation at multiple lead times and at locations where data is available, but also at a larger portfolio including wind farms at new locations. We show that all three models perform reasonably well in terms of calibration of individual density forecasts. However, we find that modelling spatial dependency is required to achieve calibrated aggregated probabilistic forecasts. This is a valuable study in the field of wind energy, since there is very little research on large spatio-temporal data sets, presumably due to the difficulty in obtaining sufficient data on wind power production. To our knowledge, this is the first work where the INLA-SPDE approach is used to obtain spatio-temporal forecasts of wind power generation and the evaluation is done with an aim of making consistent individual and aggregated probabilistic wind power forecasts.

In Paper C, we suggest a way to obtain very-short-term spatio-temporal predictions of wind power production. Forecasts at horizons of a few minutes ahead are crucial to the TSOs to optimally operate reserves. More specifically, in Denmark, the TSO has defined the 10-minute ahead prediction as the most important lead time, since power fluctuations at this time scale are the main responsible for the balance in the power system (Akhmatov, 2007). We obtain predictions using what is called kriging equations together with a latent truncated Gaussian random field. The use of the truncated Gaussian field allows us to simultaneously model both the mass at zero and also, through an anamorphosis transformation, the heavy tails of the empirical distribution. This type of model has previously been used in Baxevani and Lennartsson (2015) for modelling precipitation depth. In our application, at any individual wind farm, the wind power production is modelled using a beta distribution with shape parameters that vary spatially. The spatio-temporal dependence of the Gaussian field is modelled using three different anisotropic covariance models, two of which additionally take into account the propagation of the weather fronts that are, at least partly, responsible for the generation of wind energy. The mean of the Gaussian field is estimated locally by inverting the cumulative distribution function (cdf) of the proportion of positive wind energy. The proposed methodology is applied to a portfolio of 30 wind farms in western Denmark and results show that even the simple covariance models used in the analysis perform well in terms of calibration. This study reveals important features of the spatio-temporal correlations present in wind power data that should be extendible to other wind power time series.

1.3 Structure of the thesis

This thesis consists of a number of papers as well as seven chapters that provide a foundation for the appended papers. For brevity, in this thesis, we refer to the papers included in Appendix A, Appendix B and Appendix C only as Paper A, Paper B and Paper C, respectively.

Chapter 2 provides some background on wind power forecasting, comments on various categories of wind power prediction, and gives an overview of the problem in terms of different applications at each prediction category. Furthermore, the major challenges encountered when building models to predict wind power generation are presented.

Chapter 3 provides a description of both point and probabilistic forecast scores that are used to quantify the performance of the proposed methods. Point forecasts are assessed using the familiar root mean square error (RMSE), bias and mean absolute error (MAE). To evaluate probabilistic density forecasts, the continuous ranked probability score (CRPS), reliability and sharpness diagrams are introduced.

Chapter 4 introduces general methods for inference and prediction of spatial and spatio-temporal data sets. It starts by introducing the conventional best linear unbiased predictor (BLUP), also called spatial kriging predictor in geostatistics. The next idea is to generalize the kriging predictor to include the temporal dimension. This results in the spatiotemporal kriging predictor, which is used later on for predicting wind power production. Then, an alternative and computationally efficient methodology for spatial prediction that takes advantage of a Markov representation of the Matérn covariance family in a continuous space is introduced and a scalable inference methodology using INLA is shortly described. The extension of the spatial predictor to the spatio-temporal case is presented and it is illustrated how the INLA-SPDE methodology can be used for modelling spatio-temporal data sets.

Chapter 5 describes three different models to wind power forecasting, with techniques that handle the aforementioned numerical characteristics of wind power, such as the non-Gaussian character and the probability masses in the distribution. First, we introduce the hurdle model, which makes use of a Bernoulli-

distributed random variable to account for the probability of wind power occurrence, and a skewed continuous distribution for the non-zero values. Next, we propose the use of the generalized logit distribution and transformation to reduce the influence of the bounds and to make the assumption of a Gaussian distribution more appropriate. The last representation is based on anisotropic spatio-temporal correlation models to account for the propagation of weather fronts, and a truncated Gaussian random field to accommodate the probability masses due to chains of zeros, that can occur, for instance, when there is absence of wind, severe weather conditions or maintenance of the turbines.

Subsequently, Chapter 6 presents the main results from applying the techniques in Chapters 4 and 5 to a data set of wind power generation in Denmark. This chapter starts with the problem of purely spatial prediction. Then, it moves to the critical component of modelling the spatio-temporal correlations in wind power data. The results for individual, aggregated and spatially out-of-sample wind power forecasts are provided.

Finally, Chapter 7 reflects on what we have learned from the work reported in this thesis and concludes.

Chapter 2

Basics of wind power prediction

Before proceeding to the prediction methods and the actual models and applications, it is important to understand the different aspects that constitute wind power forecasting and the inherent limitations of the problem. This chapter comprises an introduction to some of the different paths that can be explored when forecasting wind power as well as their applicability. It also provides a discussion of the major challenges that are usually encountered when predicting wind power generation due to the variable and intermittent nature of wind, and how these issues have been addressed in the existing research.

The recognition of the difficulties involved in the forecasting of wind power can be traced back to the 1980s, when Brown et al. (1984) identified the need of forecasting wind power a few hours ahead to ensure efficient power utilization when a wind power generator is supplying power to an energy system. As one of the first proposals in the wind forecasting literature, Brown et al. (1984) used a purely temporal model consisting of an autoregressive moving average (ARMA) for predicting wind speed at lead times between a few hours and a few days. Over the following years, with the growth of wind energy in total installed power

capacity around the world, research activity in this area has rapidly increased.

An extensive literature review of the state-of-the-art in short term (i.e., less than 6 hours) prediction of wind power from 2011 was presented in Giebel et al. (2011). A review of short-term wind speed forecasting for power system operations can be found in Zhu and Genton (2012) and Foley et al. (2012). A particular important conclusion of the papers discussed in these reviews is that for short term horizons, statistical methods using local measurements as input are superior to physical models that consider numerical weather predictions.

2.1 Categories of wind power prediction

Depending on the nature of the wind power series and on the objective of the study, one could predict - or forecast, as we will use these terms indistinguishably - on different spatial and temporal scales. For instance, distribution system operators may be concerned with individual wind farms connected to their system, while a transmission system operator may be interested in the aggregated wind power generation over pre-defined areas. Likewise, some decision making problems can benefit from spatiotemporal forecasts since these require only measurements from a certain number of wind farms to predict the entire power production over a region.

While some users of wind power predictions can benefit from point forecasts, which are easy to interpret and communicate, in decision making problems such as bidding optimally in day-ahead markets, it is not sufficient to know only the expected power generation and therefore probabilistic forecasts are needed. In this thesis, we generate forecasts of wind power generation at individual and aggregated levels. Moreover, both spatial and spatio-temporal prediction are obtained. We assess the quality of predictive performance of the proposed methods using point and probabilistic forecast scores. In the following sections, we describe and motivate these different types of forecasts.

2.1.1 Individual vs aggregated prediction

Individual wind power refers to the power generated from a single wind farm, which is composed of individual turbines that can be located several kilometers apart. Based on this, aggregated wind power is defined as the sum of the power produced by all wind farms in a system. Usually, the aggregation is over a portfolio of wind farms that belong to the same power grid or over wind farms located in the same region or country.

Aggregated and individual forecasts may support different types of decisions. From the point of view of a wind farm operator, forecasts of individual farms improve decision making regarding the management of the immediate regulating and spinning reserves, which is essential given the financial penalties that are incurred for deviating from the declared power levels. Forecasting at the individual level is also important for system operators, who may be interested in the production of the individual wind farms connected to their system. On the other hand, many larger clients, e.g., the TSOs or traders, are not only concerned with the production of individual wind farms but also with their entire generation portfolio. Particularly, TSOs rely on the forecasts of the aggregate power over a region for planning the generation of the conventional units and to estimate reserves (Ortega-Vazquez and Kirschen, 2009). Also, since forecasting aggregated wind power of a set of wind farms gives the estimated total level of wind power generation available for dispatch to customers through the grid, this type of forecast is important for power generation companies.

In terms of variability of the underlying process, aggregating uncertainty across all wind farms in a certain area smooths the wind power time series and the forecasting errors, which in turn provides information that fluctuates less than those from individual wind farms. In this sense, while the marginal distribution of wind power production data at a single site possesses tails that are heavier than the Gaussian as well as chains of zeros that are usually present when observing a long period of time, aggregated data is easier to forecast because abrupt changes and intermittency are less likely to occur. Another important aspect is that the state-of-the-art wind power prediction systems mainly provide forecasts for single wind farms (Tastu et al., 2011).

One should notice that forecasts of aggregated wind power can be obtained

either by modelling the aggregated wind power as a univariate time series or by modelling the individual wind farms separately and calculating the total wind power prediction by adding up the individual predictions. Modelling the aggregated wind power time series directly has the advantage that the variability of the generated power is lower at an aggregate level than at the single wind farm level. Conversely, when modelling the wind farms separately, one can take into account the spatio-temporal correlations among individual wind farms, which may improve the results for the aggregated forecast. In addition, in this way, forecasts of the individual and aggregated power production are obtained simultaneously, so that the user can decide which information to use depending on the goal. In this work, we consider the problem of separately modelling the individual wind farms and obtaining the aggregated forecasts as the sum of the individual ones.

2.1.2 Spatial vs spatio-temporal prediction

Spatio-temporal processes involve dependencies across a large range of spatial and temporal scales. Usually, in order to understand such processes, observations or measurements are taken. In some situations, however, a temporal snapshot of a physical phenomena on the entire domain, which can be aggregated over time or a frozen temporal space-time process, gives rise to a spatial field. The spatial (and spatio-temporal) statistics literature is often divided into three main branches: geostatistics, lattice data and point patterns (Cressie, 1993), (Banerjee et al., 2014). This thesis mainly focuses on geostatistical models; also referred to as spatially continuous models. The idea behind this is that there is an underlying process that governs a particular physical phenomenon, while data is only observed at a finite set of locations. The discrete sampled data could be, for instance, measurements of wind power generation at several different wind farms.

When prediction is desired, several methods can be used to estimate the values of a target quantity at unvisited locations. One of the most common ways of obtaining predictions of spatial processes is the so called kriging method and its variants (Matheron, 1963), which is outlined in Section 4.1. Geostatistical methods such as kriging can also be applied to spatio-temporal problems

by treating time as an additional dimension. Section 4.1.2 gives an introduction to kriging in its extended spatio-temporal form. As in kriging, Gaussian processes are often used for modelling data in space and time; they have the advantage of being completely characterized by the mean and the covariance function. However, directly working with a spatio-temporal covariance function may quickly become too computationally demanding for large data sets. Alternatively to traditional covariance based modelling, Lindgren et al. (2011) use SPDEs and carry out the practical computations using Gaussian Markov random field (GMRF) representations. More details on the SPDE approach can be found in Section 4.2.

In the wind power context, spatial prediction is of relevance for a number of operational problems where wind power generation is only observed at a limited number of wind farms, while decision-making problems may require an overview of power generation at all sites over a region. The idea is based on the fact that generally there is not enough data available for all the wind farms, and even when it is available, the computational load to calculate forecasts for all of them can be very high. Several spatial interpolation techniques are available to predict the wind speed in locations where data is not available; for example kriging methods to estimate mean surfaces of wind speed and velocity are used in Cellura et al. (2008) and Luo et al. (2008), respectively. Lenzi et al. (2016) used a hierarchical model in which spatial correlation is captured by a latent Gaussian field for obtaining probabilistic predictions of wind power generation at two different time scales. Examples of spatial estimation using geographic information system (GIS) include Hossain et al. (2011), Sliz-Szkliniarz and Vogt (2011) and Etienne et al. (2010).

When the interest lies in predictions over time, as for short-term wind power prediction, models that take into account the temporal dependency are needed. For this purpose, a commonly used approach is based on time series analysis, where the process is determined by the most recent measurements only at the location where prediction is desired, and therefore, does not take the spatio-temporal dependency into consideration. However, when data is collected across time and space, treating the problem of modelling and predicting as a spatio-temporal one has been shown to have several advantages. For instance, time series models that only use historical data at a single site do not provide a straightforward way to extrapolate predictions at unmonitored locations, since

the spatial aspect is disregarded. Moreover, modelling and forecasting spatio-temporal features is essential in order to evaluate the uncertainty of aggregate generation forecasts (Lenzi et al., 2017).

The potential benefits of using spatio-temporal information were realised in Xie et al. (2014) and have been explored recently in several applications. An analysis of the spatio-temporal structure of wind power prediction errors in Denmark is done in Tastu et al. (2011) and later generalized using a larger data set in Girard and Allard (2013). For forecasting at a target single location, Tastu et al. (2014) showed that using multiple wind farms as explanatory variables improves the quality of the forecasts. In a setting for predicting at large spatial scales, Sanandaji et al. (2015) introduced a sparse spatio-temporal predictor inspired by techniques from compressive sensing. A spatial model with the dependence described through a Gaussian random field for the instantaneous estimation of wind power at a large number of unobserved sites was used in Lenzi et al. (2015). For very-short-term probabilistic forecasts, Dowell and Pinson (2016) introduced a vector-valued spatio-temporal process (sVAR) to model the location parameter of the wind power distribution. An alternative type of forecast based on the probability of the wind speed increasing, decreasing or remaining the same during the next time period was introduced in Yoder et al. (2014). Other contributions built probabilistic spatio-temporal models using sparse Gaussian random fields, which resulted in computational benefits due to factorization algorithms but are limited to data sets of modest sizes (Wytock and Kolter, 2013b), (Tastu et al., 2015), (Wytock and Kolter, 2013a).

2.1.3 Point vs probabilistic prediction

A point forecast consists of a single prediction value of some future observation. Because it has an intuitive interpretation and a non-expert can communicate point forecasts with relative ease, it is still used by several practitioners (Jones and Clark, 2011). However, in many applications, such forecasts are not informative enough, due to the inherent uncertainties, and they provide no information as to how confident one can be. More detailed information about the future observation can be obtained by using probabilistic forecasts; this type of forecast is more informative than a point forecast, since it provides an estimate of the

likelihood of a range of possible outcomes. Another aspect to consider is the fact that a point forecast can always be obtained, for instance, by taking the conditional expectation when the predictive probability distribution is available.

Probabilistic forecasts of wind power generation can be expressed in a variety of forms. Quantile forecasts, for instance, inform about the probability that the production will be less than a certain value. Similarly, an interval forecast estimates the probability that the production will fall within a certain interval. A more complete evaluation of probabilistic forecasts can be obtained using a predictive probability distribution over future observations, which can be a parametric or a non-parametric one, see Zhang et al. (2014) for a review on those techniques.

While most of the early literature focused on point forecasts, in recent years, with a clear need to quantify the inherent uncertainties in wind, more emphasis has been placed on probabilistic forecasts, see e.g. Bremnes (2004) and Pinson and Kariniotakis (2010). The additional value of using probabilistic forecasts, compared to point forecasts, has been demonstrated in several applications. Indeed, probabilistic forecasts are the optimal input to a large class of decision-making problems, where the optimal bid is not the expected power but a quantile (Pinson et al., 2007). Other examples of decision making situations that require probabilistic wind power forecasts are economic load dispatch and stochastic unit commitment (Liu, 2010), (Botterud et al., 2013), assessment of operating costs (Ortega-Vazquez and Kirschen, 2010) and reserve quantification (Bessa et al., 2012). More generally, Gneiting et al. (2008) claimed that probabilistic forecasts are a prerequisite for optimal decision-making under uncertainty.

In this thesis, while point forecasts are evaluated by taking the mean or median of the predictive probability distribution, all the methods used for predicting wind power production provide a parametric distribution as probabilistic forecasts.

2.2 Challenges in predicting wind power

From a statistical perspective, accurately predicting wind power and quantifying the uncertainties of the predictions is a challenging problem. Indeed, the statistical distribution of wind power data is characterized by the presence of complex temporal and spatial structures that are not well encompassed by stationary models. Also, the intermittent nature of wind leads to a spike at zero in the empirical distribution, which is difficult to model. In this section, we describe some of the challenges in modelling and forecasting wind power that we faced throughout this work, their relevance and how they are usually handled.

2.2.1 Boundedness and skewness

First of all, wind power is bounded below by zero when no turbines are operating, and above by the nominal capacity, when all turbines are generating their rated power output. As a result, models that exclude negative values and generate predictions that lie inside this permissible range should be preferred.

In addition, wind power series are clearly non-Gaussian. In fact, the marginal distribution of wind power production data possesses tails that are heavier than the Gaussian distribution. This is partly due to the fact that wind power depends on wind speed, which has a skewed distribution, often described by the Weibull, gamma and log-normal distributions, see e.g. Celik (2004), Chang (2011) and Gipe (2004). As a result, wind (and wind power) cannot be directly modelled using conventional Gaussian distributions. To work in the Gaussian framework and exploiting the appealing properties of the normal distribution, appropriate transformations are required. Examples of transformations used to model wind data include the square root transformation (Gneiting, 2002) and the power transformation (Ailliot et al., 2015). Otherwise, skewed distributions are usually used to model the positive part of wind data. Distributions such as skew-t or skew-normal are an extension of the normal model and therefore have the advantage of retaining several of its convenient properties. In this context, Hering and Genton (2010) used a skew-t distribution to predict short-term wind speed, while Flecher et al. (2010) built a multivariate weather generator from a skew-normal distribution. Since wind power generation is a double-bounded

process, continuous probability distributions defined on the interval $[0, 1]$ will ensure that the final predictions lie in this valid range. A comparison of the performance between the beta and the generalized logit normal distribution with potential concentration of probability mass at the bounds of the unit interval to obtain probabilistic forecast of 10-min ahead wind power generation was presented in Pinson (2012).

2.2.2 Discrete probability masses

Another potential issue when modelling wind power data is the fact that wind power production series might contain chains of zeros, which correspond either to maintenance of the turbines, calm conditions of wind speeds, or to the occurrence of severe weather conditions such that wind speeds exceed safety levels and wind turbines need to be switched off for safety reasons. This introduces a probability mass at zero wind power, which produces a discontinuity in the distribution function.

There are different approaches for dealing with applications in which data takes nonnegative values but has a substantial proportion of values at zero. One approach is to model a zero-inflation parameter that represents the probability of having zeros, given that these zero measurements come from the same distribution as the non-zero values. An example is found in Hall (2000). Alternatively, data containing an abundant amount of zeros can be modelled with two latent Gaussian processes. The first controls the probability of observing zero values, and the second governs the density distribution of the non-zero observations. Examples of this type of model, used to describe accumulated precipitation, include Berrocal et al. (2008) and Kleiber et al. (2012). In another attempt to account for the zeros, Baxevani and Lennartsson (2015) simultaneously modelled the occurrence and the intensity of rainfall using a single latent Gaussian field. In this case, the positive part of the process is considered to be observed up to a transformation of the observed data. Rainfall data shares similar features with wind power data sets, since one could have long periods of dry days with no observed rainfall.

Within the relevant literature, Messner et al. (2014) introduced a censored re-

gression model to solve the problems caused by discrete probability masses at zero for wind power data. Zhang et al. (2015) proposed a boundary kernel method to eliminate the density peak at the bounds of the wind power distribution. A post-processing technique that uses heteroscedastic censored regression to obtain predictive distributions of non negative variables, such as wind speed, was presented in Thorarinsdottir and Gneiting (2010). Pinson (2012) applied the generalized logit-normal distribution with a potential concentration of probability mass at the bounds of the unit interval $[0, 1]$ to forecast wind power fluctuations at single wind farms. A transformed latent Gaussian field model is used in Baxevani and Lenzi (2017) to accommodate the probability masses that occur in wind power distribution due to the chains of zeros.

In this work, we follow mainly three different approaches to deal with the major challenges that may arise from the marginal distribution of wind power data. In Paper A, we model wind power generation with a mixture of a degenerated distribution at zero and a skewed continuous distribution for the non-zero values. In Paper B, we use a parametric framework based on the logit-normal distribution, as in Pinson (2012). Paper C builds on a truncated Gaussian random field that handles the probability mass at zero explicitly in the model.

Chapter 3

Methods of prediction evaluation

One of the biggest challenges in forecasting problems is how to evaluate the forecast performance of different models. Depending on the particular application, different types of forecasts such as point forecasts, quantile forecasts or probabilistic density forecasts, are preferred. Ideally, a forecast is judged by a loss function that is defined by the user to quantify the loss. In a more general way, the best forecast is the one which allows the user to make optimal decisions in some sense, e.g. to minimise costs or maximise returns.

Here, some of the common evaluation measures and diagnostic tools for assessing the quality of the predictions are presented. This will provide the necessary background to understand how different models and prediction methods can be analysed and compared in terms of forecast performances. A comprehensive review of the various scores and verification methods and their interpretation can be found in Jolliffe and Stephenson (2012).

3.1 Point prediction

The so-called point forecasts are often seen from power system operators as easy to interpret and handle and therefore convenient to be used at the process of making decisions.

There are a range of prediction tools, which can be employed to evaluate the performance of point forecasts. Aiming at standardizing the methodology to evaluate point forecast performance, Madsen et al. (2005) recommend that bias, root mean square error (RMSE) and mean absolute error (MAE) should be used as a minimum set of error measures. In this work, point forecasts are used to estimate the power generation at a single location $\mathbf{s}_i$ for a given time t. Details of the evaluation of different point forecast scores are provided as follows.

Root mean square error (RMSE): It has been shown that mean forecasts are optimal under the squared loss function (Banerjee et al., 2005). As a result, from the N locations, where forecast densities are obtained in the testing data, the RMSE is given by

$$\mathrm{RMSE}_t = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y(\mathbf{s}_i,t)-\hat{y}(\mathbf{s}_i,t))^2}, \tag{3.1}$$

where $y(\mathbf{s}_i,t)$ is the power production at location $\mathbf{s}_i$ and time t and $\hat{y}(\mathbf{s}_i,t)$ is the corresponding mean of the forecast density.

Mean bias (BIAS): This metric gives insight as to whether the forecasting method tends to overestimate or underestimate the actual observation. This is particularly important during wind ramp events, when positive or negative errors indicate a need to increase or reduce system output. The mean bias over the N samples in the testing set is given by

$$\mathrm{BIAS}_t = \frac{1}{N}\sum_{i=1}^{N}(y(\mathbf{s}_i,t)-\hat{y}(\mathbf{s}_i,t)). \tag{3.2}$$

Mean absolute error (MAE): The point forecast that minimizes the MAE is the median of the predictive distribution, since MAE is a symmetric linear function in contrast to the RMSE (Fraley et al., 2010) (Pinson and Hagedorn,

2012). We evaluate median forecasts by first calculating the median from the density forecast samples as $\hat{y}$ and then we calculate the MAE as

$$\text{MAE}_t = \frac{1}{N}\sum_{i=1}^{N} |y(\mathbf{s}_i, t) - \hat{y}(\mathbf{s}_i, t)|. \tag{3.3}$$

3.2 Probabilistic prediction

Since the loss function and the nominal level of quantile forecasts to be used as evaluation score depend on the problem itself, metrics that compare density forecasts are certainly relevant for many operational decision-making problems. Generally speaking, prediction quality is related to the level of correspondence between forecasts and observations. In the case of probabilistic forecasts, one has to evaluate how similar a density forecast is to the corresponding distribution of the observations. Verification of multivariate probabilistic forecasts is an active field of research, for which new scores and diagnostic tools are being proposed and discussed, see e.g., Pinson and Tastu (2013), Scheuerer and Hamill (2015), Thorarinsdottir et al. (2016) among others.

To evaluate the performance of density forecasts generated by various models, in this work we use the continuous ranked probability score (CRPS), as well as reliability and sharpness diagrams, which are described next. Additionally, a complete overview of the proper scores available for verification of probabilistic forecasts can be found in Diebold et al. (1997).

Continuous ranked probability score (CRPS): The CRPS is appealing in practical applications, since it is a strictly proper scoring rule for the evaluation of probabilistic forecasts of a univariate quantity that assesses calibration and sharpness simultaneously (Gneiting and Raftery, 2007). A lower score indicates a better density forecast. It is defined as

$$\text{CRPS}(F, y) = \int_{-\infty}^{\infty} (F(x) - \delta_{\{x \geq y\}})^2 dx \tag{3.4}$$

where F is the predictive cumulative distribution, y is the observed realization, and δ is an indicator function taking the value 1 if the condition inside the

brackets is fulfilled and 0 otherwise. The CRPS can be equivalently calculated as

$$\mathrm{CRPS}(F, y) = E_F|Y - y| - \frac{1}{2}E_F|Y - Y'|, \tag{3.5}$$

where Y and Y' are independent random variables with distribution F. If a sample from F is available, we can approximate the mean CRPS among the wind farms in the training set for each t by

$$\mathrm{CRPS}_t(F, y) = \frac{1}{N}\sum_{i=1}^{N}\Big(\frac{1}{m}\sum_{j=1}^{m}|\hat{y}^{(j)}(\mathbf{s}_i, t) - y(\mathbf{s}_i, t)| - \frac{1}{2m^2}\sum_{j,k=1}^{m}|\hat{y}^{(j)}(\mathbf{s}_i, t) - \hat{y}^{(k)}(\mathbf{s}_i, t)|\Big), \tag{3.6}$$

where N is the number of wind farms in the training set and m is the size of the available sample.

Moreover, Gneiting and Ranjan (2011) proposed the use of different weights on the quantile scores to obtain a quantile-weighted CRPS, which is particularly useful when one is interested in a specific range of quantiles.

Reliability diagrams: Reliability represents the ability of the forecasting system to match the observation frequencies. Ideally, the nominal coverage rate and the observed frequencies would be the same, resulting in points aligning with the diagonal in the reliability diagram. For example, for a nominal coverage rate $\alpha = 0.05$, it is expected that 5% of the observations are smaller than the predictive quantile at nominal level 0.05.

To construct reliability diagrams, we start by introducing an indicator variable $\mathcal{I}^{(\alpha)}(\mathbf{s}_i, t)$, which is defined for a quantile forecast $\hat{q}^{(\alpha)}(\mathbf{s}_i, t)$ issued at wind farm $\mathbf{s}_i$ and time t, with observed value $y(\mathbf{s}_i, t)$ as follows

$$\mathcal{I}^{(\alpha)}(\mathbf{s}_i, t) = \begin{cases} 1 & \text{if} \quad y(\mathbf{s}_i, t) \leq \hat{q}^{(\alpha)}(\mathbf{s}_i, t) \\ 0, & \text{otherwise} \end{cases}$$

The indicator variable $\mathcal{I}^{(\alpha)}(\mathbf{s}_i, t)$ shows whether the actual outcome lies below the quantile forecast (hit) or not (miss). Next, let $n_{t,1}^{(\alpha)}$ denote the sum of hits

and $n_{t,0}^{(\alpha)}$ the sum of misses over all the realizations

$$n_{t,1}^{(\alpha)} = \sum_{i=1}^{N} \mathcal{I}^{(\alpha)}(\mathbf{s}_i, t) \quad \text{and} \quad n_{t,0}^{(\alpha)} = N - n_{t,1}^{(\alpha)}.$$

An estimation $\hat{a}_t^{(\alpha)}$ of the actual coverage $a_t^{(\alpha)}$ is then obtained by calculating the mean of $\mathcal{I}^{(\alpha)}(\mathbf{s}_i, t)$ over the N wind farms in the validation set

$$\hat{a}_t^{(\alpha)} = \frac{1}{N} \sum_{i=1}^{N} \mathcal{I}^{(\alpha)}(\mathbf{s}_i, t) = \frac{n_{t,1}^{(\alpha)}}{n_{t,1}^{(\alpha)} + n_{t,0}^{(\alpha)}}. \tag{3.7}$$

Since the number of observations used to calculate the reliability diagrams is often of limited size and the observed proportions are equal to the nominal ones only asymptotically (Toth et al., 2003), (Bröcker and Smith, 2007), one can use the idea of defining consistency bars for each nominal proportion. When considering probabilistic forecasting of binary variables, the method of Bröcker and Smith (2007) can be employed for generating consistency bars for reliability diagrams. For the problem of evaluating density forecasts of continuous variables where no serial correlation is present, consistency bars with a specific confidence level α can be computed by generating independent Binomial distributed random variables with parameters N equal to the corresponding number of observations and p equal to the level α (Pinson et al., 2010). In contrast, when serial correlation is present, the corresponding Bernoulli trials cannot be independent. In such a case, the sum of the N dependent Bernoulli distributed variables can be modelled with a Beta-Binomial distribution, see e.g. Ahn and Chen (1995).

Reliability diagrams allow one to easily visualize if a given method tends to systematically underestimate or overestimate the true values. The calibration of the probabilistic forecasts is performed by evaluating the quantile forecasts for all nominal levels, as shown in the example of Figure 3.1. In this example, the calibration is good, since the forecast method only slightly overestimates quantiles lower than 50% and slightly underestimated quantiles above 50%. This indicates that the corresponding predictive density is slightly too narrow.

Sharpness diagrams: In practical applications, a probabilistic forecast should be sharp, meaning that the density forecast should be as tight as possible (Gneit-

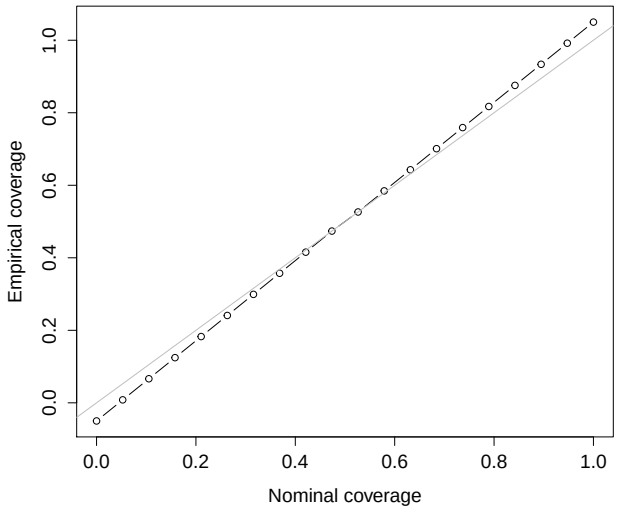


Figure 3.1: Example of a reliability diagram. The probabilistic forecasts are defined by a number of quantile predictions, with nominal levels from 5% to 95% in steps of 5%. The calibration is good, since the forecast method only slightly overestimates quantiles lower than 50% and slightly underestimated quantiles above 50%. This indicates that the corresponding density forecast is slightly too narrow.

ing et al., 2007). Indeed, a narrow predictive interval is more informative and more competitive than a wide one when subjected to calibration. For example, a forecast stating that wind power generation will be zero with probability equal to one is very sharp, even though it might not be consistent with the observed power generation.

Sharpness gives an indication of the spread of the predictive distributions. For each nominal quantile and location, a sharpness diagram can be obtained with the length of the central predicted interval, which is centered in probability around the median. For example, the value at the 60% nominal coverage is the predicted value at the 80% quantile minus the predicted value at the 20% quantile. The final predicted interval size is the average size of the predicted

intervals over the N locations in the testing set and can be written as

$$\overline{\delta}_t^{\alpha} = \frac{1}{N} \sum_{i=1}^{N} (\hat{q}(\mathbf{s}_i, t)^{(1-\alpha/2)} - \hat{q}(\mathbf{s}_i, t)^{(\alpha/2)}), \tag{3.8}$$

where $\hat{q}(\mathbf{s}_i, t)^{(\alpha)}$ is the quantile forecast with nominal coverage α estimated at wind farm $\mathbf{s}_i$ and time t.

Chapter 4

Generalities of spatial and spatio-temporal prediction methods

In this chapter, the methodologies that are along this work for predicting wind power generation are shortly introduced. The chapter starts by describing the standard spatial kriging, which is used as a benchmark in Paper A. Then it gives the details of extending the spatial kriging to the context of spatio-temporal predictions as suggested in Paper C. Next, the novel techniques to spatial modelling, the INLA (Rue et al., 2009) and the SPDE (Lindgren et al., 2011) approaches are shortly described, as they are used to perform inference and prediction in Paper A and B. Finally, motivated by Paper B, an application of the INLA-SPDE approach to model spatio-temporal data sets is given.

4.1 Kriging as a benchmark method

4.1.1 Spatial kriging

When the interest lies in modelling continuous spatial variables that are measured only at a finite set of points in a given region and in predicting their values at unobserved locations, kriging methods can be employed. Kriging is a methodology for spatial prediction developed by Matheron (1963). It has its roots in the mid nineteenth century, with the work of Krige (1951) in the field of mining engineering. Kriging is an interpolation technique in which the surrounding measured values are used in a linear combination to obtain a prediction at an unmeasured location (McDonnell, 1995). It is the optimal linear predictor of $Z(\mathbf{s}_0)$ at location $\mathbf{s}_0 \in \mathcal{D}_{\mathbf{s}}$.

At an unsampled location $\mathbf{s}_0$, $Z(\mathbf{s}_0)$ is estimated by the kriging prediction

$$Z^*(\mathbf{s}_0) = k + \sum_{i=1}^{n} \lambda_i Z(\mathbf{s}_i) = k + \boldsymbol{\lambda}^T \mathbf{Z}, \tag{4.1}$$

where λ_i is the weight assigned to the observation point $Z(\mathbf{s}_i)$ and n is the number of neighbours that are considered for the estimation of $Z(\mathbf{s}_0)$.

Values of $\boldsymbol{\lambda}$ and k are found by minimizing the mean squared prediction error,

$$\mathbb{E}\Big(Z(\mathbf{s}_0) - \boldsymbol{\lambda}^T \mathbf{Z} - k\Big)^2 = \mathrm{Var}\Big(Z(\mathbf{s}_0) - \boldsymbol{\lambda}^T \mathbf{Z} - k\Big) + \Big(\mu(\mathbf{s}_0) - \boldsymbol{\lambda}^T \boldsymbol{\mu} - k\Big)^2, \tag{4.2}$$

where $\boldsymbol{\mu} \equiv (\mu(\mathbf{s}_1), \ldots, \mu(\mathbf{s}_n))^T, \mathbf{s} \in \mathcal{D}_{\mathbf{s}}$.

The second term in (4.2), $\Big(\mu(\mathbf{s}_0) - \boldsymbol{\lambda}^T \boldsymbol{\mu} - k\Big)^2$, is minimized if k satisfies

$$k = \mu(\mathbf{s}_0) - \boldsymbol{\lambda}^T \boldsymbol{\mu}, \tag{4.3}$$

and the first term in this equation can be written as

$$\mathrm{Cov}(\mathbf{s}_0, \mathbf{s}_0) + \boldsymbol{\lambda}^T \Sigma \boldsymbol{\lambda} - 2\boldsymbol{\lambda}^T \mathbf{c}^T, \tag{4.4}$$

where $\mathbf{c} \equiv (\mathrm{Cov}(\mathbf{s}_0, \mathbf{s}_1), \ldots, \mathrm{Cov}(\mathbf{s}_0, \mathbf{s}_n))^T$ and Σ is a $n \times n$ matrix whose (i, j)th

element is $\text{Cov}(\mathbf{s}_i, \mathbf{s}_j)$.

Differentiating (4.4) with respect to $\boldsymbol{\lambda}$ and setting the result equal to $\mathbf{0}$ yields the optimal coefficients $\boldsymbol{\lambda}^* = \Sigma^{-1}\mathbf{c}$ and the optimal constant term is hence $k^* = \mu(\mathbf{s}_0) - \mathbf{c}^T\Sigma^{-1}\boldsymbol{\mu}$. The optimal linear predictor is given by

$$Z^*(\mathbf{s}_0) = \mu(\mathbf{s}_0) + \mathbf{c}^T\Sigma^{-1}(\mathbf{Z} - \boldsymbol{\mu}). \tag{4.5}$$

Finally, the minimized mean squared prediction error, also called the kriging variance, is

$$\sigma^2(\mathbf{s}_0) = \text{Cov}(\mathbf{s}_0, \mathbf{s}_0) - \mathbf{c}^T\Sigma^{-1}\mathbf{c}. \tag{4.6}$$

Notice that the kriging predictor in (4.5) requires the values of the covariances Σ and $\mathbf{c}$ to be known. In practice, these values have to be estimated through a variogram or a covariance function. Usually a parametric model $\text{Cov}(\mathbf{s}_i, \mathbf{s}_j) = C(\mathbf{s}_i, \mathbf{s}_j; \boldsymbol{\theta})$ is chosen for the covariance function, where $\boldsymbol{\theta}$ is a vector of unknown parameters to be estimated from the data. To be able to estimate the model parameters, most applications use stationary covariance functions, which depend on the observation sites only through their separation vector.

Depending on how the expectation $\mathbb{E}(Z(\mathbf{s}))$ is modelled, different versions of kriging predictions can be developed. The derivations above apply to what is called simple kriging. In the case of simple kriging, the mean of the spatial random process is supposed to be constant and known, that is, $\mathbb{E}(Z(\mathbf{s})) = \mu$, for all $\mathbf{s} \in \mathcal{D}_\mathbf{s}$. On the other hand, in ordinary kriging, the mean is considered to be constant and unknown over the neighbourhood of the value to be predicted. A third form, known as universal kriging, arises when the mean is assumed to be a function of covariates, that is $\mathbb{E}(Z(\mathbf{s})) = \mathbf{x}(\mathbf{s})^T\boldsymbol{\beta}$. For more details on ordinary and universal kriging see Cressie (1992).

4.1.2 Spatio-temporal kriging

One of the approaches for generating spatiotemporal forecasts is to obtain the kriging predictor, as described earlier for spatial predictions, except that an additional dimension, the temporal dimension, is included. Here, we describe how the spatial kriging prediction is extended for the use in spatiotemporal

prediction.

Let $Z(\mathbf{s},t)$ denote a spatiotemporal process observed at a set of sample locations $\{\mathbf{s}_1,\ldots,\mathbf{s}_n\} \in \mathcal{D}_\mathbf{s}$ and times $t = 0,1,\ldots, \in \mathcal{D}_t$. For current time t, we are interested in predicting Z at future time $t+h$, where h is the lead time, and at $m-n$ locations $\{\mathbf{s}_{n+1},\ldots,\mathbf{s}_m\}$. The $m-n$ locations can be some or all of the locations at which we already have observations, i.e. $\{\mathbf{s}_{n+1},\ldots,\mathbf{s}_m\} \subseteq \{\mathbf{s}_1,\ldots,\mathbf{s}_n\}$, they can be a completely new set of locations, i.e. $\{\mathbf{s}_{n+1},\ldots,\mathbf{s}_m\} \subseteq \mathcal{D}_\mathbf{s} - \{\mathbf{s}_1,\ldots,\mathbf{s}_n\}$, or some of the predictor locations can be locations with observations and the remaining can be new unobserved locations.

We denote by

$$\mathbf{Z}_o = (Z(\mathbf{s}_1,0),\ldots,Z(\mathbf{s}_1,t),\ldots,Z(\mathbf{s}_n,0),\ldots,Z(\mathbf{s}_n,t))^T$$

the $n(t+1)\times 1$ vector of observed values, and by

$$\mathbf{Z}_p = (Z(\mathbf{s}_{n+1},t+1),\ldots,Z(\mathbf{s}_{n+1},t+h),\ldots,Z(\mathbf{s}_m,t+1),\ldots,Z(\mathbf{s}_m,t+h))^T$$

the $(m-n)h\times 1$ vector consisting of the unobserved locations and times for which prediction is desired.

For prediction locations $\{\mathbf{s}_{n+1},\ldots,\mathbf{s}_m\}$ and lead times $h = 1,2,\ldots$, the estimator of $\mathbf{Z}_p$ takes the form of a linear combination of the past observations and can be written as

$$\mathbf{Z}_p^* = C + \Gamma\mathbf{Z}_o, \tag{4.7}$$

where C is the $(m-n)h\times 1$ vector of constants and Γ is a $(m-n)h\times n(t+1)$ matrix of weights. Both C and Γ are optimized by minimizing the mean squared prediction error (Myers, 1982). In the context of the kriging method, time is simply another dimension and the calculations can be carried out similarly to what is described in Section 4.1.1.

We shall consider the case where the mean function

$$\mathbb{E}(Z(\mathbf{s},t)) = \mu(\mathbf{s},t), \qquad \mathbf{s}\in\mathcal{D}_\mathbf{s}, \quad t\in\mathcal{D}_t,$$

is known, which underlies the theory of simple kriging. There are analogous ordinary and universal kriging equations that can be derived in the spatio-

temporal setting (Cressie and Wikle, 2015).

In the case where $Z(\mathbf{s},t)$ is a Gaussian random field, the vectors $\mathbf{Z}_o$ and $\mathbf{Z}_p$ are jointly Gaussian, with mean and covariance that are given in the next relation

$$\begin{pmatrix} \mathbf{Z}_o \\ \mathbf{Z}_p \end{pmatrix} \stackrel{d}{=} \mathcal{N}\left(\begin{pmatrix} \boldsymbol{\mu}_o \\ \boldsymbol{\mu}_p \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{oo} & \boldsymbol{\Sigma}_{op} \\ \boldsymbol{\Sigma}_{po} & \boldsymbol{\Sigma}_{pp} \end{pmatrix} \right), \tag{4.8}$$

where $\boldsymbol{\mu}_o = \mathbb{E}(\mathbf{Z}_o)$ and $\boldsymbol{\mu}_p = \mathbb{E}(\mathbf{Z}_p)$ are the mean vectors and $\boldsymbol{\Sigma}_{oo}, \boldsymbol{\Sigma}_{op}, \boldsymbol{\Sigma}_{po}$ and $\boldsymbol{\Sigma}_{pp}$ are covariance matrices with dimensions $n(t+1)\times n(t+1)$, $n(t+1)\times(m-n)h$, $(m-n)h\times n(t+1)$ and $(m-n)h\times(m-n)h$, respectively and defined as follows.

$\boldsymbol{\Sigma}_{oo}$ is a block matrix of the form

$$\boldsymbol{\Sigma}_{oo} = \begin{pmatrix} \mathcal{C}_{11} & \mathcal{C}_{12} & \dots & \mathcal{C}_{1n} \\ \mathcal{C}_{21} & \mathcal{C}_{22} & \dots & \mathcal{C}_{2n} \\ \vdots & \vdots & \ddots & \dots \\ \mathcal{C}_{n1} & \mathcal{C}_{n2} & \dots & \mathcal{C}_{nn} \end{pmatrix} \tag{4.9}$$

where $\mathcal{C}_{ij}$ is a $(t+1)\times(t+1)$ covariance matrix of the form

$$\mathcal{C}_{ij} = \begin{pmatrix} C(||\mathbf{s}_j-\mathbf{s}_i||,0) & C(||\mathbf{s}_j-\mathbf{s}_i||,1) & \dots & C(||\mathbf{s}_j-\mathbf{s}_i||,t) \\ C(||\mathbf{s}_j-\mathbf{s}_i||,-1) & C(||\mathbf{s}_j-\mathbf{s}_i||,0) & \dots & C(||\mathbf{s}_j-\mathbf{s}_i||,t-1) \\ \vdots & \vdots & \ddots & \dots \\ C(||\mathbf{s}_j-\mathbf{s}_i||,-t) & C(||\mathbf{s}_j-\mathbf{s}_i||,1-t) & \dots & C(||\mathbf{s}_j-\mathbf{s}_i||,0) \end{pmatrix} \tag{4.10}$$

for $i,j=1,\dots,n$.

The covariance matrix $\boldsymbol{\Sigma}_{pp}$ is also a block matrix consisting of $(m-n)^2$ blocks, where each block is a $h\times h$ matrix with the form of (4.10), but with $\mathbf{s}_i$ and $\mathbf{s}_j$ running through the locations of the prediction set, i.e., $i,j=n+1,\dots,m$.

The covariance matrix $\boldsymbol{\Sigma}_{po}$ is similarly defined as a matrix with $n(m-n)$ blocks

$$\boldsymbol{\Sigma}_{po} = \begin{pmatrix} \mathcal{C}'_{n+1,1} & \mathcal{C}'_{n+1,2} & \dots & \mathcal{C}'_{n+1,n} \\ \mathcal{C}'_{n+2,1} & \mathcal{C}'_{n+2,2} & \dots & \mathcal{C}'_{n+2,n} \\ \vdots & \vdots & \ddots & \dots \\ \mathcal{C}'_{m,1} & \mathcal{C}'_{m,2} & \dots & \mathcal{C}'_{m,n} \end{pmatrix}$$

where $\mathcal{C}'_{ij}$ is a $h \times (t+1)$ covariance matrix of the form

$$\mathcal{C}'_{ij} = \begin{pmatrix} C(||\mathbf{s}_j - \mathbf{s}_i||, t+1) & \dots & C(||\mathbf{s}_j - \mathbf{s}_i||, 1) \\ C(||\mathbf{s}_j - \mathbf{s}_i||, t+2) & \dots & C(||\mathbf{s}_j - \mathbf{s}_i||, 2) \\ \vdots & \ddots & \dots \\ C(||\mathbf{s}_j - \mathbf{s}_i||, t+h) & \dots & C(||\mathbf{s}_j - \mathbf{s}_i||, h) \end{pmatrix} \tag{4.11}$$

for $j = 1, \dots, n$ and $i = n+1, \dots, m$ and $\boldsymbol{\Sigma}_{op} = \boldsymbol{\Sigma}_{po}^T$.

Recall that the kriging predictor of $\mathbf{Z}_p$ takes the form $\mathbf{Z}_p^* = C + \Gamma \mathbf{Z}_o$, where C and Γ are chosen to minimize the mean squared prediction error $\mathbb{E}(\mathbf{Z}_p - C - \Gamma \mathbf{Z}_o)^2$. Under the joint assumption in (4.8), the kriging predictor of $\mathbf{Z}_p$ coincides with the conditional expectation $\mathbb{E}(\mathbf{Z}_p | \mathbf{Z}_o)$ and is given by (Cressie and Wikle, 2015)

$$\mathbf{Z}_p^* = \boldsymbol{\mu}_p + \boldsymbol{\Sigma}_{po} \boldsymbol{\Sigma}_{oo}^{-1} (\mathbf{Z}_o - \boldsymbol{\mu}_o). \tag{4.12}$$

The kriging variance is the minimized mean squared error. Under the joint Gaussian assumption of (4.8), it is the conditional variance (Cressie and Wikle, 2015), that is

$$\boldsymbol{\sigma}^2 = \boldsymbol{\Sigma}_{pp} - \boldsymbol{\Sigma}_{po} \boldsymbol{\Sigma}_{oo}^{-1} \boldsymbol{\Sigma}_{op}. \tag{4.13}$$

The kriging equations in (4.12) and (4.13) involve taking the inverse of the covariance matrix $\boldsymbol{\Sigma}_{oo}$, but for large spatio-temporal data sets, this is not always possible. In practice, some simplifications on the covariance function are usually assumed in order to overcome the computational complexity in the calculation of $\boldsymbol{\Sigma}_{oo}^{-1}$. One alternative is to assume that the space-time covariance function can be written as a product of two functions. One is only a function of space and the other only a function of time; that is

$$\boldsymbol{\Sigma}_{oo} = \boldsymbol{\Sigma}_{oo}^{(s)} \otimes \boldsymbol{\Sigma}_{oo}^{(t)}, \tag{4.14}$$

where $\otimes$ is the Kronecker product, $\boldsymbol{\Sigma}_{oo}^{(s)}$ is a $n \times n$ purely spatial covariance matrix and $\boldsymbol{\Sigma}_{oo}^{(t)}$ is a $(t+1) \times (t+1)$ covariance matrix of purely temporal covariances. When $\boldsymbol{\Sigma}_{oo}$ is given by (4.14), we have that

$$\boldsymbol{\Sigma}_{oo}^{-1} = (\boldsymbol{\Sigma}_{oo}^{(s)})^{-1} \otimes (\boldsymbol{\Sigma}_{oo}^{(t)})^{-1} \tag{4.15}$$

and therefore the matrix inverses in (4.12) and (4.13) are of much smaller order

than the $n(t+1) \times n(t+1)$ matrix $\mathbf{\Sigma}_{oo}$. A detailed survey of spatial-temporal processes with a separable covariance can be found in Mardia and Goodall (1993).

Another important point here is the choice of a covariance model in order to calculate the covariances between the observed values and the new locations and times where prediction is desired, namely $\mathbf{\Sigma}_{po}$ in (4.12). This choice should fit the data that is being modelled. Furthermore, it is important to define a spatio-temporal covariance function that is valid. Cressie and Huang (1999) proposed a method to construct valid classes of covariance models based on Bochner's theorem by considering spectral densities. Another approach, which directly builds a valid covariance model in the space-time domain is shown in Gneiting (2002). Several examples of valid space-time covariance functions are defined in Gneiting and Guttorp (2007).

4.2 INLA and the SPDE approach

4.2.1 The SPDE approach

The stochastic partial differential equation (SPDE) approach proposed by Lindgren et al. (2011) consists of representing a Gaussian random field (GRF) with Matérn covariance function as a solution of an SPDE. The original idea comes from the work in Whittle (1954) and Whittle (1963), where it has been shown that the solution to the SPDE

$$(\kappa^2 - \Delta)^{\alpha/2}(\tau x(\mathbf{s})) = \mathcal{W}(\mathbf{s}), \quad \mathbf{s} \in \mathbb{R}^d, \alpha = \nu + d/2, \kappa > 0, \nu > 0, \tag{4.16}$$

is a GRF with Matérn covariance function. The innovation process $\mathcal{W}(\mathbf{s})$ on the right hand side of (4.16) is a Gaussian white noise, κ is the spatial scale parameter, α controls the smoothness, τ controls the variance and Δ is the Laplacian. The Matérn covariance function is given by

$$\text{Cov}(\mathbf{s}, \mathbf{s}+\mathbf{h}) = \frac{\sigma^2}{\Gamma(\nu)2^{\nu-1}}(\kappa||\mathbf{h}||)^\nu K_\nu(\kappa||\mathbf{h}||), \tag{4.17}$$

where K_1 is the modified Bessel function of second kind, order ν. The parameter κ can be used to select the range and ν is a smoothness parameter determining the mean-square differentiability of the underlying process.

