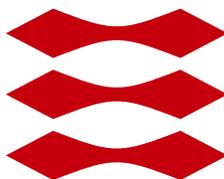


Short-term wind power forecasting: probabilistic and space-time aspects

Julija Tastu

DTU



Kongens Lyngby 2013
PhD-2013-306

Technical University of Denmark
Department of Applied Mathematics and Computer Science
Building 303B, DK-2800 Kongens Lyngby, Denmark
Phone +45 45253031
compute@compute.dtu.dk
www.compute.dtu.dk PhD-2013-306

Summary (English)

Optimal integration of wind energy into power systems calls for high quality wind power predictions. State-of-the-art forecasting systems typically provide forecasts for every location individually, without taking into account information coming from the neighbouring territories. It is however intuitively expected that owing to the inertia in meteorological systems such local approach to power forecasting is sub-optimal. Indeed, errors in meteorological forecasts might translate to fronts of imbalances, i.e. taking the form of a band of forecast errors propagating across entire regions.

The present thesis deals with the proposal and evaluation of new mathematical models and forecasting methods for short-term wind power forecasting, accounting for space-time dynamics based on geographically distributed information. Different forms of power predictions are considered, starting from traditional point forecasts, then extending to marginal predictive densities and, finally, considering multivariate space-time trajectories.

Point predictions is the most classical approach to wind power forecasting, only providing single-valued estimates of the expected future power generation. A statistical model is introduced which improves the quality of state-of-the-art prediction methods by accounting for the fact that forecasts errors made by such locally-optimized forecasting methods propagate in space and in time under the influence of prevailing weather conditions.

Subsequently, the extension from point to probabilistic forecasts is dealt with, hence requiring to describe the uncertainty associated with the point predictions previously generated. Both parametric and non-parametric approaches to form-

ing predictive densities are analysed, while ways to include space-time effects into the corresponding models are presented and evaluated.

As a final step, emphasis is placed on generating space-time trajectories: this calls for the prediction of joint multivariate predictive densities describing wind power generation at a number of distributed locations and for a number of successive lead times. A modelling approach taking advantage of the sparsity of precision matrices is introduced for the description of the underlying space-time dependence structure. Accounting for the space-time dependencies is shown to be crucial for generating high quality scenarios.

In addition to new improved approaches to wind power forecasting, a part of this thesis is devoted to problems related to the assessment of high-dimensional (multivariate) probabilistic forecasts. Namely, the work focuses on the energy score: it illustrates and discusses that this score may be difficult to use owing to its low sensitivity to changes in dependence structures and potentially high uncertainty of the estimates.

Summary (Danish)

Optimal integration af vindenergi i energinet kræver vindenergi-prognoser af høj kvalitet. Avancerede prognosemodeller giver typisk prognoser for hver individuel location uden at tage højde for informationen i de omkringliggende territorier. Det er dog intuitivt forventet på grund af inertien i meteorologiske systemer, at sådan en lokal fremgangsmåde til energiprognostisering er suboptimal. Netop fejl i meteorologiske prognoser kan blive til fronter af ubalancer, dvs. forme et bånd af prognosefejl der udbreder sig henover hele regioner.

Afhandlingen omhandler forslag og evaluering af nye matematiske modeller og prognosemetoder til kortsigtet prognostisering af vindenergi ved at tage højde for rum- og tidsdynamiske effekter, baseret på geografisk distribueret information. Forskellige typer af energiprognoser betragtes, startende fra traditionel punktprognostisering, så udvidet til marginalprognostiserings densitet, og slutteligt betragtes multivariat rumtidsbaner.

Punktprognostisering er den mest klassiske tilgang til vindenergi prognostisering der kun leverer skalare estimater af den forventede fremtidige energiproduktion. En statistisk model introduceres for at forbedre kvaliteten af avancerede prognosemetoder ved at tage højde for det faktum at prognosefejl forårsaget af sådanne lokalt optimeret prognosemetoder, udbreder sig i rum og tid under indflydelse af de fremherskende vejrforhold.