The relation between the representations in (4.16) and (4.17) is given by equalities involving the smoothness parameter $\nu = \alpha - d/2$ and the marginal variance

$$\sigma^2 = \frac{\Gamma(\nu)}{\Gamma(\alpha)(4\pi)^{d/2}\kappa^{2\nu}}. \tag{4.18}$$

Even though the formulation of Matérn fields as solution to (4.16) might seem unnecessarily complicated, Lindgren et al. (2011) showed that by using a finite basis-function representation of the infinite dimensional GRF, one can derive a local representation with Markov properties. The main advantage of this derivation is that it allows for computationally effective numerical methods, since the Markov properties imply in sparse precision matrices. Specifically, an approximation to the solution of the SPDE in (4.16) can be obtained using the finite element method (FEM), a numerical technique for solving partial differential equations. This approximation is obtained as

$$x(\mathbf{s}) = \sum_{k=1}^{n} \psi_k(\mathbf{s}) w_k, \tag{4.19}$$

for Gaussian distributed weights $\{w_k\}$ and basis functions $\{\psi_k\}$. The joint distribution for the weights in the basis function expansion in (4.19) is a GMRF and the distribution follows directly from the basis functions and a triangulation of the space. While the weights provide the values of the field at the nodes, the values in the interior of the triangles are determined by linear interpolation (Lindgren et al., 2011).

Using Neumann boundary conditions, Lindgren et al. (2011) showed that in the case where $\alpha = 2$, the precision matrix for the Gaussian weights $\mathbf{w} = \{w_1, \ldots, w_n\}$ is given by

$$\mathbf{Q} = \tau^2(\kappa^4\mathbf{C} + 2\kappa^2\mathbf{G} + \mathbf{G}\mathbf{C}^{-1}\mathbf{G}), \tag{4.20}$$

where the elements of the diagonal matrix $\mathbf{C}$ are given by $C_{ij} = \int \psi_i(\mathbf{s})\mathrm{d}\mathbf{s}$, the elements of the sparse matrix $\mathbf{G}$ are $G_{ij} = \int \nabla\psi_i(\mathbf{s})\nabla\psi_j(\mathbf{s})\mathrm{d}\mathbf{s}$, where ∇

denotes the gradient. The latent field $\mathbf{x}$ yields a GMRF with distribution $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}^{-1})$, which approximates the solution to the SPDE (in a stochastically weak sense) (Blangiardo et al., 2013). Note that we describe the case $\alpha = 2$, since this is the value that is used in our applications in Chapter 6. Other values of α have been discussed by Lindgren et al. (2011).

4.2.2 Parameter estimation and prediction using INLA

Hierarchical models belonging to the class of latent Gaussian models can be estimated using the integrated nested Laplace approximation (INLA) method proposed by Rue et al. (2009). The **INLA** package for computation in R is available for download at www.r-inla.org.

Latent Gaussian models can be represented by a hierarchical structure containing three stages

$$y_i|\mathbf{x}, \boldsymbol{\theta} \sim p(y_i|x_i, \boldsymbol{\theta}) \tag{4.21a}$$

$$\mathbf{x}|\boldsymbol{\theta} \sim p(\mathbf{x}|\boldsymbol{\theta}) \tag{4.21b}$$

$$\boldsymbol{\theta} \sim p(\boldsymbol{\theta}) \tag{4.21c}$$

where $\mathbf{y} = (y_1, \ldots, y_n)^T$ denotes the observed data, $\mathbf{x} = (x_1, \ldots, x_n)^T$ the latent field and $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_m)^T$ is the hyperparameter vector, where $x|\boldsymbol{\theta}$ follows a Gaussian distribution and the likelihood is conditionally independent in the sense that y_i only depends on x_i and $\boldsymbol{\theta}$. It follows that the likelihood is given by

$$p(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta}) = \prod_{i=1}^{n} p(y_i|\mathbf{x}, \boldsymbol{\theta}). \tag{4.22}$$

The components of the latent Gaussian field $\mathbf{x}$ are often assumed to be conditionally independent with the consequence that the precision matrix is sparse. In particular, Rue et al. (2009) considered that $\mathbf{x}$ satisfies the Markov properties, i.e., x_i and x_j are conditionally independent given all the other components $\mathbf{x}_{-ij}$, where $\mathbf{x}_{-ij}$ denotes $\mathbf{x}$ with the i, jth components omitted. This means that for a general pair i and j with $i \neq j$, the corresponding element of the precision matrix is null: $x_i \perp\!\!\!\perp x_j|\mathbf{x}_{-ij} \Leftrightarrow Q_{ij} = 0$, where $\mathbf{Q}$ is the precision matrix of $\mathbf{x}$.

The posterior distribution may be written as the product of the likelihood in (4.22), the GMRF density of $\mathbf{x}|\boldsymbol{\theta}$ and the hyperparameters prior distribution

$$p(\mathbf{x}, \boldsymbol{\theta}|\mathbf{y}) \propto p(\boldsymbol{\theta})p(\mathbf{x}|\boldsymbol{\theta}) \prod_{i=1}^{n} p(y_i|\mathbf{x}, \boldsymbol{\theta}). \tag{4.23}$$

Rue et al. (2009) focused on approximating the marginal distributions $p(x_i|\mathbf{y})$ and $p(\theta_i|\mathbf{y})$. The approximation of $p(\boldsymbol{\theta}|\mathbf{y})$, denoted by $\tilde{p}(\boldsymbol{\theta}|\mathbf{y})$, is the Laplace approximation originally proposed by Tierney and Kadane (1986),

$$\tilde{p}(\boldsymbol{\theta}|\mathbf{y}) \propto \left. \frac{p(\mathbf{x}, \boldsymbol{\theta}, \mathbf{y})}{p_G(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y})} \right|_{\mathbf{x}=\mathbf{x}^*(\boldsymbol{\theta})} \tag{4.24}$$

where $p_G(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y})$ is a Gaussian approximation of the full conditional distribution of $\mathbf{x}$ and $\mathbf{x}^*(\boldsymbol{\theta})$ is the mode of the full conditional $\mathbf{x}$ for a given $\boldsymbol{\theta}$.

There are different ways of approximating $p(x_i|\boldsymbol{\theta}, \mathbf{y})$ (Rue et al., 2009). The simplest method uses the marginal mean and the variances of the Gaussian approximation $p_G(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y})$. This approximation can afford reasonable results under certain conditions, but it can also lead to errors due to lack of skewness (Rue and Martino, 2007). A more accurate way of approximating $p(x_i|\boldsymbol{\theta}, \mathbf{y})$ is to use the Laplace approximation to obtain

$$\tilde{p}_{LA}(x_i|\boldsymbol{\theta}, \mathbf{y}) \propto \left. \frac{p(\mathbf{x}, \boldsymbol{\theta}, \mathbf{y})}{\tilde{p}_{GG}(\mathbf{x}_{-i}|x_i, \boldsymbol{\theta}, \mathbf{y})} \right|_{\mathbf{x}_{-i}=\mathbf{x}^*_{-i}(x_i, \boldsymbol{\theta})} \tag{4.25}$$

where $\mathbf{x}^*_{-i}(x_i, \boldsymbol{\theta})$ is the mode of $p(\mathbf{x}_{-i}|x_i, \boldsymbol{\theta}, \mathbf{y})$ and $\tilde{p}_{GG}(\mathbf{x}_{-i}|x_i, \boldsymbol{\theta}, \mathbf{y})$ is the Gaussian approximation of $x_{-i}|x_i, \boldsymbol{\theta}, \mathbf{y}$. A third option is the simplified Laplace approximation, which uses a Taylor expansion of (4.25) up to third order, that corrects the Gaussian approximation for location and skewness with relatively low computational cost. For more details on the Gaussian, Laplace and simplified Laplace approximations, see Rue and Martino (2007).

Given $\tilde{p}(\boldsymbol{\theta}|\mathbf{y})$ and $\tilde{p}(x_i|\boldsymbol{\theta})$, approximations to the posterior marginal distributions

$p(x_i|\mathbf{y})$ and $p(\theta_i|\mathbf{y})$ are calculated using numerical integration

$$\tilde{p}(x_i|\mathbf{y}) = \int \tilde{p}(\boldsymbol{\theta}|\mathbf{y})\tilde{p}(x_i|\boldsymbol{\theta},\mathbf{y})d\boldsymbol{\theta} \tag{4.26}$$

$$\tilde{p}(\theta_i|\mathbf{y}) = \int \tilde{p}(\boldsymbol{\theta}|\mathbf{y})d\boldsymbol{\theta}_{-j}, \tag{4.27}$$

where the integral in (4.27) can be calculated numerically through a finite weighted sum

$$\tilde{p}(x_i|\mathbf{y}) \approx \sum_j \tilde{p}(x_i|\boldsymbol{\theta}^{(j)},\mathbf{y})\tilde{p}(\boldsymbol{\theta}^{(j)}|\mathbf{y})\boldsymbol{\Delta}_j, \tag{4.28}$$

where $\boldsymbol{\theta}^{(j)}$ are the evaluation points with corresponding weights $\boldsymbol{\Delta}_j$.

4.2.3 Application to spatio-temporal inference and prediction

Consider the real-valued process $\{Y(\mathbf{s},t) : \mathbf{s} \in \mathcal{D}_s, t \in \mathcal{D}_t\}$, which is a partial realization of the process on the entire space and time domains. Next, suppose that the data stems from measurements at each location $\mathbf{s}_i$ at time t, denoted as $Y(\mathbf{s}_i,t)$, $i = 1,\ldots,n$ and $t = 1,\ldots,T$. For each location $\mathbf{s}_i$ and time t, we assume a spatio-temporal model, which is used to estimate particular matter concentration in Cameletti et al. (2013) and is given by

$$Y(\mathbf{s}_i,t) = \mathbf{z}(\mathbf{s}_i,t)^T\boldsymbol{\beta} + x(\mathbf{s}_i,t) + \epsilon(\mathbf{s}_i,t), \quad \epsilon(\mathbf{s}_i,t) \sim \mathcal{N}(0,1/\tau_\epsilon), \tag{4.29}$$

where the mean surface $\mu(\mathbf{s}_i,t) = \mathbf{z}(\mathbf{s}_i,t)^T\boldsymbol{\beta}$ is a linear function of some spatially and temporally referenced explanatory variables $\mathbf{z}(\mathbf{s}_i,t)$. The term $\epsilon(\mathbf{s}_i,t)$ is an uncorrelated Gaussian process with zero mean and variance $1/\tau_\epsilon$. Moreover, $x(\mathbf{s}_i,t)$ is the realization of a spatio-temporal process that varies in time with first order autoregressive dynamics

$$x(\mathbf{s}_i,t) = \rho x(\mathbf{s}_i,t-1) + w(\mathbf{s}_i,t), \tag{4.30}$$

for $t = 2, \ldots, T$, $|\rho| < 1$ and $x(\mathbf{s}_i, 1) \sim \mathcal{N}(0, \sigma_w^2/(1-a)^2)$. The term $w(\mathbf{s}, t)$ is a zero-mean Gaussian field with covariance function

$$\text{Cov}(w(\mathbf{s}_i, t), w(\mathbf{s}_j, t')) = \begin{cases} \sigma_w^2 C(h), & \text{if } t = t' \\ 0, & t \neq t' \end{cases}$$

for $\mathbf{s}_i \neq \mathbf{s}_j$. The correlation function C depends on the locations $\mathbf{s}_i$ and $\mathbf{s}_j$ through the distance $h = ||\mathbf{s}_i - \mathbf{s}_j||$. This means that the process is assumed to be second-order stationary and isotropic (Cressie, 1992). The marginal variance is $\text{Var}(w(\mathbf{s}_i, t)) = \sigma_w^2$ for each $\mathbf{s}_i$ and t and $C(h)$ is the Matérn correlation function in (4.17).

Using the SPDE approach outlined in Section 4.2.1, we construct a Kronecker product model starting from the basis representation

$$x(\mathbf{s}, t) = \sum_{k=1}^{d} \psi_k(\mathbf{s}, t) x_k, \tag{4.31}$$

where each basis function is the product of a spatial and a temporal basis function, $\psi_k(\mathbf{s}, t) = \psi_i^s(\mathbf{s})\psi_j^t(t)$. The space-time SPDE is given by

$$\frac{\partial}{\partial t}(\kappa(\mathbf{s})^2 - \Delta)^{\alpha/2}(\tau(\mathbf{s})x(\mathbf{s}, t)) = \mathcal{W}(\mathbf{s}, t), \quad (\mathbf{s}, t) \in \Omega \times \mathbb{R} \tag{4.32}$$

and it generates a precision matrix for the weight vector $\mathbf{x}$ as $\mathbf{Q} = \mathbf{Q}_T \otimes \mathbf{Q}_S$, where $\mathbf{Q}_S$ is the precision matrix for the purely spatial model computed using (4.20). The dimension of $\mathbf{Q}_S$ is given by the number of vertices of the domain triangulation, whereas $\mathbf{Q}_T$ is the precision of the T-dimensional precision matrix of the autoregressive process of order 1 in (4.30), that is,

$$\mathbf{Q}_T = \begin{pmatrix} 1/\sigma_w^2 & -\rho/\sigma_w^2 & & & \\ -\rho/\sigma_w^2 & (1+\rho^2)/\sigma_w^2 & & & \\ & \ddots & & & \\ & & & (1+\rho^2)/\sigma_w^2 & -\rho/\sigma_w^2 \\ & & & -\rho/\sigma_w^2 & 1/\sigma_w^2 \end{pmatrix}.$$

Writing the vector form as $\mathbf{x}_t = (x(\mathbf{s}_1, t), \ldots, x(\mathbf{s}_n, t))$, it follows that the distri-

bution of the GMRF $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_T)^T$ is

$$\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}^{-1}). \tag{4.33}$$

The basis function representation of $x(\mathbf{s}, t)$ is used to define a sparse matrix of weights $\mathbf{A}$ (also called projector matrix) with dimension $n \times d$ that selects the value of the GMRF $\mathbf{x}_t$ for each observation vector $\mathbf{y}_t$. The key point is to have a GMRF modelled at the triangulation vertices with dimension d and a response variable observed at n locations. Thus, the $\mathbf{A}$ matrix is responsible for projecting the process from the triangulation vertices to the observed locations. From this representation, (4.29) can be written as

$$Y(\mathbf{s}_i, t) = \mathbf{z}(\mathbf{s}_i, t)^T \boldsymbol{\beta} + \sum_{j=1}^{n} A_{ij} \mathbf{x}_t + \epsilon(\mathbf{s}_i, t), \tag{4.34}$$

where the row-sums of $\mathbf{A}$ are 1, since the piecewise linear basis functions that are used sum to 1 at each location (Lindgren et al., 2011).

Figure 4.1 illustrates the triangulation of the spatial domain at a specific time t for the western Denmark data, which refers to a system of 349 wind farms and it is described in more details in Section 6.1. The triangulation was done similarly for the other time points t.

Within the **R-INLA** terminology, it is possible to use the options `group` and `control.group` to specify that at each time point, the spatial locations are linked by the SPDE object, while across time the process evolves according to an AR(1) process (Blangiardo et al., 2013). The `group` feature of **R-INLA** is key for the implementation of spatio-temporal models, as it allows the user to specify that the underlying process should be replicated at each time point with a specific dependence structure.

Predictions are then obtained using the INLA framework outlined in Section 4.2.2 (for more details, refer to Rue et al. (2009)). Note that the INLA approach provides the posterior conditional distribution of the latent field $\mathbf{x}_t$ and of the response $\mathbf{y}_t$ for the d vertices in the triangulation. Predictions are then obtained from the triangulation vertices to the target locations by projecting the posterior mean of $\mathbf{x}_t$ and $\mathbf{y}_t$ using the correspondent projector matrix.

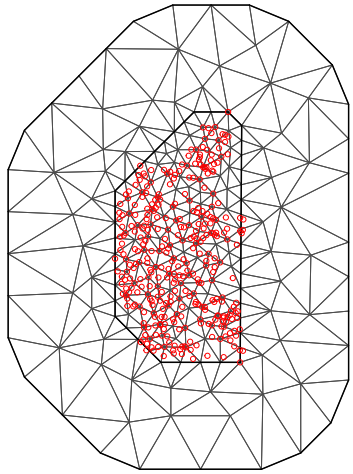


Figure 4.1: Triangulation of the domain constructed with the package **R-INLA**, where the red dots denote the observation locations. The spatial domain was extended to avoid a boundary effect.

Notice that Kronecker models generate separable covariance functions (Lindgren and Rue, 2015), which are simple but sometimes unrealistic. Even though the representation of the SPDE precision structures also allows for construction of non-separable models, in our applications using the GMRF framework we restrict ourselves to covariance structures based on the Kronecker model, which is the model that is currently implemented in the **R-INLA** package (last accessed June, 2016).

Chapter 5

Models tailored to wind power prediction

With the preliminary discussion of the various features of a typical wind power data set presented in the first chapters, the next step is to consider the potential models that are able to describe the data successfully. In this chapter, appropriate representations for different applications in wind power forecasting are constructed. Combined with the prediction methods described in Chapter 4, the approaches described here allow us to obtain spatial and spatio-temporal predictions of wind power generation together with a quantification of the uncertainties.

First, the so-called hurdle model, which is used in applications where the data consists of nonnegative values with a substantial proportion of zeros, is introduced. This approach, combined with the INLA-SPDE from Section 4.2, is used to perform inference and prediction of wind power generation in Paper A. The second representation in this chapter builds on the generalized logit-Normal distribution. This is the motivation in Paper B, where two state-of-the-art statistical techniques are combined to obtain probabilistic forecasts of wind power production: a parametric technique based on the generalized logit-normal dis-

tribution and the INLA-SPDE approach to obtain fast inference and prediction. Next, focus is placed on how to use a truncated Gaussian field to model wind power. In this case, kriging in a spatio-temporal context, as introduced in Section 4.1.2, is used to obtain probabilistic forecasts of wind power production (Paper C).

5.1 Hurdle model

The hurdle model for count data was proposed by Mullahy (1986). One part of the model is a binary model, such as a logistic or a probit regression, to determine whether the response outcome is zero or positive. To estimate the level of the positive outcomes, the second part of the model consists of a truncated model that modifies an ordinary distribution by conditioning on a positive outcome. Applications of similar models can be found in Pohlmeier and Ulrich (1995) and Gurmu (1997).

First, the probability of having power generation greater than zero is modelled as a Bernoulli distribution. In a second stage, the non-zero values are modelled with a gamma distribution, which represents the amount of wind power generated. The gamma distribution is a good choice for describing wind power values for several reasons. It provides a flexible representation of a variety of distribution shapes while utilizing only two parameters: the shape and the scale. It can range from exponential-decay forms for shape values near one to nearly normal forms for shape values beyond 20 (Wilks, 1990). In addition, a distribution that excludes negative values and is positively skewed is readily applicable for the analysis of wind power.

The hurdle model specification has the advantage of a well-behaved likelihood function that factors into two independent terms, making calculations relatively simple. The first term has only the logit model parameters and the second term involves only the parameters of the gamma distribution (see (5.5)). To overcome the so-called intermittency problem, where areas with small values lie very close to areas with large values, there is an underlying Gaussian field that is part of the linear predictor of both distributions - Bernoulli and gamma.

Let us start by defining a binary random variable $Z(\mathbf{s}_i)$ at location $\mathbf{s}_i = 1, \ldots, N$, which depends on the generation of wind power

$$Z(\mathbf{s}_i) = \begin{cases} 1, & \text{if} \quad y(\mathbf{s}_i) > 0, \\ 0, & \text{otherwise.} \end{cases}$$

where $y(\mathbf{s}_i)$ is the observed wind power generation at wind farm $\mathbf{s}_i$. We assume that $Z(\mathbf{s}_i)$ follows a Bernoulli distribution with parameter $p(\mathbf{s}_i)$

$$Z(\mathbf{s}_i) \sim \text{Bern}\left(p(\mathbf{s}_i)\right), \tag{5.1}$$

where $p(\mathbf{s}_i)$ is the probability of having wind power generation greater than zero at wind farm $\mathbf{s}_i$ and is modelled as

$$\text{logit}\left(p(\mathbf{s}_i)\right) = \alpha_z + x(\mathbf{s}_i) \tag{5.2}$$

with α_z being the intercept and $x(\mathbf{s}_i)$ an observation from a latent Gaussian random field with Matérn covariance (see (4.17)) modelled through the SPDE approach described in Section 4.2.1.

Conditional on the presence or absence of wind power, the amount of wind power generation at wind farm $\mathbf{s}_i$ is modelled as

$$Y(\mathbf{s}_i)|Z(\mathbf{s}_i) > 0 \sim \text{Gamma}\left(\phi, \frac{\phi}{\mu(\mathbf{s}_i)}\right) \tag{5.3}$$

with the expected value $\mu(\mathbf{s}_i)$ at wind farm $\mathbf{s}_i$, defined as

$$\mu(\mathbf{s}_i) = \exp\left[\alpha_y + \beta_y x(\mathbf{s}_i)\right], \tag{5.4}$$

where α_y is the intercept and β_y is the scaling parameter for $x(\mathbf{s}_i)$, which is defined in (5.2).

The joint likelihood function is given by the product of the likelihood for the

occurrence and the amount as

$$\begin{aligned}
\mathrm{L}(\boldsymbol{\theta}|\mathbf{z},\mathbf{y}) &= \prod_{z(\mathbf{s}_i)=0} \mathrm{Bern}(z(\mathbf{s}_i)|p(\mathbf{s}_i)) \\
&\quad \prod_{z(\mathbf{s}_i)=1} \mathrm{Bern}(z(\mathbf{s}_i)|p(\mathbf{s}_i))\mathrm{Gamma}(y(\mathbf{s}_i)|m(\mathbf{s}_i),\phi) \\
&= \prod_{z(\mathbf{s}_i)} \mathrm{Bern}(z(\mathbf{s}_i)|p(\mathbf{s}_i))\mathrm{Gamma}(y(\mathbf{s}_i)|m(\mathbf{s}_i),\phi)^{z(\mathbf{s}_i)} \\
&= \prod_{z(\mathbf{s}_i)} p(\mathbf{s}_i)^{z(\mathbf{s}_i)}(1-p(\mathbf{s}_i))^{1-z(\mathbf{s}_i)} \\
&\quad \Big[\frac{1}{\Gamma(\phi)}\Big(\frac{\phi}{\mu(\mathbf{s}_i)}\Big)^{\phi} y(\mathbf{s}_i)^{\phi-1}\exp\Big(-\phi\frac{y(\mathbf{s}_i)}{\mu(\mathbf{s}_i)}\Big)\Big]^{z(\mathbf{s}_i)}. \qquad (5.5)
\end{aligned}$$

The binary variables $z(\mathbf{s}_i)$ for $\mathbf{s}_i = 1, 2, .., N$ are treated as observed variables in this model. The default values are used for the prior parameter, where a log-gamma prior is assumed for κ and ϕ and a normal prior with a fixed vague precision is assumed for the fixed effects α_z, α_y and β_y.

Notice that the derivations are presented for modelling spatial data only, since that is the focus of our application when using hurdle models. However, the modification to add the temporal dimension is straightforward.

Since the hurdle model described above is essentially a spatial model with a latent Gaussian structure, it is possible to take advantage of the computational gains of the INLA-SPDE methodology described in Section 4.2 to perform inference and predictions. Combining two likelihoods in the same model is a technique that has become common while implementing advanced models in **R-INLA**, as described in the Chapter 8 of Blangiardo and Cameletti (2015). Specifically, the `copy` feature of **R-INLA** allows us to link the same GRF $\mathbf{x}$ to the occurrence and the amount of wind power production (see (5.2) and (5.4)). In this case, the hyperparameter β_y in (5.4) is responsible for copying the GRF in the linear predictor of the Bernoulli variable Z (occurrence) into the linear predictor of the gamma variable Y (amount). The motivation for using a single latent Gaussian field to control both regimes, occurrence and amount of power generation, is to prevent areas with very small values from being too close to areas with large values.

5.2 Generalized logit–normal distribution

The generalized logit transformation can be traced back to the work in Mead (1965). It was later used by Pinson (2012) to obtain predictive densities of wind power time series as an alternative to more commonly assumed normal and beta distributions. It has been shown to be a suitable candidate to model wind power, since it can account for the non-Gaussianity and the double-bounded nature of wind power generation.

Let $X(\mathbf{s},t)$ denote the normalized wind power production random variable at location $\mathbf{s} \in \mathcal{D}_s$ and time $t \in \mathcal{D}_t$ with respective observations or measurements indicated by $x(\mathbf{s},t)$. The generalized logit transformation is given by

$$y(\mathbf{s},t) = \gamma(x(\mathbf{s},t);\nu) = \ln\Big(\frac{x^\nu(\mathbf{s},t)}{1-x^\nu(\mathbf{s},t)}\Big), \quad x(\mathbf{s},t) \in (0,1) \tag{5.6}$$

with inverse

$$x(\mathbf{s},t) = \gamma^{-1}(y(\mathbf{s},t);\nu) = \left(1+\frac{1}{\exp(y(\mathbf{s},t))}\right)^{-1/\nu}, \quad \nu > 0, \quad y(\mathbf{s},t) \in \mathbb{R}. \tag{5.7}$$

Assuming that the variable $X(\mathbf{s},t)$ is generalized logit-normal distributed, it follows that the transformed variable $Y(\mathbf{s},t) = \gamma(X(\mathbf{s},t))$ is normal distributed. The density function of a generalized logit normal variable X (Pinson, 2012) is given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\frac{\nu}{x(1-x^\nu)}\exp\left[-\frac{1}{2}\left\{\frac{\gamma(x;\nu)-\mu}{\sigma}\right\}^2\right], \tag{5.8}$$

where location and scale parameters μ and σ are directly connected to the mean and variance of the variable $Y \sim \mathcal{N}(\mu,\sigma^2)$.

We now consider the more complex case where the complete distribution of $X(\mathbf{s},t)$ is a discrete-continuous mixture of the generalized logit-normal distribution $L_\nu(\mu(\mathbf{s},t),\sigma^2)$. It can be written as

$$X(\mathbf{s},t) \sim \delta_0 w^0(\mathbf{s},t) + \delta_1 w^1(\mathbf{s},t) + (1-w^0(\mathbf{s},t)-w^1(\mathbf{s},t))L_\nu(\mu(\mathbf{s},t),\sigma^2), \tag{5.9}$$

with probability masses $w^0(\mathbf{s},t)$ and $w^1(\mathbf{s},t)$ corresponding to zero (no turbines are operating) and nominal power (all turbines are generating their maximum

power), respectively, given by

$$\begin{aligned} w^0(\mathbf{s},t) &= \Phi\left\{\frac{\gamma(\epsilon;\nu)-\mu(\mathbf{s},t)}{\sigma}\right\} \quad \text{and} \\ w^1(\mathbf{s},t) &= 1-\Phi\left\{\frac{\gamma(1-\epsilon;\nu)-\mu(\mathbf{s},t)}{\sigma}\right\}, \end{aligned} \tag{5.10}$$

where Φ is the cumulative distribution function of a standard normal variable and δ_A is the indicator function that is equal to one when property A is satisfied and zero otherwise. We follow the approach by Lesaffre et al. (2007) for modelling outcome scores in $[0,1]$. For this purpose, we define the threshold value ϵ in (5.10), also known as order of measurement precision. Moreover, the applications in this thesis are restricted to the case where $\nu = 1$. This leads to the more classical logit-normal transformation, which was recently used by Lau and McSharry (2010) and Dowell and Pinson (2016) for modelling wind power production.

The key feature here is that the distribution of $X(\mathbf{s},t)$ is fully parametrized by the conditional mean and variance of $Y(\mathbf{s},t) \sim \mathcal{N}(\mu(\mathbf{s},t),\sigma^2)$. Thus, in order to calculate density forecasts of wind power generation at lead time h, $X(\mathbf{s},t+h)$, it is only necessary to model the location and scale parameters of the predictive distribution, which are the mean and variance of the transformed process $Y(\mathbf{s},t+h)$.

Dynamics representation

As already mentioned, we denote by $Y(\mathbf{s},t)$ the normalized logit-normal transformed wind power generation at location $\mathbf{s}$ and time t, which is calculated using (5.7). We assume the following distribution for $Y(\mathbf{s},t)$ at the first level of the hierarchical models

$$Y(\mathbf{s},t) \sim \mathcal{N}\left(\mu(\mathbf{s},t),\sigma_e^2\right), \tag{5.11}$$

where the variance σ^2 is assumed to be a Gaussian white noise process both serially and spatially uncorrelated. For the term $\mu(\mathbf{s},t)$, we take a Bayesian approach, which leads to other process levels giving rise to different hierarchical models that are described in what follows.

Temporal model (Model T)

The first representation is a time series model where each wind farm is considered as an independent replicate of the same random process. The independence assumption is, of course, a simplification, since the wind power production in one location is most likely dependent on the production in other locations.

We assume that the mean function $\mu(\mathbf{s}, t)$ in (5.11), which coincides with the linear predictor in the hierarchical formulation, is constant in time and can be modelled as

$$\mu(\mathbf{s}, t) = b(\mathbf{s}) + w_{\mathbf{s}}(t), \tag{5.12}$$

where $b(\mathbf{s})$ is an intercept specific for each location and $w_{\mathbf{s}}(t)$ is an autoregressive process that can be written as

$$w_{\mathbf{s}}(t) = \rho_1 w_{\mathbf{s}}(t-1) + \nu_{\mathbf{s}}(t) \tag{5.13}$$

with $t = 2, \ldots, T$ and $|\rho_1| < 1$. The term $\nu_{\mathbf{s}}$ is uncorrelated with $w_{\mathbf{s}}(t)$ and independent identically distributed as $\nu_{\mathbf{s}} \sim N(0, \sigma_\nu^2)$.

Spatio-temporal model (Model S-T)

In order to handle the spatio-temporal features present in the data, this model consists of a spatio-temporal process with temporal dynamics as in Cameletti et al. (2013). This type of representation is commonly used for modelling air quality because of its flexibility in including time and space dependency as well as covariates, see e.g. Fassò and Finazzi (2011) and Cocchi et al. (2007). The mean function $\mu(\mathbf{s}, t)$ in (5.11) is given by

$$\mu(\mathbf{s}, t) = b_0 + z(\mathbf{s}, t), \tag{5.14}$$

where b_0 is an intercept that is common to all wind farms and constant in time and space. The term $z(\mathbf{s}, t)$ is a spatio-temporal process that varies in time with first order autoregressive dynamics

$$z(\mathbf{s}, t) = \rho_2 z(\mathbf{s}, t-1) + w(\mathbf{s}, t) \tag{5.15}$$

with $t = 2, \ldots, T$ and $|\rho_2| < 1$. Moreover, $w(s,t)$ is a zero-mean Gaussian field, assumed to be temporally independent with covariance function

$$\text{Cov}(w(\mathbf{s},t), w(\mathbf{s}',t')) = \begin{cases} \sigma_w^2 C(h), & \text{if } t = t' \\ 0, & t \neq t' \end{cases}$$

for $\mathbf{s} \neq \mathbf{s}'$. The correlation function C is defined by the Matérn, given in (4.17), which depends on the locations $\mathbf{s}$ and $\mathbf{s}'$ through the distance $h = ||\mathbf{s} - \mathbf{s}'||$. This means that the process is assumed to be second-order stationary and isotropic (Cressie, 1992). The marginal variance is $\text{Var}(\mathbf{s},t) = \sigma_w^2$.

Temporal + Spatio-temporal model (Model ST+T)

This is a model defined by an autoregressive process at each location to capture the individual variability and a spatio-temporal process with temporal dynamics to take into account the spatial dependence among wind farms. Specifically, $\mu(\mathbf{s},t)$ in (5.11) is defined as

$$\mu(\mathbf{s},t) = b_0 + w_{\mathbf{s}}(t) + z(\mathbf{s},t), \tag{5.16}$$

where b_0 is a fixed unknown intercept that is common to all wind farms. The process $w_{\mathbf{s}}(t)$ is assumed to have autoregressive dynamics as defined in (5.13). Finally, $z(\mathbf{s},t)$ is a spatio-temporal component that has the structure of (5.15) and its spatio-temporal covariance function is the same as in (4.17).

For all the models described above, a log-gamma prior is assumed for the parameters in the Matérn covariance as well as for the precision parameters σ_e^2 and σ_v^2. For the fixed effect b's and for the correlations ρ_1 and ρ_2, we assume normal priors.

Essentially, the key feature of the models described above is that they can be handled within the theoretical and computational INLA-SPDE framework, which is described in Section 4.2. More details on the formulation of this approach for the spatio-temporal case, can be found in an example in Section 4.2.3.

5.3 Truncated Gaussian random field

Truncated Gaussian fields are used to model random variables that are bounded below or above by a constant c. The idea of using truncated Gaussian random fields is common in environmental science, including the study of precipitation. In these applications, the truncation is inherent to the actual process, given the natural intermittent characteristic of rainfall. Recent papers by Berrocal et al. (2008) and Kleiber et al. (2012) used a truncated model with two latent Gaussian processes to describe accumulated precipitation. The first Gaussian process controls the probability of observing zero values and the second governs the distribution of non-zero observations. A different approach was taken by Baxevani and Lennartsson (2015), who simultaneously modelled the occurrence and the intensity of rainfall using a single latent Gaussian field. Then they considered the positive part of the process to be observed up to a transformation of the observed data. In such a case, a transformed latent Gaussian field model can be formulated as

$$Y(\mathbf{s},t) = \begin{cases} f_{\mathbf{s}}(Z(\mathbf{s},t)), & Z(\mathbf{s},t) > 0 \\ 0, & Z(\mathbf{s},t) \leq 0, \end{cases} \tag{5.17}$$

where Z is a stationary spatio-temporal Gaussian random field with variance one, a mean that is chosen to make the probability of positive wind power production equal to a specified value, and $f_{\mathbf{s}}(\cdot)$, a positive monotonic function, usually referred to as anamorphosis, chosen to obtain a specified marginal distribution for the positive wind power production at the specific location $\mathbf{s}$. The idea of using models in which the marginal distribution of a Gaussian process is transformed to non-Gaussian distribution dates back to the work in Guillot (1999). This type of approach has the advantage of being able to accommodate any marginal distribution, including discontinuous ones.

The model in (5.17) is also a promising candidate to model wind power data. This is due to the fact that rainfall and precipitation share similar features with wind power, since one could have long periods of dry days without any rainfall observation. In what follows, we describe the details on how one could use a truncated Gaussian model as in (5.17) for wind power data.

Notice that the Gaussian process Z in model (5.17) serves a dual role. On one

hand, it controls the different regimes in the wind power production (equal or greater than zero) and it should, on the other hand, provide the exact value of wind power generation after the anamorphosis transformation has been applied to the positive values of wind power production. Also, in order to fully describe the model in (5.17), we only need to characterize the mean and covariance function of Z and the anamorphosis function $f_{\mathbf{s}}$, since the process Z is assumed to be Gaussian.

In other words, to obtain predictions for the wind power production field Y in (5.17), we set the power production to zero if $Z < 0$, and apply the anamorphosis $f_{\mathbf{s}}$ to the positive values of the underlying Gaussian random field Z. Therefore, models for the Gaussian random field and for the anamorphosis function that appear in (5.17) are critical for the predictive performance. In the next sections, it is presented a model for the transformation $f_{\mathbf{s}}$, which describes the continuous distribution of wind power generation given that the production is greater than zero as well as different possible models for the Gaussian process Z.

5.3.1 Model for the anamorphosis function $f_{\mathbf{s}}(\cdot)$

The anamorphosis function is generally a positive monotonic function chosen in order to obtain a specified marginal distribution for the positive wind power production. The restriction we need to have in mind when choosing the transformation $f_{\mathbf{s}}(\cdot)$ is that the distribution of wind power production is right skewed and bounded between zero and one after normalization.

Using (5.17), we see that

$$F_{Y|Y>0}(\cdot) = F_{Z|Z>0} \circ f^{-1}(\cdot) \Leftrightarrow f(\cdot) = F^{-1}_{Y|Y>0} \circ F_{Z|Z>0}(\cdot), \tag{5.18}$$

where the symbol $\circ$ denotes the composition of two functions. $F_{Y|Y>0}$ and $F_{Z|Z>0}$ (for notational simplicity we suppress the dependence on location and time), are the cdf's of Y and Z given that we observe only their positive parts.

This means that the anamorphosis function in (5.18) only depends on two components, namely $F_{Z|Z>0}$ and $F_{Y|Y>0}$. Since Z is a Gaussian random field with mean μ and variance σ^2, it follows that the distribution $F_{Z|Z>0}$ is truncated

normal with the first two moments $\tilde{\mu}$ and $\tilde{\sigma}^2$ given by the following expressions

$$\tilde{\mu} = \mu + \sigma\lambda(\alpha), \quad \tilde{\sigma}^2 = \sigma^2[1 - \delta(\alpha)], \tag{5.19}$$

with

$$\alpha = -\frac{\mu}{\sigma}, \quad \lambda(\alpha) = \frac{\phi(\alpha)}{1 - \Phi(\alpha)} \quad \text{and } \delta(\alpha) = \lambda(\alpha)[\lambda(\alpha) - \alpha],$$

where Φ and ϕ are, respectively, the cdf and the probability density function (pdf) of a standard (mean 0 and variance 1) normal random variable.

Next, we proceed to specify a model for $F_{Y|Y>0}$. Given that the wind power production data has been normalized, resulting in measurements that fall inside the interval $[0, 1]$, a reasonable model for $F_{Y|Y>0}$ is the beta distribution with parameters $\alpha_{\mathbf{s}}$ and $\beta_{\mathbf{s}}$ that are specific to each farm. That is

$$Y|Y > 0 \overset{d}{=} B(\alpha_{\mathbf{s}}, \beta_{\mathbf{s}}). \tag{5.20}$$

In general, the shape parameters of the beta distribution could be space and time dependent, but since only very short time predictions are of interest in this context, the potential dependence in time is omitted.

5.3.2 Model for the Gaussian process $Z(\mathbf{s}, t)$

With the model specification of the transformation $f_{\mathbf{s}}$, we can now proceed to specify the moments of the latent Gaussian process Z. When defining the mean function μ, one should notice that there is a clear link between the process Z and the observed wind power production Y. It is rather straightforward to show that for $Y = f_{\mathbf{s}}(Z^+)$, where Z^+ is the truncation of the field $Z \overset{d}{=} \mathcal{N}(\mu, \sigma^2)$ and $f_{\mathbf{s}}$ a positive transformation as in (5.17), the following relation holds

$$P(Y > 0) = P(Z > 0) = \Phi\Big(\frac{\mu}{\sigma}\Big). \tag{5.21}$$

It follows that an estimation of the mean value at each location can be obtained by simply inverting (5.21)

$$\hat{\mu} = \phi^{-1}(\hat{p}), \tag{5.22}$$

where $\hat{p}$ is an approximation to the left hand side of (5.21) and is calculated as the proportion of time periods where the field Y is greater than zero (i.e., the observed wind power is positive) over the observational period.

To fully characterize the latent Gaussian random field Z, what remains is to specify the covariance function. For the dependence over space and time, we follow the traditional paradigm of assuming a parametric covariance function. In the following, different covariance models for Z are presented. The choice of such a model is demonstrated for a Danish wind power data set described in Chapter 6 (see Paper C for more details). For instance, plots of the autocorrelation show that the wind power production at a single farm has a slow decay. Similarly, spatial correlation is quite strong even at longer distances. Moreover, analysis of the spatio-temporal empirical correlations indicates that there is a dependence of the correlation from west to east. This basically means that the correlation between a wind farm $\mathbf{s}_i$ and a wind farm $\mathbf{s}_j$ located at the east of $\mathbf{s}_i$ will be larger at temporal lag τ than at temporal lag $-\tau$. Since those features are assumed to be inherent to the underlying Gaussian field, they are also reflected in the covariance structure.

When constructing correlation structures for situations where the correlation depends on the orientation, the idea of a moving Lagrangian reference frame may be applied. Specifically, we can start by considering a purely spatial random field with a stationary covariance structure and suppose that the entire field moves forward in time according to some velocity field (Baxevani et al., 2011), (Baxevani and Rychlik, 2009), (Ailliot et al., 2011).

Following a similar path, Gneiting and Guttorp (2007) suggested a covariance structure that is a convex combination of a static spatio-temporal correlation structure C_{FS} and a Lagrangian correlation function C_{LGR}. This covariance was used to model an Irish wind data set, see Haslett and Raftery (1989) for a description of the data set. The resulting correlation structure can be formulated as

$$C(\mathbf{h}, \tau) = (1 - \lambda)C_{FS}(\mathbf{h}, \tau) + \lambda C_{LGR}(\mathbf{h}, \tau). \tag{5.23}$$

The authors suggested a correlation function of the form

$$C_{FS}(\mathbf{h},\tau)=\frac{1-\nu}{(1+a|\tau|^{2\alpha})}\Big(\exp\Big(-\frac{c||\mathbf{h}||}{(1+a|\tau|^{2\alpha})^{\beta/2}}\Big)+\frac{\nu}{1-\nu}\delta_{\{\mathbf{h}=0\}}\Big), \quad (5.24)$$

where a and c are nonnegative scale parameters of time and space respectively. The smoothness parameter α and the space-time interaction parameter β take values in $(0,1]$, while $\delta_{\{A\}}$ is the indicator function that is equal to one when property A is satisfied and zero otherwise. The purely spatial correlation function $C_{FS}(\mathbf{h},0)$ is of the powered exponential form, whereas the purely temporal correlation function $C_{FS}(\mathbf{0},\tau)$ belongs to the Cauchy family. The suggestion for the Lagrangian correlation function C_{LGR} is the following compactly supported function

$$C_{LGR}(\mathbf{h},\tau)=\left(\Big(1-\frac{||\mathbf{h}-\mathbf{v}\tau||}{d||\mathbf{v}||}\Big)^{\gamma}\right)^{+}, \quad (5.25)$$

where $||\mathbf{v}||$ is the speed of the weather front, $x^{+}=\max\{x,0\}$ and the parameter d controls the range of dependence. Notice that (5.25) belongs to the class of continuous homogeneous and stationary correlation functions for $\gamma\geq 2$ (Gneiting, 1999). In order to obtain a random field with realizations that are not as smooth, $\gamma=1$ is chosen, which corresponds to a non differentiable (in the mean square sense) Gaussian random field.

In a different approach, instead of altering the spatial coordinate, Lau (2011) started from a purely temporal random field and transformed the temporal lag. The suggested correlation structure has the following formulation

$$C(\mathbf{h},\tau)=C_{ST}(\mathbf{h},\tau)+\delta_{\{\mathbf{h}=0\}}C_0(\tau), \quad (5.26)$$

with the following spatio-temporal

$$C_{ST}(\mathbf{h},\tau)=c_0\exp\left(-(\alpha||\mathbf{h}||)^{2\gamma}\right)\exp\left(-(\beta\tilde{\tau})^{2\eta}\right)$$

and purely temporal covariance

$$C_0(\tau)=\exp\left(-(\tilde{\beta}\tau)^{2\tilde{\eta}}\right)-c_0\exp\left(-(\beta\tau)^{2\eta}\right).$$

The time is shifted by $(\mathbf{hv})/||\mathbf{v}||^2$, the time needed for the weather front to

propagate between the two locations, so that

$$\tilde{\tau} = \tau + \frac{\mathbf{hv}}{||\mathbf{v}||^2}.$$

The parameters $\alpha, \beta, \tilde{\beta} > 0$ are the scale parameters for space and time respectively, $0 < \gamma, \eta, \tilde{\eta} \leq 1$ are the shape parameters, $0 \leq c_0 \leq 1$ controls the nugget effect C_0. In this case, both the purely spatial $C(\mathbf{h}, 0)$ and purely temporal $C(\mathbf{0}, \tau)$ correlations are of the power exponential form.

Finally, let us compare the two previous models to a relatively simple one that consists of a separable model together with a temporal nugget effect, since the same model but with a spatial nugget effect is a part of (5.23),

$$C^t_{SP}(\mathbf{h}, \tau) = (1 - \nu)e^{-c||\mathbf{h}||}\Big(\frac{1}{(1 + a|\tau|^{2\alpha})} + \frac{\nu}{1 - \nu}\delta_{\{\tau=0\}}\Big). \tag{5.27}$$

5.3.3 Parameter estimation

Now that the structure of the model in (5.17) is specified, the details of the parameter estimation are presented. Predictions are then obtained using the spatio-temporal kriging framework outlined in Section 4.1.2.

Let us consider the estimate of the parameters in the anamorphosis function $f_{\mathbf{s}}$, in Section 5.3.1. The fitting of this function consists of two independent components: the beta distribution and the first two moments of the Gaussian random field (see (5.18)).

At each location $\mathbf{s}$, the fitting of the beta distribution is done by maximizing the log-likelihood of the positive wind power production,

$$\sum_{\{t:y(\mathbf{s},t)>0\}} \log\Big(\frac{\Gamma(\alpha_{\mathbf{s}} + \beta_{\mathbf{s}})}{\Gamma(\alpha_{\mathbf{s}})\Gamma(\beta_{\mathbf{s}})} y(\mathbf{s}, t)^{\alpha_{\mathbf{s}}-1}(1 - y(\mathbf{s}, t)^{\beta_{\mathbf{s}}-1}\Big). \tag{5.28}$$

Under the assumption that the data has been standardized to have variance one, estimation of the mean of the truncated normal distribution reduces to the estimation of the mean μ of the latent Gaussian field Z, which is estimated

using (5.22), i.e.,

$$P(y(\mathbf{s},t) > 0; t = 1, 2, \ldots, T) = \frac{\text{number of observations above level 0}}{T}. \quad (5.29)$$

Next, focus is placed on the correlation structure of the latent Gaussian field Z. To estimate the parameters in each correlation model, we need to first calculate the empirical spatio-temporal correlations of Z. This has been shown to be a challenging problem, since the Gaussian variable Z cannot be directly observed.

The fitting of the correlation models in (5.27), (5.23) and (5.26) can be achieved in the following way. Suppose that the empirically estimated anamorphosis function is given by $\hat{f}_{\mathbf{s}}$, which we regard as the true anamorphosis function. Then, from (5.17), estimated values of the truncated Gaussian random field, $\widehat{z(\mathbf{s},t)}^{+}$, are obtained by applying the inverse transformation to the corresponding strictly positive measurements of wind power production,

$$\widehat{z(\mathbf{s},t)}^{+} = \hat{f}_{\mathbf{s}}^{-1}(y(\mathbf{s},t)), \quad \text{for} \quad y(\mathbf{s},t) > 0. \quad (5.30)$$

Next, the empirical correlations of $\widehat{z(\mathbf{s},t)}^{+}$ at time lag τ can be computed by

$$\widehat{\rho_{ij}}^{+}(\tau) = \overline{\widehat{z(\mathbf{s}_i,0)}^{+} \cdot \widehat{z(\mathbf{s}_j,\tau)}^{+}} - \overline{\widehat{z(\mathbf{s}_i,0)}^{+}} \cdot \overline{\widehat{z(\mathbf{s}_j,\tau)}^{+}}, \quad (5.31)$$

where the overline indicates sample mean of the corresponding quantity.

After calculating the correlations of the truncated Gaussian random field, we are now ready to move to the correlations $\widehat{\rho_{ij}}(\tau)$ of the Gaussian random field Z.

Assume that estimates of the mean values $\hat{\mu}_i = \mathbb{E}(Z(\mathbf{s}_i))$ are available through (5.29) and that the data has been standardized to have variance one. In addition, suppose that

$$\begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \overset{d}{=} \mathcal{N}\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \right), \quad (5.32)$$

then the following result can be obtained by straightforward but lengthy calcu-

lations (Stein, 1992), (Baxevani and Lennartsson, 2015),

$$\begin{aligned}\mathrm{Cov}(Z_1^+, Z_2^+) &= (\mu_1\mu_2 + \rho\sigma_1\sigma_2)F(\alpha_1, \alpha_2; \rho) \\ &+ \frac{\mu_1\sigma_2}{(2\pi)(1/2)} e^{-\frac{\alpha_2^2}{2}} \Phi\left(\frac{\alpha_1 - \rho\alpha_2}{b}\right) + \frac{\mu_2\sigma_1}{(2\pi)(1/2)} e^{-\frac{\alpha_1^2}{2}} \Phi\left(\frac{\alpha_2 - \rho\alpha_1}{b}\right) \\ &+ \frac{b\sigma_1\sigma_2}{2\pi} \exp\left(-\frac{\alpha_2^2 + \alpha_1^2 - 2\rho\alpha_1\alpha_2}{2b^2}\right) - \mathbb{E}[Z_1^+]\mathbb{E}[Z_2^+],\end{aligned} \tag{5.33}$$

where F is the cumulative distribution function of a standard bivariate normal with correlation ρ and $\alpha_i = \mu_i/\sigma_i$ and $b = (1-\rho^2)^{1/2}$. The short-hand notation Z_i^+ has been used instead of $Z(\mathbf{s}_i, 0)^+$ and ρ instead of $\rho_{ij}(\tau)$.

Next, the correlations $\widehat{\rho_{ij}}(\tau)$ of the Gaussian random field Z can be retrieved by minimising the expression

$$\widehat{\rho_{ij}}(\tau) = \min_\rho \left| \widehat{\rho_{ij}}^+ - \mathrm{Cov}(Z(\mathbf{s}_i, 0)^+, Z(\mathbf{s}_j, \tau)^+) \right|, \quad (\sigma_i^2 = \sigma_j^2 = 1),$$

where the empirical correlations $\widehat{\rho_{ij}}^+(\tau)$ are given by (5.31) and the covariances of the truncated field, $\mathrm{Cov}(Z(\mathbf{s}_i, 0)^+, Z(\mathbf{s}_j, \tau)^+)$, are obtained with (5.33).

Finally, once we have computed the empirical correlations $\widehat{\rho_{ij}}(\tau)$, the correlation function $C(\mathbf{h}, \tau; \boldsymbol{\eta})$ with parameters denoted by $\boldsymbol{\eta}$, is fitted by minimizing the weighted squared errors

$$\min_{\boldsymbol{\eta}} \sum_\tau \sum_{i \neq j} \left(\frac{\widehat{\rho_{ij}}(\tau) - C(\mathbf{h}, \tau; \boldsymbol{\eta})}{1 - C(\mathbf{h}, \tau; \boldsymbol{\eta})} \right)^2, \tag{5.34}$$

for $\mathbf{h} = \mathbf{s}_j - \mathbf{s}_i$. In practice, the indices i and j go through all the different combinations of wind farms, but the temporal lag τ cannot go through the infinite number of time lags. A certain cut-off level τ_0 needs to be chosen so that the summation in (5.34) is truncated to $\tau = 1, 2, \ldots, \tau_0$.