Efterfølgende er udvidelsen fra punkt- til probabilistisk prognostisering behandlet, hvilket kræver en beskrivelse af usikkerheden knyttet til den foregående punktprognostisering. Både parametriske og ikke-parametriske tilgange til dannelsen af prognosedensiteter er analyseret, imens metoder til at inkludere rum-

tidseffekter i de tilsvarende modeller er præsenteret og evalueret.

Som et sidste skridt er vægten lagt på at generere rumtidsbaner; dette kræver prognosen af fælles multivariat prognosedensiteter der beskriver vindenergiproduktion for et antal distribuerede positioner og for et antal på hinanden følgende gennemløbstider. En modellingsstilgang der udnytter præcisionsmatricers tyndhed er introduceret for beskrivelsen af den underliggende rumtidsafhængighedsstruktur. Redegørelse for rumtidsafhængighederne vises at være afgørende for at generere høj kvalitetsscenerier.

Foruden nye forbedrede metoder til vindenergi prognostisering er en del af denne afhandling dedikeret til problemer relateret til evaluering af højdimensionelle (multivariate) probabilistiske prognoser. Navnlige fokuserer arbejdet sig på energimålet: det illustrerer og diskuterer at dette mål kan være vanskeligt at bruge på grund af dets lave følsomhed overfor ændringer i afhængighedsstrukturer og potentielt større usikkerhed for estimerne.

Preface

This thesis was prepared at 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 Engineering.

The thesis deals with different aspects of modelling and forecasting of wind power generation. The main focus is placed on improving the existing state-of-the-art prediction methods by additional incorporation of the space-time dynamics into the models. Special attention is given to probabilistic wind power forecasting. In addition, some theoretical aspects related to forecast verification are addressed.

The thesis consists of a summary report and a collection of 8 research papers written during the study period.

Lyngby, 21-July-2013



Julija Tastu

Acknowledgements

First and foremost I would like to thank my supervisors, prof. Henrik Madsen and prof. Pierre Pinson.

I am very grateful to Henrik for offering me to work with him in the field of wind power forecasting, for all the trust he had in me from the very start of the project.

I must admit I can not imagine how this thesis would look like if I did not have Pierre as a co-supervisor. I feel I have learnt a lot from him and I am very grateful for all the time he spent on guiding me in the field, for his patience with me and for the continuous support he gave me.

Furthermore, I would like to thank all my colleagues at the Department of Applied Mathematics and Computer Science for creating such a wonderful friendly atmosphere. In particular, I would like to thank my office mates Roland and Anne Katrine for all the fruitful discussions, for filling up the candy shelf, for keeping the plants alive and for creating such a cheerful atmosphere in the office. I would also like to thank Ewa and PiJu for their positive attitude and for becoming such an important part of my life not only inside, but also outside the DTU environment.

In addition, I would like to thank all my family for their support, whatever and whenever. In particular, I must thank Yoann for his patience.

List of Publications

Papers Included in the Thesis

- A J. Tastu, P. Pinson, E. Kotwa, H.Aa. Nielsen, H. Madsen (2011). "Spatio-temporal analysis and modeling of wind power forecast errors". *Wind Energy* 14(1), pp. 43-60.
- B J. Tastu, P. Pinson, H. Madsen (2010). "Multivariate conditional parametric models for a spatio-temporal analysis of short-term wind power forecast errors." EWEC'10, European Wind Energy Conference, Scientific Proceedings (peer-reviewed)
- C J. Tastu, P. Pinson, P.-J. Trombe, H. Madsen (2013). "Probabilistic forecasts of wind power generation accounting for geographically dispersed information". *IEEE Transactions on Smart Grid*, in press
- D J. Tastu, P. Pinson, P.-J. Trombe, H. Madsen (2013). "Spatio-temporal correction targeting Nysted Offshore. Probabilistic forecasts". Technical Report
- E J. Tastu, P. Pinson, H. Madsen. "Space-time scenarios of wind power generation produced using a Gaussian copula with parametrized precision matrix". Technical Report
- F P. Pinson, J. Tastu (2013). "Proper evaluation of neural network and learning systems based prediction intervals". submitted to *IEEE Transactions on Neural Networks and Learning Systems*
- G P. Pinson, J. Tastu (2013). "Discussion of *Prediction intervals for short-term wind farm generation forecasts* and *Combined nonparametric prediction*

intervals for wind power generation". submitted to *IEEE Transactions on Sustainable Energy*

H P. Pinson, J. Tastu. "Discrimination ability of the Energy Score". Technical Report

Other Publications not Included

The following publications have also been prepared during the course of the Ph.D. study. They are omitted in this thesis because they are covered by other papers included in the dissertation.

I J. Tastu, P. Pinson, H. Madsen (2012). Spatio-temporal correction of wind power probabilistic forecasts. 11th International Workshop on Large-Scale Integration of Wind Power and Transmission Networks, Lisbon, Portugal

J J. Tastu, P. Pinson, H. Madsen (2012). Accounting for spatio-temporal effects when forecasting wind power generation. CompSust'12, Third conference on Computational Sustainability, Copenhagen, 2012

Contents

Summary (English)	i
Summary (Danish)	iii
Preface	v
Acknowledgements	vii
List of Publications	ix
I Summary Report	1
1 Introduction	3
1.1 Thesis Objective	5
1.2 Thesis Outline	8
2 State-of-the-art in wind power forecasting	11
2.1 Basic concepts	12
2.2 Point forecasts	14
2.3 Probabilistic forecasts	17
2.4 Scenarios	20
2.5 Forecast verification	21
2.5.1 Evaluating point forecasts	22
2.5.2 Evaluating probabilistic forecasts	24
2.5.3 Evaluating scenarios	28

3	Thesis Contribution	31
3.1	Point forecasts	31
3.1.1	Prediction models with off-site data input	32
3.1.2	Space-time correction of wind power point predictions	33
3.1.3	Result using aggregated data	34
3.1.4	Result using a single wind farm	35
3.2	Probabilistic forecasts	37
3.2.1	Considering parametric predictive densities	38
3.2.2	Considering non-parametric predictive densities	39
3.2.3	Results	39
3.3	Space-time trajectories	41
3.4	Forecast verification	44
4	Conclusions and Perspectives	49
4.1	Conclusions	49
4.2	Perspectives	52
II	Papers	55
A	Spatio-temporal analysis and modeling of short-term wind power forecast errors	57
A.1	Introduction	60
A.2	Case Study and Available Data	62
A.3	Highlighting some Spatio-temporal Characteristics of Wind Power Forecast Errors	66
A.3.1	Dependency within the groups	67
A.3.2	Spatio-temporal dependencies between the groups	69
A.3.3	Dependency on wind direction	70
A.3.4	Dependency on wind speed	73
A.4	Proposal of Relevant Models	75
A.4.1	Linear models	76
A.4.2	Regime-switching models based on wind direction	77
A.4.3	Conditional parametric models with regime-switching	79
A.5	Application Results	81
A.5.1	Model fitting	82
A.5.2	Cross-Validation and generalization ability	85
A.6	Conclusions and Perspectives	86
A.7	Acknowledgments	88
B	Multivariate Conditional Parametric models for a spatio-temporal analysis of short-term wind power forecast errors	93
B.1	Introduction	95
B.2	Case study and Available Data	97

B.3	Point forecasts	99
B.3.1	The VAR model	99
B.3.2	The CP-VAR model	100
B.3.3	Estimation in the CP-VAR model	101
B.3.4	Assessment of the point forecasts	103
B.4	Probabilistic Forecasts	105
B.4.1	Assessment of the probabilistic forecasts	107
B.5	Conclusions	109
B.6	Acknowledgement	110
C	Probabilistic forecasts of wind power generation accounting for geographically dispersed information	115
C.1	Introduction	117
C.2	Wind power forecasting in a smart grid context	120
C.3	Methodology	121
C.3.1	Parametric predictive densities	122
C.3.2	Non-parametric predictive densities	127
C.4	Data	130
C.5	Empirical results	132
C.5.1	Model notation	132
C.5.2	Computational details	132
C.5.3	Overall evaluation	133
C.5.4	Conditional evaluation	135
C.6	Conclusions	137
C.7	Acknowledgement	138
D	Spatio-temporal correction targeting Nysted Offshore. Probabilistic forecasts	145
D.1	Introduction	147
D.2	Data and previous work	150
D.3	Parametric approaches to probabilistic forecasting	151
D.3.1	Censored Normal distribution	152
D.3.2	Generalized Logit-Normal distribution	154
D.3.3	Estimation of the location parameter	155
D.3.4	Estimation of the scale parameter	157
D.4	Non-parametric approach to probabilistic forecasting	166
D.5	Methods for probabilistic forecasts assessment	169
D.6	Results	172
D.6.1	Notation	172
D.6.2	Accuracy and skill assessment	174
D.6.3	Reliability assessment	181
D.6.4	Conditional operation	183
D.6.5	Demonstration of the operation of the best-performing model	189