After the model is fully specified, one can apply the spatio-temporal kriging methodology introduced in Section 4.1.2 and obtain predictions of the Gaussian random field Z. Next, given the estimates of the mean and covariance function of Z as well as the estimate of the anamorphosis $\widehat{f}_{\mathbf{s}}$, the "plug-in" predictor of Y is given by

$$\widehat{Y} = \widehat{f}_{\mathbf{s}}(\widehat{Z}). \tag{5.35}$$

Then, the resulting predicted values are truncated at zero, meaning that there is zero wind power generation if $Z \leq 0$.

Chapter 6

Application to a Danish wind power data set

After having formulated various approaches for forecasting wind power in Chapter 5, we now focus on the application of these techniques to a wind power data set. More specifically, this chapter deals with the results from applying the methodologies described in Sections 5.1, 5.2 and 5.3, together with the prediction approaches in Chapter 4 to an extensive data set of wind power generation in western Denmark, which will be described in Section 6.1. The evaluation methods for both point and probabilistic forecasts used in this chapter are outlined in Chapter 3.

As we will see, different parts of the wind power data set were used for different situations and objectives. The work in this thesis can in general be divided into two parts: spatial and spatio-temporal prediction. In the following, the main results in these two categories are summarized. First, Section 6.2 comprises an overview of our contributions to spatial prediction of wind power generation. It presents the results for spatial models in predicting wind power generation at two different time scales: annual average wind power generation and high temporal resolution. Next, Section 6.3 shows the results for spatio-temporal wind

power forecasts. It starts with the critical component of modelling the spatio-temporal correlations. Then, it moves to the results for individual, aggregated and spatially out-of-sample wind power forecasts.

6.1 Danish wind power production data

The data set considered here comes from a system of 349 wind farms located in western Denmark. Observations of wind power production between January 2006 and March 2012 were provided by the Transmission System Operator in Denmark and each measurement consists of a temporal average over a 15-min time period. In order to facilitate comparisons between wind farms with different capacities, the analyses are performed after the measurements at each wind farm has been normalised by their corresponding nominal power (i.e., when all turbines operate at their maximum capacity), so that they are within the range $[0, 1]$. The locations of the wind farms are plotted in Figure 6.1.

6.2 Spatial prediction

In order to have a better understanding of a spatio-temporal process throughout a country, such as the wind power production in Denmark, let us begin with the examination of the spatial aspects underlying the process. In this section we present the results from a case study based on wind farms located in the western part of Denmark (see Section 6.1), where we, for example, use the model described in Section 5.1 to obtain spatial prediction of wind power production. For the purpose of illustrating the proposed methods, the year of 2010 has been chosen.

In a first step, the focus is on a model describing spatial variation of annual average wind power generation. This type of scenario is of relevance for system operation as well as for electricity companies, since it provides an average of the wind energy supply throughout the year. Annual average wind power generation is obtained by averaging the 15-min power output at each wind farms in 2010, resulting in purely spatial data. Although the distribution of annual average

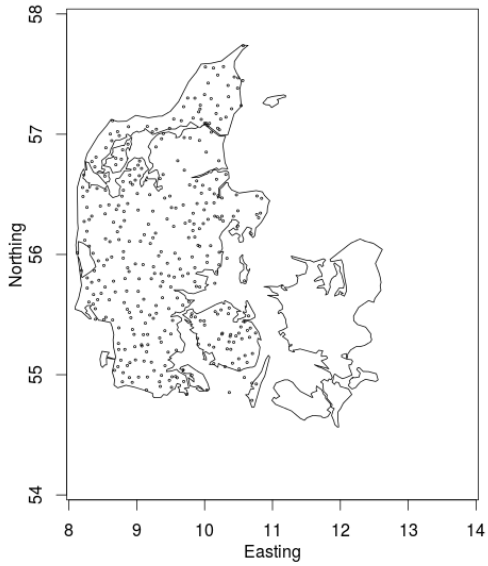


Figure 6.1: Locations of 349 wind farms in western Denmark.

wind power does not have the problem of probability mass at zero and is less skewed than individual power, there is still the challenge that it is bounded below by zero and above by one after normalization. We model this setting with a beta distribution with a stochastic mean that has a log-normal distribution including both covariates and a spatial structure, which is captured by a latent Gaussian random field with Matérn covariance. To issue probabilistic predictions, we use the INLA-SPDE approach introduced in Section 4.2, and directly compute approximations to the posterior marginals. More details of this representation can be found in Paper A.

Next, we illustrate a methodology for modelling wind power generation at high temporal resolution. Power fluctuations at this time scale are the ones that most seriously affect the balance in the power systems. Hence, having a method that predicts the power generation at a high temporal resolution helps to ensure the continuous balance of the power system. Our method makes it possible to predict the wind power production at a high resolution at all relevant sites over

a country, even when historical observations are only available from a limited number of farms. To obtain prediction of wind power at a high resolution, the hurdle model in Section 5.1 is fitted separately to the data from each time period during the first day of every month in 2010. Then, we present the results as the average of the scores from the individual prediction cases. Inference and prediction are obtained with the INLA-SPDE approach.

To assess the quality of the predictive distributions for the annual average and high temporal resolution of wind power generation, a k-fold cross validation with $k = 4$ is used and the results from our models are compared to those obtained with the benchmark kriging in Section 4.1.1. We repeat the k-fold cross validation 50 times to reduce sampling bias and variance. The prediction error is obtained by combining the estimates from the 50 data sets. To formally test for a significant difference between the predictions made for two spatial fields, we apply the spatial prediction comparison test (SPCT) introduced in Hering and Genton (2011), where the null hypothesis to be tested is that of equal predictive ability on average in terms of a loss function.

In the following, the main results from the spatial probabilistic prediction of wind power generation at annual average as well as high temporal resolution are presented. The spatial predictions generated by the fitted spatial models described above are compared with the benchmark kriging. This mainly summarizes the work presented in Paper A.

6.2.1 Annual average wind power generation

Table 6.1 shows summary measures of predictive performance for annual average wind power generation. The large p-values in this table indicate that one does not reject the hypothesis of equal predictive ability on average in terms of RMSE, CRPS, Bias and MAE between the beta model and kriging. This is not surprising, given that here, since the data is averaged over an entire year, the individual noises are smoothed out and the distribution becomes closer to Gaussian.

The results for modelling the annual average of wind power generation indicates that kriging should be preferred over the beta model with covariates fitted with

the INLA-SPDE approach. Both methods present similar results, while kriging is easier to set up and has lower computational cost.

	Kriging	Beta model	p-values
RMSE	7.518	7.377	0.97
CRPS	4.235	4.319	0.934
Bias	0.012	0.164	0.923
MAE	5.547	5.665	0.936

Table 6.1: RMSE, CRPS, bias and MAE (as % of nominal power) of annual average wind power prediction using kriging and the beta model with covariate. The last column shows the p-values for the differences in RMSE, CRPS, bias and MAE between kriging and the beta model.

6.2.2 High temporal resolution of wind power generation

Table 6.2 compares the summary measures of predictive performance for wind power greater than zero obtained using kriging and the hurdle model. The hierarchical hurdle model produces significantly better predictions on average in terms of RMSE, CRPS and MAE than kriging. The superiority of the hierarchical spatial method over kriging may stem from a lack of flexibility of the latter, as it does not consider the point mass at zero in the wind power distribution and is optimal only when data is Gaussian distributed. In contrast, the hurdle model attempts to accommodate wind occurrences with a Bernoulli distribution and wind power magnitude using the gamma distribution, where a shared latent process is included to handle spatial correlation between wind farms in both distributions.

The hierarchical hurdle model additionally gives predictions of the Bernoulli-distributed random variable that maps the occurrence of wind power generation (see (5.1)). We assess the reliability of the probability predictions for the occurrence of wind power generation through the diagram in the bottom line of Figure 6.2. This plot shows the empirically observed frequency of wind power occurrence as a function of the binned forecast probability. The actual observed relative frequency is well approximated by the predicted probability, as the line in this plot lies close to the diagonal. The top plots in the same figure correspond to histograms of the empirical probability (left) and the predicted probability

	Kriging	Hurdle	p-values
RMSE	13.689	7.214	0.05
CRPS	5.827	2.455	0.039
Bias	0.56	-0.222	0.201
MAE	8.665	2.917	0.044

Table 6.2: RMSE, CRPS, bias and MAE (as % of nominal power) of wind power prediction at high temporal resolution when the wind power output is greater than zero using kriging and the hierarchical hurdle model. The last column shows the p-values for the differences in RMSE, CRPS, bias and MAE between kriging and the hurdle model.

(right) of wind power occurrence. As we can see from the left plot, there are almost five times as many time periods with generated power greater than zero than equal to zero. The histogram of the predicted probabilities on the right side shows the same tendencies, since most of the estimated probabilities of wind power occurrence are close to one.

The analysis shows that the hierarchical hurdle model produces significantly better predictions on average than the more traditional kriging methodology. However, the trade-off between a method that offers more accurate predictions with a sharper predictive density and a method that is simpler to set up and requires less computational effort will most likely depend on the type of application.

6.3 Spatio-temporal prediction

This section summarizes the results presented in Paper B and C. These papers deal with techniques for application to short and very-short-term wind power forecasting. We consider time forward predictions at the locations of the training set and at new locations where no historical data is available, which we call spatially out-of-sample forecasts. Spatially out-of-sample forecasts are relevant for a number of decision making problems, which require an overview of power generation at all farms over a region, while measurements are available only at a limited number of sites.

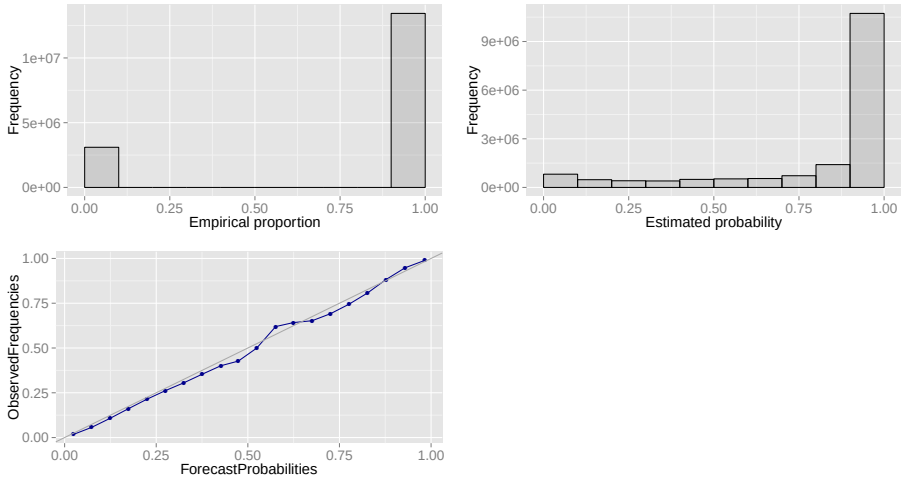


Figure 6.2: *First line*: Histograms of the empirical probability (left) and the predicted probability of wind power occurrence using the hierarchical hurdle model (right). *Second line*: Reliability diagram for the probability of wind power occurrence using the hierarchical hurdle model.

Paper B builds on a framework based on the latent truncated Gaussian random field from Section 5.3 together with a methodology based on spatio-temporal kriging, as in Section 4.1.2. These approaches have been validated on a portfolio of 30 wind farms, which is part of a larger network described in Section 6.1. The prediction performances of the three different correlation structures in Section 5.3.2 are evaluated and compared in Section 6.3.1. To obtain spatially out-of-sample forecasts, we initially randomly select 24 wind farms for training, leaving the remaining 6 for validation purposes. We repeat this procedure 15 times to reduce bias. The forecasts are obtained at time horizons ranging from 15 minutes up to 2 hours ahead. The goal of Paper C is to show the cases where forecasts can benefit from having a spatio-temporal model rather than a simple autoregressive one. For this purpose, we study the performance of the three models in Section 5.2 to forecast wind power from individual and aggregated farms from 15 minutes up to 5 hours ahead, which are summarized in Sections 6.3.2 and 6.3.3, respectively. The results from spatially out-of-sample forecasts are based on k-fold cross-validation with $k = 5$ and are presented in Section 6.3.4.

6.3.1 Comparison of covariance models

Figure 6.3 shows the RMSE and CRPS scores for the spatially out-of-sample predictions at lead times $h = 1, \ldots, 8$ (i.e., from 15 minutes up to 2 hours) for the three correlation models described in Section 5.3.2. Notice that Paper C also provides the evaluation of time forward predictions at the locations of the training set with comparable results. The model in (5.23) results in smaller RMSE and CRPS scores for all lead times except one. For lead time one, the model in (5.27) results in a slightly smaller CRPS value. The plots in Figure 6.4 correspond to reliability diagrams for prediction at lead times $h = 1$ (left) and $h = 8$ (right). All covariance models perform reasonably well in terms of forecast densities. However, the forecast densities corresponding to the correlation model given by (5.23) outperform the other two models in terms of calibration at lead time 1. At lead time 8, the correlation model given in (5.26) is better calibrated than the other two for quantiles larger than 0.7.

It appears that correlations that account for the weather front dynamics improve the forecast quality as measured by both deterministic and probabilistic scores for lead times of 30 minutes to 2 hours ahead. For lead times of less than 30 minutes, the simpler covariance model generates better forecasts than the more complicated ones.

6.3.2 Individual forecast evaluation

Here, the performance of the models in Section 5.2, when we consider time forward forecasts at individual wind farms of the training set, are evaluated and discussed. The results show that even the simpler autoregressive model is able to produce calibrated short-term forecasts for individual wind farms.

Figure 6.5 summarizes the spatio-temporal forecast performance of the three models introduced in Section 5.2 in terms of RMSE and CRPS. As it is apparent from Figure 6.5, Model T and Model ST+T outperformed Model S-T with respect to RMSE and CRPS when forecasting individual wind farms at lead times 1-8 (i.e, from 15 minutes up to 2 hours ahead). For higher lead times, the three models have similar performance. Reliability diagrams for each model

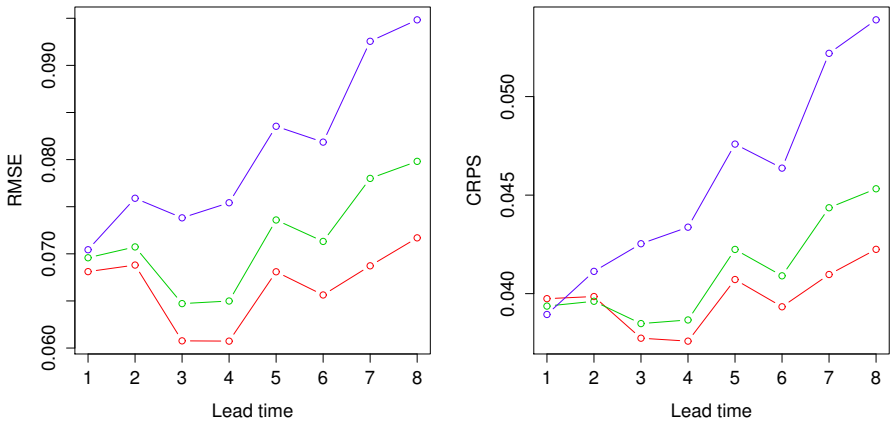


Figure 6.3: RMSE and CRPS (as % of nominal power) of spatio-temporal wind power forecasts at lead times $1, \ldots, 8$ (i.e., from 15 minutes up to 2 hours) for the correlation models given in (5.27) (blue), (5.23) (red) and (5.26) (green).

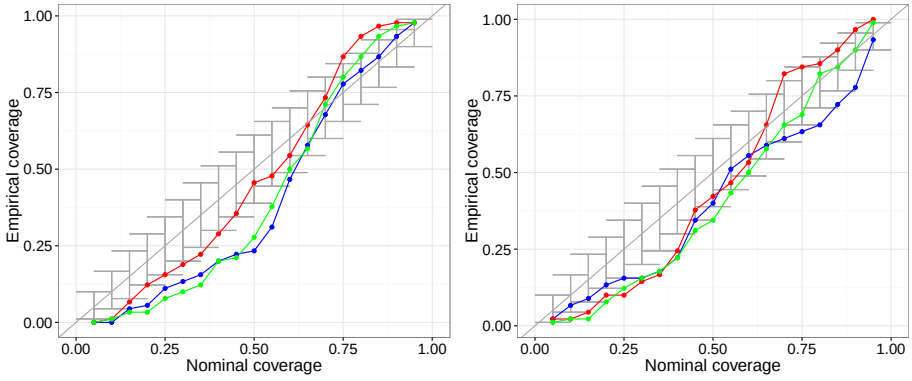


Figure 6.4: Reliability diagrams of spatio-temporal wind power forecasts at lead time 1 (*Left*), and 8 (*Right*). The diagrams were calculated using the correlation models given in (5.27) (blue), (5.23) (red) and (5.26) (green).

at lead times $h = 1, 7, 13$ and 19 are presented in Figure 6.6. The forecasts at individual wind farms produced by the three models perform similarly well in terms of reliability, with points close to the diagonal for most quantiles, see

Figure 6.6.

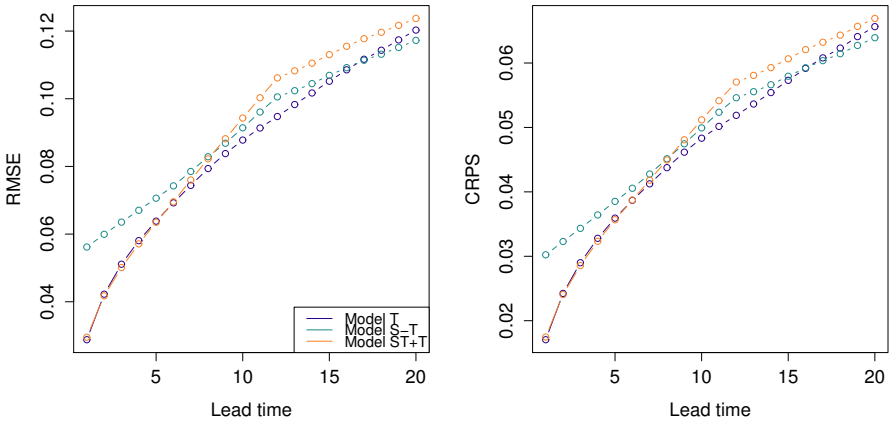


Figure 6.5: RMSE and CRPS (as % of nominal power) of spatio-temporal wind power forecasts for individual wind farms at lead times $1, \ldots, 20$ (i.e., from 15 minutes up to 5 hours) for Model T (blue), Model S-T (green) and Model ST+T (orange).

6.3.3 Aggregated forecast evaluations

Next, the aggregated forecast performances of the models in Section 5.2 are considered. Aggregated wind power is obtained by adding up the power from all the wind farms and normalizing by the total capacity.

Figure 6.7 illustrates that Model T performs similar to Model ST+T in terms of point forecast (RMSE), but has poor performance according to CRPS values. The aggregated forecasts provided by Model ST+T are the best calibrated among the three models for most of the quantiles at all lead times, followed by Model S-T, as shown in Figure 6.8.

In summary, results from aggregated forecasts show that even though the performance of Model T is comparable to the performance of the other models in terms of aggregated forecast density mean (RMSE), this model does not produce reliable probabilistic forecasts for the aggregated data. This fact is more obvious

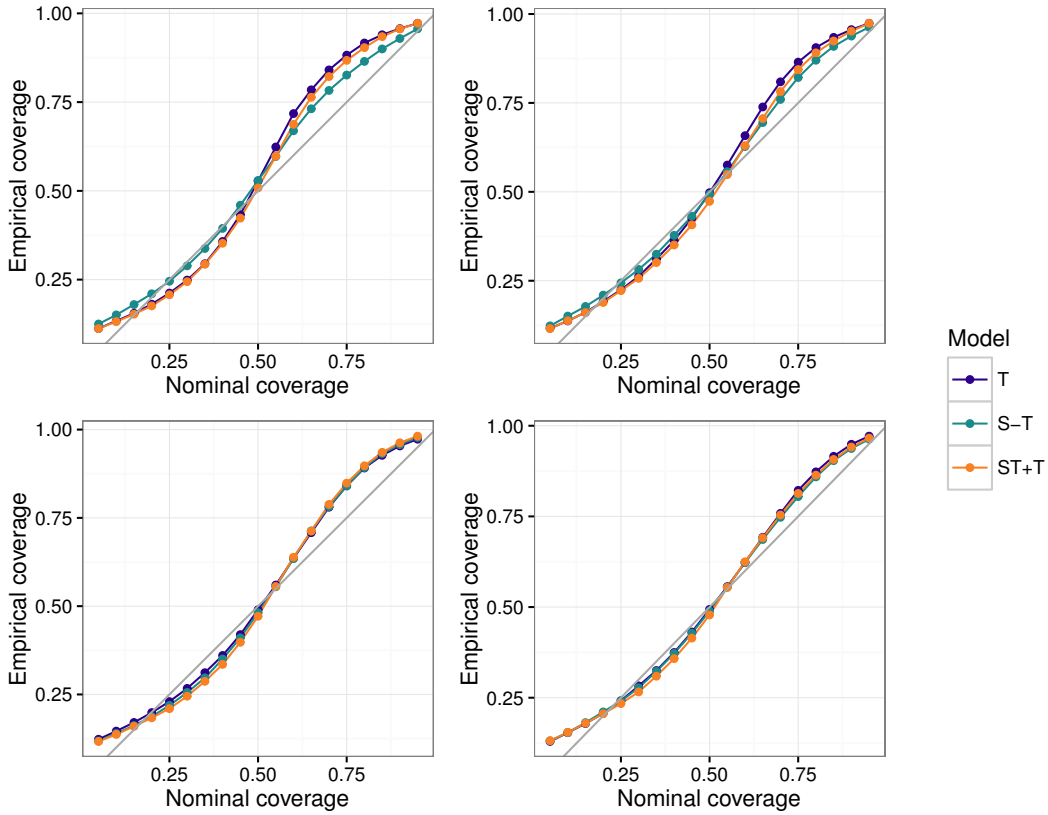


Figure 6.6: Reliability diagrams of spatio-temporal wind power forecasts for individual wind farms at lead time 1 (*Top left*), 7 (*Top right*), 13 (*Bottom left*) and 19 (*Bottom right*). The diagrams were calculated using Model T (blue), Model S-T (green) and Model ST+T (orange).

for the lower quantiles; more than 50% of the observed aggregated forecasts are below the nominal 5% quantile at lead times $h = 7, 13$ and 19, as illustrated by Figure 6.8.

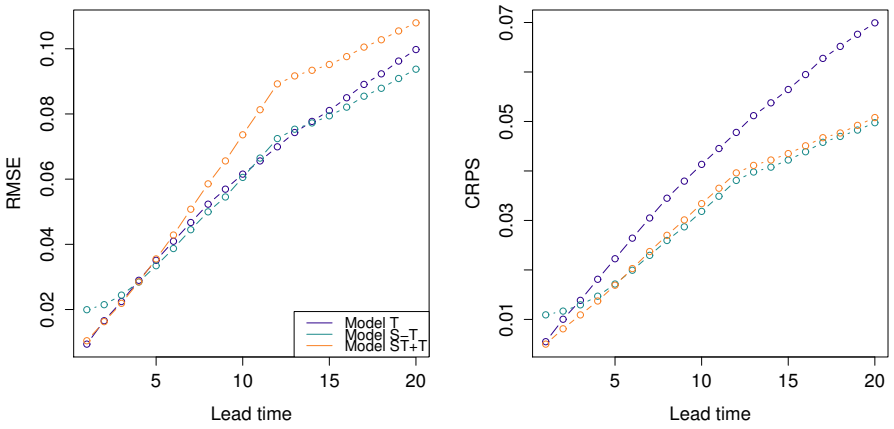


Figure 6.7: RMSE and CRPS (as % of nominal power) of spatio-temporal wind power forecasts for aggregated wind farms at lead times $1, \ldots, 20$ (i.e., from 15 minutes up to 5 hours) for Model T (blue), Model S-T (green) and Model ST+T (orange).

6.3.4 Spatially out-of-sample forecast performances

The spatio-temporal models, Model S-T and Model ST+T, have the advantage of being able to provide forecasts at locations where recent observations are not available. Based on this, using the methodologies as described, we obtain spatially out-of-sample forecasts at the individual and aggregated levels.

Figure 6.9 shows the forecast performances in terms of RMSE and mean CRPS for individual wind farms (a) and aggregated wind power (b). They are computed as the mean of the RMSE and CRPS from the 5-fold cross validations. It can be seen that Model ST+T outperforms Model S-T at all lead times when predicting wind power at individual wind farms under RMSE and CPRS. When looking at aggregated out-of-sample forecasts, while for shorter lead times than 2 hours Model S-T is better than Model ST+T in terms of RMSE, for longer horizons, Model ST+T out-performs Model S-T under the same score. In terms of CRPS, Model ST+T produces better aggregated forecasts at lead times 1-20 (i.e., from 15 minutes to 5 hours ahead).

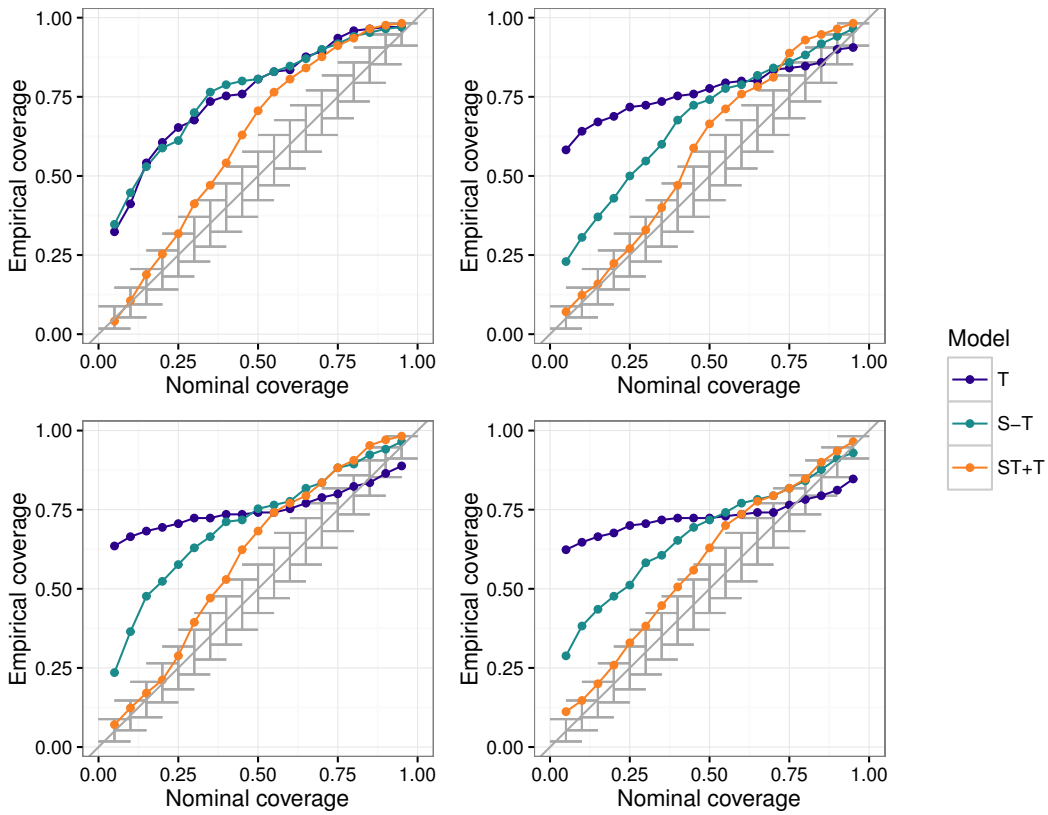
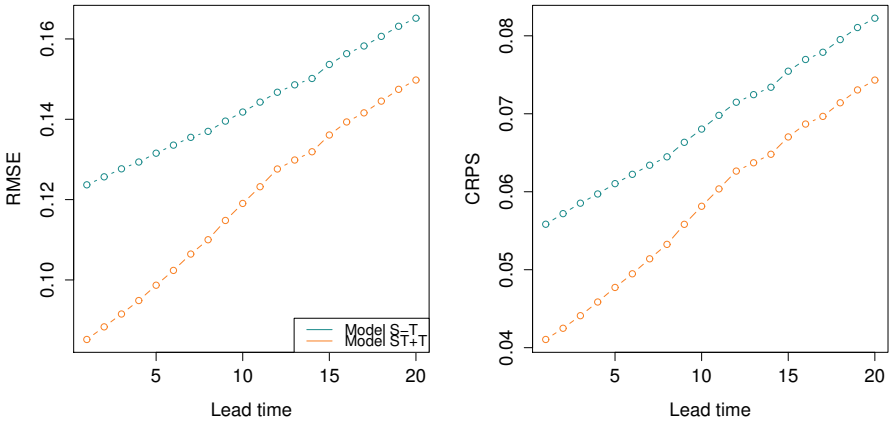
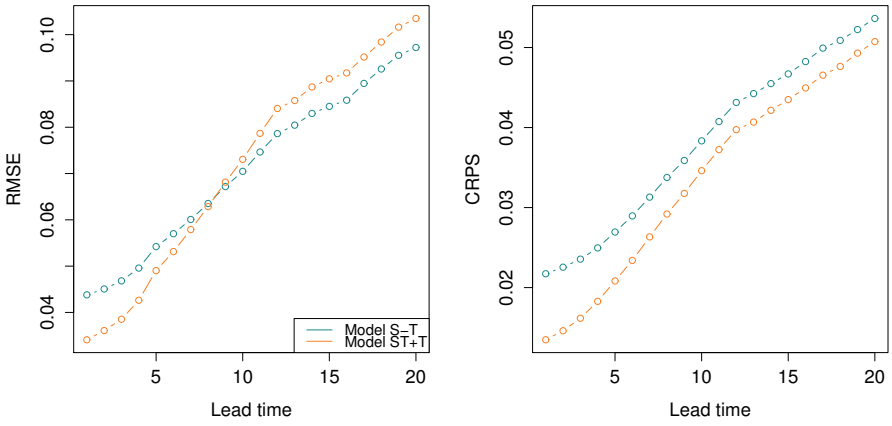


Figure 6.8: Reliability diagrams of spatio-temporal wind power forecasts for aggregated wind farms at lead time 1 (*Top left*), 7 (*Top right*), 13 (*Bottom left*) and 19 (*Bottom right*). The diagrams were calculated using Model T (blue), Model S-T (green) and Model ST+T (orange).

Reliability diagrams at lead times $h = 1, 7, 13$ and 19 are presented in Figure 6.10. We observe from Figure 6.10 (a) that Model S-T and Model ST+T provide relatively well calibrated forecast densities for individual farms. In terms of aggregated forecasts, we can see from Figure 6.10 (b) that Model ST+T is calibrated, since the line remains within the consistency bars. On the other hand, aggregated forecast densities obtained with Model S-T are poorly calibrated for quantiles lower than 0.75. Indeed, 20% of the observations are below the 5% forecast quantile at lead times 1, 7, 13 and 19.

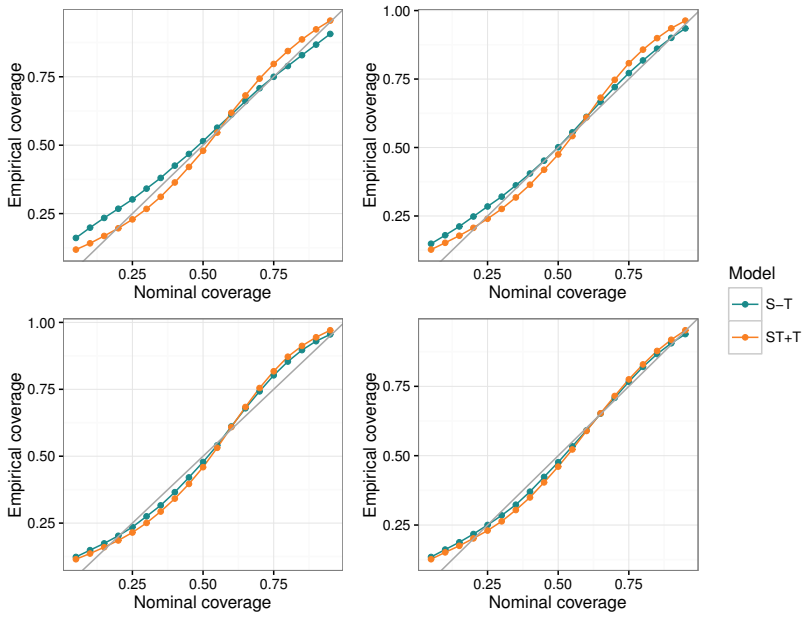


(a)

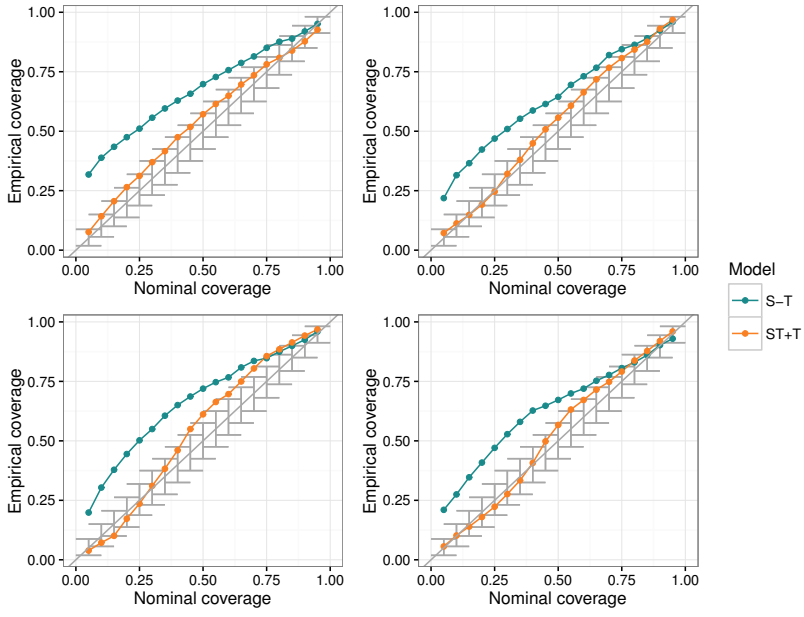


(b)

Figure 6.9: RMSE and CRPS (as % of nominal power) of spatially out-of-sample wind power forecasts at lead times $1, \ldots, 20$ (i.e., from 15 minutes up to 5 hours) for Model T (blue), Model S-T (green) and Model ST+T (orange). (a) Forecasts for individual wind farms. (b) Forecasts for aggregated wind farms.



(a)



(b)

Figure 6.10: Reliability diagrams of spatially out-of-sample wind power forecasts at lead time 1 (*Top left*), 7 (*Top right*) , 13 (*Bottom left*) and 19 (*Bottom right*). The diagrams were calculated using Model T (blue), Model S-T (green) and Model ST+T (orange). (a) Forecasts for individual wind farms. (b) Forecasts for aggregated wind farms.

Chapter 7

Conclusions and perspectives

7.1 Conclusions

This thesis was motivated by the growing need for reliable short-term wind power forecasting in order to successfully integrate wind energy into the existing power systems. The long-term positive impact of this integration could be substantial for global warming and climate change. Most of the operational wind power prediction systems issue forecasts for each location individually. In this work, we argue that there is a need to better utilize the spatial structure between geographically separate locations. We have shown by analysing a real wind power data set, that spatial dependence among different wind farms is present and that its inclusion in the modelling stage increases forecast performance in several cases. Following this, new methods that capture the spatial and spatio-temporal dependencies between wind farms have been proposed and compared to traditional benchmark methods.

In order to understand the wind power production in a region, we analysed

the various features of wind power data, such as distributional characteristics, appropriate transformations, occurrences of probability masses and spatio-temporal correlations. Considering the challenges and features just mentioned, we demonstrated how short-term wind power forecasts at wind farms in monitored sites, as well as at new unmonitored locations, can be obtained through kriging predictors, spatio-temporal correlation models and truncated Gaussian models. Moreover, from aggregated forecasts, we saw that it is possible to benefit from having a spatio-temporal model instead of a time series model optimized for each location individually.

An important part of this work has been devoted to the exploration of new models for wind power data that can be formulated in a hierarchical framework. This kind of model is well suited to represent many environmental phenomena thanks to its flexibility in modelling the effect of relevant covariates (e.g. meteorological and geographical variables), as well as time and space dependence. However, performing Bayesian inference - through Markov chain Monte Carlo (MCMC) techniques - can be a challenging task due to convergence problems and a heavy computational load. In this thesis we have explored different ways of dealing with spatio-temporal modelling while casting the model in a computationally tractable framework. Towards this goal, we implemented a scalable inference methodology using the INLA-SPDE approach and illustrated how this methodology can be used for obtaining probabilistic forecasts of wind power generation. To the best of our knowledge, this is the first attempt to date to use the INLA-SPDE approach to obtain forecasts of wind power generation. We have shown that complex hierarchical spatial and spatio-temporal models can be implemented seamlessly under the INLA-SPDE approach that is available in the **R-INLA** library providing satisfactory results in reasonable computational time. However, a limitation of this approach lies in the structure of the covariance function, which is currently implemented only for separable Kronocker-based models.

Within the hierarchical framework, we first proposed spatial models for predicting wind power generation at two different time scales: for annual average wind power generation and for a high temporal resolution. The distribution of the annual average wind power generation was modelled with a beta distribution and wind power at a high temporal resolution was modelled with a hurdle model consisting of a degenerated distribution at zero and a skewed continu-

ous distribution for the non-zero values. In both cases, spatial correlation was captured by a latent Gaussian field with Matérn correlation function. The proposed methods were tested on data from 349 wind farms in western Denmark and the results were compared to those obtained from standard kriging methods. We have shown that complex hierarchical spatial models are well suited for wind power data and can be implemented seamlessly in reasonable computation time in order to obtain predictions on a country level. This is essential when considering the realistic case, where the power production data for most wind farms is only available with a delay of up to a month. More specifically, while the benchmark kriging generated competitive predictions for the average wind power generation, as it is easier to model the relatively smooth average data, for wind power at a high temporal resolution, predictions generated by the hurdle model were a significant improvement to those generated by kriging methods. In the last case, the more sophisticated hurdle model has shown to be necessary for handling the occurrence of zeros and the skewed distribution of the positive values. In fact, we saw that the kriging method predicted less values close to zero in comparison to the hurdle model and in comparison with the true distribution of the data.

Next, hierarchical models for obtaining probabilistic forecasts of wind power generation at multiple locations and lead times have been developed. The simplest model is an autoregressive process with a location specific intercept. We saw that even though the results for individual probabilistic forecasts were satisfactory in terms of skill scores and reliability, the aggregated probabilistic forecasts were not calibrated. Then, after finding the unsatisfactory results for the reliability of aggregated forecasts, two different spatio-temporal models were introduced. The first has a common intercept for all farms and a spatio-temporal model that varies in time with first order autoregressive dynamics and has spatially correlated innovations given by a zero mean Gaussian process with Matérn covariance. The second model has a common intercept, an autoregressive process to capture the local variability and the spatio-temporal term. In all cases, to deal with the non-Gaussianity of wind power series, a parametric framework for distributional forecasts based on the logit-normal transformation was used. We demonstrated that modelling spatial dependency is required to achieve calibrated aggregated probabilistic forecasts. Indeed, a case study and simulations showed that spatial dependency is important for aggregated properties, and individual forecasts did not reveal this.

In the wind power literature, most of the proposed techniques are evaluated only at locations with available monitored measurements. This means that these models lack testing for out-of-sample robustness. In this work, in addition to out-of-sample forecasts in terms of time, that are obtained for wind farms inside the training set, we explore spatially out-of-sample forecasts generated by the proposed spatio-temporal models. For these cases, temporal models that require local information for the parameter estimation cannot be used to obtain forecasts. Thus, it is important to have a method of forecasting that is as robust as possible, so that parameters estimated using only part of the portfolio can readily be used to forecast a larger data set, including wind farms at unmonitored locations. In the presented case study, the approaches used to obtain spatially out-of-sample forecast performances performed well in terms of individual forecasts, however, we noticed that having an autoregressive term and a spatio-temporal component was necessary to obtain calibrated aggregated forecasts.

Another critical topic tackled was the modelling of spatio-temporal correlations. As opposed to data driven methods, working directly with a parametric covariance (or precision) has the advantage of keeping the number of parameters in the model low (usually around 6) independently of the amount of data used to fit the model. We also studied the characteristics of the spatio-temporal correlations in wind power and explored the anisotropy feature in the correlation structure. These anisotropies are supported by the corresponding meteorological references that wind in general blows in the direction of the propagation of the weather front. With this in mind, we proposed three different anisotropic covariance models, two of which additionally take into account the propagation of the weather fronts that are, at least partly, responsible for the generation of wind energy. We obtained very-short-term spatio-temporal predictions of wind power production using spatio-temporal kriging equations together with a latent truncated Gaussian random field to accommodate the probability mass at zero. All covariance models performed reasonably well in terms of predictive densities when tested in a Danish wind power data set. However, the use of the richer and more realistic covariance structures that take into account the weather front dynamics outperformed the simple separable model in terms of RMSE and CRPS.

7.2 Perspectives

In this section, we address the limitations of the methodologies presented in this thesis and introduce potential alternatives and ideas for future work.

One could consider extending the spatial hurdle model presented in Paper A to the spatio-temporal domain by incorporating an extra term for the temporal effect, such as an autoregressive component, or by the introduction of a non-separable spatio-temporal structure.

In Paper B, we assumed that the precision matrix in the Gaussian random field can be written as a Kronecker product of two standard covariance matrices, which results in a separable space-time model. This assumption of separability is somewhat unrealistic due to the complex physical process of wind, which is largely determined by both geographical and seasonal properties. We would therefore like to accommodate more sophisticated spatio-temporal correlation functions that account for the non-separability in space and time (and possibly anisotropy). Within the SPDE framework, this can be achieved by representing a Gaussian random field as a solution to the non-separable SPDE transport and diffusion equation. While the current version of the **R-INLA** does not allow for non-separable covariance models, a strategy to try is to solve the SPDE equation using, for instance, Fourier functions or numerical finite difference methods.

A possible extension of the models described in this work is to include weather forecast information issued by meteorological services or satellite images as additional explanatory variable in the linear predictor. This could give a better insight into the patterns of weather front propagation, however it has the drawback of usually requiring ensemble forecasts to be generated from sophisticated numerical weather prediction (NWP) models.

When deriving the spatio-temporal kriging predictors, if the anamorphosis function $\widehat{f}_{\mathbf{s}}$ in (5.35) is nonlinear, the kriging predictors are biased. A possible improvement could be obtained with a bias correction obtained from the Taylor expansion of $\widehat{f}_{\mathbf{s}}$ around the mean using only the first two derivatives. Another possibility that might be worth considering is based on a bootstrap method to produce asymptotically unbiased predictors.

There is a potential to generalize the ideas from this work to different applications such as temperature, solar power, air pollution and energy consumption. As with wind power, several environmental problems can also be naturally modelled as a realization of a latent Gaussian random field since, in such cases, it is reasonably to assume that exists an underlying spatio-temporal process that governs a particular physical phenomenon, but it is not directly observed.

Another challenging and promising approach is to use vector autoregressive (VAR) models with some kind of regularization to impose sparsity on the coefficient matrix, in order to reduce numerical complexity due to the large number of parameters. The regularization could, for instance, be the elastic net penalty, which incorporates the variable selection feature of the lasso and the shrinkage of correlated predictors like the Ridge regression. This kind of approach has the flexibility of being completely data driven, so that only a limited number of assumptions need to be made.

Furthermore, in paper B, we use a parametric probabilistic framework based on the logit-normal distribution. A natural extension of this method would be to try the generalised logit-normal transformation and distribution instead. However, a method for determining the optimal shape parameter for modelling different locations and spatio-temporal dependencies would have to be developed.

Bibliography

Ackermann, T. (2005). *Wind power in power systems*. John Wiley & Sons.

Ahn, H. and J. J. Chen (1995). Generation of over-dispersed and under-dispersed binomial variates. *Journal of Computational and Graphical Statistics 4*(1), 55–64.

Ailliot, P., A. Baxevani, A. Cuzol, V. Monbet, and N. Raillard (2011). Space-time models for moving fields with an application to significant wave height fields. *Environmetrics 22*, 354–369.

Ailliot, P., J. Bessac, V. Monbet, and F. Pene (2015). Non-homogeneous hidden markov-switching models for wind time series. *Journal of Statistical Planning and Inference 160*, 75–88.

Akhmatov, V. (2007). Influence of wind direction on intense power fluctuations in large offshore windfarms in the north sea. *Wind Engineering 31*(1), 59–64.

Banerjee, A., X. Guo, and H. Wang (2005). On the optimality of conditional expectation as a Bregman predictor. *IEEE Transactions on Information Theory 51*(7), 2664–2669.

Banerjee, S., B. P. Carlin, and A. E. Gelfand (2014). *Hierarchical modeling and analysis for spatial data*. Crc Press.

Baxevani, A. and J. Lennartsson (2015). A spatiotemporal precipitation generator based on a censored latent Gaussian field. *Water Resources Research 51*(6), 4338–4358.

Baxevani, A. and A. Lenzi (2017). Very short-term spatio-temporal wind power prediction using a censored gaussian field. *Stochastic Environmental Research and Risk Assessment*, submitted.

Baxevani, A., K. Podgórski, and I. Rychlik (2011). Dynamically evolving Gaussian spatial fields. *Extremes 14*(2), 223–251.

Baxevani, A. Caires, S. and I. Rychlik (2009). Spatio-temporal statistical modelling of significant wave height. *Environmetrics 20*, 14–31.

Berrocal, V. J., A. E. Raftery, and T. Gneiting (2008). Probabilistic quantitative precipitation field forecasting using a two-stage spatial model. *The Annals of Applied Statistics*, 1170–1193.

Bessa, R. J., M. A. Matos, I. C. Costa, L. Bremermann, I. G. Franchin, R. Pestana, N. Machado, H. Waldl, and C. Wichmann (2012). Reserve setting and steady-state security assessment using wind power uncertainty forecast: A case study. *IEEE Transactions on Sustainable Energy 3*(4), 827–836.

Blangiardo, M. and M. Cameletti (2015). *Spatial and spatio-temporal Bayesian models with R-INLA*. John Wiley & Sons.

Blangiardo, M., M. Cameletti, G. Baio, and H. Rue (2013). Spatial and spatio-temporal models with R-INLA. *Spatial and Spatio-Temporal Epidemiology 7*, 39–55.

Botterud, A., Z. Zhou, J. Wang, J. Sumaili, H. Keko, J. Mendes, R. J. Bessa, and V. Miranda (2013). Demand dispatch and probabilistic wind power forecasting in unit commitment and economic dispatch: A case study of Illinois. *IEEE Transactions on Sustainable Energy 4*(1), 250–261.

Bremnes, J. B. (2004). Probabilistic wind power forecasts using local quantile regression. *Wind Energy 7*(1), 47–54.

Bröcker, J. and L. A. Smith (2007). Increasing the reliability of reliability diagrams. *Weather and Forecasting 22*(3), 651–661.

Brown, B. G., R. W. Katz, and A. H. Murphy (1984). Time series models to simulate and forecast wind speed and wind power. *Journal of Climate and Applied Meteorology 23*(8), 1184–1195.

Cameletti, M., F. Lindgren, D. Simpson, and H. Rue (2013). Spatio-temporal modeling of particulate matter concentration through the SPDE approach. *AStA Advances in Statistical Analysis 97*(2), 109–131.

Celik, A. N. (2004). A statistical analysis of wind power density based on the Weibull and Rayleigh models at the southern region of Turkey. *Renewable Energy 29*(4), 593–604.

Cellura, M., G. Cirrincione, A. Marvuglia, and A. Miraoui (2008). Wind speed spatial estimation for energy planning in Sicily: A neural kriging application. *Renewable Energy 33*(6), 1251–1266.

Chang, T. P. (2011). Estimation of wind energy potential using different probability density functions. *Applied Energy 88*(5), 1848–1856.

Cocchi, D., F. Greco, and C. Trivisano (2007). Hierarchical space-time modelling of PM 10 pollution. *Atmospheric Environment 41*(3), 532–542.

Cressie, N. (1992). Statistics for spatial data. *Terra Nova 4*(5), 613–617.

Cressie, N. and H.-C. Huang (1999). Classes of nonseparable, spatio-temporal stationary covariance functions. *Journal of the American Statistical Association 94*(448), 1330–1339.

Cressie, N. and C. K. Wikle (2015). *Statistics for spatio-temporal data*. John Wiley & Sons.

Cressie, N. A. (1993). Statistics for spatial data: Wiley series in probability and mathematical statistics. *Find this article online*.

Diebold, F. X., T. A. Gunther, and A. S. Tay (1997). Evaluating density forecasts.

Dowell, J. and P. Pinson (2016). Very-short-term probabilistic wind power forecasts by sparse vector autoregression. *IEEE Transactions on Smart Grid 7*(2), 763–770.

Energinet.dk (2014, sep). Energinet.dk's analysis assumptions 2014-2035, update september 2014. Technical report.

Etienne, C., A. Lehmann, S. Goyette, J.-I. Lopez-Moreno, and M. Beniston (2010). Spatial predictions of extreme wind speeds over Switzerland using

generalized additive models. *Journal of Applied Meteorology and Climatology 49*(9), 1956–1970.

Fassò, A. and F. Finazzi (2011). Maximum likelihood estimation of the dynamic coregionalization model with heterotopic data. *Environmetrics 22*(6), 735–748.

Flecher, C., P. Naveau, D. Allard, and N. Brisson (2010). A stochastic daily weather generator for skewed data. *Water Resources Research 46*(7).

Focken, U., M. Lange, K. Mönnich, H.-P. Waldl, H. G. Beyer, and A. Luig (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*(3), 231–246.

Foley, A. M., P. G. Leahy, A. Marvuglia, and E. J. McKeogh (2012). Current methods and advances in forecasting of wind power generation. *Renewable Energy 37*(1), 1–8.