D.7	Conclusions	193
E	Space-time scenarios of wind power generation produced using a Gaussian copula with parametrized precision matrix.	197
E.1	Introduction	199
E.2	Data	202
E.3	Methodology	203
	E.3.1 Preliminaries and definitions	203
	E.3.2 Copulas for wind power data	205
	E.3.3 Gaussian Copula	207
	E.3.4 Modelling as a conditional autoregression	208
E.4	Parametrization of the precision matrix	210
	E.4.1 Structure of the diagonal elements	211
	E.4.2 Structure of the standardized precision matrix	213
	E.4.3 Final parametrization of the precision matrix	217
E.5	Estimation	220
	E.5.1 The valid parameter space	220
	E.5.2 Choosing an appropriate optimization criterion	221
	E.5.3 Parameter estimation using maximum likelihood	221
E.6	Results	223
	E.6.1 Assessing global model fit	223
	E.6.2 Assessing predictive model performance	225
	E.6.3 Scenario generation	228
E.7	Conclusions	229
F	Proper Evaluation of Neural Network and Learning Systems based Prediction Intervals	237
F.1	Introduction	239
F.2	Proper scores for prediction intervals	240
F.3	The Coverage Width-based Criterion is not a proper score	242
	F.3.1 Definition of the Coverage Width-based Criterion	242
	F.3.2 Why is the CWC score not proper?	243
F.4	Illustrating the consequences of not using a proper score	244
F.5	Conclusions	245
G	Discussion of "Prediction intervals for short-term wind farm generation forecasts" and "Combined nonparametric prediction intervals for wind power generation"	249
H	Discrimination ability of the Energy score	257
H.1	Introduction	259
H.2	Discrimination ability of the Energy score	261
	H.2.1 General setup	261
	H.2.2 From propriety of scoring rules to their discrimination ability	261

H.2.3	Characterizing the discrimination ability of the Energy score	262
H.2.4	Illustrating the discrimination ability of the Energy score for multivariate Gaussian processes	263
H.3	Some theoretical results on the discrimination ability of the En- ergy score	268
H.3.1	An upper bound on the discrimination ability of the En- ergy score in the multivariate Gaussian case	271
H.4	Discussion	272
H.4.1	Necessary calculations to evaluate $\overline{\text{Es}}_i$	273
H.4.2	Necessary calculations to evaluate $\overline{\text{Es}}^*$	276
H.4.3	Some results on relevant distributions and integrals	278
Bibliography		283

Part I

Summary Report

Introduction

Renewable energy in general and wind energy in particular have been growing rapidly in the last decade, becoming a more and more important component of the global energy supply. A catalyst for such growth has been growing energy demand, spiralling fossil fuel prices and an acute necessity to reduce carbon dioxide emissions. From the current perspective all the factors which led to the growth of wind energy in the last decades will continue advancing it in the future. Global energy demand is expected to keep growing, even under the declared intentions to increase energy efficiency (use energy wiser)[1]. Fossil fuel prices are expected (under a relatively optimistic scenario) to stay at least as high in the future [2]. Global warming continues calling for significant reductions in carbon dioxide emissions. In addition, the recent Fukushima disaster has led to a new wave of serious debate on the safety of nuclear energy, making it somewhat undesirable in the forward-looking policies. All this makes wind energy a very attractive alternative, which is expected to keep growing significantly in the years to come [3].