Fraley, C., A. E. Raftery, and T. Gneiting (2010). Calibrating multimodel forecast ensembles with exchangeable and missing members using Bayesian model averaging. *Monthly Weather Review 138*(1), 190–202.

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

Gipe, P. (2004). Wind power. *Wind Engineering 28*(5), 629–631.

Girard, R. and D. Allard (2013). Spatio-temporal propagation of wind power prediction errors. *Wind Energy 16*(7), 999–1012.

Girard, R., K. Laquaine, and G. Kariniotakis (2013). Assessment of wind power predictability as a decision factor in the investment phase of wind farms. *Applied Energy 101*, 609–617.

Gneiting, T. Genton, M. and P. Guttorp (2007). *Statistics of Spatio-Temporal Systems*, Chapter Geostatistical Space-Time Models, Stationarity, Separability, and Full Symmetry, pp. 151–175. Monographs in Statistics and Applied Probability. Chapman and Hall/CRC Press,.

Gneiting, T. (1999). Correlation functions for atmospheric data analysis. *Quart. J. R. Meteorol. Soc. 125*, 2449–2464.

Gneiting, T. (2002). Nonseparable, stationary covariance functions for space–time data. *Journal of the American Statistical Association 97*(458), 590–600.

Gneiting, T., F. Balabdaoui, and A. E. Raftery (2007). Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 69*(2), 243–268.

Gneiting, T. and A. E. Raftery (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association 102*(477), 359–378.

Gneiting, T. and R. Ranjan (2011). Comparing density forecasts using threshold-and quantile-weighted scoring rules. *Journal of Business & Economic Statistics 29*(3), 411–422.

Gneiting, T., L. I. Stanberry, E. P. Grimit, L. Held, and N. A. Johnson (2008). Assessing probabilistic forecasts of multivariate quantities, with an application to ensemble predictions of surface winds. *Test 17*(2), 211.

Guillot, G. (1999). Approximation of Sahelian rainfall fields with meta-Gaussian random functions Part 1: model definition and methodology. *Stochastic Environmental Research and Risk Assesment 13*, 100–112.

Gurmu, S. (1997). Semi-parametric estimation of hurdle regression models with an application to Medicaid utilization. *Journal of Applied Econometrics 12*(3), 225–242.

Hall, D. B. (2000). Zero-inflated poisson and binomial regression with random effects: a case study. *Biometrics 56*(4), 1030–1039.

Haslett, J. and A. Raftery (1989). Space-time modelling with long-memory dependence: assessing Ireland's wind power resource (with discussion). *Applied Statistics 38*, 1–50.

Hering, A. S. and M. G. Genton (2010). Powering up with space-time wind forecasting. *Journal of the American Statistical Association 105*(489), 92–104.

Hering, A. S. and M. G. Genton (2011). Comparing spatial predictions. *Technometrics 53*(4), 414–425.

Hossain, J., V. Sinha, and V. Kishore (2011). A GIS based assessment of potential for windfarms in India. *Renewable Energy 36*(12), 3257–3267.

Jolliffe, I. T. and D. B. Stephenson (2012). *Forecast verification: a practitioner's guide in atmospheric science*. John Wiley & Sons.

Jones, L. and C. Clark (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*, Volume 2526.

Katzenstein, W., E. Fertig, and J. Apt (2010). The variability of interconnected wind plants. *Energy Policy 38*(8), 4400–4410.

Kleiber, W., R. W. Katz, and B. Rajagopalan (2012). Daily spatiotemporal precipitation simulation using latent and transformed gaussian processes. *Water Resources Research 48*(1).

Krige, D. G. (1951). A statistical approach to some basic mine valuation problems on the Witwatersrand. *Journal of the Southern African Institute of Mining and Metallurgy 52*(6), 119–139.

Landberg, L. (1999). Short-term prediction of the power production from wind farms. *Journal of Wind Engineering and Industrial Aerodynamics 80*(1), 207–220.

Lau, A. (2011). *Probabilistic wind Power Forecasts: From Aggregated Approach to Spatiotemporal Models*. Ph. D. thesis, Mathematical Institute, University of Oxford.

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

Lenzi, A., C. P. de Souza, R. Dias, N. L. Garcia, and N. Heckman (2016). Analysis of aggregated functional data from mixed populations with application to energy consumption. *Environmetrics 28*(2).

Lenzi, A., G. Guillot, and P. Pinson (2015). A spatial model for the instantaneous estimation of wind power at a large number of unobserved sites. *Procedia Environmental Sciences 26*, 131–134.

Lenzi, A., P. Pinson, L. H. Clemmensen, and G. Guillot (2016). Spatial models for probabilistic prediction of wind power with application to annual-average and high temporal resolution data. *Stochastic Environmental Research and Risk Assessment*, 1–17.

Lenzi, A., I. Steinsland, and P. Pinson (2017, April). Benefits of spatio-temporal modelling for short term wind power forecasting at both individual and aggregated levels. *ArXiv e-prints*.

Lesaffre, E., D. Rizopoulos, and R. Tsonaka (2007). The logistic transform for bounded outcome scores. *Biostatistics 8*(1), 72–85.

Lindenberg, S., S. B. O. K. and E. DeMeo (2008). 20 Percent Wind Energy by 2030: Increasing Wind energy's contribution to U.S. electricity supply. Technical report, U.S. Department of Energy.

Lindgren, F. and H. Rue (2015). Bayesian spatial modelling with R-INLA. *Journal of Statistical Software 63*(19).

Lindgren, F., H. Rue, and J. Lindström (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 73*(4), 423–498.

Liu, X. (2010). Economic load dispatch constrained by wind power availability: A wait-and-see approach. *IEEE Transactions on Smart Grid 1*(3), 347–355.

Luo, W., M. Taylor, and S. Parker (2008). A comparison of spatial interpolation methods to estimate continuous wind speed surfaces using irregularly distributed data from England and Wales. *International Journal of Climatology 28*(7), 947–959.

Madsen, H., P. Pinson, G. Kariniotakis, H. A. Nielsen, and T. S. Nielsen (2005). Standardizing the performance evaluation of short-term wind power prediction models. *Wind Engineering 29*(6), 475–489.

Mardia, K. V. and C. R. Goodall (1993). Spatial-temporal analysis of multivariate environmental monitoring data. *Multivariate Environmental Statistics 6*(76), 347–385.

Matheron, G. (1963). Principles of geostatistics. *Economic Geology 58*(8), 1246–1266.

McDonnell, R. (1995). *International GIS dictionary*. John Wiley & Sons.

Mead, R. (1965). A generalised logit-normal distribution. *Biometrics 21*(3), 721–732.

Messner, J. W., A. Zeileis, J. Broecker, and G. J. Mayr (2014). Probabilistic wind power forecasts with an inverse power curve transformation and censored regression. *Wind Energy 17*(11), 1753–1766.

Morales, J. M., A. J. Conejo, H. Madsen, P. Pinson, and M. Zugno (2013). *Integrating renewables in electricity markets: Operational problems*, Volume 205. Springer Science & Business Media.

Mullahy, J. (1986). Specification and testing of some modified count data models. *Journal of Econometrics 33*(3), 341–365.

Myers, D. E. (1982). Matrix formulation of co-kriging. *Mathematical Geology 14*(3), 249–257.

Ortega-Vazquez, M. A. and D. S. Kirschen (2009). Estimating the spinning reserve requirements in systems with significant wind power generation penetration. *IEEE Transactions on Power Systems 24*(1), 114–124.

Ortega-Vazquez, M. A. and D. S. Kirschen (2010). Assessing the impact of wind power generation on operating costs. *IEEE Transactions on Smart Grid 1*(3), 295–301.

Pinson, P. (2012). Very-short-term probabilistic forecasting of wind power with generalized logit–normal distributions. *Journal of the Royal Statistical Society: Series C (Applied Statistics) 61*(4), 555–576.

Pinson, P., C. Chevallier, and G. N. Kariniotakis (2007). Trading wind generation from short-term probabilistic forecasts of wind power. *Power Systems, IEEE Transactions on 22*(3), 1148–1156.

Pinson, P. and R. Hagedorn (2012). Verification of the ECMWF ensemble forecasts of wind speed against analyses and observations. *Meteorological Applications 19*(4), 484–500.

Pinson, P. and G. Kariniotakis (2010). Conditional prediction intervals of wind power generation. *IEEE Transactions on Power Systems 25*(4), 1845–1856.

Pinson, P., P. McSharry, and H. Madsen (2010). Reliability diagrams for non-parametric density forecasts of continuous variables: Accounting for serial correlation. *Quarterly Journal of the Royal Meteorological Society 136*(646), 77–90.

Pinson, P. and J. Tastu (2013). Discrimination ability of the energy score. Technical report, Technical University of Denmark.

Pohlmeier, W. and V. Ulrich (1995). An econometric model of the two-part decision making process in the demand for health care. *Journal of Human Resources*, 339–361.

Rue, H. and S. Martino (2007). Approximate Bayesian inference for hierarchical Gaussian Markov random field models. *Journal of Statistical Planning and Inference 137*(10), 3177–3192.

Rue, H., S. Martino, and N. Chopin (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series b (Statistical Methodology) 71*(2), 319–392.

Sanandaji, B. M., A. Tascikaraoglu, K. Poolla, and P. Varaiya (2015). Low-dimensional models in spatio-temporal wind speed forecasting. In *American Control Conference (ACC), 2015*, pp. 4485–4490. IEEE.

Scheuerer, M. and T. M. Hamill (2015). Variogram-based proper scoring rules for probabilistic forecasts of multivariate quantities. *Monthly Weather Review 143*(4), 1321–1334.

Sliz-Szkliniarz, B. and J. Vogt (2011). GIS-based approach for the evaluation of wind energy potential: A case study for the Kujawsko–Pomorskie Voivodeship. *Renewable and Sustainable Energy Reviews 15*(3), 1696–1707.

Stein, M. (1992). Prediction and Inference for Truncated Spatial Data. *Journal of Computational and Graphical Statistics*, 91–110.

Tastu, J., P. Pinson, E. Kotwa, H. Madsen, and H. A. Nielsen (2011). Spatio-temporal analysis and modeling of short-term wind power forecast errors. *Wind Energy 14*(1), 43–60.

Tastu, J., P. Pinson, and H. Madsen (2015). Space-time trajectories of wind power generation: Parametrized precision matrices under a Gaussian copula approach. In *Modeling and Stochastic Learning for Forecasting in High Dimensions*, pp. 267–296. Springer.

Tastu, J., P. Pinson, P.-J. Trombe, and H. Madsen (2014). Probabilistic forecasts of wind power generation accounting for geographically dispersed information. *IEEE Transactions on Smart Grid 5*(1), 480–489.

Thorarinsdottir, T. L. and T. Gneiting (2010). Probabilistic forecasts of wind speed: ensemble model output statistics by using heteroscedastic censored regression. *Journal of the Royal Statistical Society: Series A (Statistics in Society) 173*(2), 371–388.

Thorarinsdottir, T. L., M. Scheuerer, and C. Heinz (2016). Assessing the calibration of high-dimensional ensemble forecasts using rank histograms. *Journal of Computational and Graphical Statistics 25*(1), 105–122.

Tierney, L. and J. B. Kadane (1986). Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association 81*(393), 82–86.

Toth, Z., O. Talagrand, G. Candille, and Y. Zhu (2003). Probability and ensemble forecasts.

Whittle, P. (1954). On stationary processes in the plane. *Biometrika*, 434–449.

Whittle, P. (1963). Stochastic-processes in several dimensions. *Bulletin of the International Statistical Institute 40*(2), 974–994.

Wilks, D. S. (1990). Maximum likelihood estimation for the gamma distribution using data containing zeros. *Journal of Climate 3*(12), 1495–1501.

Wytock, M. and J. Z. Kolter (2013a). Large-scale probabilistic forecasting in energy systems using sparse Gaussian conditional random fields. In *Decision and Control (CDC), 2013 IEEE 52nd Annual Conference on*, pp. 1019–1024. IEEE.

Wytock, M. and J. Z. Kolter (2013b). Sparse Gaussian conditional random fields: Algorithms, theory, and application to energy forecasting. In *ICML (3)*, pp. 1265–1273.

Xie, L., Y. Gu, X. Zhu, and M. G. Genton (2014). Short-term spatio-temporal wind power forecast in robust look-ahead power system dispatch. *IEEE Transactions on Smart Grid 5*(1), 511–520.

Yoder, M., A. S. Hering, W. C. Navidi, and K. Larson (2014). Short-term forecasting of categorical changes in wind power with Markov chain models. *Wind Energy 17*(9), 1425–1439.

Zhang, Y., J. Wang, and X. Luo (2015). Probabilistic wind power forecasting based on logarithmic transformation and boundary kernel. *Energy Conversion and Management 96*, 440–451.

Zhang, Y., J. Wang, and X. Wang (2014). Review on probabilistic forecasting of wind power generation. *Renewable and Sustainable Energy Reviews 32*, 255–270.

Zhu, X. and M. G. Genton (2012). Short-term wind speed forecasting for power system operations. *International Statistical Review 80*(1), 2–23.

Appendix A

Spatial models for probabilistic prediction of wind power with application to annual-average and high temporal resolution data

---

Lenzi, A., Pinson, P., Clemmensen, L. H., Guillot, G. (2016). Spatial models for probabilistic prediction of wind power with application to annual-average and high temporal resolution data. *Stoch Environ Res Risk Assess* DOI 10.1007/s00477-016-1329-0, pp. 1-17

Stoch Environ Res Risk Assess
DOI 10.1007/s00477-016-1329-0

ORIGINAL PAPER

Spatial models for probabilistic prediction of wind power with application to annual-average and high temporal resolution data

Amanda Lenzi[1] · **Pierre Pinson**[2] · **Line H. Clemmensen**[1] · **Gilles Guillot**[1]

Abstract Producing accurate spatial predictions for wind power generation together with a quantification of uncertainties is required to plan and design optimal networks of wind farms. Toward this aim, we propose spatial models for predicting wind power generation at two different time scales: for annual average wind power generation, and for a high temporal resolution (typically wind power averages over 15-min time steps). In both cases, we use a spatial hierarchical statistical model in which spatial correlation is captured by a latent Gaussian field. We explore how such models can be handled with stochastic partial differential approximations of Matérn Gaussian fields together with Integrated Nested Laplace Approximations. We demonstrate the proposed methods on wind farm data from Western Denmark, and compare the results to those obtained with standard geostatistical methods. The results show that our method makes it possible to obtain fast and accurate predictions from posterior marginals for wind power generation. The proposed method is applicable in scientific areas as diverse as climatology, environmental sciences, earth sciences and epidemiology.

Keywords Wind power · Spatial prediction · Latent Gaussian field · Integrated nested Laplace approximation

✉ Amanda Lenzi
amle@dtu.dk

[1] Applied Mathematics and Computer Science Department, Technical University of Denmark, 2800 Kgs Lyngby, Denmark

[2] Electrical Engineering Department, Technical University of Denmark, 2800 Kgs Lyngby, Denmark

1 Introduction

In a society increasingly concerned with sustainability, the share of wind energy in total installed power capacity has grown rapidly in recent years around the world. For example, Denmark has the largest proportion of wind energy capacity relative to the volume of electricity consumption and the Danish government aims at having 50 % of the energy demand met by wind power by 2025 (Tastu et al. 2011). The main expected benefit from using wind power as a source of energy instead of fossil fuels is the reduction of carbon emissions. However, advanced forecasting methodologies are necessary to address issues related to the limited predictability of wind power generation. Increasing the quality of wind energy forecasts is not only important in order to efficiently handle the energy demand (Katzenstein et al. 2010), but it also increases the revenues from the electricity market, with the optimization of bidding strategies (Pinson et al. 2007).

From a statistical perspective, accurately predicting wind power and quantifying the uncertainties of the predictions at a regional scale is a challenging problem. Indeed, the statistical distribution of wind power data is characterized by the presence of complex temporal and spatial trends that are not well encompassed by stationary models. Also, the intermittent nature of wind leads to a spike at zero in the empirical distribution for high temporal resolutions (e.g. 15 min interval), which is difficult to model.

Studies on the medium-term and short-term forecasting of wind speed and wind power have received a lot of attention lately. Predictions of wind speed are ultimately in order to predict power; thus, there is a strong link to power, even when assessing the quality of wind speed predictions. The reader should note that it is common to have an

Published online: 14 October 2016

overlap between wind speed and wind power in the literature. Based on this loose behaviour, we talk interchangeably here about wind speed and wind power. A common approach when the focus is on a single wind farm based on local measurements only, is related to a time-series framework that usually assumes a Gaussian distribution for the wind speed response. One of the first proposals in the literature was published in Brown et al. (1984) and uses auto-regressive moving average (ARMA) models for wind speed observations at lead times of between a few hours and a few days. The following year, Bossanyi (1985) used a Kalman Filter to predict wind speed at a one-minute resolution with the last six values as input, and observed an improvement in the RMS error over persistence for the prediction of the next time step of up to 10 %. Some years later, Daniel and Chen (1991), on the other hand, used stochastic simulation and forecasting models of hourly averages of wind speeds, taking into account the autocorrelation, non-Gaussian distribution and diurnal non-stationarity, and fitted an ARMA process to wind speed data. On a regional scale, Shih (2008) assessed periodic diurnal components and prevailing wind directions of wind speed time series in Taiwan using spectral analysis.

Recently, in a more general set up, Gneiting et al. (2006) introduced the regime-switching space-time model that identifies the atmospheric regime at the wind energy site and fits a conditional predictive model for each regime providing probabilistic forecasts of wind speed data. This approach deals with non-Gaussianity and with the discontinuity at zero by making use of a truncated normal distribution. To deal with discrete probability masses and the fact that normalized wind power is bounded between zero and one, Pinson (2012) applied the generalized logit-normal distribution with a potential concentration of probability mass at the bounds of the unit interval [0, 1] to forecast wind power fluctuations at single wind farms.

The methods mentioned above use only historical data at a single site. Because the spatial aspect of the problem is disregarded, these methods do not provide a straightforward way to extrapolate predictions at un-monitored locations. Moreover, wind power forecasting models developed for one location do not match the other sites due to, for example, change in terrain, different wind speed patterns and atmospheric factors. It is therefore not straightforward to transpose the results to other locations. In this sense, developing a portable and general model that mimics the spatial dependence structure and gives an overview of power generation at all wind power generation sites over a region is a timely objective.

Several spatial interpolation techniques are available to predict the wind speed in locations where data is not available. Luo et al. (2008) studied seven methods to estimate the daily mean wind velocity surface showing that kriging methods produce more accurate results than deterministic techniques on a country level. Joyner et al. (2015) compared the number of high-error stations produced when interpolating stations from wind data using ordinary kriging and cokriging. A geographic information system (GIS) based approach is used to assess wind resources in India and Poland in Hossain et al. (2011) and Sliz-Szkliniarz and Vogt (2011), respectively. Cellura et al. (2008) dealt with spatial estimation of the wind fields in Sicily by using neural kriging modeling. Etienne et al. (2010) predicted extreme wind speed with a combination of GIS techniques and Generalized Additive Models.

The purpose of the present paper is to propose statistical models for wind power generation that incorporate the spatial features of all the wind farm locations and yield calibrated predictive distributions with a minimum amount of computational effort. Reliability, also referred to as calibration, of probabilistic forecasts is assessed with reliability diagrams. A calibrated forecast should have the observed levels matching the nominal levels for specific quantile forecasts. Reliability is considered as the main required property of probabilistic forecasting (Gneiting et al. 2007) since it is used as input for decision problems and a probabilistic bias in the forecasts would yield poor operational decisions.

We use conventional kriging as a benchmark method for predicting the annual average wind power generation and high temporal resolution of wind power generation. This is one of the standard techniques for spatial interpolations, as described in Cressie (1988). Although kriging is an optimal method when the data follows a Gaussian distribution, it has proved to be a robust method under a range of conditions (Deutsch and Journel 1992). It can provide an efficient way to linearly interpolate nearby observations and thus lead to an estimation of wind power generation in each station; exclusively based on the mean and covariance structure of the Gaussian field.

In a first step, we focus on a model describing spatial variation of annual average wind power generation. We use a hierarchical spatial model based on a skewed continuous distribution with a stochastic, spatially structured mean that depends on the covariate. The spatial structure is captured by a latent Gaussian random field with a Matérn covariance function. In a second step, we propose a model tailored for wind power generation data with high temporal resolution. This type of scenario is of relevance for a number of operational problems where wind power generation is only observed at a limited number of wind farms, while decision-making problems may require an overview of power generation at all sites over a region. This setting is modelled with a mixture of degenerated distributions at zero and a skewed continuous distribution for the non-zero

values. The distributions share a Gaussian random field with a Matérn covariance function.

Note that for these models, the posterior marginals are not available in closed form due to the non-Gaussian response variables. Departure from normality can easily be handled, but comes at a computational cost (Diggle et al. 1998). To issue probabilistic forecasts, we use an integrated nested Laplace approximation (INLA) (Rue et al. 2009) as an alternative to MCMC methods, and directly compute approximations to the posterior marginals. The resulting predictions are evaluated on a case study based on wind farms located in the western part of Denmark, while comparing the results from our approach to those from the ordinary kriging method.

Although this paper is motivated by the problem of spatial prediction of wind power generation, the solution developed here is relevant to many spatial prediction problems in earth and environmental sciences involving non-Gaussian data.

The remainder of this article is organized as follows. In Sect. 2, we give details of the data wind power data set used as a case study in order to show a realistic view of the methods proposed in this paper. In Sect. 3, we first describe the kriging method used here as benchmark. Then we present the hierarchical spatial model for predicting annual average wind power generation as well as the model tailored for wind power generation with high temporal resolution. We eventually explain how to perform inference and prediction with such models. A detailed explanation of the methods used for evaluation is given in Sect. 4. In Sect. 5, we present results for the probabilistic prediction of annual average and high temporal resolution of wind power generation, and compare our method with kriging. Section 6 contains a discussion of the limitations and possible extensions of our method and draws conclusions to our work.

2 Western Denmark wind power data

We consider a data set consisting of wind power generation measurements in wind farms located in the western part of Denmark. Each measurement is the temporal average power over a 15-min time step. The period covered ranges from January 2006 to March 2012. The distances between the wind farms range between 1 and 310 km.

The amount of wind power produced at a wind farm depends on its capacity, which is the maximum output when all turbines operate at their maximum nominal power. Since most of the wind farms have different capacities, and in order to facilitate comparisons between data sets, we normalize the wind power data by dividing all the measurements by the maximum nominal power value of each specific wind farm.

We start by modelling the annual average wind power generation in the year of 2010, where average wind power is obtained by averaging the 15-min normalized power output at each of 349 wind farms in 2010, resulting in purely spatial data. The year of 2010 was chosen for no specific reason to illustrate the proposed methods and the other years gave similar results. Note that here, in contrast with the scenario for the wind power at high temporal resolution, the large amount of zero measurements is not present, since these are averaged out with all the measurements at that specific station over 2010. A map of the normalized annual average wind power generation data set for 2010 is shown in Fig. 1. Thus, a value of 0.4 indicates that the annual average wind power generation for that specific wind farm is 40 % of the highest measurement obtained for that wind farm in 2010.

Next, to illustrate the methodology for modelling wind power generation at high temporal resolution, we fit the model separately to the data of each time step from the first day of every month in 2010. Please note that the remaining days and years gave similar results. Taking measurements from the first day of each month during 2010 results in $4 \times 24 \times 12 = 1152$ time steps. Since 165 of the 1152 time steps contain a large number of zero measurements, which results in problems during the estimation when using the R-INLA package, we used the remaining 987 time steps in our analysis, so that, in total, we have measurements from 349 wind farms over 987 time steps during the year of 2010.

Figure 2 shows the observed relative frequency of observations with wind power generation greater than zero

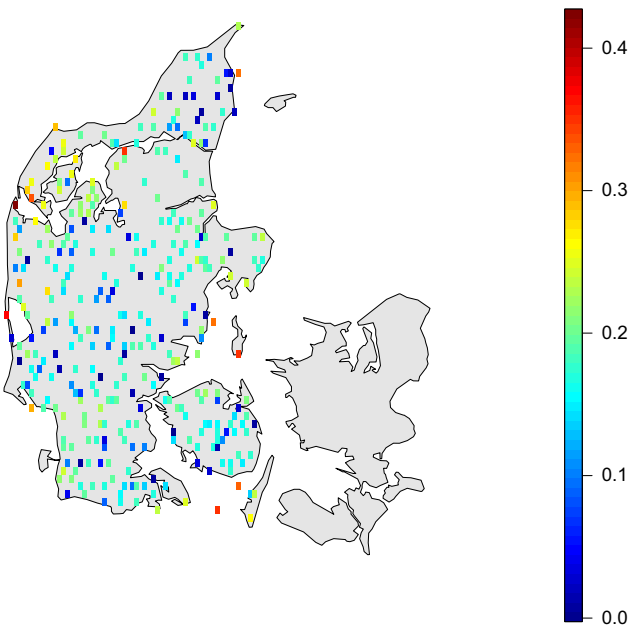


Fig. 1 Normalized annual average wind power in 2010

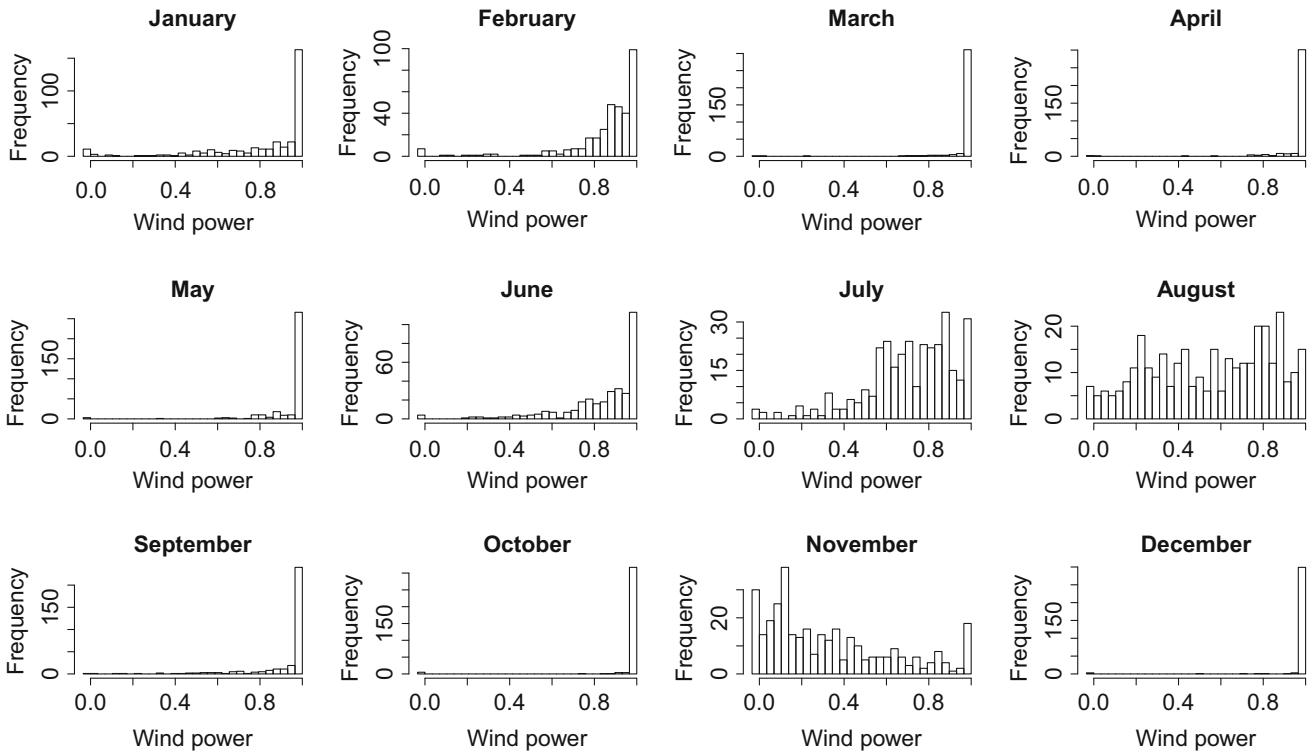


Fig. 2 Histogram of the relative frequency of wind power generation greater than zero over the 349 wind farms in Denmark in the first day of each month of 2010. The relative frequency of wind power generation greater than zero for each wind farm is calculated as the number of observations with wind power generation greater than zero divided by the total number of observations in a day

for the first day of each month of 2010. In each plot, we calculate the relative frequency for wind farms $1, \ldots, 349$, by dividing the number of observations with power produced greater than zero by the total number of observations in a day. As we can see from this plot, in 2010, July, August and November had lower empirical probability of producing wind power, meaning that the data sets for these months contain a larger number of zeros.

3 Models and prediction methods

In this section we start by describing the standard benchmark kriging. Next, we introduce two different spatial models, one for the annual average wind power generation and another for high temporal resolution of wind power generation. The section ends with the methods used to perform inference and obtain probabilistic forecasts.

3.1 Kriging as a benchmark model

Kriging is just the usual name to describe the best linear unbiased predictor for a spatial process Z_i at location i (Matheron 1963). Although the kriging equations also hold for non-Gaussian processes, the kriging predictor coincides with the best linear unbiased predictor only when the process is Gaussian. Nevertheless, the kriging method is an attractive benchmark due to its robustness and simplicity in obtaining predictions. Examples of models for precipitation and air quality monitoring using kriging methods can be found in Atkinson and Lloyd (1998) and Ignaccolo et al. (2014), respectively.

Depending on how the mean of Z_i is modelled, different versions of kriging can be derived. While in ordinary kriging the mean is assumed to be constant and unknown over the neighbourhood of the value to be predicted, in universal kriging models, the mean is a function of covariates.

Two types of kriging methods are used in this paper—ordinary kriging and universal kriging. For the universal kriging, we define the covariate d_i as the distance from wind farm i to the closest neighbouring coordinate on the border of the west coast of Denmark. The initial idea of including this covariate was inspired by maps of Danish wind speed developed to assist the Danish municipalities in their planning work for wind-turbine installation. These maps show that the prevailing wind directions in Denmark are west and southwest. Since the covariate improved the predictions only for the annual average of wind power generation, we use ordinary kriging to model wind power generation at a high temporal resolution and universal kriging to model the annual average of wind power

generation. For simplicity of notation, we refer to both ordinary and universal kriging as just kriging.

To obtain predictions at unobserved locations, we fit a parametric variogram to the data. Before fitting the parametric variogram, we need to divide the data into several bins by breaking up the distances between each of the points based on a lag size between the distances, and afterwards the actual semi-variogram value is calculated for the bins. We start by calculating a sample variogram from the data that depends only on the distance between the wind farms. Next, we fit a parametric variogram model to the binned data from the sample variogram with the sill being equal to the maximum estimate of the sample variogram. The estimation of the variogram model parameters is done by iteratively reweighted least squares, with weights equal to N_h/h^2 (Cressie 1985), where N_h is the number of point pairs and h is the distance between the locations. The range is set to be equal to the maximum of the distance from the sample variogram divided by two, and a parametric Matérn covariance structure is assumed. The Matérn covariance function is a flexible model that contains as special cases many of the covariance functions used in spatial statistics and is given by

$$\Sigma(s,s') = \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)}(\kappa||s-s'||)^{\nu}K_{\nu}(\kappa||s-s'||) \tag{1}$$

where K_ν is the modified Bessel function of second kind of order $\nu > 0$, κ can be used to select the range and σ to achieve the desired marginal variance. The parameter ν is a smoothness parameter determining the mean-square differentiability of the underlying process. Although this parameter is fixed to 1 for computational reasons, it remains flexible enough to handle a broad class of spatial variation (Rue et al. 2009). Applications with fixed parameter ν include (Guillot et al. 2015; Cameletti et al. 2013; Munoz et al. 2013; Musenge et al. 2013). Detailed information on the Matérn covariance model can be found in Guttorp and Gneiting (2006) and Stein (1999).

We perform kriging using the R package *gstat*.

3.2 A spatial model for annual average wind power generation

Annual average wind power generation is obtained by averaging the power produced over 2010 at each wind farm. Although the distribution of annual average wind power does not have the problem of probability mass at zero and is less skewed than individual power, there is still the challenge that it is bounded below by zero and above by the maximum total capacity of the turbines. As a result, it is reasonable to generate predictions that lie inside this permissible range. We propose a hierarchical spatial model for annual average wind power generation. To ensure that the final predictions lie in the valid range, we use a Beta distribution with a stochastic mean that we model using a log-normal distribution including both covariates and a spatial structure which is captured by a latent Gaussian random field.

Let $Y_1, \ldots, Y_N$ be the annual average normalized wind power, where N is the number of spatial points. We use the following parametrization for the Beta distribution with parameters a and b,

$$m = \frac{a}{a+b}, \quad 0 < m < 1 \quad \text{and} \quad \phi = a + b, \quad \phi > 0, \tag{2}$$

which implies that $a = m\phi$ and $b = (1-m)\phi$. The distribution of Y_i can be written, in the new parametrization, as

$$Y_i \sim \text{Beta}(m_i\phi, (1-m_i)\phi). \tag{3}$$

We define a linear predictor for the log of the mean of Y_i, i. e.,

$$\log\,(m_i) = f(d_i) + x_i \tag{4}$$

where d_i is the distance from wind farm i to the closest neighbouring coordinate on the border of the west coast of Denmark, to be thoroughly described in the following and x_i is a value of a Gaussian random field.

First of all, the spatial correlation of the random field formed by the set of x_i's in (4) is incorporated through a zero mean Gaussian random field $\mathbf{x}$

$$\mathbf{x} \sim \mathcal{N}(\boldsymbol{0}, \Sigma). \tag{5}$$

The covariance function Σ belongs to the Matérn family. Instead of the parametrization given in (1), we redefine the covariance function depending on the range, $r = \frac{\sqrt{8\nu}}{\kappa}$

$$\Sigma(s,s') = \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)}\left(\frac{\sqrt{8\nu}}{r}||s-s'||\right)^{\nu}K_{\nu}\left(\frac{\sqrt{8\nu}}{r}||s-s'||\right) \tag{6}$$

The range parameter r introduced in the correlation function is interpreted as the minimum distance for which the correlation between two locations becomes negligible.

Now, we turn towards the component $f(d_i)$ in (4). Let $\mathbf{d} = (d_1, \ldots, d_N)$, the vector of distances from each wind farm to the closest neighbouring coordinate on the border of the west coast of Denmark. We define $\tilde{\mathbf{d}}$ as a grouped version of $\mathbf{d}$ with groups indexed by $c = 1, \ldots, 25$ and components $\tilde{d}_c$'s. We obtain the groups by first ordering the values of $\mathbf{d}$ from the smallest to the largest, and then using bins of equal length with the groups set to the median of the covariates belonging to that group. Next, the effect of the covariate d_i is modelled as a smooth function f, defined as

$$f(d_i) = \sum_{c=1}^{m} w_i[c]\, l_c(i) \tag{7}$$

where $w_i[c]$ is the cth component of the vector $\mathbf{w}_i$. We suppose that the vector $\mathbf{w}_i$ is a set of serially randomly

correlated regression coefficients, normally distributed with mean 0 and covariance matrix Q. The vector $\mathbf{l}(i)$ forms the series of basis functions. The basis function $\mathbf{l}(i)$ could be chosen, for example, as the so-called spline or B-spline basis (Lindgren et al. 2011). However, here we explore the use of explanatory variables as the basis function. We define $l_c(i)$ equal to an indicator function equal to one if the covariate d_i belongs to group $\tilde{d}_c$, and zero otherwise.

Moreover, we assume that the coefficients over the range of the covariate values have a first order random walk prior. The random walk model of order 1 for a vector $(u_1, \ldots, u_n)$ is constructed assuming independent increments:

$$\Delta u_i = u_i - u_{i+1} \sim N(0, \tau^{-1}) \tag{8}$$

In order to test the significance of the covariate $\tilde{d}_i$ in the model, we compare the error of prediction with and without the covariate. Since including the covariate results in a smaller error of prediction, we choose to include it in the model.

Using Y_i's Beta distribution in (3), it follows that the likelihood is

$$\begin{aligned} \mathrm{L}\ (\boldsymbol{\theta}|\mathbf{y}) &= \prod_{y_i=1}^{N} \mathrm{Beta}\ (y_i|m_i, \phi) \\ &= \prod_{y_i=1}^{N} \frac{\Gamma(\phi)}{\Gamma(m_i\phi)\Gamma(\phi - m_i\phi)} y_i^{m_i\phi-1}(1-y_i)^{\phi(1-m_i)-1} \end{aligned} \tag{9}$$

The smoothness parameter ν in the Matérn covariance is set to 1 as in Sect. 3.1. Moreover, the function f in (7) only depends on the parameter τ in (8). It follows that the vector of hyperparameters is given by

$$\boldsymbol{\theta} = \{\phi, \tau, \kappa, \sigma\}.$$

Default log-Gamma priors are assumed for all the hyperparameters in the model. We obtain predictions for the model just described with the INLA methodology implemented in the R-INLA package to be described in Sect. 3.4.

3.3 A spatial model for high temporal resolution of wind power generation

This model is tailored for wind power generation at high temporal resolution. It has a competitive advantage over using a truncated Gaussian process to handle the bound at zero, as explained in Stein (1992). Instead, our model uses a Bernoulli distribution to model the large amount of zero measurements in the data set, and it makes use of a Gamma distribution to model the asymmetric distribution of the positive values. This specification has the advantage of a well-behaved likelihood function that factors into two independent terms, making calculations relatively simple. The first term has only the logit model parameters and the second term involves only the parameters of the Gamma distribution [see Eq. (14)]. To overcome the so-called intermittency problem, where areas with small values lie very close to areas with large values, there is an underlying Gaussian field that is part of the linear predictor of both distributions—Bernoulli and Gamma.

Different approaches exist to deal with applications in which data take nonnegative values but have a substantial proportion of values at zero. One approach is to model a zero-inflation parameter that represents the probability of having zeros, given that these zero measurements come from the same distribution as the non-zero values. An example is found in Hall (2000).

Alternatively, data containing an abundant amount of zeros can be modelled with two latent Gaussian processes. The first controls the probability of observing zero values, and the second governs the density distribution of non-zero observations. Examples of this type of model used to describe accumulated precipitation include Berrocal et al. (2008) and Kleiber et al. (2012). On the other hand, Baxevani and Lennartsson (2015) model simultaneously the occurrence and intensity of rainfall using a single latent Gaussian field and then the positive part of the process is considered to be observed up to a transformation of the observed data. Rainfall and precipitation share similar features with wind power data sets, since one could have long periods of dry days with no rainfall observation. Here, we consider that wind power generation is driven only by one latent Gaussian field that controls both the occurrence and the intensity. Moreover, the probability of having power generation greater than zero is modelled as a Bernoulli distribution with probability that depends on the latent Gaussian field.

The model used in this paper is often called a hurdle model. The hurdle model for count data was proposed in Mullahy (1986). One part of the model is a binary model, such as a logistic or a probit regression, for whether the response outcome is zero or positive. To estimate the level of the positive outcomes, the second part of the model consists of a truncated model that modifies an ordinary distribution by conditioning on a positive outcome. Applications of similar models can be found in Pohlmeier and Ulrich (1995) and Gurmu (1997). Within the INLA framework, Serra et al. (2014) used a hurdle model to predict the occurrence of wildfires with point mass at zero followed by a truncated Poisson distribution for the non-zero observations.

In the second stage of our model, the distribution is a Gamma density for the non-zero values, which represents the amount of wind power generated. The Gamma distribution is a good choice for describing wind power values

for several reasons. It provides a flexible representation of a variety of distribution shapes while utilizing only two parameters: the shape and the scale. It can range from exponential-decay forms for shape values near one, to nearly normal forms for shape values beyond 20 (Wilks 1990). In addition, a distribution that excludes negative values and is positively skewed is readily applicable for the analysis of wind power.

We start by defining a binary random variable Z_i at location $i = 1, \ldots, N$ which depends on the generation of wind power

$$Z_i = \begin{cases} 1, & \text{if } y_i > 0, \\ 0, & \text{otherwise .} \end{cases}$$

where y_i is the observed wind power generation at wind farm i. We assume that Z_i follows a Bernoulli distribution with parameter p_i

$$Z_i \sim \text{Bern}(p_i) \tag{10}$$

where p_i is the probability of having wind power generation greater than zero at wind farm i and is modeled as

$$\text{logit } (p_i) = \alpha_z + x_i \tag{11}$$

with α_z being the intercept and x_i an observation from a latent Gaussian random field with Matérn covariance as defined in (5). Then, conditional on the presence or absence of wind power, we model the amount of wind power generation at station i

$$Y_i | Z_i > 0 \sim \text{Gamma}\left(\phi, \frac{\phi}{m_i}\right) \tag{12}$$

with the expected value m_i at wind farm i, defined as

$$m_i = \exp\left[\alpha_y + \beta_y x_i\right] \tag{13}$$

Finally, α_y is the intercept and β_y the scaling parameter for x_i, which is defined in (11). The vector of the parameters to be estimated is given by

$$\boldsymbol{\theta} = (\alpha_z, \alpha_y, \beta_y, \phi, \kappa, \sigma)$$

The joint likelihood function is given by the product of the likelihood for the occurrence and the amount as

$$\begin{aligned} \text{L } (\boldsymbol{\theta}|\mathbf{z}, \mathbf{y}) &= \prod_{z_i=0} \text{Bern } (z_i|p_i) \\ &\quad \prod_{z_i=1} \text{Bern } (z_i|p_i) \text{ Gamma } (y_i|m_i, \phi) \\ &= \prod_{z_i} \text{Bern } (z_i|p_i) \text{ Gamma } (y_i|m_i, \phi)^{z_i} \\ &= \prod_{z_i} p_i{}^{z_i} (1 - p_i)^{1 - z_i} \\ &\quad \left[\frac{1}{\Gamma(\phi)} \left(\frac{\phi}{m_i}\right)^{\phi} y_i{}^{\phi - 1} \exp\left(-\phi \frac{y_i}{m_i}\right)\right]^{z_i} \end{aligned} \tag{14}$$

The binary variables z_i for $i = 1, 2, .., N$ are treated as observed variables in this model. We use the default values for the prior parameter, where a log-Gamma prior is assumed for κ and ϕ and a Normal prior with a fixed vague precision is assumed for the fixed effects α_z, α_y and β_y.

Once again, we obtain predictions for the model just described with the INLA methodology implemented in the R-INLA package to be described in Sect. 3.4.

3.4 Inference and prediction

Recall that the ultimate goal here is to obtain prediction and the corresponding uncertainty of wind power generation at unobserved locations. In this section, we explain how to do the parameter inference and obtain probabilistic prediction using the models described in Sects. 3.2 and 3.3.

When the focus is on prediction, latent Gaussian models can easily become computationally expensive as the cost of inverting dense covariance matrices increases cubically with the number of observed locations. A recent breakthrough alternative was proposed in Rue et al. (2009) and Lindgren et al. (2011). The former develops a framework for Bayesian inference in a broad class of models enjoying a latent Gaussian structure. The latter bridges a gap between Gaussian Markov random fields (GMRF) and Gaussian random fields theory and makes it possible to combine the flexibility of Gaussian random fields for modelling and the computational efficiency of GMRF for inference. The method of Lindgren et al. (2011) specifies a spatial model from a stochastic partial differential equation (SPDE) formulation instead of explicitly defining the covariance function. The key point of the SPDE approach is the finite element representation of the Matérn field that establishes the link between the Gaussian random field and the GMRF defined by the Gaussian weights to which a Markovian structure can be given. In particular, it is possible to find an explicit mapping of the Matérn covariance function of the Gaussian random field to the elements of the precision matrix Q of the GMRF with a computational cost of $O(n)$. References on the accuracy of the INLA/SPDE approach in spatial statistics, which has been widely validated, can be found in Lindgren et al. (2011), Simpson et al. (2012) and Martins et al. (2013).

Specifically, the Gaussian field $\mathbf{x}$ with Matérn covariance is a solution to the linear SPDE

$$(\kappa^2 - \Delta)^{\alpha/2} \mathbf{x} = \mathcal{W}, \quad \alpha = \nu + D/2, \kappa > 0, \nu > 0, \tag{15}$$

where D is the spatial dimension and $(\kappa^2 - \Delta)^{\alpha/2}$ is a pseudo-differential operator defined in terms of its spectral properties (Lindgren et al. 2011). The random Gaussian field $\mathbf{x}$ is then approximated by a linear combination of basis functions and random weights

$$\mathbf{x} = \sum_k \psi_k w_k^* \tag{16}$$

The random weights $\mathbf{w}^* = (w_1^*, \ldots, w_m^*)$ in (16) determine the values of the field at the vertices, and the values in the interior of the triangles are determined by linear interpolation. The full distribution of the solution to the SPDE in (15) is determined by the joint distribution of the weights in (16) (Lindgren et al. 2011).

Subsequently, the posterior estimates of parameters and hyperparameters are computed using INLA (Rue et al. 2009). INLA approximates the integral involved in the calculation of the marginal posterior distributions of the hyperparameters by Laplace approximation while the latent field is calculated using a Gaussian approximation evaluated at the mode of the posterior distribution.

We use the R-INLA package to perform inference and prediction. For more information on the package see http://www.r-inla.org.

Model fitting and prediction of the spatial random effect are done simultaneously on a grid of locations. The grid with the prediction locations, usually called *mesh*, is a partition of the region into triangles that discretizes the random field at m nodes. The *mesh* is constructed by defining the basis function $\psi_1, \ldots, \psi_m$ in (16) for every node in the triangulation so that they are equal to one in the *mesh* nodes and zero in all other nodes. The advantage of the triangulation over a regular grid is the possibility to have smaller triangles where there is the need for higher accuracy of the field representation, so the observation locations are dense, and larger triangles where there is no data and spending computational resources would be wasteful (Lindgren 2012).

Figure 3 shows the triangulation using the western Denmark data described in Sect. 2. The red dots denote the 349 wind farms in our data set. Note that the area of Denmark where we have data includes several islands. In order to construct the triangulation, we form one single polygon that represents the global boundary of the western part of Denmark (blue line at Fig 3).

We use a copula-based correction for the INLA (Ferkingstad and Rue 2015). The correction is especially useful for generalized linear mixed models that involve the Binomial or Poisson distributions, where inaccuracies in the Laplace approximation can occur because of the very low degree of smoothing in some models. Following the Bayesian framework in Rue et al. (2009), it is necessary to approximate the full joint distribution of the latent field given the parameters and the observations in order to compute the posterior marginal distributions of the parameters and the latent field given only the data. This approximation is usually done in INLA using a Gaussian approximation found by matching the mode and the curvature at the mode of the posterior marginal distribution of the latent field. However, approximations using skew normal densities based on a second Laplace approximation are shown to be more accurate than the Gaussian approximation (Rue et al. 2009). Ferkingstad and Rue (2015) shows how to use the Gaussian copula to construct an approximation to the full joint distribution that retains the dependence structure of the Gaussian approximation, while having the improved marginals from skew normal densities. The correction has been added as part of the R-INLA and adds minimally to the running time of the algorithm.

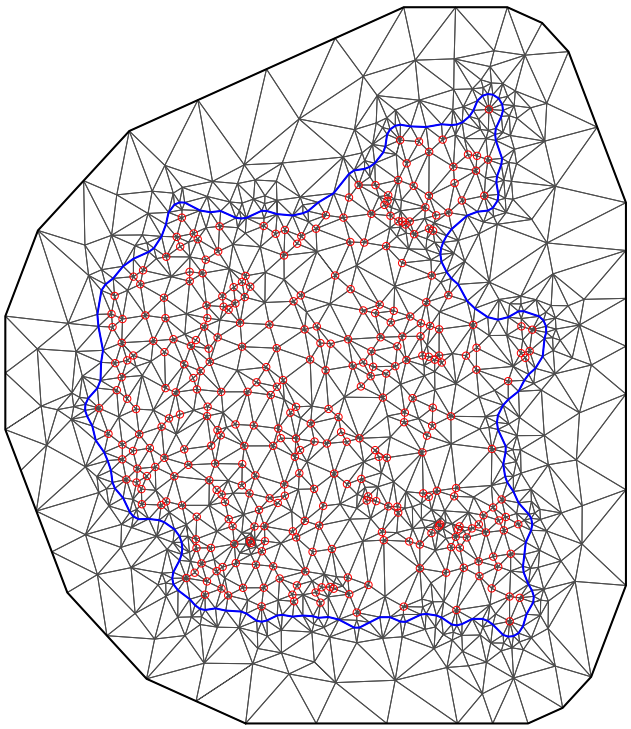


Fig. 3 The western Denmark triangulation. *Red* denotes the 349 wind farms where we have data. The *blue line* is the boundary that covers all the islands in the western part of Denmark

4 Evaluation framework

To assess the quality of the predictive distributions for the annual average and high temporal resolution of wind power generation, we use k-fold cross validation with $k = 4$ for the three methods described in Sect. 3. The idea of the k-fold cross validation is to split the data into k roughly equal-sized parts. For each split, we fit the model to the remaining $k - 1$ parts of the data and calculate the prediction error of the fitted model when predicting the k-th part of the data. We repeat this procedure 50 times to reduce sampling bias and variance. The prediction error is obtained by combining the four estimates from the 50 data

sets. Overall, 5 to 10-fold cross-validation is recommended as a good compromise between bias and variance (Breiman and Spector 1992; Kohavi 1995). However, here we choose 4-fold cross-validation since we have a large sample size and want to reduce computational costs.

The cross-validation estimate of the prediction error is measured through the continuous ranked probability score (CRPS), which is a strictly proper scoring rule for the evaluation of probabilistic forecasts of a univariate quantity (Gneiting and Raftery 2007). The CRPS is negatively oriented, i.e., the smaller the better, and is defined as

$$\text{CRPS}\ (P,x) = \int_{-\infty}^{\infty} (P(y) - I(y \geq x))^2 dy \tag{17}$$

where P is the cumulative distribution function of the density forecast and x is the normalized observed wind power.

In addition, we obtain the point forecast for the normalized wind power at a specific location as the mean of the predictive distribution. We calculate the root mean squared error (RMSE) and bias of the point forecasts to compare prediction performances. Note that since RMSE is a quadratic loss function, the mean of the predictive distribution is the optimal point predictor (Banerjee et al. 2005). The RMSE and bias are defined as follows

$$\text{RMSE} = \sqrt{\frac{1}{NR}\sum_{i=1}^{N}\sum_{j=1}^{R}(y_{ij} - \hat{y}_{ij}^{-k(i)})^2} \tag{18}$$

$$\text{bias} = \frac{1}{NR}\sum_{i=1}^{N}\sum_{j=1}^{R}(y_{ij} - \hat{y}_{ij}^{-k(i)}) \tag{19}$$

where N is the number of data points, R is the number of replicates in the k-fold cross-validation, $k(i)$ is the function that maps the observation i to the group k and $\hat{y}_{ij}^{-k(i)}$ is the mean of the posterior predictive distribution when the model was fitted with the data set excluding the $k(i)$ part.

Similarly, using the median of the density forecasts as the point forecast, we compute the mean absolute error (MAE). Here, the point forecast that minimizes the MAE is the median of the predictive distribution, since MAE is a symmetric linear function in contrast to the RMSE (Fraley et al. 2010; Pinson and Hagedorn 2012).

$$\text{MAE} = \frac{1}{NR}\sum_{i=1}^{N}\sum_{j=1}^{R}|y_{ij} - \hat{y}_{ij}^{-k(i)}| \tag{20}$$

To formally test for a significant difference between the predictions made for two spatial fields, we apply the spatial prediction comparison test (SPCT) introduced in Hering and Genton (2011). The null hypothesis to be tested is that of equal predictive ability on average in terms of a loss function, such as absolute differences, squared differences, as well as skill score functions. Gilleland (2013) proposed two new loss functions to the SPCT, the first is based on distance maps and the second on image warping.

The SPCT yields a statistical test that accounts for spatial correlation without imposing any assumption on the underlying data or on the resulting loss differential field. The loss differential field is a field giving the straight difference between the two loss functions calculated for each of two forecasts. We are interested in the average of the loss differential field, which is asymptotically Normal distributed (Hering and Genton 2011). Because the loss differential field is likely to have a strong spatial correlation, the standard error for its mean is calculated from the variogram. Here, we use an isotropic exponential variogram to fit the data. Moreover, if a trend on the data is suspected, it should be removed before applying the SPCT, since the mis-specification of the trend can result in a test for prediction comparison that is undersized or oversized. We perform ordinary least squares (OLS) to estimate and, if necessary, remove the trend from our data. Based on the large p-values of the regression coefficients in the OLS that were obtained when we fitted the data from the annual average power output, we conclude that there is no significant trend to be removed in this case.

The SPCT is implemented in the R package *SpatialVx*, which is used here to obtain p-values for the difference in RMSE, CRPS, bias and MAE between the models described in Sect. 3.

We assess the predictive performance with reliability and sharpness diagrams. Reliability represents the ability of the forecasting system to match the observation frequencies. Ideally, the nominal coverage rates, which is the proportion of times that the cumulative distribution of a random variable is below a threshold, and the observed frequencies would be the same, resulting in points aligning with the diagonal. For example, for a nominal coverage rate $\alpha = 0.05$, it is expected that 5 % of the observations are smaller than the predictive quantile at nominal level 0.05. However, a reliable forecast is not useful if it is not informative. The sharpness diagram gives an indication of the spread of the predictive distributions. It is measured by the average interval size in the case of predictive intervals, which should be as tight as possible for a sharp forecast (Gneiting et al. 2007).

In order to construct reliability diagrams, we start by introducing an indicator variable $\mathcal{I}_{i,j}^{(\alpha)}$, which is defined for a quantile forecast $\hat{p}_{i,j}^{(\alpha)}$ made at wind farm i and replicate j with observed value p_{ij} as follows

$$\mathcal{I}_{i,j}^{(\alpha)} = \begin{cases} 1 & \text{if } p_{ij} \leq \hat{p}_{i,j}^{(\alpha)} \\ 0, & \text{otherwise} \end{cases}$$

The indicator variable $\mathcal{I}_{i,j}^{(\alpha)}$ tells whether the actual outcome lies below the quantile forecast (hits) or not (miss). Next, denote $n_1^{(\alpha)}$ the sum of hits and $n_0^{(\alpha)}$ the sum of misses over all the realizations

$$n_1^{(\alpha)} = \sum_{i=1}^{N}\sum_{j=1}^{R} \mathcal{I}_{i,j}{}^{(\alpha)} \quad \text{and} \quad n_0^{(\alpha)} = N - n_1^{(\alpha)}$$

An estimation $\hat{a}^{(\alpha)}$ of the actual coverage $a^{(\alpha)}$ is then obtained by calculating the mean of $\mathcal{I}_{i,j}{}^{(\alpha)}$ over the N wind farms and R replicates in the validation set.

$$\hat{a}^{(\alpha)} = \frac{1}{NR}\sum_{i=1}^{N}\sum_{j=1}^{R} \mathcal{I}_{i,j}{}^{(\alpha)} = \frac{n_1^{(\alpha)}}{n_1^{(\alpha)} + n_0^{(\alpha)}} \tag{21}$$

This approach to the evaluation of prediction intervals was proposed in Baillie and Bollerslev (1992) and McNees and Fine (1995). Consistency bars for the reliability diagram with 95 % confidence level are computed by simulating a Binomial distribution with parameters N equal to the corresponding number of wind farms, and p, the probability of success, equal to the nominal level α.

In the sharpness diagram, for each nominal quantile and each station, we can obtain the length of the central predicted interval. Central predictive intervals are centered in probability around the median. For example, the value at the 60 % nominal coverage is the predicted value at the 80 % quantile minus the predicted value at the 20 % quantile. The final predicted interval size is the average size of the predicted intervals over all the wind farms and replicates as follows

$$\bar{\delta}^{(\alpha)} = \frac{1}{NR}\sum_{i=1}^{N}\sum_{j=1}^{R} \delta_{i,j}^{(\alpha)} = \frac{1}{NR}\sum_{i=1}^{N}\sum_{j=1}^{R} \left(\hat{p}_{i,j}^{(1-\alpha/2)} - \hat{p}_{i,j}^{(\alpha/2)}\right) \tag{22}$$

We complete this section by explaining how to assess the reliability of the induced probability forecasts for the occurrence of wind power in the model shown in Sect. 3.3. We compute the observed relative frequencies in 21 equally spaced bins between 0 and 1. For each predicted value, it is established which of the bins the value falls into. The corresponding observed relative frequency is the number of times the event happens, given that the predicted probability belongs to a specific bin, divided by the total number of predicted values in that bin. Each bin is represented by a single forecast probability, which is chosen to be the average of the predicted values over the bin. Bröcker and Smith (2007) mentions the advantages of considering the average instead of the popular choice of the arithmetic center of the bin.

5 Results

We now present verification results for the probabilistic prediction of annual average and high temporal resolution of wind power generation after k-fold cross-validation based on the evaluation framework described in Sect. 4.

5.1 Verification results for annual average wind power generation

In this section, we show results obtained from modelling the annual average wind power generation in the year of 2010, where average wind power is obtained by averaging the power output at each of the 349 wind farms in the western part of Denmark. We compare the predictions obtained with the Beta model with covariates fitted with the INLA/SPDE approach described in Sect. 3.2 with kriging in Sect. 3.1. Table 1 shows summary measures of predictive performance for annual average wind power generation, which are fully described in Sect. 4. The large p-values in this table indicate that we do not reject the hypothesis of equal predictive ability on average in terms of RMSE, CRPS, Bias and MAE between the Beta model with covariate fitted with the INLA/SPDE approach and kriging. This is not surprising given that here, since the data is averaged over an entire year, the individual noises are smoothed out and the distribution becomes closer to Gaussian.

We assess reliability and sharpness of the predicted annual average wind power generation through the diagrams in Fig 4. In this figure we can see a comparison between the Beta model with covariates fitted with the INLA/SPDE approach and kriging of the reliability diagram together with the respective consistency bars (left) and sharpness diagram (right). The Beta model tends to overestimate the annual average wind power generation for quantiles larger than 0.25, which results in a reliability curve below the diagonal. However, this model also has a considerably smaller prediction interval, as shown in the right plot of the same figure. We remark that the consistency bars in this figure are constructed without considering the replicates of the k-fold validation as part of the total number of observations, since they are not independent realizations of the process. The method of combining replicates of predictions when building consistency bars for

Table 1 Root mean square error (RMSE), continuous ranked probability score (CRPS), bias and mean absolute error (MAE) with respect to the maximum capacity (in percentage) for annual average wind power generation using kriging and the Beta model with covariate.

	Kriging	Beta model	p-values
RMSE	7.518	7.377	0.97
CRPS	4.235	4.319	0.934
Bias	0.012	0.164	0.923
MAE	5.547	5.665	0.936

The last column shows the p-values for the differences in RMSE, CRPS, bias and MAE between kriging and the Beta model

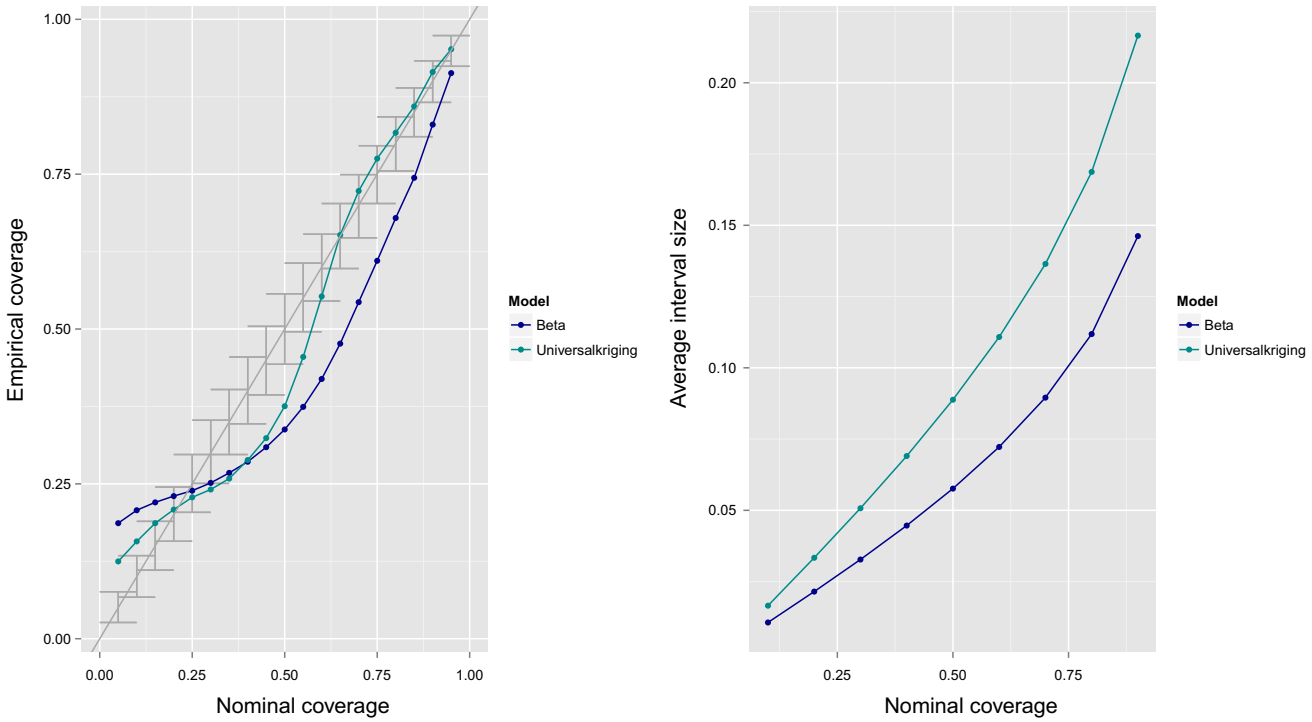


Fig. 4 Reliability diagram with respective consistency bars (*left*) and sharpness diagrams (*right*) for the annual average wind power generated in 2010

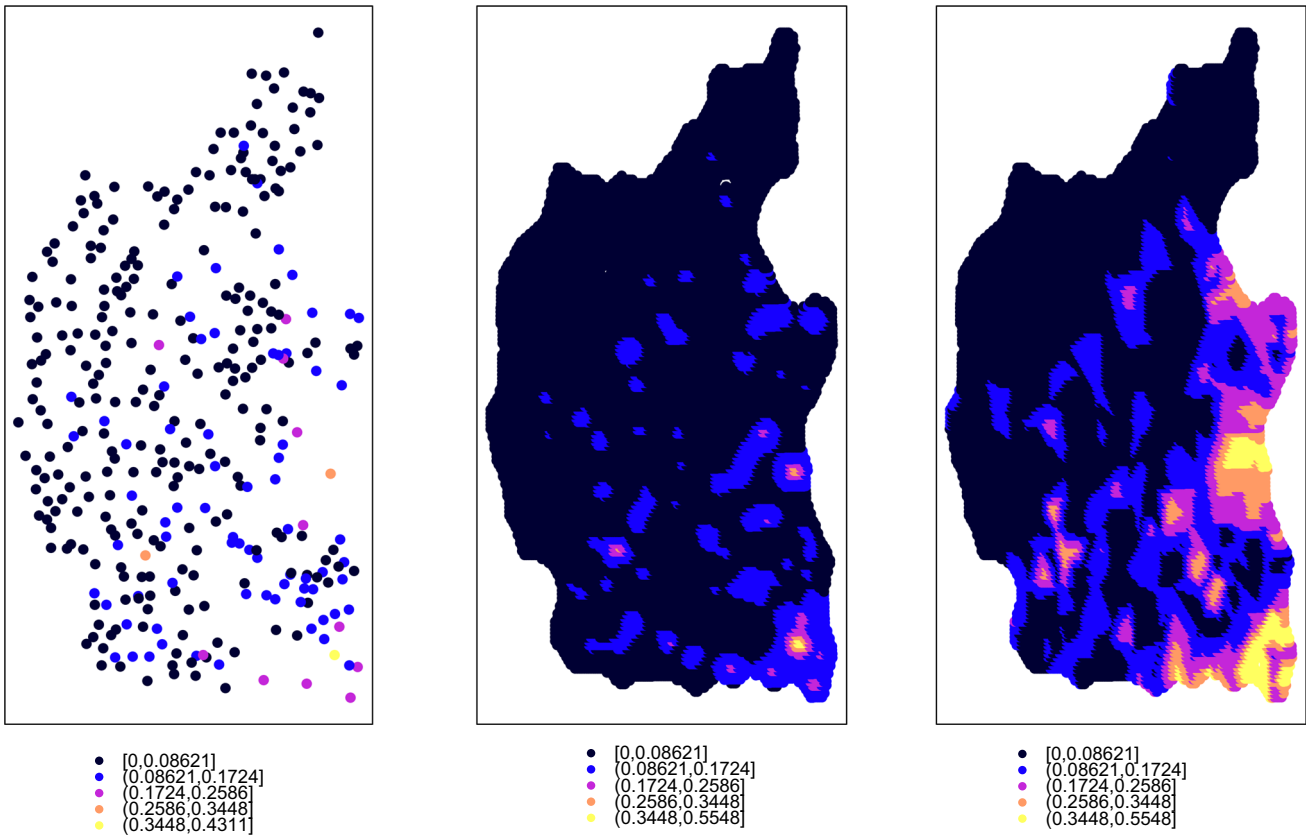


Fig. 5 Example of normalized wind power data generated at a fixed 15 min interval from 2010 (*left*). Predicted mean of the normalized wind power obtained with kriging (*middle*) and with the hierarchical hurdle Gamma model (*right*)

the reliability diagram should be investigated further to have consistency bars of correct size in Fig. 4.

In summary, the results for modelling the annual average of wind power generation show that kriging should be preferred over the Beta model with covariates fitted with the INLA/SPDE approach. Both methods present similar results, while kriging is easier to set up and has lower computational cost. While kriging takes approximately 0.05 s to fit and get point predictions at one replication and 1-fold in a single processor Intel Core i7-4600U/2.10 GHz machine, the INLA method takes approximately 20.08 s to get inference and density of the predictions in the same machine.

5.2 Verification results for high temporal resolution of wind power generation

We now present verification results for predicting wind power at a high temporal resolution, considering data from 2010 in the western part of Denmark. We fit the model separately to each of the 987 time steps spread over 12 days, each day belonging to a different month in the 2010 calendar year. Then, we calculate the average of the results and scores from the individual prediction cases. Here, we compare the models described in Sects. 3.3 and 3.1.

Figures 5 and 6 show an example of normalized wind power data at high temporal resolution in 2010, together with the predictive maps of the mean and the standard deviation for the western part of Denmark. In Fig 5, the left plot is the observed wind power after normalization, the middle plot corresponds to the predicted mean obtained with kriging and the plot on the right corresponds to the hierarchical hurdle gamma model fitted with the INLA/SPDE approach. From Fig 5, we can see that the Gamma model is able to predict larger mean values of normalized wind power than the kriging. The predicted standard deviations are shown in Fig 6. In this figure, the left plot is produced with kriging and the plot on the right corresponds to the hierarchical hurdle Gamma model fitted with the INLA/SPDE approach. While the standard deviation from the Gamma model is larger where the predicted mean value is also larger, the kriging has a large standard deviation everywhere except where the observations are placed.

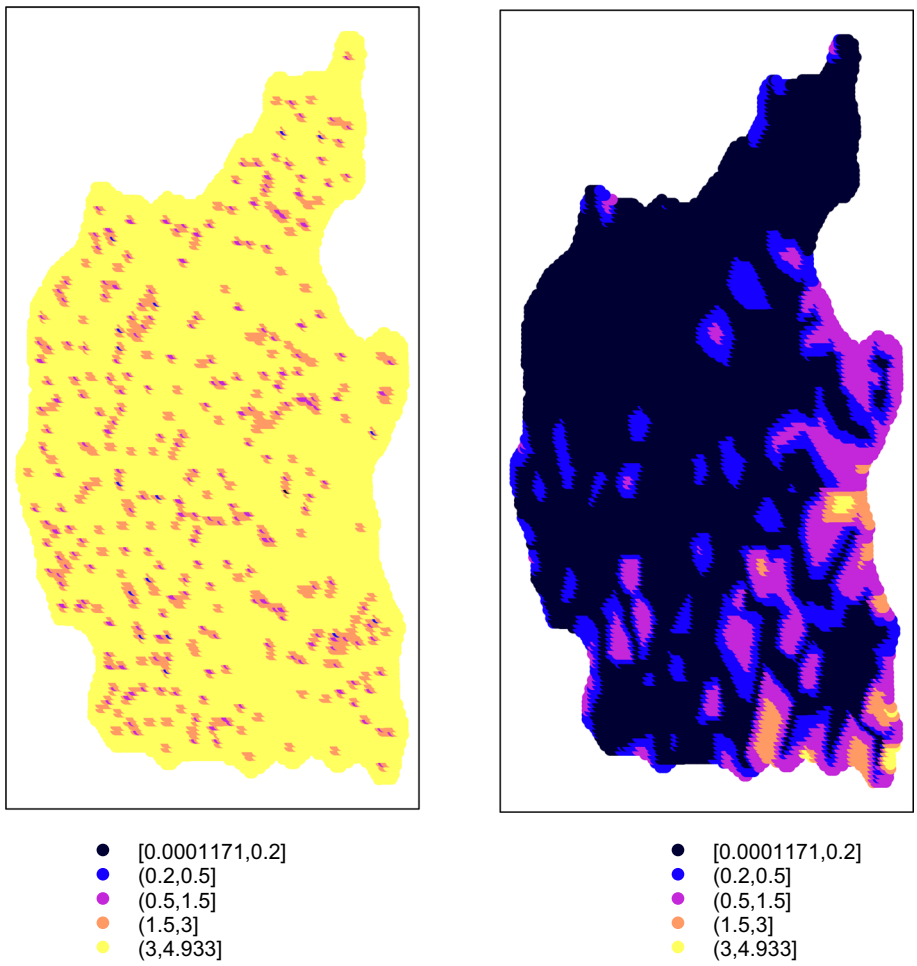


Fig. 6 Predicted standard deviation of the normalized wind power obtained with kriging (*left*) and with the hierarchical hurdle Gamma model (*right*)

Table 2 Root mean square error (RMSE), continuous ranked probability score (CRPS), bias and mean absolute error (MAE) with respect to the maximum capacity (in percentage) of wind power generation at high temporal resolution when the wind power output is greater than zero using kriging and the hierarchical hurdle Gamma model

	Kriging	Gamma	*p*-values
RMSE	13.689	7.214	0.05
CRPS	5.827	2.455	0.039
Bias	0.56	−0.222	0.201
MAE	8.665	2.917	0.044

The last column shows the *p*-values for the differences in RMSE, CRPS, bias and MAE between kriging and the Gamma model

We start by presenting the verification results for the observed measurements that are greater than zero. In Table 2, we compare summary measures of predictive performance for wind power greater than zero obtained using the hurdle gamma model fitted with the INLA/SPDE approach which is described in Sect. 3.3 and kriging described in in Sect. 3.1. The hierarchical hurdle Gamma model produces significantly better predictions on average in terms of RMSE, CRPS and MAE than kriging. The superiority of the hierarchical spatial method over the kriging may stem from a lack of flexibility of the latter, as it does not consider the point mass at zero in the wind power distribution and is optimal only when data is Gaussian distributed. In contrast, the hurdle Gamma model attempts to accommodate wind occurrences with a Bernoulli distribution and wind power magnitude using the Gamma distribution, where a shared latent process is included to handle spatial correlation between wind farms in both distributions.

On the other hand, kriging is considerably faster than the model fitted with the INLA/SPDE approach. It takes approximately 0.06 s to estimate one replication and 1 fold, while the hurdle Gamma model fitted with INLA takes approximately 44.4 s to get inference and density of the predictions in a single processor Intel Core i7-4600U/2.10 GHz machine. The trade-off between a method that offers more accurate predictions with a sharper predictive density and a method that is simpler to set up and requires less computational effort will most likely depend on the type of application.

The plots in Fig. 7 show a comparison between the hierarchical hurdle Gamma model and kriging in terms of reliability (left) and sharpness (right) for a power output greater than zero. Kriging has a curve close to the diagonal, while the hierarchical hurdle Gamma model has a sigmoid-shaped curve, which could be due to a violation of the model assumptions. In this scenario, we do not need consistency bars because of the large sample size, since we have 349 wind farms and 50 replicates for each of the 987 time steps.

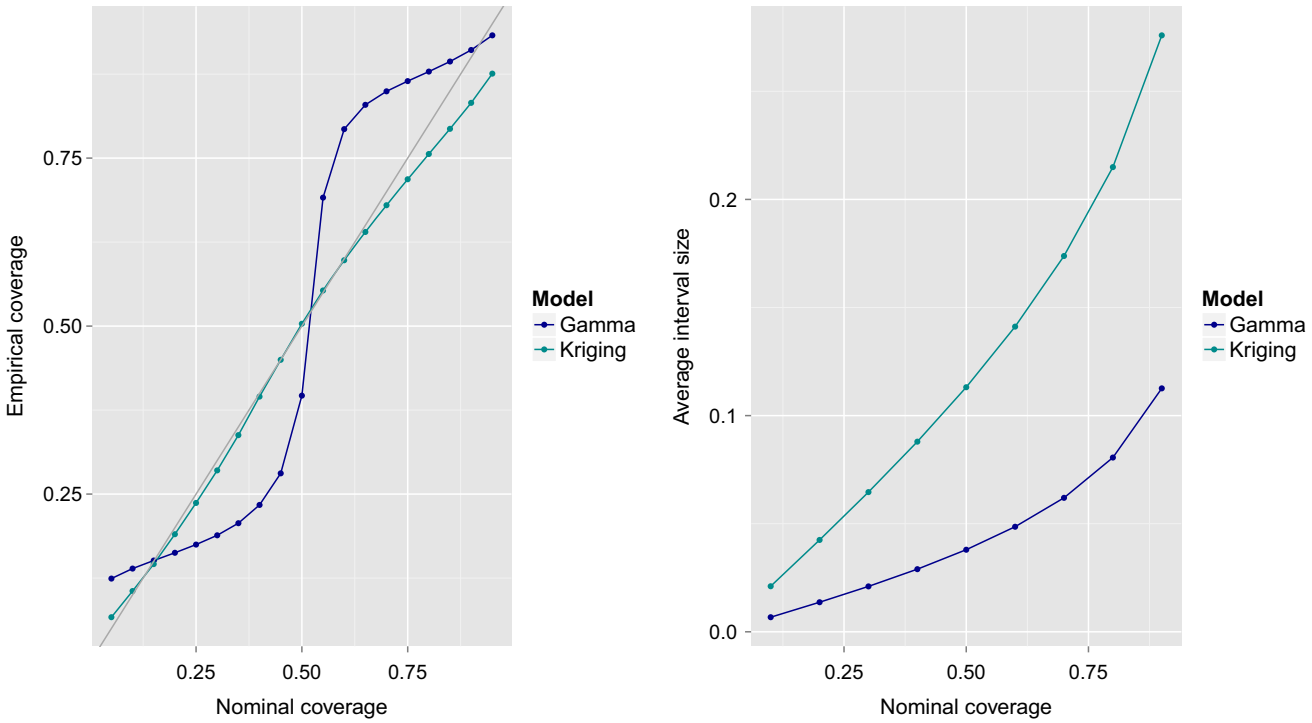


Fig. 7 Reliability (*left*) and sharpness (*right*) diagrams for wind power generation at high temporal resolution when the wind power output is greater than zero using kriging and the hierarchical hurdle Gamma model

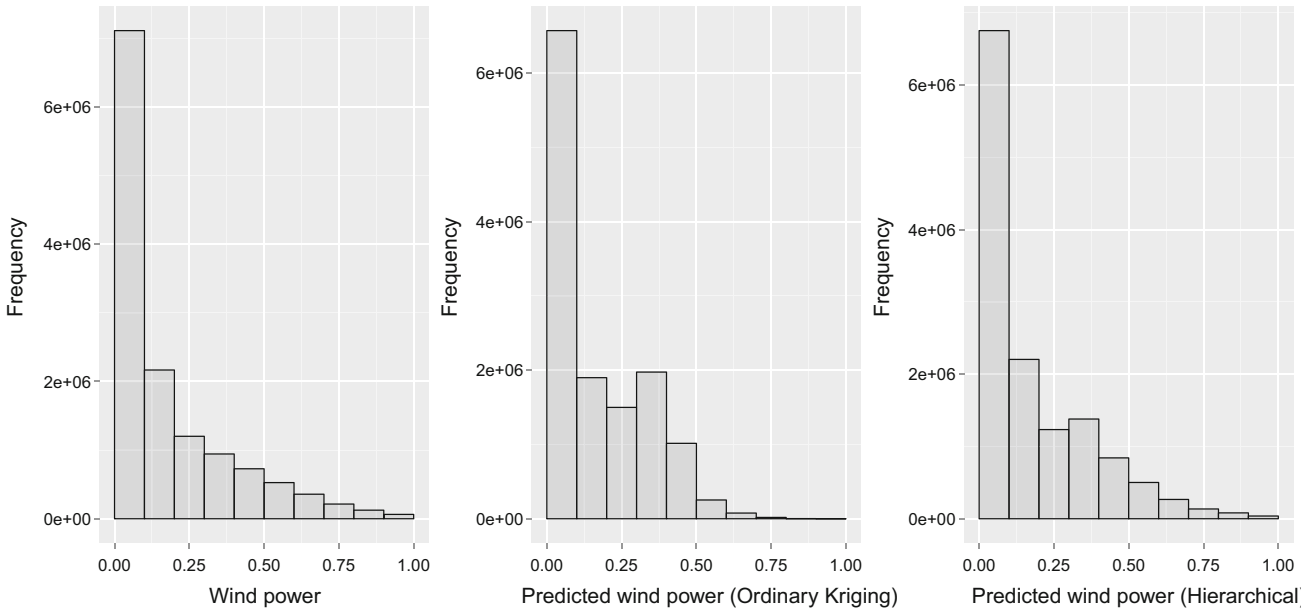


Fig. 8 Histograms of the normalized wind power greater than zero (*left*), of the predicted wind power using kriging (*middle*) and of the predicted wind power using the hierarchical hurdle Gamma model (*right*)

Figure 8 shows the histogram for the wind power measurements greater than zero together with their predictions, which is the mean value of the posterior distribution, using the two methods of comparison: kriging and the hierarchical hurdle Gamma model. The positive values of wind power are clearly not Gaussian distributed. It can also be observed that the kriging method (second plot) predicts less values close to one in comparison to the hurdle Gamma model (third plot) and in comparison with the true distribution (first plot). In fact, the proportions of observations, kriging predictions and Gamma model predictions that exceed 0.75 are 0.0207, 0.0006 and 0.0131 respectively.

Recall that the hierarchical hurdle Gamma model fitted with the INLA/SPDE approach additionally gives predictions of the Bernoulli-distributed random variable that maps the occurrence of wind power generation. We assess the reliability of the probability predictions for the occurrence of wind power generation through the diagram in the bottom line of Fig 9. This plot shows the empirically observed frequency of wind power occurrence as a function of the binned forecast probability. The actual observed relative frequency is well approximated by the forecast probability, as the line in this plot lies close to the diagonal. The top plots in the same figure correspond to histograms of the empirical probability (left) and the predicted probability (right) of wind power occurrence. As we can see from the left plot, there are almost five-times as many time steps with generated power greater than zero than equal to zero. The histogram of the predicted probabilities on the right side shows the same tendencies, since most of the estimated probabilities of wind power occurrence are close to one.

6 Discussions

We have presented statistical methods for obtaining probabilistic predictions of wind power generation at annual average as well as high temporal resolution (15 min averages) and we have compared the results from these methods with the benchmark kriging. In the first scenario, at any individual wind farm, the distribution of the annual average wind power generation is modelled with a Beta distribution, where the distance to the west coast of Denmark is used as a covariate and the spatial dependence between different locations is captured by a spatial Gaussian process with Matérn covariance. The second scenario builds on a hierarchical hurdle Gamma model, which is similar to well-known models for precipitation such as Berrocal et al. (2008) and Baxevani and Lennartsson (2015) in the sense that it is also a two-stage model with a Gaussian field to account for spatial correlation. However, our approach introduces a new Bernoulli-distributed random variable to account for the probability of wind power occurrence. The parameter p of this distribution depends on the Gaussian field and can be fully estimated. The continuous part of our two-stage model has a Gamma distribution where the mean depends on the same Gaussian field that is used for modelling p.

To perform inference and prediction, instead of using MCMC, we have used the novel INLA approach. We have

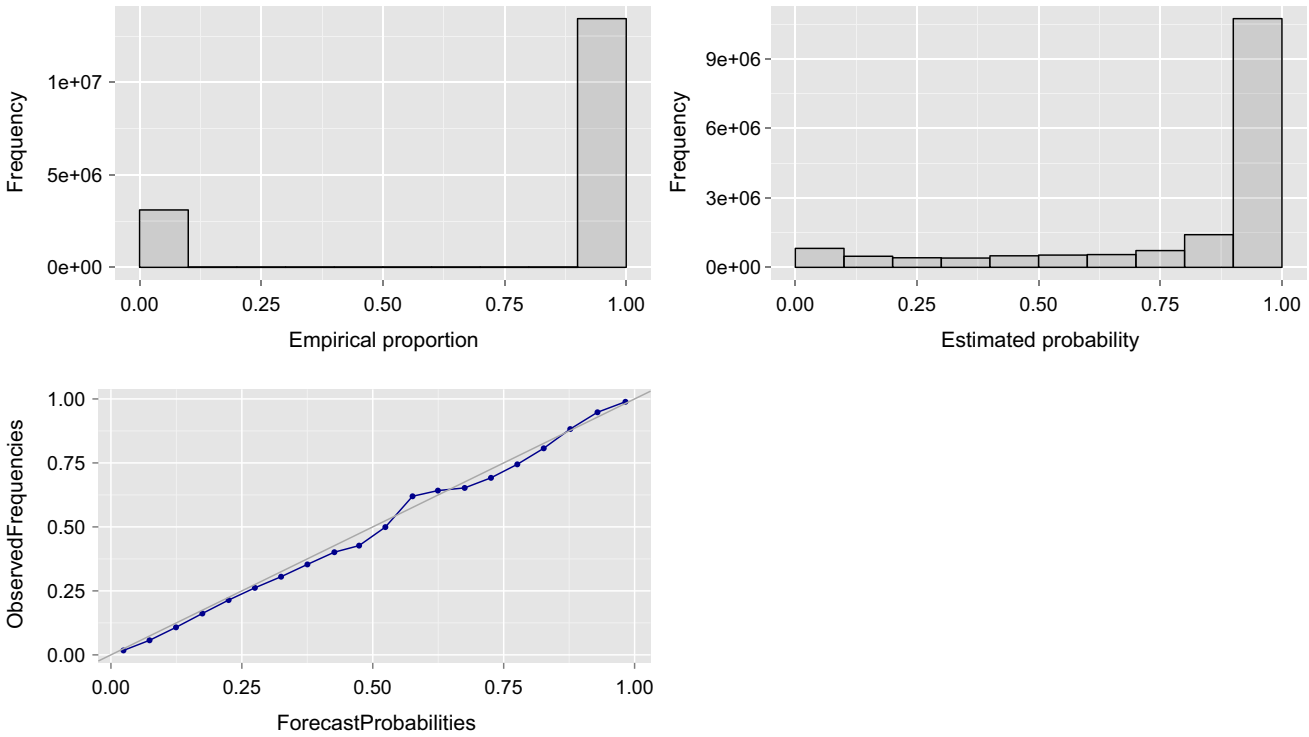


Fig. 9 *First line* Histograms of the empirical probability (*left*) and the predicted probability (*right*) of wind power occurrence using the hierarchical hurdle Gamma model. *Second line* Reliability diagram for the probability of wind power occurrence using the hierarchical hurdle Gamma model

shown that complex hierarchical spatial models are well suited to wind power data and can be implemented seamlessly under the SPDE approach that is implemented in the R-INLA library providing results in reasonable computational time.

We show results from case studies on the probabilistic prediction of annual average and high temporal resolution of wind power generation with wind farms from the western part of Denmark. While the Beta model approach showed similar results to the benchmark method, the hierarchical hurdle Gamma model resulted in predictive distributions that consistently outperformed the benchmark model in terms of the validation measures CRPS, RMSE, bias and MAE. The predictive distributions obtained from the hurdle Gamma model have increased sharpness, resulting in less reliability as compared to the kriging method. Therefore, a more accurate method for generating consistency bars in the reliability diagram should be considered to draw more solid conclusions about the reliability of the predictions from our method.

The models presented in this work could also be extended to the spatiotemporal domain by incorporating an extra term for the temporal effect such as an autoregressive component or by the introduction of a non-separable spatiotemporal structure. This would come with a computational cost, which would have to be assessed.

Acknowledgments The data used here was kindly provided by Energinet.dk (system operator in Denmark) which is hereby acknowledged, then quality was checked and prepared by Robin Girard at Mines Paristech, France. The authors also thank the Danish Strategic Council for Strategic Research through the project 5s—Future Electricity Markets (No. 12-132636/DSF), the Danish e-Infrastructure Cooperation, DeIC and CAPES for support. Moreover, we thank the Associated Editor and the two reviewers who provided valuable comments.

References

Atkinson PM, Lloyd CD (1998) Mapping precipitation in Switzerland with ordinary and indicator kriging. special issue: spatial interpolation comparison 97. J Geogr Info Decis Anal 2(1–2):72–86

Baillie RT, Bollerslev T (1992) Prediction in dynamic models with time-dependent conditional variances. J Econ 52(1):91–113

Banerjee A, Guo X, Wang H (2005) On the optimality of conditional expectation as a bregman predictor. IEEE Trans Info Theory 51(7):2664–2669

Baxevani A, Lennartsson J (2015) A spatiotemporal precipitation generator based on a censored latent gaussian field. Water Resour Res 51(6):4338–4358

Berrocal VJ, Raftery AE, Gneiting T (2008) Probabilistic quantitative precipitation field forecasting using a two-stage spatial model. Ann Appl Stat 2(4):1170–1193
Bossanyi E (1985) Short-term wind prediction using Kalman filters. Wind Eng 9(1):1–8
Breiman L, Spector P (1992) Submodel selection and evaluation in regression. the x-random case. Inter Stat Rev/Revue Internationale de Statistique 60(3):291–319
Bröcker J, Smith LA (2007) Increasing the reliability of reliability diagrams. Weather Forecasting 22(3):651–661
Brown BG, Katz RW, Murphy AH (1984) Time series models to simulate and forecast wind speed and wind power. J Climate Appl Meteorol 23(8):1184–1195
Cameletti M, Lindgren F, Simpson D, Rue H (2013) Spatio-temporal modeling of particulate matter concentration through the SPDE approach. AStA Adv Stat Anal 97(2):109–131
Cellura M, Cirrincione G, Marvuglia A, Miraoui A (2008) Wind speed spatial estimation for energy planning in Sicily: a neural kriging application. Renew Energy 33(6):1251–1266
Cressie N (1985) Fitting variogram models by weighted least squares. J Inter Assoc Math Geol 17(5):563–586
Cressie N (1988) Spatial prediction and ordinary kriging. Math Geol 20(4):405–421
Daniel A, Chen A (1991) Stochastic simulation and forecasting of hourly average wind speed sequences in Jamaica. Solar Energy 46(1):1–11
Deutsch CV, Journel AG (1992) Geostatistical software library and users guide. Oxford University Press, New York
Diggle PJ, Tawn J, Moyeed R (1998) Model-based geostatistics. J R Stat Soc 47(3):299–350
Etienne C, Lehmann A, Goyette S, Lopez-Moreno JI, Beniston M (2010) Spatial predictions of extreme wind speeds over Switzerland using generalized additive models. J Appl Meteorol Climatol 49(9):1956–1970
Ferkingstad E, Rue H (2015) Improving the INLA approach for approximate Bayesian inference for latent Gaussian models. arXiv preprint arXiv:1503.07307
Fraley C, Raftery AE, Gneiting T (2010) Calibrating multimodel forecast ensembles with exchangeable and missing members using Bayesian model averaging. Mon Weather Rev 138(1):190–202
Gilleland E (2013) Testing competing precipitation forecasts accurately and efficiently: the spatial prediction comparison test. Mon Weather Rev 141(1):340–355
Gneiting T, Balabdaoui F, Raftery AE (2007) Probabilistic forecasts, calibration and sharpness. J R Stat Soc 69(2):243–268
Gneiting T, Larson K, Westrick K, Genton MG, Aldrich E (2006) Calibrated probabilistic forecasting at the stateline wind energy center: the regime-switching space-time method. J Am Stat Assoc 101(475):968–979
Gneiting T, Raftery AE (2007) Strictly proper scoring rules, prediction, and estimation. J Am Stat Assoc 102(477):359–378
Guillot G, Jónsson H, Hinge A, Manchih N, Orlando L (2016) Accurate continuous geographic assignment from low-to high-density snp data. Bioinformatics 32(7):1106–1108
Gurmu S (1997) Semi-parametric estimation of hurdle regression models with an application to Medicaid utilization. J Appl Econ 12(3):225–242
Guttorp P, Gneiting T (2006) Studies in the history of probability and statistics XLIX on the Matérn correlation family. Biometrika 93(4):989–995
Hall DB (2000) Zero-inflated Poisson and binomial regression with random effects: a case study. Biometrics 56(4):1030–1039
Hering AS, Genton MG (2011) Comparing spatial predictions. Technometrics 53(4):414–425
Hossain J, Sinha V, Kishore V (2011) A GIS based assessment of potential for windfarms in India. Renew Energy 36(12):3257–3267
Ignaccolo R, Mateu J, Giraldo R (2014) Kriging with external drift for functional data for air quality monitoring. Stoch Env Res Risk Assess 28(5):1171–1186
Joyner TA, Friedland CJ, Rohli RV, Treviño AM, Massarra C, Paulus G (2015) Cross-correlation modeling of European windstorms: a cokriging approach for optimizing surface wind estimates. Spat Stat 13:62–75
Katzenstein W, Fertig E, Apt J (2010) The variability of interconnected wind plants. Energy Policy 38(8):4400–4410
Kleiber W, Katz RW, Rajagopalan B (2012) Daily spatiotemporal precipitation simulation using latent and transformed gaussian processes. Water Resour Res 48(1):1–17
Kohavi R et al (1995) A study of cross-validation and bootstrap for accuracy estimation and model selection. Ijcai 14:1137–1145
Lindgren F (2012) Continuous domain spatial models in r-inla. ISBA Bull 19(4):14–20
Lindgren F, Rue H, Lindström J (2011) An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. J R Stat Soc 73(4):423–498
Luo W, Taylor M, Parker S (2008) A comparison of spatial interpolation methods to estimate continuous wind speed surfaces using irregularly distributed data from England and Wales. Inter J Climatol 28(7):947–959
Martins TG, Simpson D, Lindgren F, Rue H (2013) Bayesian computing with INLA: new features. Comput Stat Data Anal 67:68–83
Matheron G (1963) Principles of geostatistics. Econ Geol 58(8):1246–1266
McNees S, Fine L (1995) Forecast uncertainty: can it be measured. Federal Reserve Bank of New York Discussion Paper
Mullahy J (1986) Specification and testing of some modified count data models. J Econ 33(3):341–365
Munoz F, Pennino MG, Conesa D, López-Quílez A, Bellido JM (2013) Estimation and prediction of the spatial occurrence of fish species using Bayesian latent Gaussian models. Stoch Env Res Risk Assess 27(5):1171–1180
Musenge E, Chirwa TF, Kahn K, Vounatsou P (2013) Bayesian analysis of zero inflated spatiotemporal HIV/TB child mortality data through the INLA and SPDE approaches: applied to data observed between 1992 and 2010 in rural North East South Africa. Inter J Appl Earth Obs Geoinf 22:86–98
Pinson P (2012) Very-short-term probabilistic forecasting of wind power with generalized logit-normal distributions. Journal R Stat Soc 61(4):555–576
Pinson P, Chevallier C, Kariniotakis GN (2007) Trading wind generation from short-term probabilistic forecasts of wind power. Power Syst IEEE Trans 22(3):1148–1156
Pinson P, Hagedorn R (2012) Verification of the ECMWF ensemble forecasts of wind speed against analyses and observations. Meteorol Appl 19(4):484–500
Pohlmeier W, Ulrich V (1995) An econometric model of the two-part decision making process in the demand for health care. J Human Res. 30:339–361
Rue H, Martino S, Chopin N (2009) Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. J R Stat Soc 71(2):319–392
Serra L, Saez M, Juan P, Varga D, Mateu J (2014) A spatio-temporal Poisson hurdle point process to model wildfires. Stoch Env Res Risk Assess 28(7):1671–1684
Shih DCF (2008) Wind characterization and potential assessment using spectral analysis. Stoch Env Res Risk Assess 22(2):247–256
Simpson D, Lindgren F, Rue H (2012) Think continuous: markovian gaussian models in spatial statistics. Spat Stat 1:16–29
Sliz-Szkliniarz B, Vogt J (2011) GIS-based approach for the evaluation of wind energy potential: a case study for the

Kujawsko-Pomorskie Voivodeship. Renew Sustain Energy Rev 15(3):1696–1707

Stein M (1999) Spatial interpolation, some theory for kriging. Spring-Verlag, New York

Stein ML (1992) Prediction and inference for truncated spatial data. J Comput Gr Stat 1(1):91–110

Tastu J, Pinson P, Kotwa E, Madsen H, Nielsen HA (2011) Spatio-temporal analysis and modeling of short-term wind power forecast errors. Wind Energy 14(1):43–60

Wilks DS (1990) Maximum likelihood estimation for the gamma distribution using data containing zeros. J Clim 3(12):1495–1501

Appendix B

Benefits of spatio-temporal modelling for short term wind power forecasting at both individual and aggregated levels

---

Lenzi, A., Steinsland, I., Pinson, P. (2017). Benefits of spatio-temporal modelling for short term wind power forecasting at both individual and aggregated levels. Submitted to *Environmetrics* (under review)

Environmetrics 00, 1–25

DOI: 10.1002/env.XXXX

Benefits of spatio-temporal modelling for short term wind power forecasting at both individual and aggregated levels

Amanda Lenzi[a]* and Ingelin Steinsland[b] and Pierre Pinson[c]

Summary: The share of wind energy in total installed power capacity has grown rapidly in recent years around the world. Producing accurate and reliable forecasts of wind power production, together with a quantification of the uncertainty, is essential to optimally integrate wind energy into power systems. We build spatio-temporal models for wind power generation and obtain full probabilistic forecasts from 15 minutes to 5 hours ahead. Detailed analysis of the forecast performances on the individual wind farms and aggregated wind power are provided. We show that it is possible to improve the results of forecasting aggregated wind power by utilizing spatio-temporal correlations among individual wind farms. Furthermore, spatio-temporal models have the advantage of being able to produce spatially out-of-sample forecasts. We evaluate the predictions on a data set from wind farms in western Denmark and compare the spatio-temporal model with an autoregressive model containing a common autoregressive parameter for all wind farms, identifying the specific cases when it is important to have a spatio-temporal model instead of a temporal one. This case study demonstrates that it is possible to obtain fast and accurate forecasts of wind power generation at wind farms where data is available, but also at a larger portfolio including wind farms at new locations.
The results and the methodologies are relevant for wind power forecasts across the globe as well as for spatial-temporal modelling in general.

Keywords: wind power; aggregated forecast; probabilistic forecast; integrated nested Laplace approximation.

[a] *Applied Mathematics and Computer Science Department, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark*
[b] *Department of mathematical sciences, Norwegian University of Science and Technology, N-7491 Trondheim, Norway*
[c] *Electrical Engineering Department, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark*
* *Correspondence to: Applied Mathematics and Computer Science Department, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark. E-mail: amle@dtu.dk*

This paper has been submitted for consideration for publication in *Environmetrics*

1. INTRODUCTION

Wind power is a clean, renewable and widely available source of energy and electricity generated from wind power is increasing world wide. A challenge for utilizing wind power is that the generated amount of energy varies much and relatively fast over time due to variations in wind. An important tool for efficiently integrating wind power in a system with energy sources that can be controlled, e.g. thermal energy and hydro power, is high quality probabilistic forecasts for short term wind power production Ackermann [2005]. Moreover, accurate forecasting of wind power generation makes wind more competitive in the energy market, since it reduces the imbalance costs to producers Girard et al. [2013]. Recently, there has been an increasing amount of research in wind speed and wind power forecasts. Most of the developments are for point forecasts (e.g. Louka et al. [2008], Catalão et al. [2011]), i.e. the forecast consists of one value for each wind farm or location. To make better decisions one also needs to quantify the uncertainty of the forecast, and provide a probability density function (pdf) instead of a point forecast. This is called a probabilistic forecast. For a probabilistic forecast to be useful it needs to be calibrated and sharp Gneiting et al. [2007]. Calibrated refers to a forecast that is reliable: in the long term, 90% of the observed wind production should be within a 90% forecast interval, 80% of the observations within a 80% forecast interval and so forth. Sharpness refers to the spread of the predictive distribution, a sharper forecast is more concentrated and better when subject to calibration.

In recent years, more emphasis has been placed on probabilistic forecasts in order to quantify the inherent uncertainties in wind, see Pinson and Kariniotakis [2010] and Bremnes [2004]. From the point of view of a wind farm operator, probabilistic forecasts improve decision making regarding the management of the immediate regulating and spinning reserves, which is essential given the financial penalties that are incurred for deviating from the declared power levels. From the point of view of a system operator, the aggregated wind power generation over pre-defined areas is of particular importance. Some recent

contributions to the modelling and forecasting of aggregated wind power energy are Lau and McSharry [2010] and Focken et al. [2002], which do, however, not account for spatio-temporal dependencies.

To illustrate the challenge of forecasting individual and aggregated wind power simultaneously, we consider a toy example of two wind farms at one lead time and denote their forecasts X_1 and X_2 (these are random variables). The aggregated forecast for the system is $Y = X_1 + X_2$. We know from basic probability, see e.g Ross [2015], that the expected value for the system is $\mathrm{E}(Y) = \mathrm{E}(X_1) + \mathrm{E}(X_2)$ and the variance is $\mathrm{Var}(Y) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2) + 2\mathrm{Cov}(X_1, X_2)$. Hence, to obtain a forecast for the system Y we also need to model the dependency between the wind farms. This calls for a spatio-temporal model for wind power production. If the productions at the two farms in our toy example are dependent and have a positive covariance, but are assumed independent in the forecast, the variance of Y gets too small and the forecast for Y is not calibrated. Verification of multivariate probabilistic forecasts is an active field of research, for which new scores and diagnostic tools are being proposed and discussed, see, e.g., Pinson and Tastu [2013], Scheuerer and Hamill [2015], Thorarinsdottir et al. [2016] among others. A pragmatic approach is to evaluate relevant univariate probabilistic forecasts derived from the multivariate probabilistic forecast.

Wind speed, and hence wind power production, has temporal and spatial dependencies. In Section 2 we will see that this is also the case for western Denmark. Indeed, our approach of basing the forecast on recent observations relies on the temporal dependency. As demonstrated with our toy example, the spatial dependency also needs to be considered for the aggregated forecasts to be calibrated. Furthermore, borrowing information by utilizing the spatial correlation among individual wind farms has been shown to reduce the errors in point forecasts significantly Tastu et al. [2011], and has the advantage of producing models that are able to generate forecasts at locations that are not within the observation samples.

Several characteristics in a typical wind power series make it a challenging problem to generate accurate forecasts. First of all, wind power is bounded below by zero, when no turbines are operating, and above by the nominal capacity, when all turbines are generating their rated power output. In addition, wind power series are clearly non-Gaussian. In fact, the marginal distribution of wind power production data possesses tails that are heavier than the Gaussian distribution. Instead of using a classical Gaussian distribution, truncated Gaussian, censored Gaussian and generalized logit-normal distributions have been proposed to model the conditional density of wind power Gneiting et al. [2006] Pinson [2012]. Our approach is based on the logistic function, which has shown to be a suitable transformation to normalize wind power data Dowell and Pinson [2016].