Already now wind energy meets 5.5% of the EU's electricity consumption. Denmark is the leader with about 30 % of national electricity demand coming from wind. According to a new Danish energy agreement this number should raise up to 50 % by 2020 [4] and up to 100% by 2035. In the long run Denmark has set an ambitious plan to become independent of fossil fuels (not only in terms of electricity) by 2050. This calls for significant changes to the existing energy

system.

Historically electricity was produced in large central plants located close to the available resources, cities or industrial areas and then transported on to the consumers through the transmission and distribution networks. This has already changed, as thousands of wind turbines have been installed throughout Denmark. Thus, power systems are moving from the traditional centralized systems towards distributed power generation.

Electricity is a highly perishable commodity – it has to be used at the same instant as it is produced and cannot be stored in the grid. Thus, the core principle of power system management is to ensure the balance between power generation and consumption at all grid points and at any time. In conventional power systems this is achieved by letting the power supply follow the demand. That is, the electricity users increase and decrease their consumption whenever they need to do so. Central power stations are then dispatched to provide the power to meet the demand.

In this respect, wind is different from the conventional energy sources. Wind energy cannot be scheduled at will – it is produced when the wind blows and cannot be produced when it does not. Wind energy, thus, inherits the variable, stochastic nature of wind. Due to that, wind power generation (like solar, wave, tidal) is often referred to as stochastic.

When there is a relatively small penetration of wind energy into power systems, the uncertainties in the corresponding power generation, instead of being modelled and accounted for in any special manner, are simply treated as additional uncertainties on the demand side. That is, the conventional power stations are dispatched to cover for the additional variability. However, this calls for more reserves, which reduces the potential environmental and economic benefits offered by wind power.

A number of ideas can be considered to mitigate the increased uncertainty in power systems stemming from large scale integration of wind energy:

1. *A strong power grid.* One way to balance the power supply and the demand is to export/import power to/from an interconnected grid [5], not disregarding a possibility of that grid also being subject to the same power balance issue. An ambitious and innovative proposal is given by the idea of creating a *supergrid*. It is presented in [6] where G. Szisch claims that even if only currently available technologies at their current prices are used, a High Voltage Direct Current transmission grid across Europe, once installed, would guarantee a 100% renewable electricity supply with costs of

such electricity lying not far above today's costs.

2. *Energy storage.* For the moment there are no cheap large batteries available for wind energy storage. However, electricity can be transformed to other energy sources and stored correspondingly. Traditionally, this is done via pumped storage or heat pumps. Research into other technologies (compressed air energy storage, flow batteries, hydrogen,) is ongoing. Information on the existing storage systems can be found in [7] and references therein.
3. *Reserves.* Following [8], when wind power penetration is 10% of gross consumption, the extra reserves needed are in the order of 2-8% of the installed wind capacity. The total requirement depends on the grid particularities and quality of wind power forecasts. At higher penetration levels more reserves are needed.
4. *Demand-side management* could make it possible to transfer energy consumption to the times when renewable energy sources are plentiful. Electric vehicles, domestic micro Combined Heat and Power (CHP) units and heat pumps could act as a vast electricity storage facility.
5. *Wind power forecasting.* Using forecasts as input to decision making problems in power grid operations is not a new concept. In particular, TSOs have a long history of using load forecasts in their decision making [9, 10, 11, 12]. Thus, introduction of wind power forecasts has been relatively smooth and well accepted. Already today wind power forecasts are widely used by many electrical utilities and are acknowledged to reduce operating costs [13, 8, 14].

Wind power forecasts do not provide the solution by themselves. However, being used as a key input to various decision making processes related to power system operations and participation in electricity markets, they comprise a necessary and cost-effective element required for the optimal integration of wind power into energy systems [15, 16]. Quality of the forecasts is very important [16] and thus improving prediction systems' performance has been set as one of the priorities in wind energy research needs for the period 2000-2020 [17].

1.1 Thesis Objective

The aim of this thesis is to contribute to advancing the frontier of wind power forecasting by improving the quality of the existing state-of-the-art prediction

systems. The principal approach for achieving this goal is to consider the spatio-temporal characteristics of wind power forecast errors.