We propose statistical models that yield calibrated probabilistic forecasts of wind power generation at multiple sites and lead times simultaneously. We define three different models that share the same data process, or likelihood, but differ in the process model. We start with a model consisting of a location specific intercept and an autoregressive component that captures the local variability without considering the dependency between the farms. This model is well suited for individual forecasts, but it is not calibrated for aggregated forecasts. To obtain reliable aggregated forecasts, we introduce two different models that capture the spatio-temporal features present in the data. The first has a common intercept and a spatio-temporal process, in which spatial and temporal dependency is modelled by a latent Gaussian field. The second is a combination of the previous two models, with a common intercept, an autoregressive process and a spatio-temporal term that varies in time with first order autoregressive dynamics. To meet the computational requirements the stochastic partial differential equations (SPDE) approach to spatial and temporal-spatial modelling is taken Lindgren et al. [2011] Blangiardo and Cameletti [2015], for which fast Bayesian inference can be performed using integrated nested Laplace approximations (INLA).

Moreover, we study the performance of the proposed models in forecasting wind power from

individual and aggregated farms under two different scenarios. In a first stage, we consider out-of-sample forecasts in terms of time, that is, they are obtained for wind farms inside the training set. However, there are situations where not enough data is available for all the wind farms, and even when it is available, the computational load to calculate forecasts for all of them can be very high. In those cases, it is important to have a method of forecasting that is as robust as possible, so that parameters estimated using only part of the portfolio can readily be used to forecast a larger data set, including wind farms at new locations. In such cases, temporal models that require local information for the parameter estimation cannot be used to obtain forecasts. Based on this, in a second stage, we consider spatially out-of-sample forecasts generated by the proposed spatio-temporal models. We develop and evaluate the forecasts for wind power production in western Denmark based on a data set for 349 wind farms with energy production observations every 15 minutes from 2006 to 2012.

In Section 2, we provide a short description of the wind power data that we use in our study and the data treatment. The hierarchical models used to generate probabilistic forecasts of wind power generation, as well as the framework for producing probabilistic forecasts with such models, are outlined in Section 3. In Section 4, we give details of the probabilistic forecasting scheme and outline the scores and the scenarios used for forecast evaluation. In Section 5, we show the results of a case study where we obtain spatio-temporal forecasts and spatially out-of-sample forecasts on the individual and aggregated level. Section 5 also contains the results of a simulation study, whereas conclusions of the work are drawn in Section 6.

2. DANISH WIND POWER PRODUCTION DATA

This project is based on a system of 349 wind farms in western Denmark. Observations of wind power production between January 2006 and March 2012 were provided by the

Transmission System Operator in Denmark and each measurement consists of temporal average over a 15-min time period.

The measurements at each site have been normalised by the nominal power of the corresponding wind farm, so that they are within the range [0,1]. Moreover, to avoid including long chains of zeros that come from temporary shutdown of the turbines for maintenance or missing data that are reflected as unreasonably long periods of zero wind power production, we choose to analyze only wind farms containing at most 10% of zero observations. The evaluation of the predictive performance of individual wind farms and aggregated wind power is done as % of nominal power, which is a common practice in the wind power field (e.g., Pinson [2012], Tastu et al. [2011], Dowell and Pinson [2016]).

Figure 1 (a) shows the spatial correlation of wind power production between one wind farm located in the southern part of Denmark and the remaining wind farms of the portfolio. The higher correlations come from farms that are closer, while the correlations of wind farms far from it are almost zero. Next, we check the dependency of the temporal correlation at fixed locations. Figure 1 (b) shows the mean autocorrelation function of wind power production among wind farms located in western Denmark. The autocorrelation function of the normalized wind power production at a single farm has a slow decay and on average, it drops down to zero after about 40 hours.

[Figure 1 about here.]

Wind power generated by a farm over a period of time is non-Gaussian and bounded between zero and one after the normalization. In fact, wind power distribution has a sharper peak than the Gaussian distribution and is also significantly right-skewed. In all the approaches to be described next, we apply the logit-normal transformation to the normalized wind power data following the procedure in Pinson [2012].

Let $X(\mathbf{s}, t)$ denote the normalized wind power production at location $\mathbf{s} \in \mathcal{D}_s$ and time $t \in \mathcal{D}_t$, with respective observations or measurements indicated by $x(\mathbf{s}, t)$. The logit-normal

transformation is given by

$$y(\mathbf{s},t) = \gamma(x(\mathbf{s},t)) = \ln\Big(\frac{x(\mathbf{s},t)}{1-x(\mathbf{s},t)}\Big), \quad x(\mathbf{s},t) \in (0,1), \tag{1}$$

with inverse

$$x(\mathbf{s},t) = \gamma^{-1}(y(\mathbf{s},t)) = (1+e^{-y(\mathbf{s},t)})^{-1}, \quad y(\mathbf{s},t) \in \mathbb{R}. \tag{2}$$

To represent the logit-normal transformation in the cases where measurements are equal to zero and one, we follow the approach by Lesaffre et al. [2007] for modelling outcome scores in $[0,1]$.

Moreover, to evaluate the performance of aggregated wind power forecasts, we obtain the normalized aggregated wind power at lead time h by

$$x_A(t+h) = \frac{\sum_{j=1}^{N} c_j x(\mathbf{s}_j, t+h)}{\sum_{j=1}^{N} c_j}, \tag{3}$$

where c_j is the capacity of wind farm at location $\mathbf{s}_j$ and N is the total number of wind farms in the portfolio.

3. MODELS AND FITTING SCHEME

In this section, we introduce three different statistical models for wind power production. We start with a simpler autoregressive model, where each wind farm is considered as an independent replicate of the same process. Next, we describe two versions of a spatio-temporal model, in which spatial correlation is captured by a latent Gaussian field with a Matérn covariance function. The simplest version has only a spatio-temporal component, while the other has both, an autoregressive process and a spatio-temporal model. The section ends with the estimation procedure and how we obtain probabilistic forecasts.

3.1. Likelihood

We denote by $Y(\mathbf{s},t)$ the normalized logit-normal transformed wind power generation at location $\mathbf{s}$ and time t, which is calculated using (1). We assume the following distribution for $Y(\mathbf{s},t)$ at the first level of the hierarchical models considered in this section

$$Y(\mathbf{s},t) \sim \text{Normal}\left(\mu(\mathbf{s},t), \sigma_e^2\right), \tag{4}$$

with σ_e^2 being the variance of the measurement error, defined by a Gaussian white noise process both serially and spatially uncorrelated. The term $\mu(\mathbf{s},t)$ is the mean of the random process and can be defined by other process levels giving rise to different hierarchical models that are described in the following sections.

3.2. Latent Gaussian structure

3.2.1. Temporal model (Model T)

We start with a time series model where each wind farm is considered as an independent replicate of the same random process. The independence assumption is of course a simplification, since the wind power production in one location is probably dependent on the production in other locations. We assume that $\mu(\mathbf{s},t)$, in (4), is constant in time and can be modelled as

$$\mu(\mathbf{s},t) = b(\mathbf{s}) + w_{\mathbf{s}}(t), \tag{5}$$

where $b(\mathbf{s})$ is an intercept specific for each location and $w_{\mathbf{s}}(t)$ is an autoregressive process that can be written as

$$w_{\mathbf{s}}(t) = \rho_1 w_{\mathbf{s}}(t-1) + \nu_{\mathbf{s}}(t), \tag{6}$$

with $t = 2, \ldots, T$ and $|\rho_1| < 1$. The term $\nu_{\mathbf{s}}$ is uncorrelated with $w_{\mathbf{s}}(t)$ and independent identically distributed as $\nu_{\mathbf{s}} \sim N(0, \sigma^2_\nu)$.

3.2.2. Spatio-temporal model (Model S-T)

This model is a spatio-temporal process with temporal dynamics as in Cameletti et al. [2013]. This type of model is commonly used for modelling air quality because of its flexibility in including time and space dependency, as well as the effect of covariates (see e.g. Fassò and Finazzi [2011] and Cocchi et al. [2007]). The mean function $\mu(\mathbf{s}, t)$ in (4) is given by

$$\mu(\mathbf{s}, t) = b_0 + z(\mathbf{s}, t), \tag{7}$$

where b_0 is an intercept that is common to all wind farms and constant in time and space. The term $z(\mathbf{s}, t)$ refers to a spatio-temporal process that varies in time with first order autoregressive dynamics

$$z(\mathbf{s}, t) = \rho_2 z(\mathbf{s}, t-1) + w(\mathbf{s}, t), \tag{8}$$

with $t = 2, \ldots, T$ and $|\rho_2| < 1$. Moreover, $w(s, t)$ is a zero-mean Gaussian field, assumed to be temporally independent with covariance function

$$\mathrm{Cov}(w(\mathbf{s}, t), w(\mathbf{s}', t')) = \begin{cases} \sigma^2_w C(h), & \text{if } t = t' \\ 0, & t \neq t' \end{cases}$$

for $\mathbf{s} \neq \mathbf{s}'$. The correlation function C depends on the locations $\mathbf{s}$ and $\mathbf{s}'$ through the distance $h = ||\mathbf{s} - \mathbf{s}'||$. This means that the process is assumed to be second-order stationary and isotropic (see Cressie [1992]). The marginal variance is $\mathrm{Var}(\mathbf{s}, t) = \sigma^2_w$ and $C(h)$ is the correlation function defined by the Matérn, given by

$$C(h) = \frac{1}{\Gamma(\nu)2^{\nu-1}}(\kappa h)^{\nu} K_{\nu}(\kappa h), \tag{9}$$

where K_1 is the modified Bessel function of second kind, order ν. The parameter κ can be used to select the range, while ν is a smoothness parameter determining the mean-square differentiability of the underlying process. More precisely, the range is defined to be $r = \sqrt{8\nu}/\kappa$. Although the parameter ν is fixed to 1 for computational reasons, it remains flexible enough to handle a broad class of spatial variation Rue et al. [2009]. Applications with fixed parameter ν include Ingebrigtsen et al. [2014], Cameletti et al. [2013] and Munoz et al. [2013].

3.2.3. Temporal + Spatio-temporal model (Model ST+T)

This is a model defined by an autoregressive process at each location to capture the individual variability and a spatio-temporal process with temporal dynamics to take into account the spatial dependence among wind farms. Specifically, $\mu(\mathbf{s}, t)$ from (4) is defined as

$$\mu(\mathbf{s}, t) = b_0 + w_{\mathbf{s}}(t) + z(\mathbf{s}, t), \tag{10}$$

where b_0 is a fixed unknown intercept that is shared by all wind farms. The process $w_{\mathbf{s}}(t)$ is assumed to have autoregressive dynamics as defined in (6). Finally, $z(\mathbf{s}, t)$ is a spatio-temporal component that has the structure of (8) and its spatio-temporal covariance function is the same as in (9).

For all the models described above, a log-Gamma prior is assumed for the parameters in the Matérn covariance as well as for the precision parameters σ_e^2 and σ_ν^2. For the fixed effect b's we assume Gaussian priors. The correlations ρ's are specified over the parametrization $\log(\frac{1+\rho}{1-\rho})$ with prior Gaussian distributions.

3.3. Inference and prediction

The key feature of the models described above is that they can be handled within the theoretical and computational framework developed by Rue et al. [2009] and Lindgren et al. [2011]. The approach by Rue et al. [2009] allows us to directly compute accurate and fast approximations of the posterior marginals. In addition, the method by Lindgren et al. [2011] is computationally efficient for inferential purposes: instead of using a Gaussian random fields (GRF) with dense covariance matrix, the computations are carried out with a Gaussian Markov random field (GMRF) with sparse precision matrix. The original idea comes from the work of Whittle [1954] and Whittle [1963], where it is shown that the solution to the SPDE

$$(\kappa^2 - \Delta)^{\alpha/2} x(\mathbf{u}) = \mathcal{W}(\mathbf{u}), \quad \mathbf{u} \in \mathbb{R}^d, \alpha = \nu + D/2, \kappa > 0, \nu > 0, \tag{11}$$

is a GRF with Matérn covariance function. The innovation process $\mathcal{W}$ on the right hand side of (11) is Gaussian white noise and Δ is the Laplacian.

An approximation to the solution of the SPDE in (11) can be obtained using the finite element method (FEM), a numerical technique for solving partial differential equations Lindgren et al. [2011]. This is done by representing the infinite dimensional GRF by a linear combination of finite basis function

$$x(\mathbf{u}) = \sum_k \psi_k(\mathbf{u}) w_k \tag{12}$$

where the w_k's are random weights chosen so that the representation in (12) approximates the distribution of the solution to the SPDE in (11). The ψ_k's are basis functions defined on a triangulation of the domain, i.e. a subdivision into non-intersecting triangles. Figure 2 shows the triangulation of western Denmark data set described in Section 2.

[Figure 2 about here.]

Next, the posterior estimates of parameters and hyperparameters are computed using INLA Rue et al. [2009]. This method approximates the integral involved in the calculation of the marginal posterior distributions of the hyperparameters by Laplace approximation, making use of the Markov structure of the latent variables in the computation. We use the R-INLA package to perform inference and prediction. For more information on the package see http://www.r-inla.org.

4. FORECAST EVALUATION

4.1. Probabilistic forecasting scheme

We evaluate the predictive performance of the models described in Section 3, using a time moving window approach with data from western Denmark in 2009, so that each training set consists of $L = 2 \times 96 = 192$ observations, i.e., two days. In total, the model is fit to $364/2 = 182$ different data sets. We obtain forecasts for lead times $h = 1, \ldots, 20$, that is, from 15 minutes up to 5 hours following the training data. Notice that we have compared different lengths of data window L with respect to the root mean squared error (RMSE) and continuous ranked probability score (CRPS). The temporal model presented in Section 3.2.1 is very sensitive to the window length, such that less than two days of observations in the training set resulted in poor estimation at all lead times. On the other hand, the spatio-temporal models in Section 3.2.2 and 3.2.3 showed to be robust for different values of L, with small changes in the forecast performance for different training sets.

Moreover, because of the high-time resolution of the Danish wind power time series (15-minutes) and the dependency structure in space and time of Model S-T and Model ST+T, the fitting can be very computationally expensive. One way to deal with high-time resolution data is to define the model on a set of knots instead of all time points. Knot-based linear combinations are widely used to tackle computational problems in large data sets (e. g.

Paciorek [2007] and Wikle and Cressie [1999]). To fit the spatio-temporal component $z(\mathbf{s}, t)$ in (8), we define a set of equally spaced knots at every 12 data points (3 hours), such that the points in time are reduced to only 17 knots, instead of the original 192 observations. Note that the component $w_{\mathbf{s}}(t)$ in models Model T and Model ST+T is fitted to the complete training data, since it does not involve spatio-temporal interactions.

We evaluate probabilistic forecasts of wind power production from individual wind farms and aggregated.

Let $\hat{X}(\mathbf{s}_j, t+h)$ denote the random variable of the wind power forecast at wind farm $\mathbf{s}_j$ and lead time h. The aggregated forecast of wind power generation is taken as

$$\hat{X}_{A(t+h)} = \frac{\sum_{j=1}^{N} c_j \hat{X}(\mathbf{s}_j, t+h)}{\sum_{j=1}^{N} c_j}, \tag{13}$$

where c_j is the capacity of wind farm $\mathbf{s}_j$ and N is the number of wind farms. To find the pdf of the aggregated forecasts, $\hat{f}_{X_{A(t+h)}}$, the joint distribution for all wind farms $\{\hat{X}(\mathbf{s}_1, t+h), \hat{X}(\mathbf{s}_2, t+h), \dots \hat{X}(\mathbf{s}_N, t+h)\}$ needs to be assessed. Finally, point forecast of aggregated wind power production is obtained as the mean (or median) of $\hat{f}_{X_{A(t+h)}}$.

4.2. Point and probabilistic forecast scores

We assess the quality of predictive performance of the models proposed in Section 3 using both point and probabilistic forecast scores. We obtain point forecast at a specific location as the mean of the forecast density. For each lead time, point forecast of individual power is assessed using the root mean squared error (RMSE), where the mean is taken over all wind farms and data sets,

$$\text{RMSE}(t+h) = \sqrt{\frac{1}{DN} \sum_{i=1}^{D} \sum_{j=1}^{N} (x(\mathbf{s}_{ij}, t+h) - \hat{x}(\mathbf{s}_{ij}, t+h))^2} \tag{14}$$

where D is the number of data sets, N is the number of wind farms and $\hat{x}(\mathbf{s}_{ij}, t+h) = \gamma^{-1}(\hat{y}(\mathbf{s}_{ij}, t+h))$ is the predicted value of $x(\mathbf{s}_{ij}, t+h)$.

To evaluate the performance of forecast densities, we use the continuous ranked probability score (CRPS). Gneiting and Raftery [2007] showed that CRPS is a strictly proper scoring rule for the evaluation of probabilistic forecasts of a univariate quantity that assesses calibration and sharpness simultaneously Gneiting and Raftery [2007]. A lower score indicates a better density forecast. It is defined as

$$\text{CRPS}(F, x) = \int_{-\infty}^{\infty} (F(y) - \delta_{\{y \geq x\}})^2 dy \tag{15}$$

where F is the cumulative distribution function of the density forecast and y is the observation. With the available samples, we can approximate the mean CRPS at each lead time by

$$\begin{aligned} \text{CRPS}_{F,x}(t+h) = &\frac{1}{DN} \sum_{i=1}^{D} \sum_{j=1}^{N} \Big(\frac{1}{n} \sum_{k=1}^{n} |\hat{x}^{(k)}(\mathbf{s}_{ij}, t+h) - x(\mathbf{s}_{ij}, t+h)| \\ &- \frac{1}{2n^2} \sum_{k,l=1}^{n} |\hat{x}^{(k)}(\mathbf{s}_{ij}, t+h) - \hat{x}^{(l)}(\mathbf{s}_{ij}, t+h)| \Big), \end{aligned} \tag{16}$$

where n is the number of samples. Again, the mean CRPS is taken over all the wind farms and data sets in the training set.

Reliability, also referred to as calibration, of probabilistic forecasts is assessed with reliability diagrams. In a calibrated forecast, the observed levels should match the nominal levels for specific quantile forecasts, which results in points aligning with the diagonal in the reliability diagram. To construct reliability diagrams, we start by introducing an indicator variable $\mathcal{I}^{(\alpha)}(\mathbf{s}_{ij}, h)$, which is defined for a quantile forecast $\hat{q}^{(\alpha)}(\mathbf{s}_{ij}, t+h)$ issued at lead time

h and wind farm $\mathbf{s}_i$ of the training data j, with observed value $x(\mathbf{s}_{ij}, t+h)$ as follows

$$\mathcal{I}^{(\alpha)}(\mathbf{s}_{ij}, h) = \begin{cases} 1 & \text{if} \quad x(\mathbf{s}_{ij}, t+h) \leq \hat{q}^{(\alpha)}(\mathbf{s}_{ij}, t+h) \\ 0, & \text{otherwise} \end{cases}$$

The indicator variable $\mathcal{I}^{(\alpha)}(\mathbf{s}_{ij}, h)$ shows whether the actual outcome lies below the α quantile forecast (hits) or not (miss). Next, $n_{h,1}^{(\alpha)}$ denotes the sum of hits and $n_{h,0}^{(\alpha)}$ the sum of misses over all the realizations

$$n_{h,1}^{(\alpha)} = \sum_{i=1}^{D} \sum_{j=1}^{N} \mathcal{I}^{(\alpha)}(\mathbf{s}_{ij}, h) \quad \text{and} \quad n_{h,0}^{(\alpha)} = DN - n_{h,1}^{(\alpha)}.$$

An estimation $\hat{a}_h^{(\alpha)}$ of the actual coverage $a_h^{(\alpha)}$ is then obtained by calculating the mean of $\mathcal{I}^{(\alpha)}(\mathbf{s}_{ij}, h)$ over the N wind farms in the D validation sets

$$\hat{a}_h^{(\alpha)} = \frac{1}{DN} \sum_{i=1}^{D} \sum_{j=1}^{N} \mathcal{I}^{(\alpha)}(\mathbf{s}_{ij}, h) = \frac{n_{h,1}^{(\alpha)}}{n_{h,1}^{(\alpha)} + n_{h,0}^{(\alpha)}}. \tag{17}$$

Here, we use nominal levels from 5% to 95% in steps of 5%. Since the number of observations used to calculate the reliability diagrams is of limited size and the observed proportions are equal to the nominal ones only asymptotically Toth et al. [2003] Bröcker and Smith [2007], we follow the idea of Bröcker and Smith [2007] of generating consistency bars for reliability diagrams.

4.3. Evaluation scheme

We evaluate probabilistic forecasts of Danish wind power production from two different scenarios. First, we consider time forward forecast performances at the locations of the training set. The spatio-temporal models, i.e, Model S-T and Model ST+T, have the advantage of being able to provide forecasts where recent observations are not available.

Based on this, in a second evaluation scheme, we study the performances of spatially out-of-sample forecasts, which are based on k-fold cross-validation with $k = 5$. Notice that overall, 5 to 10-fold cross-validation is recommended as a good compromise between bias and variance (Breiman and Spector [1992]; Kohavi et al. [1995]). The forecast performance measures from the second scenario are obtained by combining the estimates from the 182 data sets in the training set.

Finally, we validate our results with a simulation study consisting of 200 simulated spatio-temporal data sets. In each data set, logit transformed wind power production measurements, $y(\mathbf{s}_i, t)$, are "observed" at 200 wind farms belonging to the wind power data set (see left plot of Figure 1). To mimic the case study based on the Danish wind data set, we simulate data every 15 minutes for 2 days and 5 hours. In total, there are $2 \times 96 + 20 = 212$ measurements taken at each location. All data sets are generated according to Model ST+T directly using the SPDE model construction. We use the set of parameters found for one specific data set of the training set from fitting Model ST+T to the logit transformed Danish wind power data.

5. RESULTS

In this section we show the results from a case study, where we use the models described in Section 3 to forecast individual and aggregated wind power in Denmark. As described in Section 4.3, we evaluate and discuss the performances of our models when we consider time forward forecasts at the locations of the training set. We call these spatio-temporal forecasts, and we also show the case of spatially out-of-sample forecasts, i.e, for wind farms that are not in the training set. Furthermore, we illustrate the results from a simulation study based on our case study. Details of the probabilistic forecasting scheme can be found in Section 4.1, while the methodology used to rank point and probabilistic forecasts is in Section 4.2.

5.1. Spatio-temporal forecast performance

Figure 3 summarizes the spatio-temporal forecast performances of the three models introduced in Section 3 in terms of RMSE and CRPS. As we can see from Figure 3 (a), Model T and Model ST+T outperformed Model S-T with respect to RMSE and CRPS when forecasting individual wind farms at lead times 1-6 (i.e, from 15 minutes up to 2 hours ahead). For higher lead times, the three models have similar performance. In terms of aggregated wind power production, Model T performed similar to Model ST+T in terms of point forecast (RMSE), but it has poor performance according to CRPS values, as shown in Figure 3 (b).

Reliability diagrams for each model at lead times $h = 1, 7, 13$ and 19 are presented in Figure 4. These diagrams compare the theoretical and the observed proportions of a set of quantiles from forecasts made at all wind farms and data sets in the training set. The forecasts at individual wind farms produced by the three models presented in Section 3 perform similarly well in terms of reliability, with points close to the diagonal for most quantiles, see Figure 4 (a). Since the number of observations used to calculate reliability diagrams is relatively small (182 data sets in the training set), consistency bars for the evaluation of forecasts from aggregated farms are also plotted, as shown in Figure 4 (b). The aggregated forecasts provided by Model ST+T are the best calibrated among the three models for most of the quantiles at all lead times, followed by Model S-T. Even though the performance of Model T is comparable with the performance of the other models in terms of aggregated forecast density mean (RMSE), we can see that this model does not produce reliable probabilistic forecasts for the aggregated data. This fact is more obvious for the lower quantiles; more than 50% of the observed aggregated forecasts are below the nominal 5% quantile at lead times $h = 7, 13$ and 19.

[Figure 3 about here.]

[Figure 4 about here.]

We further explore aggregated probabilistic forecasts from models in Section 3 with plots containing the 5%, 50% and 95% quantiles of the aggregated forecast densities together with the actual observed aggregated power produced at four different data sets in the training set, as shown in Figure 5. We noticed that Model T results in forecast densities that are consistently too narrow. On the other hand, Model ST+T provides the widest aggregated forecast densities among the three models in most of the data sets, which produces calibrated forecasts at all lead times. This is confirmed in Figure 4 (b) and will be further explored in the simulation studies in Section 5.3.

[Figure 5 about here.]

5.2. Spatially out-of-sample forecast performance

Figure 6 shows the out-of-sample forecast performances in terms of RMSE and mean CRPS for individual wind farms (a) and aggregated wind power (b). They are computed as the mean of the RMSE and CRPS from the 5-fold cross validations as described in Section 4.3. It can be seen that Model ST+T outperforms Model S-T at all lead times when predicting wind power at individual wind farms under RMSE and CPRS. When looking at aggregated out-of-sample forecasts, while for shorter lead times than 2 hours, Model S-T is better than Model ST+T in terms of RMSE, for longer horizons, Model ST+T out-performs Model S-T under the same score. In terms of CRPS, Model ST+T produces better aggregated forecasts at lead times 1-20 (i.e., from 15 minutes to 5 hours ahead).

Reliability diagrams at lead times $h = 1, 7, 13$ and 19 are presented in Figure 7. We observe from Figure 7 (a) that Model S-T and Model ST+T provide relatively well calibrated forecast densities for individual farms. In terms of aggregated forecasts, we can see from Figure 7 (b) that Model ST+T is calibrated, since the line is always within the consistency bars. On the other hand, aggregated forecast densities obtained with Model S-T are poorly calibrated for

quantiles lower than 0.75. Indeed, 20% of the observations are below the 5% forecast quantile at lead times 1, 7, 13 and 19.

[Figure 6 about here.]

[Figure 7 about here.]

5.3. Simulation study

From the data analysis in Section 5.1 and 5.2, we see that Model ST+T is the only model among the three that produces individual and aggregated calibrated forecasts. In this section, we simulate 200 data sets according to this model. We set the parameters equal to the estimates given by the fit of this model to one of the training data sets from our case study. More details on the evaluation scheme can be found in Section 4.3.

RMSE and mean CRPS of the three different models for forecasting simulated data at lead times 1-20 are shown in Figure 8. According to RMSE and CRPS for individual wind farms, the three models perform similarly, while, according to CRPS for aggregated forecasts, Model ST+T out-performs Model T and Model S-T.

Results from individual and aggregated forecasts calibration are shown in Figure 9. The forecasts from all three models are calibrated for individual wind farms, as shown in Figure 9 (a). We observe from Figure 9 (b) that the aggregated forecasts produced by the Model ST+T are better in terms of calibration than the forecasts from the other models, which is in agreement with the results from the analysis of the aggregated Danish wind power data, as shown in Figure 4 (b). In fact, the aggregated forecasts produced by Model ST+T are well calibrated at lead times $h = 1, 13$, and 19, since the line is always inside the consistency bars.

The simulations show that when we fit simulated data from Model ST+T using Model S-T, the spatial range r (see (9)) is underestimated. In fact, when data is generated with $r = 62.1$, the first and third quartiles of the 200 estimates of this parameter from Model

ST+T are 27.7 and 164.6, while with Model S-T the estimated quartiles are 25.2 and 28.0, respectively. Thus, a larger estimated spatial dependency results in a larger variance to the aggregated forecasts and makes it possible to borrow more information from close wind farms when doing out-of-sample predictions, causing the variance of a sum to increase. Hence, this explains both why the aggregated forecasts are not calibrated for Model S-T as well as why Model ST+T gives better spatially out-of-sample predictions than model S-T.

[Figure 8 about here.]

[Figure 9 about here.]

6. CONCLUSIONS

In this article we have presented hierarchical spatio-temporal models for obtaining probabilistic forecasts of wind power generation at multiple locations and lead times. We started with a time series model consisting of an autoregressive process with a location specific intercept. The results for individual probabilistic forecasts were satisfactory in terms of skill scores and reliability, however, the aggregated probabilistic forecasts were not calibrated. After finding the unsatisfactory results for the reliability of aggregated forecasts, we introduced two different spatio-temporal models. The first has a common intercept for all farms and a spatio-temporal model that varies in time with first order autoregressive dynamics and has spatially correlated innovations given by a zero mean Gaussian process with Matérn covariance. The second model has a common intercept, an autoregressive process to capture the local variability and the spatio-temporal term. To deal with the non-Gaussianity of wind power series, a parametric framework for distributional forecasts based on the logit-normal transformation was used.

In a case study, the proposed models have been used to produce probabilistic forecasts of wind power at wind farms in western Denmark from 15 minutes up to 5 hours ahead for a test

period of one year. Using the SPDE approach that is implemented in the R-INLA library, we obtained fast and accurate forecasts of wind power generation at wind farms where data is available, but also at a larger portfolio including wind farms at locations that are not included in the training set. We provided detailed analysis on the forecast performances based on appropriate metrics tailored for probabilistic forecasts. To better understand the properties of our methods, we analysed artificial data sets from a simulation study.

Our results showed that all the proposed approaches produce calibrated short-term forecasts for individual wind farms. However, we found that modeling spatial dependency is required to achieve calibrated aggregated probabilistic forecasts. Indeed, our case study showed that spatial dependency is important for aggregated properties, and individual forecasts do not reveal this. Moreover, when we simulated from the spatio-temporal model containing an autoregressive term (Model ST+T), we obtained results that are in accordance with our case study, where the proposed models performed equally well for individual forecasts, while aggregated probabilistic forecasts benefit from having a spatio-temporal model with the autoregressive term. Model ST+T was introduced due to unsatisfactory reliability for the aggregated forecasts. Hence, evaluating aggregated forecasts can be a tool for investigating and improving models, even when spatially out-of-sample forecasts are the purpose of the modelling. Indeed, results from spatially out-of-sample forecast performances showed that when predicting wind power at new locations that are not included in the training set, having the autoregressive term in the spatio-temporal model improved the forecast performance.

This work was motivated by the need to produce accurate short term probabilistic forecasts at multiple wind farms and lead times, which will ultimately be applied on a national scale. A possible extension of the models described in this work is to include weather forecast information in the linear predictor. This approach usually requires ensemble forecasts to be generated from sophisticated numerical weather prediction (NWP) models and has shown

to produce reliable wind power forecasts up to 10 days ahead Taylor et al. [2009].

ACKNOWLEDGEMENTS

The authors are grateful to Energinet.dk (system operator in Denmark) for providing the data and to Robin Girard at MinesParistech, France for checking the quality of the data. The authors also thank the Danish Strategic Council for Strategic Research through the project 5s-Future Electricity Markets (No. 12-132636/DSF), Research Concil of Norway, project 250362 and CAPES for support.

REFERENCES

Thomas Ackermann. *Wind power in power systems*. John Wiley & Sons, 2005.

Marta Blangiardo and Michela Cameletti. *Spatial and spatio-temporal bayesian models with R-INLA*. John Wiley & Sons, 2015.

Leo Breiman and Philip Spector. Submodel selection and evaluation in regression. the x-random case. *International Statistical Review/Revue Internationale de Statistique*, pages 291–319, 1992.

John Bjørnar Bremnes. Probabilistic wind power forecasts using local quantile regression. *Wind Energy*, 7 (1):47–54, 2004.

Jochen Bröcker and Leonard A Smith. Scoring probabilistic forecasts: The importance of being proper. *Weather and Forecasting*, 22(2):382–388, 2007.

Michela Cameletti, Finn Lindgren, Daniel Simpson, and Håvard Rue. Spatio-temporal modeling of particulate matter concentration through the spde approach. *AStA Advances in Statistical Analysis*, 97(2):109–131, 2013.

João Paulo da Silva Catalão, Hugo Miguel Inácio Pousinho, and Víctor Manuel Fernandes Mendes. Short-term wind power forecasting in portugal by neural networks and wavelet transform. *Renewable Energy*, 36(4):1245–1251, 2011.

Daniela Cocchi, Fedele Greco, and Carlo Trivisano. Hierarchical space-time modelling of pm 10 pollution. *Atmospheric Environment*, 41(3):532–542, 2007.

Noel Cressie. Statistics for spatial data. *Terra Nova*, 4(5):613–617, 1992.

Jethro Dowell and Pierre Pinson. Very-short-term probabilistic wind power forecasts by sparse vector autoregression. *IEEE Transactions on Smart Grid*, 7(2):763–770, 2016.

Alessandro Fassò and Francesco Finazzi. Maximum likelihood estimation of the dynamic coregionalization model with heterotopic data. *Environmetrics*, 22(6):735–748, 2011.

Ulrich Focken, Matthias Lange, Kai Mönnich, Hans-Peter Waldl, Hans Georg Beyer, and Armin Luig. Short-term prediction of the aggregated power output of wind farmsa statistical analysis of the reduction of the prediction error by spatial smoothing effects. *Journal of Wind Engineering and Industrial Aerodynamics*, 90(3):231–246, 2002.

Robin Girard, K Laquaine, and Georges Kariniotakis. Assessment of wind power predictability as a decision factor in the investment phase of wind farms. *Applied Energy*, 101:609–617, 2013.

Tilmann Gneiting and Adrian E Raftery. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378, 2007.

Tilmann Gneiting, Kristin Larson, Kenneth Westrick, Marc G Genton, and Eric Aldrich. Calibrated probabilistic forecasting at the stateline wind energy center: The regime-switching space–time method. *Journal of the American Statistical Association*, 101(475):968–979, 2006.

Tilmann Gneiting, Fadoua Balabdaoui, and Adrian E Raftery. Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(2):243–268, 2007.

Rikke Ingebrigtsen, Finn Lindgren, and Ingelin Steinsland. Spatial models with explanatory variables in the dependence structure. *Spatial Statistics*, 8:20–38, 2014.

Ron Kohavi et al. A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Ijcai*, volume 14, pages 1137–1145. Stanford, CA, 1995.

Ada Lau and Patrick McSharry. Approaches for multi-step density forecasts with application to aggregated wind power. *The Annals of Applied Statistics*, pages 1311–1341, 2010.

Emmanuel Lesaffre, Dimitris Rizopoulos, and Roula Tsonaka. The logistic transform for bounded outcome scores. *Biostatistics*, 8(1):72–85, 2007.

Finn Lindgren, Håvard Rue, and Johan Lindström. An explicit link between gaussian fields and gaussian markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(4):423–498, 2011.

Petroula Louka, Georges Galanis, Nils Siebert, Georges Kariniotakis, Petros Katsafados, I Pytharoulis, and G Kallos. Improvements in wind speed forecasts for wind power prediction purposes using kalman filtering. *Journal of Wind Engineering and Industrial Aerodynamics*, 96(12):2348–2362, 2008.

Facundo Munoz, M Grazia Pennino, David Conesa, Antonio López-Quílez, and José M Bellido. Estimation and prediction of the spatial occurrence of fish species using bayesian latent gaussian models. *Stochastic Environmental Research and Risk Assessment*, 27(5):1171–1180, 2013.

Christopher J Paciorek. Bayesian smoothing with gaussian processes using fourier basis functions in the spectralgp package. *Journal of Statistical Software*, 19(2):nihpa22751, 2007.

Pierre Pinson. Very-short-term probabilistic forecasting of wind power with generalized logit–normal distributions. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 61(4):555–576, 2012.

Pierre Pinson and George Kariniotakis. Conditional prediction intervals of wind power generation. *IEEE Transactions on Power Systems*, 25(4):1845–1856, 2010.

Pierre Pinson and Julija Tastu. Discrimination ability of the energy score. Technical report, Technical University of Denmark, 2013.

Sheldon Ross. *A first course in probability*. Pearson, 2015.

Håvard Rue, Sara Martino, and Nicolas Chopin. Approximate bayesian inference for latent gaussian models by using integrated nested laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(2):319–392, 2009.

Michael Scheuerer and Thomas M Hamill. Variogram-based proper scoring rules for probabilistic forecasts of multivariate quantities. *Monthly Weather Review*, 143(4):1321–1334, 2015.

Julija Tastu, Pierre Pinson, Ewelina Kotwa, Henrik Madsen, and Henrik Aa Nielsen. Spatio-temporal analysis and modeling of short-term wind power forecast errors. *Wind Energy*, 14(1):43–60, 2011.

James W Taylor, Patrick E McSharry, and Roberto Buizza. Wind power density forecasting using ensemble predictions and time series models. *IEEE Transactions on Energy Conversion*, 24(3):775–782, 2009.

Thordis L Thorarinsdottir, Michael Scheuerer, and Christopher Heinz. Assessing the calibration of high-dimensional ensemble forecasts using rank histograms. *Journal of Computational and Graphical Statistics*, 25(1):105–122, 2016.

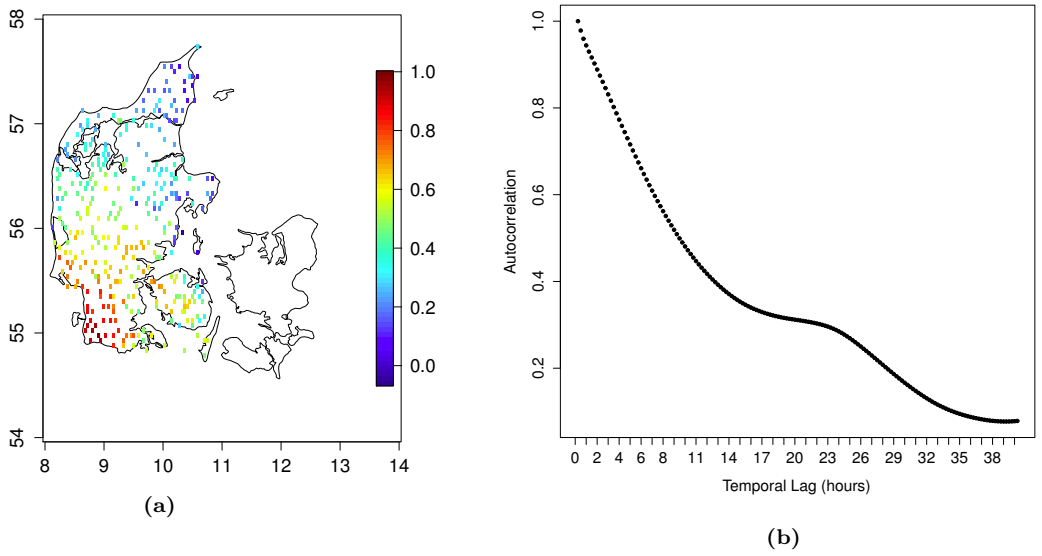
Zoltan Toth, Oliver Talagrand, Guillem Candille, and Yuejian Zhu. Forecast verification: A practitioners guide in atmospheric science, 2003.

Peter Whittle. On stationary processes in the plane. *Biometrika*, pages 434–449, 1954.

Peter Whittle. Stochastic-processes in several dimensions. *Bulletin of the International Statistical Institute*, 40(2):974–994, 1963.

Christopher K Wikle and Noel Cressie. A dimension-reduced approach to space-time kalman filtering. *Biometrika*, 86(4):815–829, 1999.

FIGURES



(a) **(b)**

Figure 1. (a) Map of spatial correlation of wind power production between one wind farm located in the southern part of western Denmark and the remaining wind farms. The correlations between wind farms in a closer proximity are clearly higher than between wind farms that are farther apart. (b) Mean autocorrelation function of wind power production at wind farms located in western Denmark. The autocorrelations decay slowly.

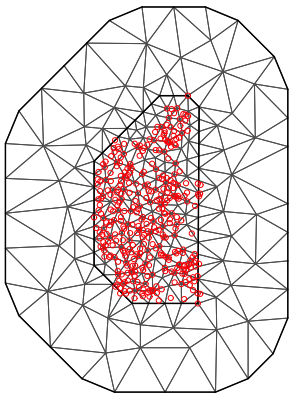
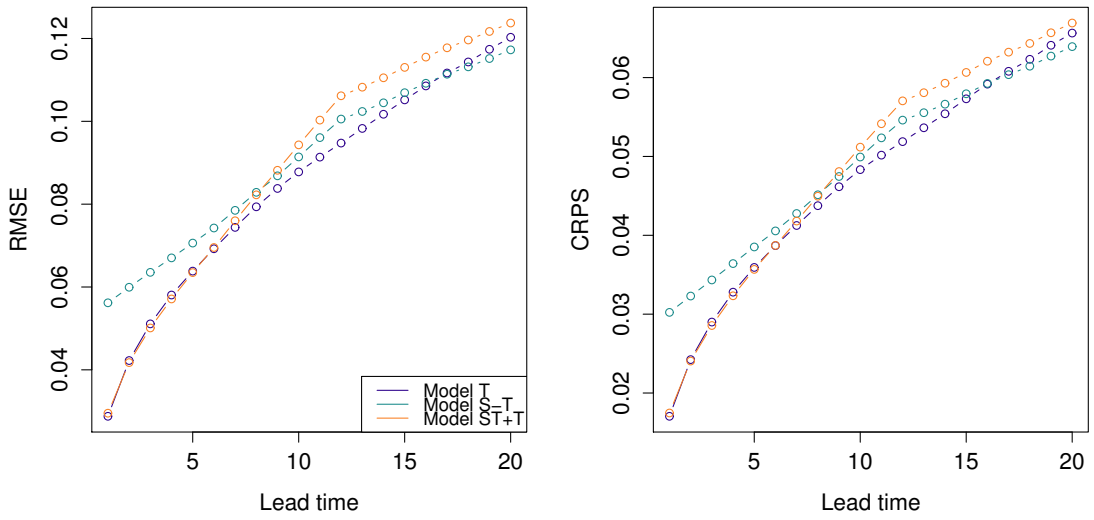
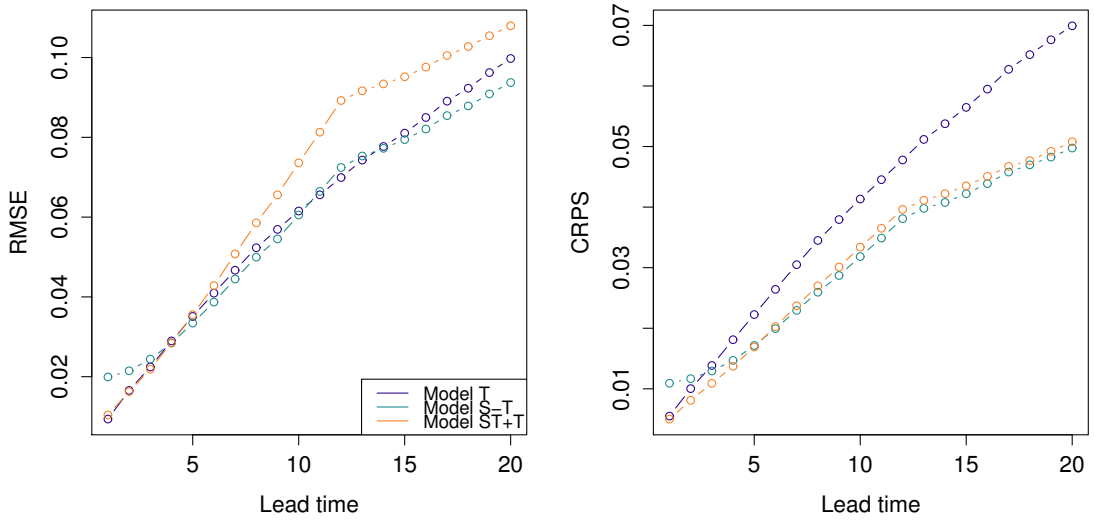


Figure 2. The western Denmark triangulation. The red dots denote the observation locations of the wind power production data.

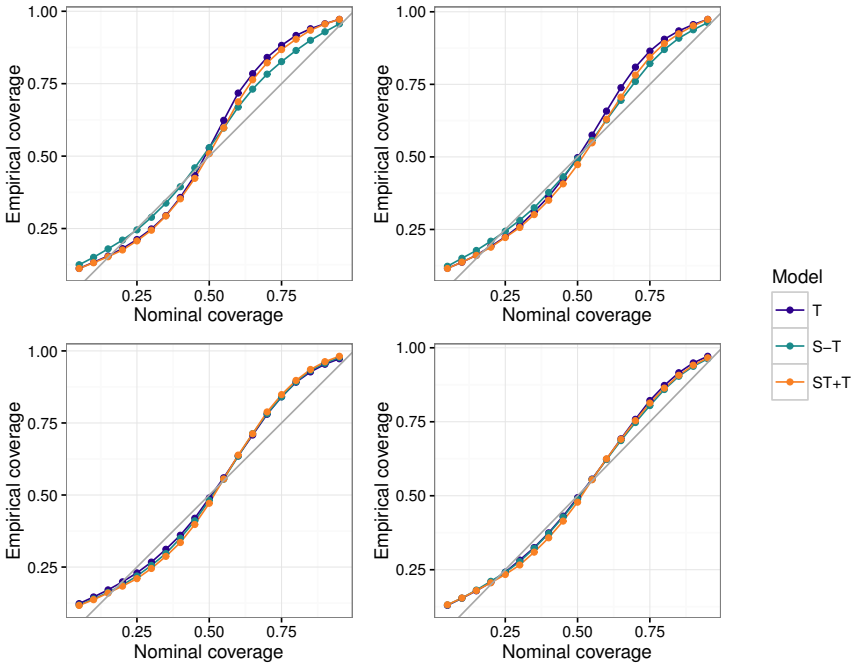


(a)

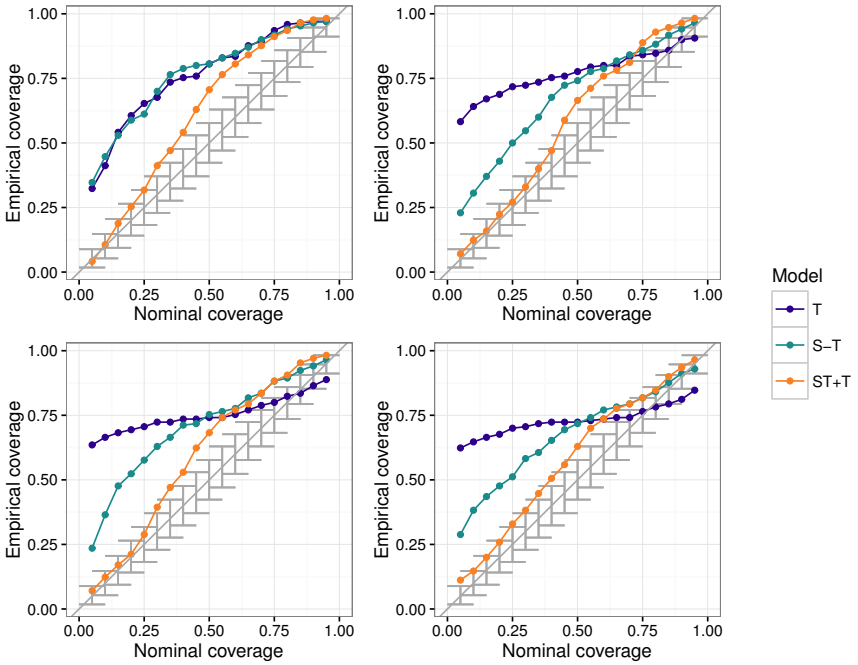


(b)

Figure 3. RMSE and CRPS (as % of nominal power) of spatio-temporal wind power forecasts at lead times $1, \ldots, 20$ (i.e., from 15 minutes up to 5 hours) for Model T (blue), Model S-T (green) and Model ST+T (orange). (a) Forecasts for individual wind farms. (b) Forecasts for aggregated wind farms.



(a)



(b)

Figure 4. Reliability diagram of spatio-temporal wind power forecasts at lead time 1 (*Top left*), 7 (*Top right*) , 13 (*Bottom left*) and 19 (*Bottom right*). The diagrams were calculated using Model T (blue), Model S-T (green) and Model ST+T (orange). (a) Forecasts for individual wind farms. (b) Forecasts for aggregated wind farms.

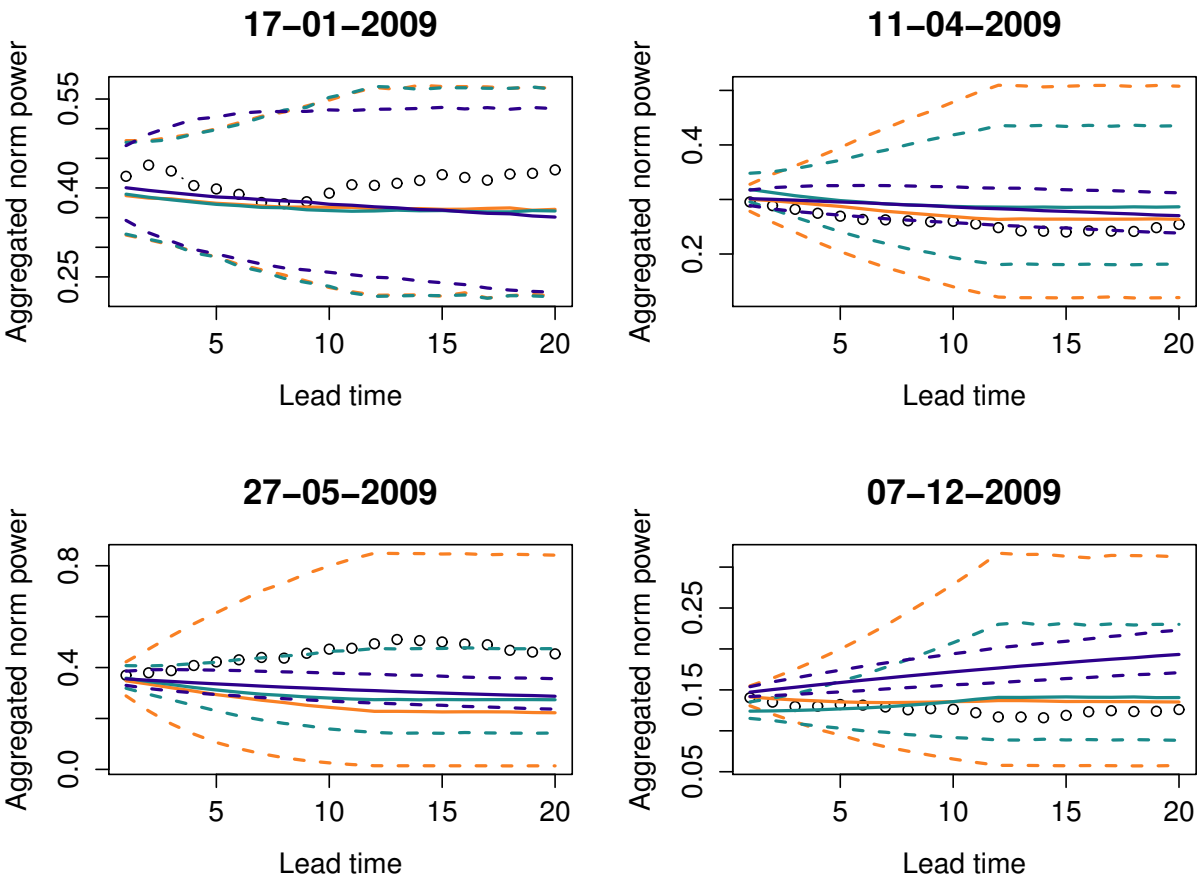
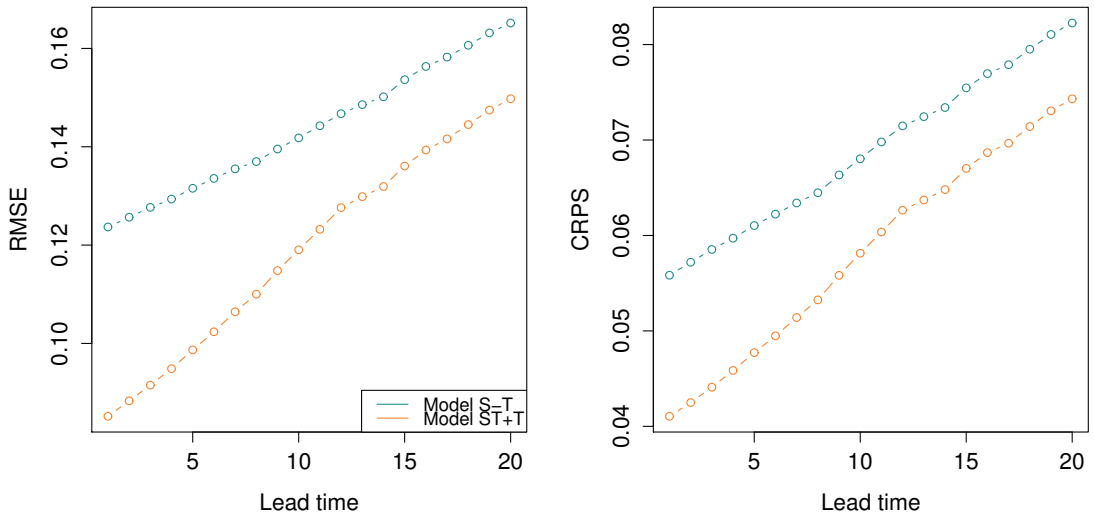
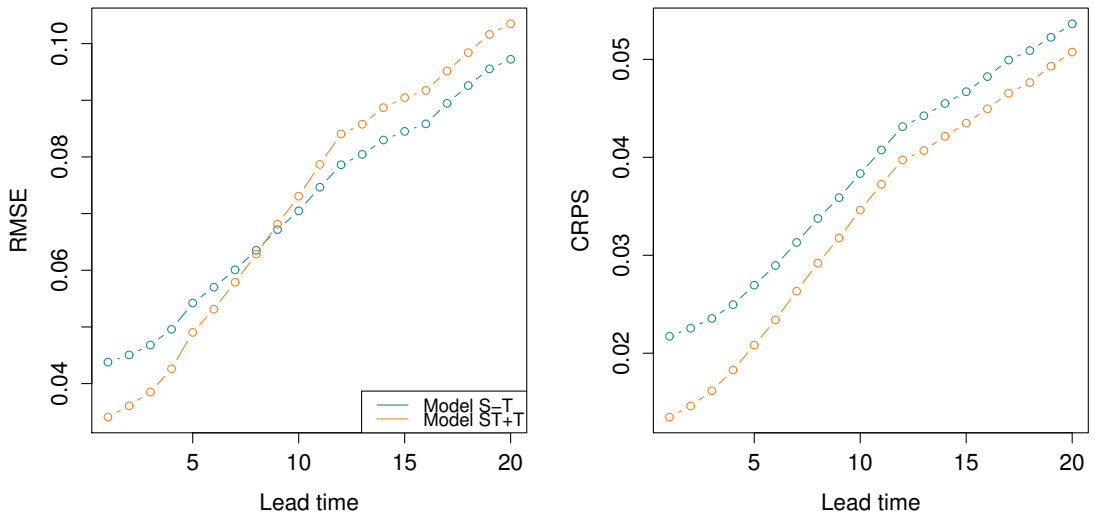


Figure 5. 5% and 95% quantiles (dashed lines), as well as the median (solid lines) of the aggregated forecast densities from four different data sets in the training set, together with the actual observed aggregated power produced (circles) at lead times 1-20 (i.e., from 15 minutes up to 5 hours). The forecast densities correspond to Model T (blue), Model S-T (green) and Model ST+T (orange). An example of a data set where all the models have forecast densities that cover the actual aggregated production is shown in the *Top left* plot. In the *Top right* plot, the observations lie close to the median of the forecast densities from Model S-T and Model ST+T, but close to the 5% quantile of the forecast density from Model T. *Bottom left* and *Bottom right* plots illustrate cases where Model T has forecast densities that are too narrow and fail to predict the aggregated wind power, while the forecasts from Model ST+T provide densities that are wide enough to cover the true value at all lead times.

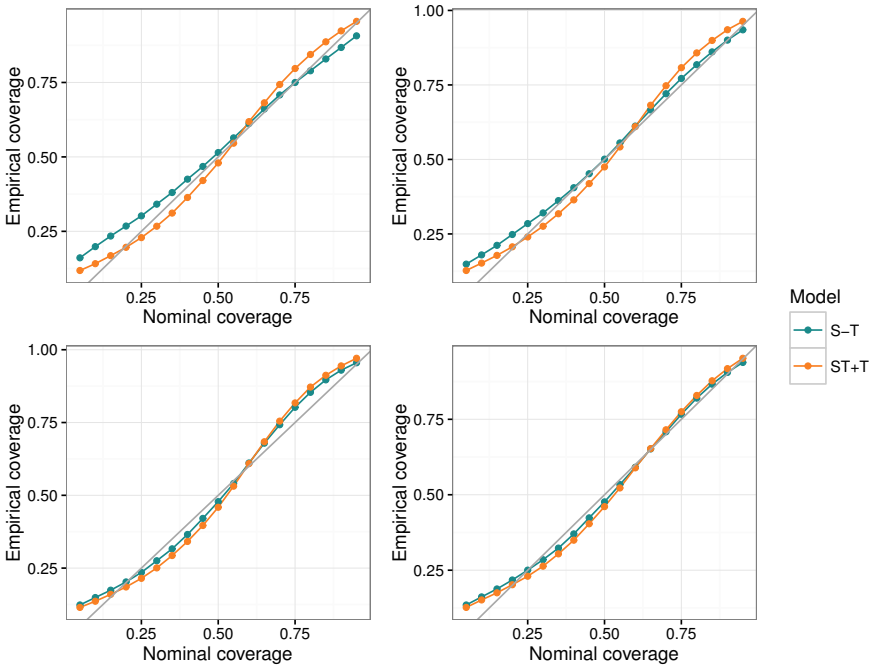


(a)

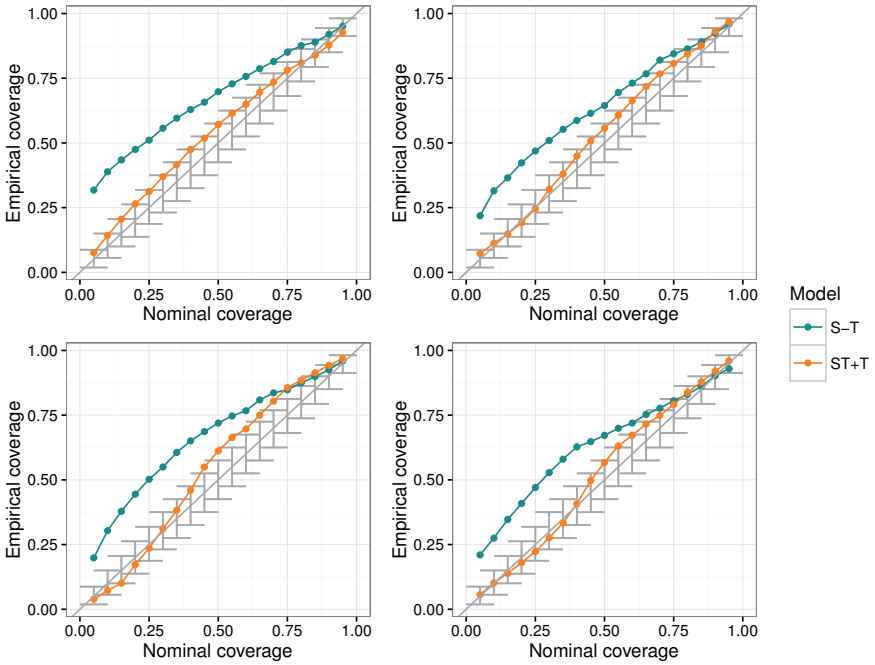


(b)

Figure 6. RMSE and CRPS (as % of nominal power) of spatially out-of-sample wind power forecasts at lead times $1, \ldots, 20$ (i.e., from 15 minutes up to 5 hours) for Model T (blue), Model S-T (green) and Model ST+T (orange). (a) Forecasts for individual wind farms. (b) Forecasts for aggregated wind farms.

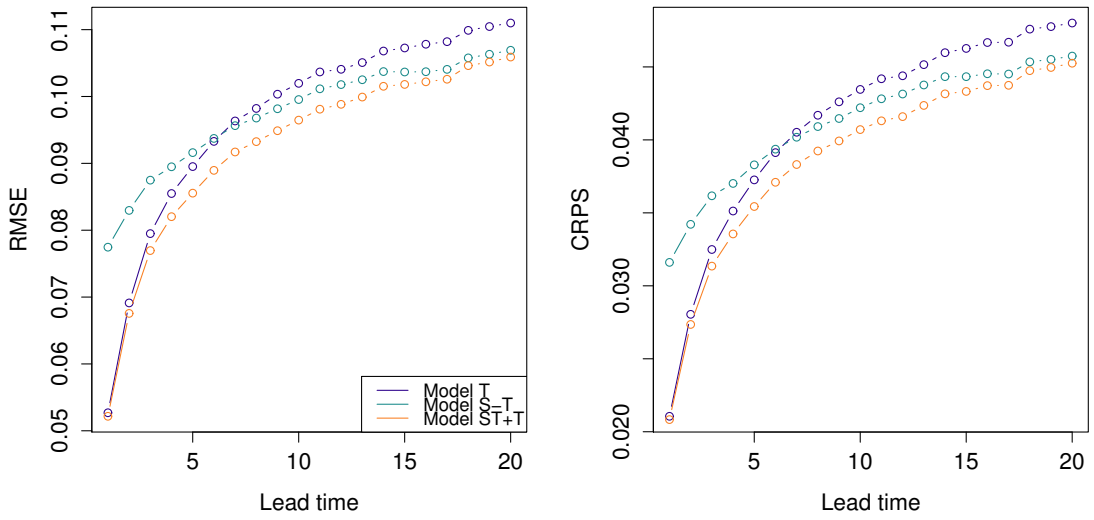


(a)

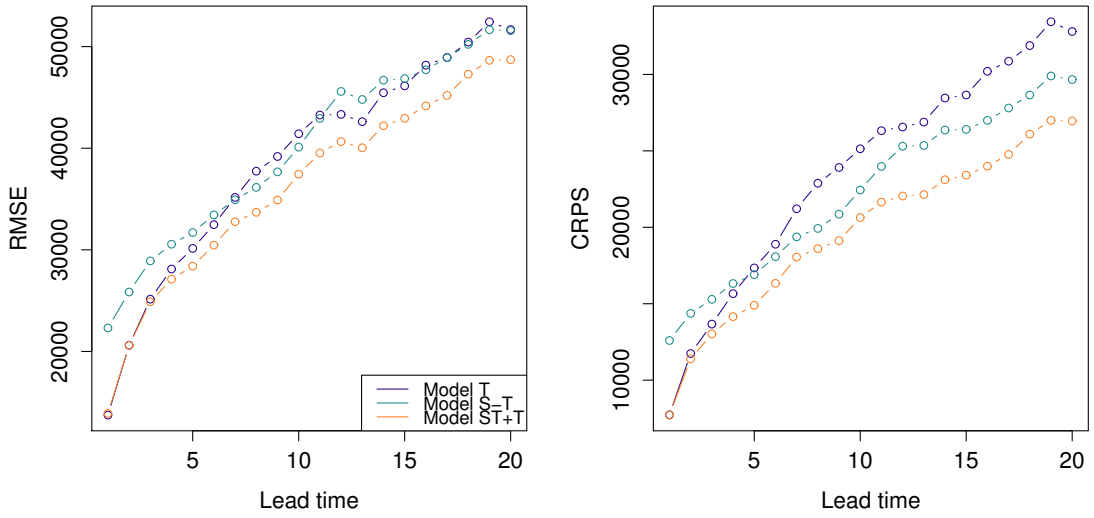


(b)

Figure 7. Reliability diagram of spatially out-of-sample wind power forecasts at lead time 1 (*Top left*), 7 (*Top right*) , 13 (*Bottom left*) and 19 (*Bottom right*). The diagrams were calculated using Model T (blue), Model S-T (green) and Model ST+T (orange). (a) Forecasts for individual wind farms. (b) Forecasts for aggregated wind farms.

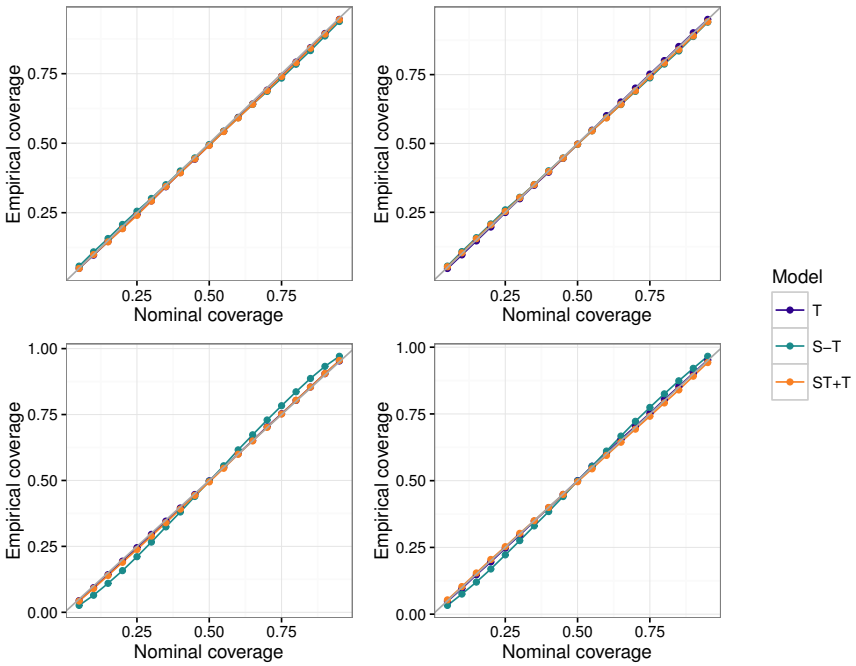


(a)

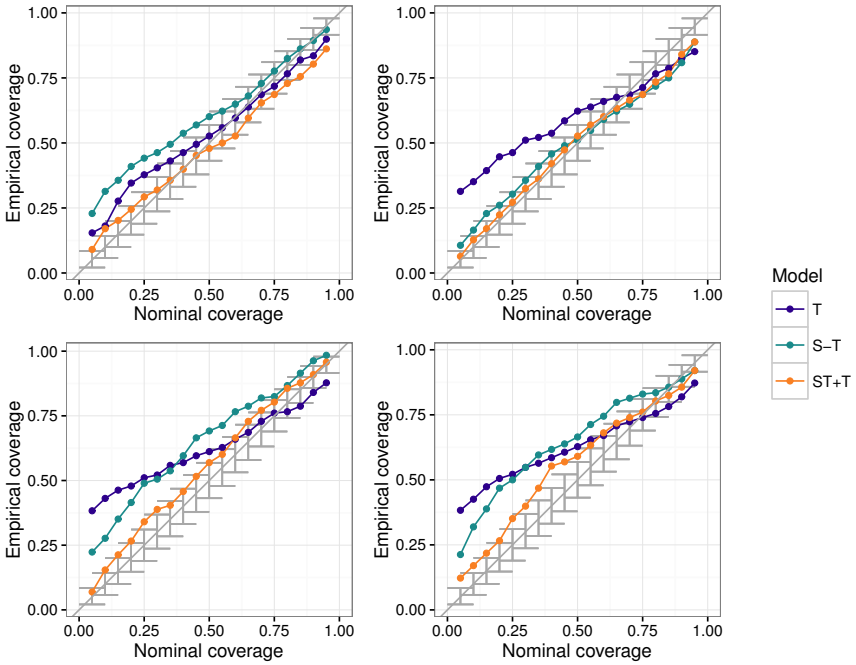


(b)

Figure 8. RMSE and CRPS (as % of nominal power) of forecasts from simulated data at lead times $1, \ldots, 20$ (i.e., from 15 minutes up to 5 hours) for Model T (blue), Model S-T (green) and Model ST+T (orange). (a) Forecasts for individual wind farms. (b) Forecasts for aggregated wind farms.



(a)



(b)

Figure 9. Reliability diagram for forecasts from simulated data at lead time 1 (*Top left*), 7 (*Top right*), 13 (*Bottom left*) and 19 (*Bottom right*). The diagrams were calculated using Model T (blue), Model S-T (green) and Model ST+T (orange). (a) Forecasts for individual wind farms. (b) Forecasts for aggregated wind farms.

Appendix C

Very short-term spatio-temporal wind power prediction using a censored Gaussian field

---

Baxevani, A., Lenzi, A. (2017). Very short-term spatio-temporal wind power prediction using a censored Gaussian field. Submitted to *Stoch Environ Res Risk Assess* (under review)

Noname manuscript No.
(will be inserted by the editor)

Very short-term spatio-temporal wind power prediction using a censored Gaussian field

Anastassia Baxevani · Amanda Lenzi

Received: date / Accepted: date

Abstract Wind power is a renewable energy resource, that has relatively cheap installation costs and it is highly possible that will become the main energy resource in the near future. Wind power needs to be integrated efficiently into electricity grids, and to optimize the power dispatch, techniques to predict the level of wind power and the associated variability are critical. Ideally, one would like to obtain reliable probability density forecasts for the wind power distributions. We aim at contributing to the literature of wind power prediction by developing and analysing a spatio-temporal methodology for wind power production, that is tested on wind power data from Denmark. We use anisotropic spatio-temporal correlation models to account for the propagation of weather fronts, and a transformed latent Gaussian field model to accommodate the probability masses that occur in wind power distribution due to chains of zeros. We apply the model to generate multi-step ahead probability predictions for wind power generated at both locations where wind farms already exist but also to nearby locations.

Keywords Wind power · spatio-temporal model · kriging equations · Gaussian transformed model · covariance function

Anastassia Baxevani
Department of Mathematics and Statistics,
University of Cyprus
Cyprus Tel.: +357-22892817
Fax: +357-22895353
E-mail: baxevani@ucy.ac.cy

Amanda Lenzi
Applied Mathematics and Computer Science Department
Technical University of Denmark
Denmark

1 Introduction

Over the last years there is a clear energy policy shift with the goal of achieving a reduction of carbon emissions and other pollutants. Wind power, as an alternative to burning fossil fuels, is renewable, clean, produces no greenhouse gas emissions and consumes no clean water. European Union aims at having 20% of the power coming from renewable sources by 2020, and mostly likely the United States will receive 20% of its electrical energy from wind by 2030, [24].

This means that national power grids need to be connected to wind farms and incorporate a significant part of the generated wind power. Wind power is driven by wind which depends on a lot of factors such as geometry of the surrounding landscape, location of the site (coastal or inland), season of the year etc, i.e., wind power depends on a complicated physical process that is largely determined by both geographical and seasonal properties. Therefore, it is only natural to model the wind power data as a realization of a spatio-temporal random field.

Since wind power cannot be stored efficiently, and wind turbines may need to shut down when wind speeds are too high, it is extremely important to quantify the uncertainties of wind power production for optimal planning, which calls for new methods and simulation tools that can assist electric utilities in analysing the impact of stochastic wind power production on power system operation, maintenance of wind power units and energy reserve scheduling, see [9] and [22]. Such analyses usually require a probabilistic approach. Managing the variability of wind power production is the key to optimal integration of that renewable energy into electricity grids, [12].

Broadly speaking, there are two approaches to wind power prediction: statistical models and physical models. The former uses only historical wind speed and power data to build statistical models, which since wind varies rapidly with time are effective only for very short-term predictions (1-4 hours ahead). This type of predictions are required for applications including power system's balancing and the optimal operation of reserves [33], [2], and wind farm control [19], [34]. On the other hand, the inclusion of spatial and temporal factors in a full fluid-dynamics model increase the validity of physical models to longer prediction horizons (from several hours up to a few days).

Statistical approaches for the stochastic modelling of wind power production fall mainly into two categories: the wind speed (indirect) approach and the wind power (direct) approach. In the first approach, focus is on modelling the wind speed and then transform the wind speed into wind power through a power curve, see [31]. Wind speed is modelled using artificial neural networks in [23]. In [28], adaptive neuro-fuzzy inference systems are used to anticipate the wind speed and direction frequency dispersion. Non-parametric copulas are used in [7] to study the influence of wind directions on the wind power production, and a process for assessing the wind speed using spectral analysis is given in [32].

The present paper aims at contributing in the literature of the statistical direct approach, where models are built for wind power production using directly the wind power measurements, see [29] and [25] amongst others. So far, most wind power prediction systems provide with temporal predictions for a wind turbine or a wind farm without taking the spatio-temporal dependency into consideration. Utilizing the spatial correlation amongst the different wind farms has been shown to significanlty reduce the errors in point forecasts [36], and has the advantage of producing models that are able to generate forecasts at locations that are not within the observation samples. Moreover, many forecast users are concerned with the production of their entire generation portfolio. In this case it is desirable to model and forecast spatio-temporal features in order to evaluate the uncertainty of aggregate generation forecasts. A few spatial models have been proposed to capitalise the spatio-temporal relation between wind power generation at wind farms. [38] use multiple wind farms as spatial sensors to forecast wind power at a specific location. Spatial correlation of wind speed and direction has been studied in [17]. Recent contributions build probabilistic spatial models with sparse Gaussian random fields, but are limited to small spatial dimension and only predicts one time step ahead, [37], [39], [11]. We suggest a model that although stationary in space, is valid over longer areas and can easily be extended to a non-stationary one and allows for several step ahead predictions.

When modelling directly the wind power data difficulties usually arise from the highly non-linear and clearly non-Gaussian character of the wind power series. In particular, at individual wind farms the wind power production series contain long chains of zeros which correspond to either calm conditions or wind speeds that have reached above a safety level and wind turbines need to be switched off for safety reasons. Additionally, the marginal distribution of wind power production data possesses tails that are heavier than the Gaussian, making direct use of the Gaussian model not feasible. Alternatively, some transformation of the data can be modelled as a truncated Gaussian random field. The truncation is inhered to the actual process since there are locations or times or both with no wind power production.

We suggest using the following model

$$Y(\mathbf{s},t) = \begin{cases} f_{\mathbf{s}}(Z(\mathbf{s},t)), & Z(\mathbf{s},t) > 0 \\ 0, & Z(\mathbf{s},t) \leq 0, \end{cases} \tag{1}$$

where Z is a stationary spatio-temporal Gaussian random field with variance one, mean that is chosen to make the probability of positive wind power production equal to a specified value and $f_{\mathbf{s}}(\cdot)$ a positive monotonic function, usually referred to as anamorphosis and chosen to obtain a specified marginal distribution for the positive wind power production at the specific location $\mathbf{s}$. The model in (1) has already been used in [4] for the precipitation depth at location $\mathbf{s}$ and day t.

Predictions - or forecasts as they are also called indistinguishably - are usually obtained by kriging, a common geostatistical technique, which in the case

of a Gaussian random field with known expectation coincides with the conditional expectation. Predictions of CO_2 emission and air quality monitoring using kriging methods can be found in [6] and [18], respectively.

This article is organised as follows: In Section 2 we present the models that will be used for the first two moments of the underlying random field as well as the anamoprhosis function and we discuss the fitting procedure. In Section 3, we provide a short introduction to kriging in a spatio-temporal context for the model presented in (1). The kriging equations are based on the assumption the mean and the covariance structure of the underlying Gaussian random field are known. A short account of the data set used in this article and a discussion on the model fitting are given in Section 4. Finally, in Section 5 we discuss the performance of the obtained short-term predictions.

In this work we shall reserve capital letters for the models and small letters for the observations, i.e. a random field $Z(\mathbf{s},t)$ has observations or measurements indicated by $z(\mathbf{s},t)$.

2 Modelling Procedure

Throughout the rest of this article we shall assume that $Z(\mathbf{s},t)$ is a spatio-temporal Gaussian random field with mean $\mu(\cdot)$ and covariance $C(\cdot,\cdot)$. We shall also assume that wind power production $Y(\mathbf{s},t)$ is measured at location $\mathbf{s}\in\mathcal{D}_{\mathbf{s}}$ and time $t\in\mathcal{D}_t$ and is modelled by $Y(\mathbf{s},t)=f_{\mathbf{s}}(Z(\mathbf{s},t))\delta_{\{Z(\mathbf{s},t))>0\}}$, with $\delta_{\{A\}}$ being the indicator function that is equal to one when property A is satisfied and zero otherwise. Hence, we assume the existence of an unobserved Gaussian random field Z, with excesses above some specified level (here taken to be 0) that coincides with the marginal distributions of the wind power production field Y after some transformation is applied to them. The random field Z is fully characterised by the first two moments and the distributional characterization of Y further requires models for the anamorphosis transformation f. We initiate this analysis by discussing different possible models for the above mentioned quantities and their fitting procedure.

2.1 Covariance models

Wind power production exhibits non-trivial spatio-temporal correlation structure. When restricted to a fixed location, the temporal correlation is quite strong. As it can be seen in Fig 2 (*Left*), the corresponding autocorrelation drops down to zero in about 20 hours. Similarly, there are interesting spatial dependence patterns. For example, Fig. 2 (*Right*) shows the simultaneous spatial correlation between the 24 wind farms of the portfolio, described in section 4.1. As can be observed, spatial dependence is quite strong even at distances of about 300 kilometers. This can be partly explained by the relatively smooth Danish landscape. Moreover, Denmark is a rather small country so wind farms lie close to each other with the biggest distance between the

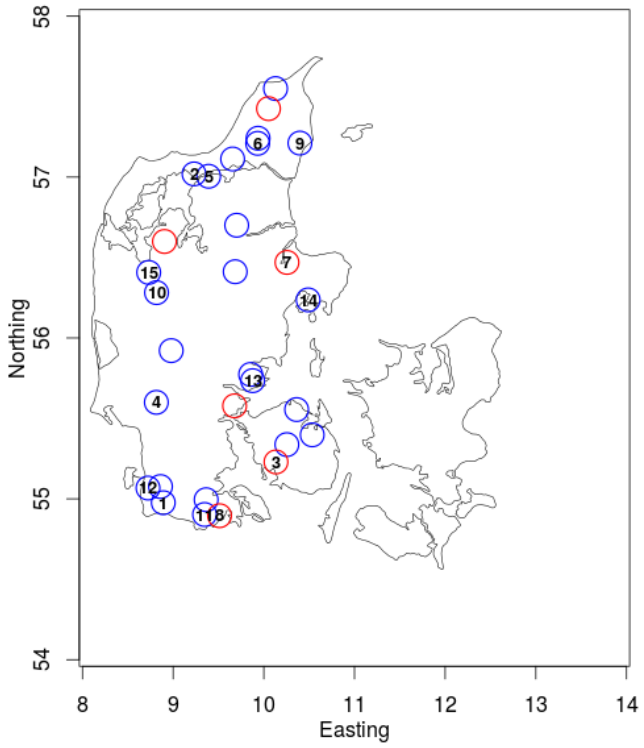


Fig. 1: Locations of the 15 stations the cross-correlations of which are presented in Fig.3. The 24 wind farms that constitute the training set are marked as blue circles and the remaining 6 farms are marked as red circles.

farms being 310 km. A spatial nugget effect, i.e. a discontinuity of the correlation function at $\mathbf{h} = 0$, can also be detected, which could be due either to the sampling resolution or to some measurement errors. When it comes to prediction to new unobesrved locations at future times, the spatio-temporal correlations are of great importance since they contain most of the information related to both the temporal and spatial dependence structure. In Fig. 3, the cross-correlations between wind power generated at the 15 wind farms indicated by numbers in Fig. 1 are presented.

As was already mentioned, wind power is driven by wind and is affected by a number of physical factors that are collectively represented in the weather patterns. In most places on Earth, one can observe prevailing wind patterns, which are due to the movement of weather fronts across that region. These weather fronts usually propagate along certain directions. In the case of Den-

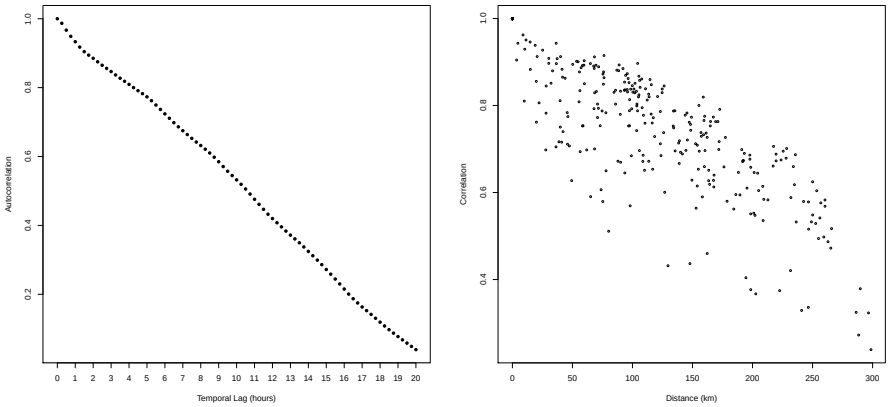


Fig. 2: (*Left:*) Autocorrelation function of wind power production at a single farm labelled as 4 in Fig. 1. (*Right:*) Spatial correlations of wind power production generated at the 24 wind farms of the portfolio, see section 4.1, at temporal lag equal to zero.

mark this direction is from west to east. This basically means that the weather on a westerly location is significantly correlated to the weather at a location to the east of the first location at a future time. As a result, the covariance structure is anisotropic since space is no longer homogeneous over all directions.

Fig. 3 (*Left*) shows the empirical correlations of contemporary wind power production, and Fig. 3 (*Right*) shows the wind power production for a temporal lag of 6 hours, between the fifteen stations indicated in Fig. 1. There is a dependence of the correlation on the direction west to east. For example wind farm 1 is at the west of 14 and the maximum of their correlation is not attained at temporal lag 0, which is the case when the correlation is symmetric in time. This basically means that the correlation between a wind farm $\mathbf{s}_i$ and a wind farm $\mathbf{s}_j$ located at the east of $\mathbf{s}_i$ will be larger at temporal lag τ than at temporal lag $-\tau$. Indeed, the correlation between the wind power production at the farm 10 and farm 3 that is at the east of farm 10, after 6 hours is less than the corresponding correlation between the wind power production at the farm 3 and farm 10 after the same time lag, i.e $(\mathbb{C}ov(Y(\mathbf{s}_{10}, t), Y(\mathbf{s}_3, t+24)) = 0.69 \neq 0.8 = \mathbb{C}ov(Y(\mathbf{s}_3, t), Y(\mathbf{s}_{10}, t+24)))$. This is mainly due to the time needed for the weather fronts to travel from station $\mathbf{s}_3$ to station $\mathbf{s}_{10}$.

It is reasonable to assume that this type of behaviour is also inherited by the underlying Gaussian random field. This suggests that the covariance structure for the underlying field needs to be anisotropic in the spatial dimension and hence since it is not fully symmetric anymore it can no longer be separable, see [14].

When constructing correlation structures for this type of situations, where correlation depends on the orientation, the general idea of a moving Lagrangian reference frame may be applied. Specifically, we may start by considering a purely spatial random field with a stationary correlation structure and suppose that the entire field moves time forward according to some velocity field, see [5], [3] and [1].

In this spirit, [14] suggest for the Irish wind data, see [16] for a description of the data set, a correlation structure that is a convex combination of a static spatio-temporal correlation structure C_{FS} and a Lagrangian correlation function C_{LGR}. The resulting correlation can be formulated as

$$C(\mathbf{h},\tau) = (1-\lambda)C_{FS}(\mathbf{h},\tau) + \lambda C_{LGR}(\mathbf{h},\tau). \tag{2}$$

The suggested choice of correlation functions, that look compatible with the plots in Fig. 3, is

$$C_{FS}(\mathbf{h},\tau) = \frac{1-\nu}{(1+a|\tau|^{2\alpha})}\Big(\exp\Big(-\frac{c||\mathbf{h}||}{(1+a|\tau|^{2\alpha})^{\beta/2}}\Big) + \frac{\nu}{1-\nu}\delta_{\{\mathbf{h}=0\}}\Big), \tag{3}$$

where a and c are nonnegative scale parameters of time and space respectively, the smoothness parameter α and the space-time interaction parameter β take values in $(0,1]$. The purely spatial correlation function $C_{FS}(\mathbf{h},0)$ is of the powered exponential form, and the purely temporal correlation function $C_{FS}(\mathbf{0},\tau)$ belongs to the Cauchy family. The suggestion for the Langrangian correlation function C_{LGR} is the following compactly supported function

$$C_{LGR}(\mathbf{h},\tau) = \left(\Big(1-\frac{||\mathbf{h}-\mathbf{v}\tau||}{d||\mathbf{v}||}\Big)^{\gamma}\right)^{+}, \tag{4}$$

where $||\mathbf{v}||$ is the speed of the weather front, $x^+ = \max\{x,0\}$ and the parameter d controls the range of dependence. Notice that (4) belongs to the class of continuous homogeneous and stationary correlation functions for $\gamma \geq 2$, see [13]. To obtain a random field with realizations that look not as smooth, we have fixed $\gamma = 1$ which corresponds to a non differentiable (in the mean square sense) Gaussian random field. Compactly supported correlation are preferred correlation models since they lead to sparse matrices.

On a different approach, instead of altering the spatial coordinate, [21] started from a purely temporal random field and transformed the temporal lag instead of the spatial. The suggested correlation structure has the following formulation:

$$C(\mathbf{h},\tau) = C_{ST}(\mathbf{h},\tau) + \delta_{\{\mathbf{h}=0\}}C_0(\tau), \tag{5}$$

with the following spatio-temporal

$$C_{ST}(\mathbf{h},\tau) = c_0 \exp\left(-(\alpha||\mathbf{h}||)^{2\gamma}\right)\exp\left(-(\beta\tilde{\tau})^{2\eta}\right)$$

and purely temporal correlation

$$C_0(\tau) = \exp\left(-(\tilde{\beta}\tau)^{2\tilde{\eta}}\right) - c_0\exp\left(-(\beta\tau)^{2\eta}\right).$$

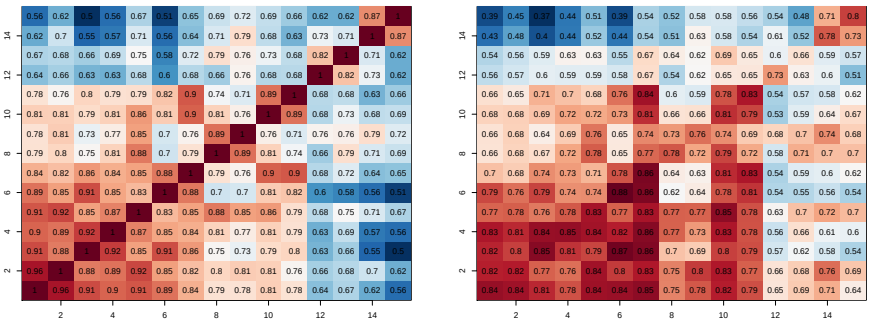


Fig. 3: (*Left:*) Contemporary empirical cross-correlations between the 15 wind farms in Fig. 1. (*Right:*) Empirical cross-correlations with a temporal delay of 6 hours. The correlations were calculated using a data set of length 18 days.

The time is shifted by $(\mathbf{h}\cdot\mathbf{v})/||\mathbf{v}||^2$, the time needed for the weather front to propagate between the two locations, so that

$$\tilde{\tau} = \tau + \frac{\mathbf{h}\cdot\mathbf{v}}{||\mathbf{v}||^2}.$$

The parameters $\alpha, \beta, \tilde{\beta} > 0$ are the scale parameters for space and time respectively, $0 < \gamma, \eta, \tilde{\eta} \leq 1$ are the shape parameters, $0 \leq c_0 \leq 1$ controls the nugget effect C_0. In this case, both the purely spatial $C(\mathbf{h}, 0)$ and purely temporal $C(\mathbf{0}, \tau)$ correlations are of the power exponential form.

Finally, we compare the two previous models that contain a dynamic component, to a relatively simple model that consists of a separable correlation together with a temporal nugget effect, since the same model but with a spatial nugget effect is a part of the model in (2),

$$C^t_{SP}(\mathbf{h}, \tau) = (1-\nu)e^{-c||\mathbf{h}||}\Big(\frac{1}{(1+a|\tau|^{2\alpha})} + \frac{\nu}{1-\nu}\delta_{\{\tau=0\}}\Big). \tag{6}$$

Note that we discuss and fit correlation rather than covariance structures, i.e. we consider the variance is constant and equal to one.

2.1.1 Correlation fitting

The fitting of the correlation models can be achieved in the following way. Suppose that the empirically estimated anamorphosis function is given by $\hat{f}_{\mathbf{s}}$, which we regard as the true anamorphosis function. Using (1), we can obtain estimates of the truncated Gaussian random field, $\widehat{z(\mathbf{s},t)}^{+}$, by applying the

inverse of $\hat{f}_{\mathbf{s}}$ to the observed wind power data. Then, the empirical correlations of the truncated field Z^+ between the two farms indicated by the subscripts and time lag τ can be computed by

$$\widehat{\rho_{ij}}^{+}(\tau) = \overline{\widehat{z(\mathbf{s}_i,0)}^{+} \cdot \widehat{z(\mathbf{s}_j,\tau)}^{+}} - \overline{\widehat{z(\mathbf{s}_i,0)}^{+}} \cdot \overline{\widehat{z(\mathbf{s}_j,\tau)}^{+}}, \tag{7}$$

where the overline indicates sample mean of the corresponding quantity. Next, we assume that estimates of the mean values $\hat{\mu}_i = \mathbb{E}(Z(\mathbf{s}_i))$ are available, and the data have been standardized to have variance one. The correlations $\widehat{\rho_{ij}}(\tau)$ of the Gaussian random field Z can now be retrieved by minimising the expression

$$\widehat{\rho_{ij}}(\tau) = min_{\rho} \left| \widehat{\rho_{ij}}^{+}(\tau) - \mathbb{C}ov(Z(\mathbf{s}_i,0)^+, Z(\mathbf{s}_j,\tau)^+) \right|, \tag{8}$$

where the empirical correlations $\widehat{\rho_{ij}}^{+}(\tau)$ are given by 7. On the other hand, $\mathbb{C}ov(Z(\mathbf{s}_i,0)^+, Z(\mathbf{s}_j,\tau)^+)$ are related to $\rho_{ij}(\tau)$ (the correlation of the random field Z) through the following relation that can be obtained by straightforward but lengthy calculations, see [35] and [4],

$$\begin{aligned} \mathbb{C}ov(Z_i^+, Z_j^+) &= (\mu_i\mu_j + \rho\sigma_i\sigma_j)F(\alpha_i,\alpha_j;\rho) \\ &+ \frac{\mu_i\sigma_j}{(2\pi)^{1/2}} e^{-\frac{\alpha_j^2}{2}} \Phi\left(\frac{\alpha_i - \rho\alpha_j}{b}\right) + \frac{\mu_j\sigma_i}{(2\pi)^{1/2}} e^{-\frac{\alpha_i^2}{2}} \Phi\left(\frac{\alpha_j - \rho\alpha_i}{b}\right) \\ &+ \frac{b\sigma_i\sigma_j}{2\pi} \exp\left(-\frac{\alpha_j^2 + \alpha_i^2 - 2\rho\alpha_i\alpha_j}{2b^2}\right) - \mathbb{E}[Z_i^+]\mathbb{E}[Z_j^+]. \end{aligned} \tag{9}$$

F is the cumulative distribution function (cdf) of a standard bivariate normal with correlation ρ, and $\alpha_i = \mu_i/\sigma_i$ and $b = (1-\rho^2)^{1/2}$. The short-hand notation Z_i^+ has been used instead of $Z(\mathbf{s}_i,0)^+$ and ρ instead of $\rho_{ij}(\tau)$. Since (9) involves the cdf of a normal random variable evaluated at the target correlation, numerical methods need to be employed for the calculations.

Once the correlation estimates are obtained, the model correlation function $C(\mathbf{h},\tau;\boldsymbol{\eta})$ with parameters denoted by $\boldsymbol{\eta}$, is fitted by minimizing the weighted squared errors

$$min_{\boldsymbol{\eta}} \sum_{\tau} \sum_{i \neq j} \left(\frac{\widehat{\rho_{ij}}(\tau) - C(\mathbf{h},\tau;\boldsymbol{\eta})}{1 - C(\mathbf{h},\tau;\boldsymbol{\eta})} \right)^2, \tag{10}$$

for $\mathbf{h} = \mathbf{s}_j - \mathbf{s}_i$. In practice, the indices i and j go through all the different combinations of wind farms, but the temporal lag τ cannot go through the infinite number of time lags, a certain cut-off level τ_0 needs to be chosen so that the summation in (10) is truncated to $\tau = 1, 2, \ldots, \tau_0$.

Alternatively, once could fit the correlation function based on some cross-validation method, i.e., by choosing an empirical cross-validation criterion and then estimating the correlation parameters based on the prediction results of the specific data set and according to some objective each time, see [27]. This direction has not been explored any further.

2.2 Anamorphosis function

In this section we discuss how to model the anamorphosis transformation $f(\cdot)$ that appears in (1). The anamorphosis function in general is a positive monotonic function chosen to obtain a specified marginal distribution for the positive wind power production. The restriction that one needs to have in mind when choosing the transformation $f(\cdot)$ is that the distribution of wind power production is right skewed and bounded between zero and one after normalization.

Using (1), we see that

$$F_{Y|Y>0}(\cdot) = F_{Z|Z>0} \circ f^{-1}(\cdot) \Leftrightarrow f(\cdot) = F^{-1}_{Y|Y>0} \circ F_{Z|Z>0}(\cdot),$$

where the symbol $\circ$ denotes the composition of two functions. $F_{Y|Y>0}$ and $F_{Z|Z>0}$ (for notational simplicity the dependence on location and time has been supressed), are the cdf's of the distribution of Y and Z conditionally that we observe only their positive parts. For Z being a Gaussian random field with mean μ and variance σ^2, the distribution $F_{Z|Z>0}$ is a truncated normal distribution with the first two moments, $\tilde{\mu}$ and $\tilde{\sigma}^2$ given by the following expressions

$$\tilde{\mu} = \mu + \sigma\lambda(\alpha), \quad \tilde{\sigma}^2 = \sigma^2[1 - \delta(\alpha)], \tag{11}$$

where

$$\alpha = -\frac{\mu}{\sigma}, \quad \lambda(\alpha) = \frac{\phi(\alpha)}{1 - \Phi(\alpha)} \quad \text{and } \delta(\alpha) = \lambda(\alpha)[\lambda(\alpha) - \alpha].$$

Given that wind power production data have been normalized, resulting to measurements that fall inside the interval $[0, 1]$, a reasonable model for $F_{Y|Y>0}$ is the beta distribution. In general, transformation $f(\cdot)$ and therefore the shape parameters of the Beta distribution could be space and time dependent, but since only very short time predictions are of interest in this work, the dependence in time is dropped.

2.2.1 Anamorphosis fitting

Anamorphosis fitting consists of two independent components, fitting the Beta distribution and the first two moments of the Gaussian random field, as the moments of the truncated normal distribution depend on them, see (11).

The fitting of the Beta distribution can be achieved at each wind farm located at $\mathbf{s}$, by maximizing the log-likelihood of the positive wind power production,

$$\sum_{\{t:y(\mathbf{s},t)>0\}} \log\Big(\frac{\Gamma(\alpha_{\mathbf{s}} + \beta_{\mathbf{s}})}{\Gamma(\alpha_{\mathbf{s}})\Gamma(\beta_{\mathbf{s}})} y(\mathbf{s},t)^{\alpha_{\mathbf{s}}-1}(1 - y(\mathbf{s},t)^{\beta_{\mathbf{s}}-1}\Big). \tag{12}$$

Under the assumption that data have been standardized to have variance one, estimation of the mean of the truncated normal distribution reduces to

estimation of the mean of the normal distribution. To estimate the mean μ, it is enough to notice that the following relation holds

$$P(Y > 0) = P(Z > 0) = \Phi(\mu), \tag{13}$$

where Φ is the cumulative distribution of a standard (mean 0 and variance 1) normal random variable.

Hence, to estimate the mean μ we numerically invert (13) and observe that the left-hand side of (13) can be approximated as the proportion of time the field $Y(\mathbf{s}, t)$ stays above level 0 over the whole observation time, i.e.,

$$P(y(\mathbf{s}, t) > 0; t = 1, 2, \ldots, T) = \frac{\text{number of observations above level 0}}{T}. \tag{14}$$

3 Kriging

The goal of this section is to provide with background for predicting the wind power production at new unobserved locations and future time points. The *natural* predictor of $Y(\cdot, \cdot)$ would be the conditional distribution of $Y(\mathbf{s}, t)$ given the observations of Y, which can be computed assuming that the distribution of Y is specified. Even if this is the case, this conditional probability density is expressed in terms of high-dimensional integrals, which usually could be calculated using Monte-Carlo methods, see [35]. This is an approach that we will not pursue any further in this article but we leave it for further study.

Instead, we suggest the use of kriging, see [10]. The idea is to transform observations to a Gaussian random field through the "known" anamorphosis function and then standard kriging theory, or as is also known unbiased prediction theory, is applied to the Gaussian random field. The resulting predicted values are transformed back by using the inverse anamorphosis transformation and then truncated. This sounds like a reasonable prediction of the original process except from being biased since in general $\mathbb{E}(f(X)) \neq f(\mathbb{E}((X))$ for nonlinear f.

In the following we shall derive the prediction formulation for discrete space and discrete time domains but it is equally appropriate for continuous space-time formulations also. Suppose we have a set of locations $\{\mathbf{s}_1, \ldots, \mathbf{s}_n\} \subset \mathcal{D}_\mathbf{s}$ where the random field is observed at times $\{0, \Delta t, \ldots, t \cdot \Delta t\} \subset \mathcal{D}_t$. Suppose also that we want to predict the random field Y at a set of $m - n$ locations $\{\mathbf{s}_{n+1}, \ldots, \mathbf{s}_m\}$ at times $\{(t+1) \cdot \Delta t, \ldots, (t+h)\Delta t\}$. The $m - n$ locations can be some or all of the locations at which we have already observations, i.e, $\{\mathbf{s}_{n+1}, \ldots, \mathbf{s}_m\} \subseteq \{\mathbf{s}_1, \ldots, \mathbf{s}_n\}$ or it can be a totally new set of locations, i.e. $\{\mathbf{s}_{n+1}, \ldots, \mathbf{s}_m\} \subseteq \mathcal{D}_\mathbf{s} - \{\mathbf{s}_1, \ldots, \mathbf{s}_n\}$ or even some of the predictor locations are locations with observations and the rest are new unobserved locations.

For simplicity in notation, from now on we shall write $x_{i,j} = (\mathbf{s}_i, j\Delta t)$ and

$$\mathbf{Y}_o = (Y(x_{1,0}), \ldots, Y(x_{1,t}), \ldots, Y(x_{n,0}), \ldots, Y(x_{n,t}))^T$$

the $n(t+1) \times 1$ vector of observed values. the superscript T denotes the transpose. Similarly we define the vector $\mathbf{Z}_o$.

For current time t the target is to obtain predictions $Y(\mathbf{s}, t+h)$ at the prediction locations $\{\mathbf{s}_{n+1}, \ldots, \mathbf{s}_m\}$ and future times $t+h$, where h is the prediction or forecast horizon, also known as lead time. To do so, we shall obtain predictions for the Gaussian random field Z, and then transform them back to predictions for Y. For this, let us denote by

$$\mathbf{Y}_p = \left(Y(x_{n+1,t+1}), \ldots, Y(x_{m,t+h})\right)^T,$$

the $(m-n)h \times 1$ vector consisting of the random field Y at the unobserved locations and times where prediction is desired and by $\mathbf{Z}_p$ the corresponding vector for the Gaussian random field Z.

We shall consider the case of a "known mean" which underlies the theory of simple kriging. In case of a Gaussian random field, the kriging estimator coincides with the conditional expectation $\mathbb{E}(\mathbf{Z}_p|\mathbf{Z}_o)$, which is the ideal estimator of $\mathbf{Z}_p$ in the mean square sense and the error is uncorrelated with every random variable in $\mathbf{Z}_p$ and with the predictor itself, see [8].

Since Z is a Gaussian random field, the vectors $\mathbf{Z}_o$ and $\mathbf{Z}_p$ are jointly Gaussian, with mean and covariance that are given in the next relation

$$\begin{pmatrix} \mathbf{Z}_o \\ \mathbf{Z}_p \end{pmatrix} \stackrel{d}{=} \mathcal{N}\left(\begin{pmatrix} \boldsymbol{\mu}_o \\ \boldsymbol{\mu}_p \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{oo} & \boldsymbol{\Sigma}_{op} \\ \boldsymbol{\Sigma}_{po} & \boldsymbol{\Sigma}_{pp} \end{pmatrix} \right), \tag{15}$$

where $\boldsymbol{\mu}_o = \mathbb{E}(\mathbf{Z}_o)$ and $\boldsymbol{\mu}_p = \mathbb{E}(\mathbf{Z}_p)$ are the mean vectors and $\boldsymbol{\Sigma}_{oo}, \boldsymbol{\Sigma}_{op}, \boldsymbol{\Sigma}_{po}$ and $\boldsymbol{\Sigma}_{pp}$ are covariance matrices with dimensions $n(t+1) \times n(t+1)$, $n(t+1) \times (m-n)h$, $(m-n)h \times n(t+1)$ and $(m-n)h \times (m-n)h$ respectively and defined as follows.

$\boldsymbol{\Sigma}_{oo}$ is a block matrix of the form

$$\boldsymbol{\Sigma}_{oo} = \begin{pmatrix} \mathcal{C}_{11} & \mathcal{C}_{12} & \ldots & \mathcal{C}_{1n} \\ \mathcal{C}_{21} & \mathcal{C}_{22} & \ldots & \mathcal{C}_{2n} \\ \vdots & \vdots & \ddots & \ldots \\ \mathcal{C}_{n1} & \mathcal{C}_{n2} & \ldots & \mathcal{C}_{nn} \end{pmatrix},$$

where $\mathcal{C}_{ij}$ is a $(t+1) \times (t+1)$ covariance matrix of the form

$$\mathcal{C}_{ij} = \begin{pmatrix} C(||\mathbf{s}_j - \mathbf{s}_i||, 0) & C(||\mathbf{s}_j - \mathbf{s}_i||, 1) & \ldots & C(||\mathbf{s}_j - \mathbf{s}_i||, t) \\ C(||\mathbf{s}_j - \mathbf{s}_i||, -1) & C(||\mathbf{s}_j - \mathbf{s}_i||, 0) & \ldots & C(||\mathbf{s}_j - \mathbf{s}_i||, t-1) \\ \vdots & \vdots & \ddots & \ldots \\ C(||\mathbf{s}_j - \mathbf{s}_i||, -t) & C(||\mathbf{s}_j - \mathbf{s}_i||, 1-t) & \ldots & C(||\mathbf{s}_j - \mathbf{s}_i||, 0) \end{pmatrix} \tag{16}$$

for $i, j = 1, \ldots, n$.

The covariance matrix $\boldsymbol{\Sigma}_{pp}$ is also a block matrix consisting of $(m-n)^2$ blocks each one of which is a $h \times h$ matrix of the form of the matrix in (16) but with the $\mathbf{s}_i$ and $\mathbf{s}_j$ running through the locations of the prediction set, i.e. $i, j = n+1, \ldots, m$.

The covariance matrix $\boldsymbol{\Sigma}_{po}$ is defined similarly as a matrix with $n(m-n)$ blocks

$$\boldsymbol{\Sigma}_{po} = \begin{pmatrix} \mathcal{C}'_{n+1,1} & \mathcal{C}'_{n+1,2} & \dots & \mathcal{C}'_{n+1,n} \\ \mathcal{C}'_{n+2,1} & \mathcal{C}'_{n+2,2} & \dots & \mathcal{C}'_{n+2,n} \\ \vdots & \vdots & \ddots & \dots \\ \mathcal{C}'_{m,1} & \mathcal{C}'_{m,2} & \dots & \mathcal{C}'_{m,n} \end{pmatrix},$$

where $\mathcal{C}'_{ij}$ is a $h \times (t+1)$ covariance matrix of the form

$$\mathcal{C}'_{ij} = \begin{pmatrix} C(||\mathbf{s}_j - \mathbf{s}_i||, t+1) & \dots & C(||\mathbf{s}_j - \mathbf{s}_i||, 1) \\ C(||\mathbf{s}_j - \mathbf{s}_i||, t+2) & \dots & C(||\mathbf{s}_j - \mathbf{s}_i||, 2) \\ \vdots & \ddots & \dots \\ C(||\mathbf{s}_j - \mathbf{s}_i||, t+h) & \dots & C(||\mathbf{s}_j - \mathbf{s}_i||, h) \end{pmatrix} \tag{17}$$

for $j = 1, \dots, n$ and $i = n+1, \dots, m$ and $\boldsymbol{\Sigma}_{op} = \boldsymbol{\Sigma}_{po}^T$.

Since the field Z is assumed to be Gaussian the best linear unbiased predictor of $\mathbf{Z}_p$ given the observed vector $\mathbf{Z}_o$ coincides with the conditional expectation $\mathbb{E}(\mathbf{Z}_p|\mathbf{Z}_o)$ and is given by

$$\tilde{\mathbf{Z}}_p = \boldsymbol{\mu}_p + \boldsymbol{\Sigma}_{po}\boldsymbol{\Sigma}_{oo}^{-1}(\mathbf{Z}_o - \boldsymbol{\mu}_o) \tag{18}$$

with kriging covariance matrix

$$\boldsymbol{\sigma}^2 = \boldsymbol{\Sigma}_{pp} - \boldsymbol{\Sigma}_{po}\boldsymbol{\Sigma}_{oo}^{-1}\boldsymbol{\Sigma}_{op}, \tag{19}$$

which depends on the locations and times of the different sampling points. Obviously when the random field is considered stationary this dependence is through the distances in space and temporal lags of the sampling points.

Once the optimal unbiased predictor $\tilde{\mathbf{Z}}_p$ is obtained, then, the "plug-in" but biased predictor of $\mathbf{Y}_p$ would be given by

$$\tilde{\mathbf{Y}}_p = \mathbf{f}_p(\tilde{\mathbf{Z}}_p), \tag{20}$$

where $\mathbf{f}_p = (f_{\mathbf{s}_{n+1}} \times \mathbf{1}, \dots, f_{\mathbf{s}_m} \times \mathbf{1})$ with $\mathbf{1}$ denoting a vector of 1's of size h. We understand the symbol $\mathbf{f}_p(\tilde{\mathbf{Z}}_p)$ in (20) as a pointwise application of the components of the vector transformation $\mathbf{f}_p$, i.e. as $f_{\mathbf{s}_{n+1}}(\tilde{Z}(x_{n+1,t+j}))$, for $j = 1, \dots, h$.

Although the mean μ, the covariance C and the anamorphosis f are unknown, we shall suppose that the estimators $\widehat{\mu}_Z, \widehat{C}_Z$ and $\widehat{\mathbf{f}}_p$ are given (see Sections 2.1.1 and 2.2.1) so that

$$\widehat{\mathbf{Y}}_p = \widehat{\mathbf{f}}_p(\widehat{\mathbf{Z}}_p), \tag{21}$$

where $\widehat{\mathbf{Z}}_p$ is the $\tilde{\mathbf{Z}}_p$ in (18) with the mean and covariance function being replaced by their corresponding estimates.

When $f_{\mathbf{s}}$ and therefore $\mathbf{f}_p$ are nonlinear the predictors in (20) and (21) are biased. Often this bias is significant. [10] suggests, for anamorphosis transformation that is at least twice differentiable, a bias correction obtained from the Taylor expansion of $f_{\mathbf{s}}(\cdot)$ around μ using only the first two derivatives.[30] has shown that this bias correction works well for low order polynomials but fails to correct the bias of most non-linear functions with nonzero higher order derivatives and suggest instead the use of a bias-corrected predictor based on a bootstrap method that produces asymptotically unbiased predictors. Since, in this work the transformation is highly nonlinear (a composition of a beta distribution with the empirical cdf) and the interest is on very short-term predictions, both of the above correction methods are not of any assistance. We have not pursued this any further.

4 Model application

4.1 Danish wind power data

We shall apply the methodology presented in this article to a wind power portfolio consisting of 30 wind farms spread throughout the western part of Denmark, see Fig 1. The portfolio, that is part of a large network consisting of 349 wind farms, was provided by the Transmission System Operator in Denmark, and covers the period January 2006 - March 2012, with the wind power measurements being recorded every fifteen minutes. The wind farms are located at distances that range from 1 up to 310 km.

Different wind farms may have different capacities, so in order to facilitate the analysis, wind power production data has been normalised by dividing individual measurements by the maximum nominal power value of the specific wind farm, resulting in measurements that fall within $[0, 1]$.

4.2 Model fitting

To allow for out-of-sample evaluation of predictive performance, we split the data set into a training set of 24 wind farms and a test set consisting of the remaining 6 farms, see Fig. 1.

Beta distribution: The estimates $\alpha_{\mathbf{s}}$ and $\beta_{\mathbf{s}}$ of the Beta distribution are obtained for each farm separately, by numerically maximizing (12) using the Nelder-Mead method, see [26]. Direct estimates of the Beta parameters are available only at the locations of the wind farms and over the observation time period. When the goal is to perform temporal kriging, it is not unreasonable to assume that these parameters are constant over the short time horizon of the predictions.

In a spatio-temporal setting, to obtain the spatial dependence of the beta shape parameters, we regress both $\alpha_{\mathbf{s}}$ and $\beta_{\mathbf{s}}$ on covariates. Coastal areas are

expected to be more windy than offshore areas, and some of this information is encoded in the geographic coordinates of each location. Hence, a natural choice of covariates is the spatial coordinates of the location of each wind farm. In particular, we have

$$\begin{aligned} \log(\alpha_{\mathbf{s}}) &= \xi_{\alpha,0} + \xi_{\alpha,1} x_1(\mathbf{s}) + \xi_{\alpha,2} x_2(\mathbf{s}) + \varepsilon(\mathbf{s}) \\ \log(\beta_{\mathbf{s}}) &= \xi_{\beta,0} + \xi_{\beta,1} x_1(\mathbf{s}) + \xi_{\beta,2} x_2(\mathbf{s}) + \varepsilon(\mathbf{s}), \end{aligned} \tag{22}$$

where $\mathbf{s} = (x_1(\mathbf{s}), x_2(\mathbf{s}))$ and $\varepsilon(\mathbf{s})$ is an error term distributed as $\varepsilon(\mathbf{s}) \sim N(0, \tau_\alpha^2)$. The log transformation guarantees the parameters take on positive values. The regression coefficients in (22) are estimated by maximum likelihood. The estimated parameters of the beta distribution can be found in Fig. 4. Note that we have also performed regression with more covariates than just the spatial coordinates, such as distance from the coast but it did not improve the performance of the predictions. Another option would be to consider these parameter estimates as a realization of a random field and obtain estimates at new locations in a hierarchical framework or even use kriging to obtain parameters estimates at new locations. An example of the Beta distribution with its mean and precision parameter modelled through regression structures involving fixed and random effects can be found in [20].

Mean function: The mean function $\hat{\mu}(\mathbf{s})$ is obtained at each farm separately, by numerically inverting (13), i.e., $\hat{\mu}(\mathbf{s}) = \Phi^{-1}(\hat{p}_{\mathbf{s}})$, where $\hat{p}_{\mathbf{s}} = P(y(\mathbf{s}) > 0)$ is the proportion of time process Y stays above the level 0, and is computed using (14). The mean $\mu(\mathbf{s})$ depends on the wind farm but it is considered as constant over the times covered by the prediction and observation periods. Direct estimates of the mean are available, as in the case of beta distribution, only at the locations of the farms in the training set.

In a spatio-temporal setting the estimates of the mean function at the locations in the test set, are also obtained by fitting a linear regression with covariates as expressed by the longitude and latitude of each wind farm,

$$\mu(\mathbf{s}) = \xi_{\mu,0} + \xi_{\mu,1} x_1(\mathbf{s}) + \xi_{\mu,2} x_2(\mathbf{s}) + \varepsilon(\mathbf{s}). \tag{23}$$

Correlation structure: Estimates of the inverse anamorphosis transformation are applied to the positive wind power production data to obtain estimates $\widehat{z(\mathbf{s},t)}^{+}$. The correlation functions discussed in Section 2.1 were fitted, using the methodology described in Section 2.1.1 to the data from the training set for $|\tau| \leq 20$ (i.e. 5 hours). The value of $\tau_0 = 20$ has been selected having the following two aspects in mind. First, the computational burden of inverting the $n(t+1) \times n(t+1)$ covariance matrices in (18) for $t \leq \tau_0$ when τ_0 is large. Moreover, when τ_0 is large, the covariance model does not result in reasonable fits for all the possible values of τ_0. To minimize the error function in (10), the R function *optim* with method L-BFGS-B, which is a modification of the quasi-Newton method, has been used. This function allows specification of a lower and upper bound for each variable.

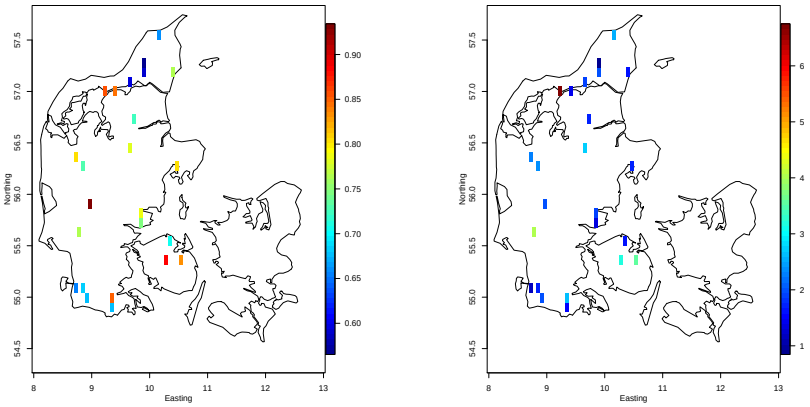


Fig. 4: (*Left:*) The α parameter of the Beta distribution. (*Right:*) The β parameter of the Beta distribution. Both parameters are plotted as functions of the location.

To fit the correlation model in (2), a sequential fitting procedure has been applied, see [14]. The model component (3) was fitted and the parameters were estimated. Then the model component (4) was fitted with the values of the parameters in 3 equal to their estimated values. [14] claim that the fitting quality is the same using this sequential procedure or fitting the whole model simultaneously to the Irish wind data. We have decided to use the sequential fittin due to convergence issues. We have applied this approach for different values of λ (the contribution of each model component to the final model) and the resulting parameters were robust.

For the velocity vector, we have considered three different directions, the horizontal, vertical and the one of the main diagonal, denoted v_x, v_y and v_d respectively. The resulting estimated velocities for (2) are $v_x = 1.2673, v_y = 1.0082$ and $v_d = 1$. The estimated values for the different choices of the λ parameter are almost indistinguishable. The corresponding estimates for (5) are $v_x = 5.0358, v_y = 5.8179$ and $v_d = 5.0376$. As it can be noticed, the velocity estimates for the different directions are almost the same, therefore the prediction results will be presented only for velocity in the horizontal direction. Hence, we shall only consider the case the weather fronts are moving from west to east, which is in accordance to the prevalence of westerly winds over Denmark. The velocity estimates are approximately 1m/s for the model given in (2) and 5m/s for the model given in (5), which are reasonable estimates since the annual average wind speed across Denmark is 5.8 m/s.

The graphs in Fig. 5 illustrate the goodness of fit for the different correlation models and for v_x, the velocity along the west to east direction. In each of the graphs, the fitted correlations are plotted versus the empirical ones, for

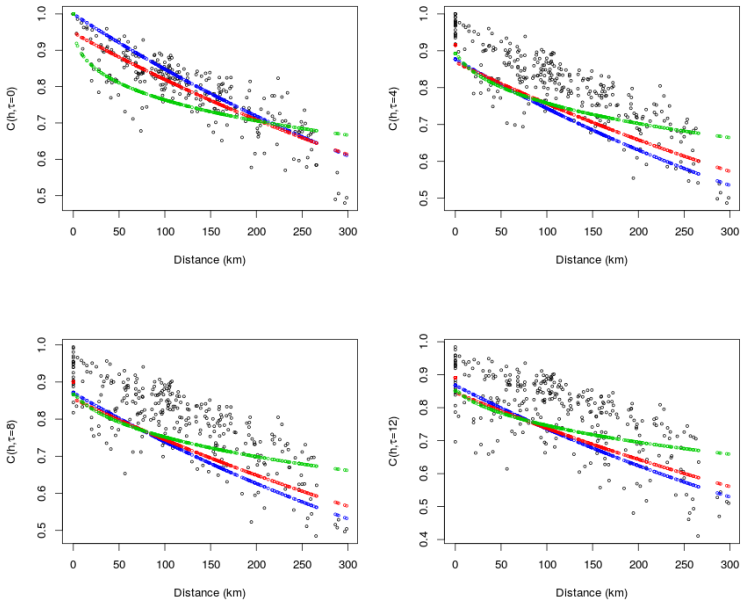


Fig. 5: (*Top, Left:*) Empirical versus fitted correlations at temporal lag zero, (*Top, Right:*) temporal lag equal to three, i.e 1 hour, (*Bottom, Left:*) temporal lag eight, i.e 2 hours and (*Bottom, Right:*) temporal lag equal to twelve, i.e. 3 hours. The lines represent the fitted correlation models given in (6) (blue), (2) (red) and (5) (green).

Eq 2	ν	c	a	α	β	λ	v_x	d
	0.0506	0.0015	0.0705	0.1102	1.0000	0.0019	1.0024	1.0576
Eq 5	α	γ	β	$\tilde{\beta}$	η	$\tilde{\eta}$	c_0	v_x
	0.0001	0.5637	0.0022	0.2315	0.0001	0.1487	1.0000	5.0358
Eq 6	ν	c	a	α				
	0.0506	0.0015	0.0705	0.1102				

Table 1: Estimated parameters of the correlation functions.

all pairs of wind farms and temporal lag zero (*Top, Left*), 1 hour (*Top, Right*), 2 hours (*Bottom, Left*) and 4 hours (*Bottom, Right*).

The estimated parameters of the three correlation structures are gathered in Table 1.

Kriging: As it is obvious from (18) and was mentioned earlier, the kriging predictor coincides with the expectation of $\mathbf{Z}_p$ conditionally on $\mathbf{Z}_o = \mathbf{z}_o$. Estimates $\widehat{\mathbf{z}}_o$ are obtained by the inverse of the estimated anamorphosis transformation

only in the case the corresponding $\mathbf{y}_o > 0$. When $\mathbf{y}_o < 0$ we generate non-positive values of $\widehat{\mathbf{z}}_o$ from a normal distribution with mean $\widehat{\boldsymbol{\mu}}_o$ and variance 1 since we assume from the beginning that the data has been normalized to give variance one, see (15).

5 Prediction performance

To assess the performance of predictions based on the three different correlation structures, we consider time forward predictions at the locations of the training set.

5.1 Temporal kriging

We initiate this analysis by considering $1-8$ time steps forward kriging predictions at the 24 wind farms that constitute the training set. For the correlation function we used $\tau_o = 20$ and in (15) we have used the past 12 observations, i.e. $t = 12$. This choice is based on the fact that the autocorrelation of the wind power production at an individual wind farm drops to 0.1 in about 20 hours and the covariance between the wind power produced at two wind farms of distance 282 km drops to 0.55 in 12 time steps, i.e 3 hours. So $t = 12$ seems like a good compromise between numerical efficiency and making the most of the spatio-temporal correlation. At each location of the training set simple kriging predictors for the wind power production at times $t+1, \ldots, t+8$ were obtained following the procedure that was described explicitly in section 4. The associated kriging variance matrix is given by (19).

In addition to the point predictions provided by (18), we obtain probabilistic prediction by drawing 5000 samples from the multivariate Gaussian probability distribution, $\mathcal{N}(\widehat{\mathbf{Z}}_p, \widehat{\boldsymbol{\sigma}^2})$ with predictive mean vector $\widehat{\mathbf{Z}}_p$ and predictive covariance structure $\widehat{\boldsymbol{\sigma}^2}$ respectively and then translating them to samples for the wind power production field Y following the procedure described in section 4.

Both point and probabilistic forecast scores are used to assess the performance of the proposed method. To quantify and rank the point forecasts at lead times $t = 1, \ldots, 8$, we use the root mean squared error (RMSE), mean absolute error (MAE) and bias which are computed for every time t by

$$\mathrm{RMSE}_t = \sqrt{\frac{1}{24}\sum_{i=1}^{24}\Big(y(\mathbf{s}_i,t) - \hat{Y}(\mathbf{s}_i,t)\Big)^2} \tag{24}$$

$$\mathrm{bias}_t = \frac{1}{24}\sum_{i=1}^{24}\Big(y(\mathbf{s}_i,t) - \hat{Y}(\mathbf{s}_i,t)\Big) \tag{25}$$

$$\mathrm{MAE}_t = \frac{1}{24}\sum_{i=1}^{24}|y(\mathbf{s}_i,t) - \hat{Y}(\mathbf{s}_i,t)|. \tag{26}$$

where $\hat{Y}(\mathbf{s}_i, t)$ is the translation of the kriging predictor $\widehat{\mathbf{Z}}_p$ to the wind power production at location $\mathbf{s}_i$ and time t and $y(\mathbf{s}_i, t)$ is the observed value.

To evaluate the predictive distributions we consider the continuous ranked probability score (CRPS), which is a strictly proper scoring rule for the evaluation of probabilistic forecasts of a univariate quantity [15]. The CRPS is negatively oriented, i.e., the smaller the better, and is defined as

$$\mathrm{CRPS}(F, y) = \int_{-\infty}^{\infty} (F(x) - \delta_{\{x \geq y\}})^2 dx \tag{27}$$

where F is the cumulative distribution function of the density forecast, y is the observed realization.

With the samples available, we can approximate the mean CRPS among the wind farms in the training set for each t by

$$\mathrm{CRPS}_t(F, y) = \frac{1}{n} \sum_{i=1}^{n} \Big(\frac{1}{N} \sum_{j=1}^{N} |\hat{Y}^{(j)}(\mathbf{s}_i, t) - y(\mathbf{s}_i, t)| - \frac{1}{2N^2} \sum_{j,k=1}^{N} |\hat{Y}^{(j)}(\mathbf{s}_i, t) - \hat{Y}^{(k)}(\mathbf{s}_i, t)| \Big), \tag{28}$$

for $n = 24$ and $N = 5000$.

Fig. 6 compares the RMSE, MAE and bias of point forecast and CRPS values using the correlation models provided in section 2.1, at horizons from 15 minutes up to 2 hours ahead. The results are the mean of the performances across all 24 wind farms in the portfolio. In terms of these performance measures, the kriging predictions based on model (5) performed better than the rest, since it has the smallest RMSE, MAE and CRPS and its bias is closer to zero at all forecast horizons. The ranking of the covariance performance was sustained over different data sets. Moreover, a small experiment showed that the data set length does not affect the ranking of the covariance functions according to those criteria. However, the exact impact of the length of the data set used in the covariance estimation on the prediction performance should be investigated further in future work.

We access the predictive distributions in terms of probabilistic reliability with reliability diagrams. Reliability measures how the predictive distributions correspond to the actual observations. For instance, a calibrated prediction with nominal proportion α should cover the observation $\alpha\%$ of the times. The diagrams in Fig. 7 compare the theoretical proportions of a set of quantiles with the observed proportions of the same set using the correlation models described in section 2.1. We use 19 quantiles, from the 5% to the 95% in steps of 5%. The expected fluctuations in the observed frequencies are plotted as consistency bars using the binomial density.

The plots in Fig. 7 correspond to reliability diagrams for prediction at times $t + 1$ and $t + 8$. For perfect calibration, the empirical coverage should lie in the straight line along the diagonal. We can see that all the covariance models perform similarly in terms of reliability, with most of the quantiles within the consistency bars, which indicates relatively well calibrated prediction densities.

The plots Fig. 8 compare the density forecast of each of the 24 wind farms in the training set, which are computed using 5000 samples of $\hat{Y}(\mathbf{s}, t)$ with

the actual observed values at times $t+1$ and $t+8$, respectively. Each curve represents one correlation model: (6) is the curve in blue (2) in red and (5) in green. The vertical line is the observed value of normalized wind power. We can see that at times $t+1$ and $t+8$ the observed normalized wind power is well estimated, since in most of the wind farms, it lies close to the mean of the predictive distributions.

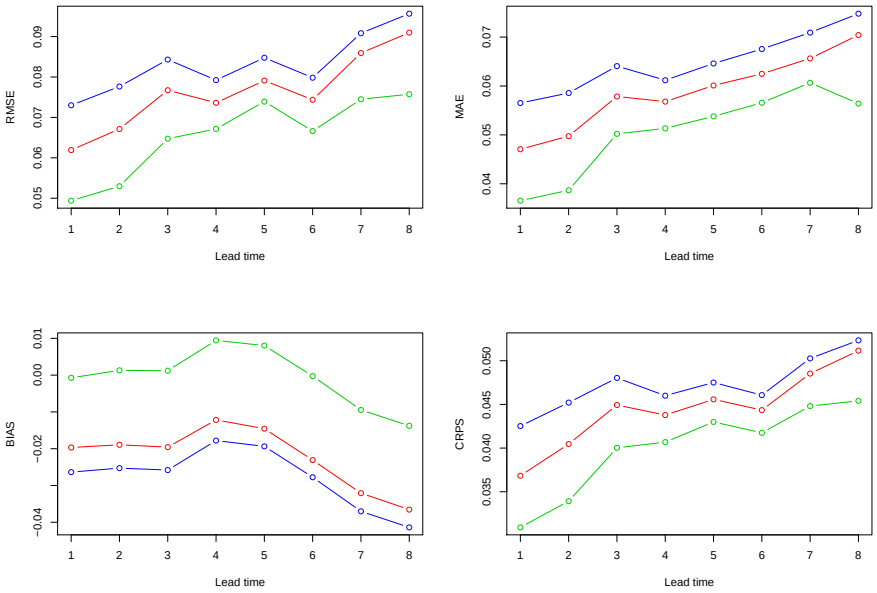


Fig. 6: RMSE, MAE, bias and CRPS for 1 up to 8 time steps ahead kriging predictions of the wind power production at the training set using the correlation models given in (2) (red), (5) (green) and (6) (blue).

5.2 Spatio-temporal kriging

All the predictions considered in section 5.1 are for the wind farms inside the training set. In order to assess the performance of the spatial component of the covariance function, we perform out-of-sample predictions for the locations in the test set. Ideally, parameters estimated using data from the wind farms inside the training set could be readily used to predict the wind power production at a larger portfolio including wind farms with no historical data available.

To obtain out-of-sample predictions, we randomly select 6 wind farms out of the 30 wind farms in the data set, which are indicated as red circles in Fig. 1.

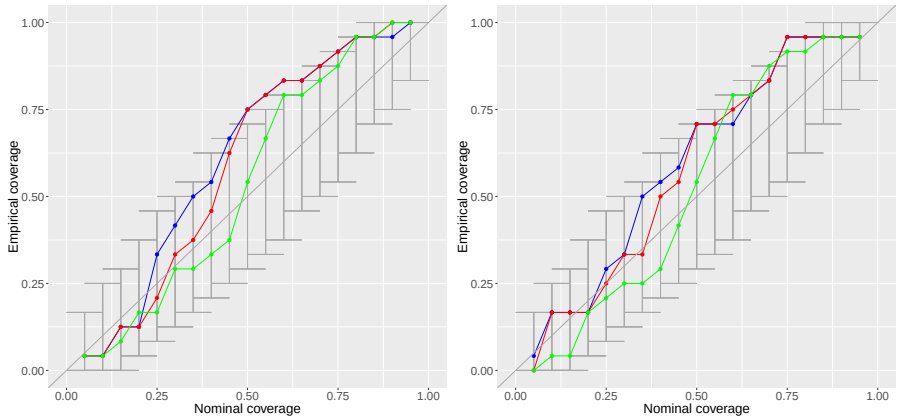


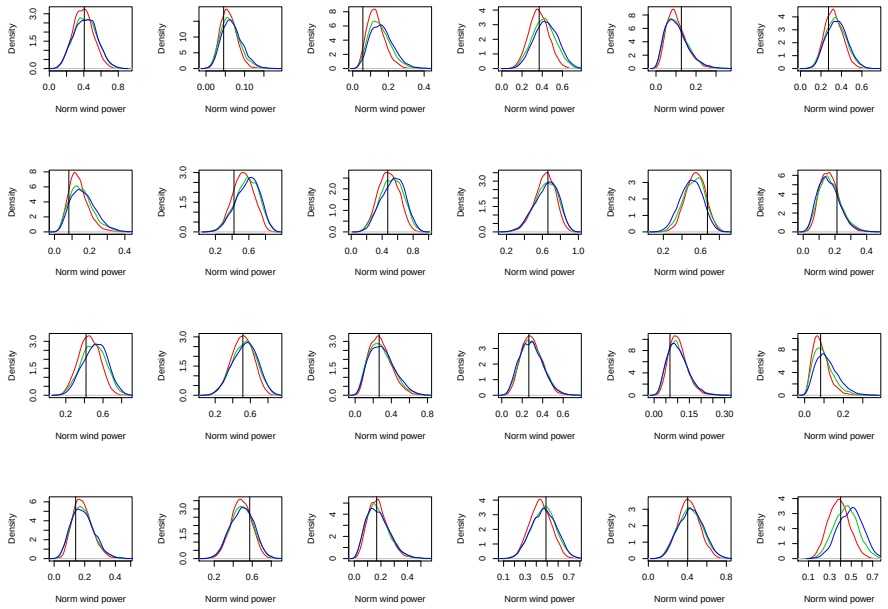
Fig. 7: (*Left:*) Reliability diagram with respective consistency bars for 1 time step ahead kriging predictions of wind power production at the training set. (*Right*) Reliability diagram with respective consistency bars for 8 time steps ahead kriging predictions of the wind power production at the training set. The diagrams were calculated using the correlation models given in (2) (red), (5) (green) and (6) (blue).

We repeat this procedure 15 times. The resulting $15 \times 6 = 90$ out-of-sample predictions are estimated at time horizons ranging from 15 minutes up to 2 hour ahead using the spatio-temporal kriging predictor described in section 4.

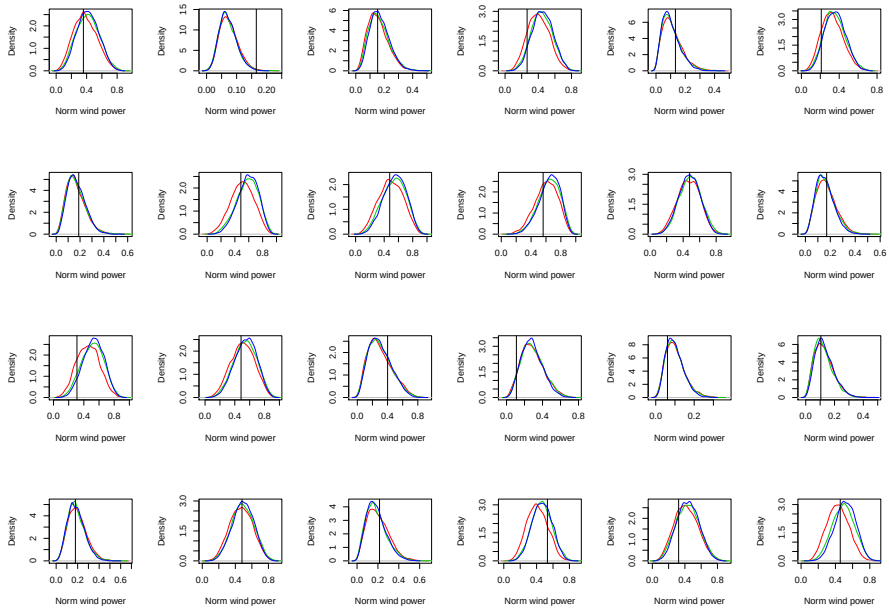
For each prediction horizon, the point prediction performances are ranked under the RMSE, MAE and bias criteria, with the mean taken over the 90 randomly drawn selections of wind farms in the test set. Using the same random samples, we evaluate the performance of probabilistic prediction under mean CRPS, which is described in (28) and reliability diagrams. The out-of-sample prediction densities are obtained as in section 5.1, i.e., by drawing 5000 samples from the multivariate Gaussian probability distribution $\mathcal{N}(\widehat{\mathbf{Z}}_p, \widehat{\boldsymbol{\sigma}^2})$ with predictive mean vector $\widehat{\mathbf{Z}}_p$ and predictive covariance structure $\widehat{\boldsymbol{\sigma}^2}$.

Fig. 9 shows the RMSE, MAE, bias and CRPS scores for the out-of-sample predictions at lead times 1-8 for the three correlation models described in section 2.1. The model in (2) resulted in smaller RMSE, MAE and CRPS scores for all lead times but one. For lead time 1 the model in (6) resulted in a slightly smaller CRPS value. The plots in Fig. 10 correspond to reliability diagrams for prediction at times $t+1$ and $t+8$. The prediction densities corresponding to the correlation model given by (2) outperformed the other two models in terms of calibration at lead time 1. At lead time 8, the correlation model given in (5) seems to be better calibrated than the other two for quantiles larger than 0.7.

It can be argued that the bounded nature of wind power generation may cause the shape of predictive distributions to vary depending on the power level. From Fig. 10, we see that the predictive distributions especially for



(a)



(b)

Fig. 8: Predicted density obtained with the transformed values from the 5000 samples drawn from $\mathcal{N}(\widehat{\mathbf{Z}}_p, \boldsymbol{\sigma}^2)$ at times $t+1$ in (a) and $t+8$ in (b) of the 24 wind farms in the training set. Each plot corresponds to one wind farm and each curve represents one correlation model: (2) (red), (5) (green) and (6) (blue).

models (5) and (6) seem to be too wide on average, since the observed coverages are smaller than the nominal coverages for nominal proportions below 50%. A possible explanation is that the variance is overestimated when wind generation is at a low level, which leads to a slightly underestimated coverage for power values close to zero.

The plots in Fig. 11 show the predictive distribution of the 6 wind farms (indicated in red in Fig. 1) at $t+1$ and $t+8$, respectively, using the correlation models in section 2.1. The vertical line is the observed value of normalized wind power at each wind farm. These plots show that the three models considered in the analysis have similar predictive distributions at times $t+1$ and $t+8$. While the predictive densities of the models given in (2) and (6) have a similar shape, the densities of the predictions generated using model (5) are more narrow around the mode. Moreover, for all three correlation models and for most of the wind farms in the test set, the true normalized wind power generation is well-estimated, since the highest density coincides with the observed values.

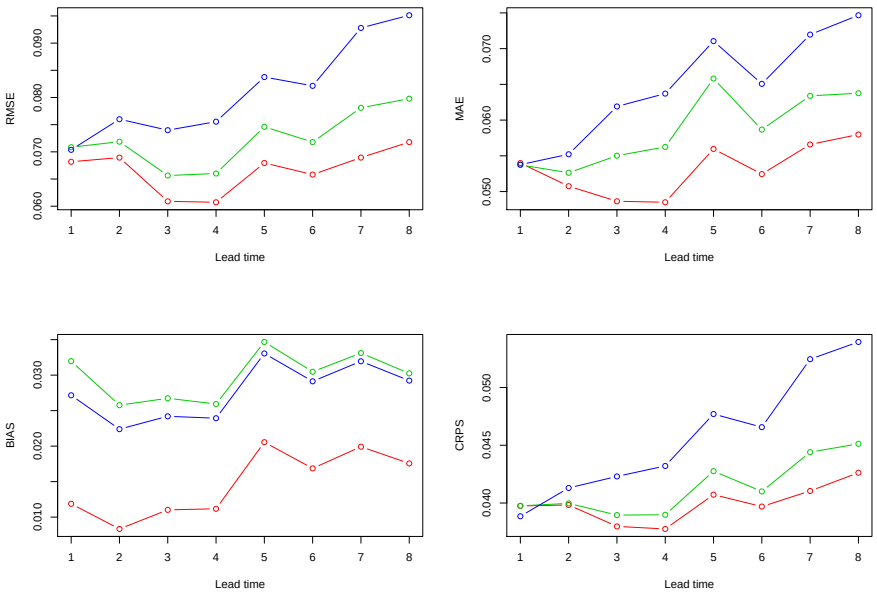


Fig. 9: RMSE, MAE, bias and CRPS for 1 up to 8 time steps ahead kriging predictions of the wind power production at the training set using the correlation models given in (2) (red), (5) (green) and (6) (blue).

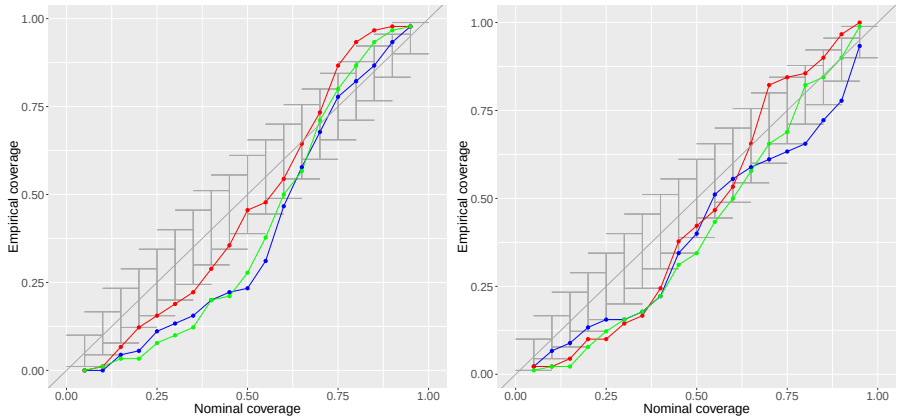
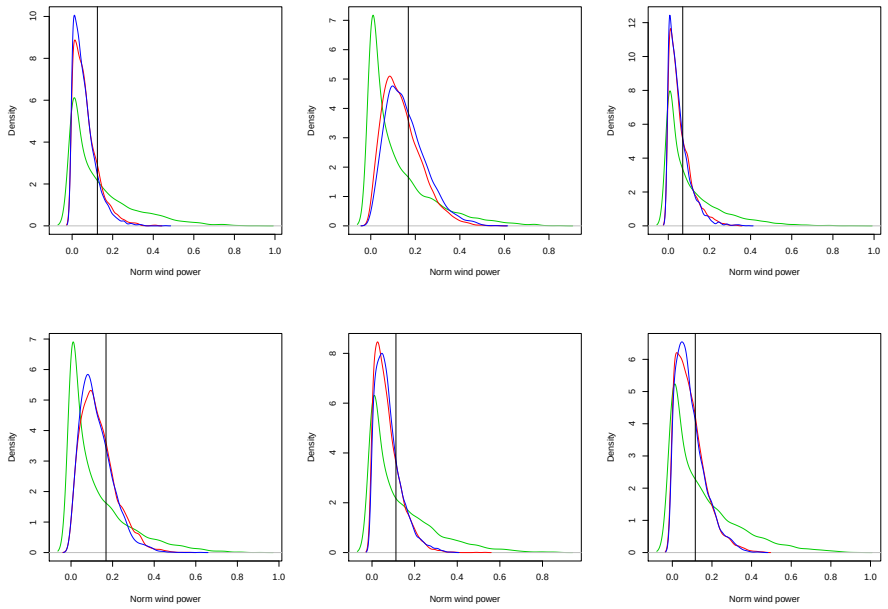


Fig. 10: (*Left:*) Reliability diagram with respective consistency bars for 1 time step ahead kriging out-of-sample predictions of wind power production. (*Right*) Reliability diagram with respective consistency bars for 8 time steps ahead kriging out-of-sample predictions of the wind power production. The diagrams were calculated using the correlation models given in (2) (red), (5) (green) and (6) (blue).

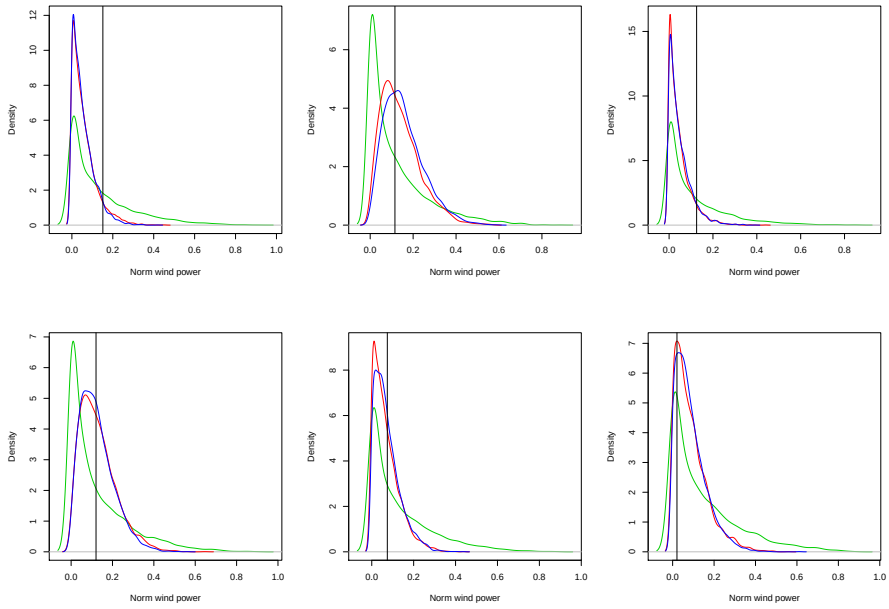
6 Summary and Discussion

In this article we suggest a way to obtain very short-term spatio-temporal predictions of wind power production using what is called kriging equations together with a latent truncated Gaussian random field. Use of the truncated Gaussian field allows to model simulatneously both the mass at zero, and also through an anamorphosis transformation, the heavy tails of the empirical distribution. At any individual wind farm, the wind power production is modelled using a Beta distribution with shape parameters that vary spatially. The spatio-temporal dependence of the Gaussian field is modelled using three different anisotropic covariance models, two of which take additionally into account the propagation of the weather fronts that are, at least partly, responsible for the generation of wind energy. The mean of the Gaussian field is estimated locally by inverting the cumulative distribution function of the proportion of positive wind energy.

The proposed methodology has been applied to a portfolio of 30 wind farms located in western Denmark. 15 minutes up to 2 hours ahead probabilistic predictions of wind power production were obtained at the locations of wind farms and the results were validated against historical data, using different statistical scores, reliability diagrams and predictive distribution. Same type of predictions were also obtained for wind farms that did not contribute to the estimation data set, with comparable results. All covariance models used in the analysis performed reasonably well in terms of predictive densities. However, in terms of both deterministic and probabilistic scores and reliability diagrams,



(a)



(b)

Fig. 11: Predicted density obtained with the transformed values from the 5000 samples drawn from $\mathcal{N}(\widehat{\mathbf{Z}}_p, \boldsymbol{\sigma}^2)$ at times $t+1$ in (a) and $t+8$ in (b) of the 6 wind farms in the test set. Each plot corresponds to one wind farm and each curve represents one correlation model: (2) (red), (5) (green) and (6) (blue).

the use of the richer and more realistic covariance structures that take into account the weather front dynamics outperformed the simple separable model.

For a country like Denmark, that aims making wind power one of the main sources of clean energy in the coming years, the ability to provide with reliable short-term predictions of wind power in a spatio-temporal setting is of utmost importance. This article is a first attempt towards this direction.

Acknowledgements The authors would like to thank Energinet.dk (system operator in Denmark) for kindly offering the data set used in this work. Moreover, we thank the Associated Editor and the two reviewers who provided valuable comments.

References

1. P. Ailliot, A. Baxevani, A. Cuzol, V. Monbet, and N. Raillard. Space-time models for moving fields with an application to significant wave height fields. *Environmetrics*, 22:354–369, 2011.
2. V. Akhmatov. Influence of wind direction on intense power fluctuations in large offshore windfarms in the north sea. *Wind Engineering*, 31(1):59–64, 2007.
3. A. Baxevani, S. Caires, and I. Rychlik. Spatio-temporal statistical modelling of significant wave height. *Environmetrics*, 20:14–31, 2009.
4. A. Baxevani and J. Lennartsson. A spatiotemporal precipitation generator based on a censored latent Gaussian field. *Water Resources Reseach*, 51:4338–4358, 2015.
5. A. Baxevani, K. Podgórski, and I. Rychlik. Dynamically evolving Gaussian spatial fields. *Extremes*, 14(2):223–251, 2011.
6. J. B. Boisvert and C. V. Deutsch. Modeling locally varying anisotropy of CO2 emissions in the United States. *Stochastic Environmental Research and Risk Assessment*, 25(8):1077–1084, 2011.
7. J. A. Carnicero, M. C. Ausín, and M. P. Wiper. Non-parametric copulas for circular–linear and circular–circular data: an application to wind directions. *Stochastic Environmental Research and Risk assessment*, 27(8):1991–2002, 2013.
8. J.-P. Chilés and P. Delfiner. *Geostatistics: Modeling Spatial Uncertainty*. Wiley Series in Probability and Statistics, 2nd edition, 2012.
9. A. Costa, A. Crespo, J. Navarro, G. Lizcano, H. Madsen, and E. Feitosa. A review on the young history of the wind power short-term prediction. *Renewable and Sustainable Energy Reviews*, 12:1725–1744, 2008.
10. N. Cressie. *Statistics for Spatial Data*. Wiley Series in Probability and Statistics, 1991.
11. J. Dowell and P. Pinson. Very-short-term probabilistic wind power forecasts by sparse vector autoregression. *IEEE Transactions on Smart Grid*, 7(2):763–770, 2016.
12. H. Farhangi. The path of the smart grid. *IEEE Power and Energy Magazine*, 8(1):18–28, 2010.
13. T. Gneiting. Correlation functions for atmospheric data analysis. *Quarterly Journal of the Royal Meteorological Society*, 125:2449–2464, 1999.
14. T. Gneiting, M.G. Genton, and P. Guttorp. *Statistics of spatio-temporal systems*, chapter Geostatistical Space-Time Models, Stationarity, Separability, and Full Symmetry, pages 151–175. Monographs in statistics and applied probability. Chapman and Hall/CRC Press,, 2007.
15. T. Gneiting and A. E. Raftery. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378, 2007.
16. J. Haslett and A. E. Raftery. Space-time modelling with long-memory dependence: assessing Ireland's wind power resource (with discussion). *Applied Statistics*, 38:1–50, 1989.
17. A. S. Hering and M. G. Genton. Powering up with space-time wind forecasting. *Journal of the American Statistical Association*, 105(489):92–104, 2010.

18. R. Ignaccolo, J. Mateu, and R. Giraldo. Kriging with external drift for functional data for air quality monitoring. *Stochastic Environmental Research and Risk Assessment*, 28(5):1171–1186, 2014.
19. J. R. Kristoffersen and P. Christiansen. Horns Rev offshore windfarm: its main controller and remote control system. *Wind Engineering*, 27(5):351–359, 2003.
20. B. M. Lagos-Álvarez, R. Fustos-Toribio, J. Figueroa-Zúñiga, and J. Mateu. Geostatistical mixed beta regression: a Bayesian approach. *Stochastic Environmental Research and Risk Assessment*, pages 1–14, 2016.
21. A. Lau. *Probabilistic Wind Power Forecasts: From Aggregated Approach to Spatiotemporal Models*. PhD thesis, Mathematical Institute, University of Oxford, 2011.
22. M. Lei, L. shiyan, J. Chuanwen, L. Hongling, and Z. Yan. A review on the forecasting of wind speed and generated power. *Renewable and Sustainable Energy Reviews*, 13(4):915–920, 2009.
23. G. Li and J. Shi. On comparing three artificial neural networks for wind speed forecasting. *Applied Energy*, 87(7):2313–2320, 2010.
24. S. Lindenberg, B. Smith, K. O'Dell, and E. DeMeo. 20 Percent wind energy by 2030: Increasing wind energy's contribution to U.S. electricity supply. Technical report, U.S. Department of Energy, 2008.
25. M. Milligan, M. Schwartz, and Y. Wan. Statistical wind power forecasting models: Results for U.S. wind farms. In Washington American Meteorological Society Annual Meeting, Seattle, editor, *The 17th Conference on Probability and Statistics in the Atmospheric Sciences*, volume 3, 2004.
26. J.A Nelder and R. Mead. A simplex method for function minimization. *The Computer Journal*, 7(4):308–313, 1965.
27. D. Olivier. Cross validation of kriging in a unique neighborhood. *Journal of the International Association for Mathematical Geology*, 15(6):687–699, 1983.
28. D. Petković, S. Shamshirband, N. B. Anuar, S. Naji, M. L. M. Kiah, and A. Gani. Adaptive neuro-fuzzy evaluation of wind farm power production as function of wind speed and direction. *Stochastic Environmental Research and Risk Assessment*, 29(3):793–802, 2015.
29. P. Pinson. Very-short-term probabilistic forecasting of wind power with generalized logit–normal distributions. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 61(4):555–576, 2012.
30. K Rister and S.N Lahiri. Bootstrap based trans-Gaussian kriging. *Statistical Modelling*, 13(5, 6):509–539, 2013.
31. I. Sanchez. Short-term prediction of wind energy production. *International Journal of Forecasting*, 22:43–56, 2006.
32. D. C.-F. Shih. Wind characterization and potential assessment using spectral analysis. *Stochastic Environmental Research and Risk Assessment*, 22(2):247–256, 2008.
33. P. Sorensen, N. A. Cutululis, A. Vigueras-Rodríguez, L. E. Jensen, J. Hjerrild, M. H. Donovan, and H. Madsen. Power fluctuations from large wind farms. *IEEE Transactions on Power Systems*, 22(3):958–965, 2007.
34. P. E. Sørensen, A. D. Hansen, K. Thomsen, T. Buhl, P. E. Morthorst, L. H. Nielsen, F. Iov, F. Blaabjerg, H. Aa. Nielsen, H. Madsen, et al. Operation and control of large wind turbines and wind farms. final report. Technical report, 2005.
35. M. Stein. Prediction and inference for truncated spatial data. *Journal of Computational and Graphical Statistics*, pages 91–110, 1992.
36. J. Tastu, P. Pinson, E. Kotwa, H. Madsen, and H.A. Nielsen. Spatio-teporal analysis and modeling of short-term wind power forecast errors. *Wind Energy*, 14(1):43–60, 2011.
37. J. Tastu, P. Pinson, and H. Madsen. Space-time trajectories of wind power generation: Parametrized precision matrices under a Gaussian copula approach. In *Modeling and Stochastic Learning for Forecasting in High Dimensions*, pages 267–296. Springer, 2015.
38. J. Tastu, P. Pinson, P.-J. Trombe, and H. Madsen. Probabilistic forecasts of wind power generation accounting for geographically dispersed information. *IEEE Transactions on Smart Grid*, 5(1):480–489, 2014.
39. M. Wytock and J. Z. Kolter. Large-scale probabilistic forecasting in energy systems using sparse gaussian conditional random fields. In *Decision and Control (CDC), 2013 IEEE 52nd Annual Conference on*, pages 1019–1024. IEEE, 2013.