More specifically, the initial idea of the work stems from the fact that operational state-of-the-art wind power prediction systems are optimized for each and every location individually (let it be a single wind farm or an aggregated portfolio of wind farms), without properly accounting for the space-time interdependence structure in the wind power generation field. That is, traditional inputs to prediction models consist of on-site observations (wind power measurements, wind speed and direction) and/or meteorological forecasts. Information coming from the neighbouring territories is not adequately considered as it is assumed that the space-time dynamics are captured by the meteorological predictions used as input. However, given a wide geographical spread of wind farms and owing to the inertia of meteorological systems, it is expected that the errors of such locally optimized forecasts would exhibit a certain dependence pattern in space and in time. This renders the state-of-the-art forecasting systems suboptimal.

The purpose of this research is to analyse such underlying dependence patterns and to account for them when deriving and examining new improved models and methods for wind power forecasting. Wind power forecasts of different types are of interest: starting with classical point predictions, then moving towards univariate probabilistic forecasts describing wind power generation at a single location for a given lead time and, finally, considering multivariate space-time trajectories.

The first step in this work consists of a preliminary examination of data to illustrate that wind power forecast errors do indeed have spatio-temporal characteristics. This is presented in Paper A. The results based on a conditional cross-correlation analysis show that forecast errors propagate in space and in time under the influence of forecasted wind speed and direction.

Following the results presented in Paper A, further research aims at deriving models which could capture the spatio-temporal dependence structure in order to improve the quality of the related wind power forecasts.

First, focus is on point forecasts which comprise a classical form of wind power predictions given by a single-value estimate of the expected future power generation for each location and each look-ahead time. Paper B proposes a methodology for improving the quality of the state-of-the-art point predictions by capturing the residual interdependence structure observed between forecast errors made by the locally optimized systems at a number of distributed locations. Conditional Parametric Vector AutoRegressive (CP-VAR) models are considered in the study. This is a new type of model based on the extension of ordinary Conditional Parametric (CP) models to a multivariate framework. CP

models comprise a class of linear models for which the coefficients are replaced by smooth functions of other variables. In our case the coefficients are described as non-parametric functions of wind direction, hence accounting for weather-driven pattern of error propagation. Existing estimation techniques (adaptive recursive least squares) are extended to CP-VAR models, therefore accounting for slow variations in process dynamics.

Owing to the complexity of the decision-making tasks related to integration of wind energy into power systems, primary interest has recently moved from classical point forecasts to probabilistic ones. For continuous stochastic variables (such as wind power generation), probabilistic forecasts are optimally given in the form of predictive densities. If focus is placed on a univariate stochastic process only, i.e. if the interest is in describing wind power generation at a single location for a particular lead time, then marginal predictive densities are required. In a more general case, if aiming to describe wind power generation at a number of locations over a period of time, then probabilistic forecasts are optimally issued in the form of multivariate (joint) predictive densities which describe both the marginal densities and the dependence structure.

Further in this study focus is placed on marginal predictive densities for wind power generation, hence requiring description of the uncertainty associated with the point predictions previously generated. This problem is addressed in Papers C and D where both parametric and non-parametric approaches to shaping the uncertainty are analysed, while ways to include space-time effects into the corresponding models are presented and evaluated.

Subsequently, emphasis is placed on generating space-time trajectories (also referred to as scenarios), which calls for prediction of multivariate densities describing wind power generation at a number of distributed locations and for a number of successive lead times. The main feature of scenarios, which distinguishes them from ordinary probabilistic forecasts, is given by the fact that, in addition to appropriate probabilistic description of power generation at each location and each look ahead time, the scenarios ought to respect spatial and temporal dependencies in the power generation field. For instance, if the power generation at a given time at a chosen location exceeds the expected value, then it is very likely that the corresponding power measurements at nearby sites around the same time are also higher than expected. One of the goals of the thesis is to propose a methodology for issuing space-time trajectories for wind power generation. The related task is to examine the structure of the underlying space-time dependence and to propose an adequate parametrization for describing it. This task is addressed in Paper E.

An important aspect to mention is that all the presented models are developed with their practical applicability in mind. Case studies have been conducted

considering real-life limitations and conditions, so that the performance of the resulting models reflects the performance that would be achieved in real world operations. To ease the computational load and to account for slow changes in the process dynamics which are hard to model deterministically (e.g., dirtiness of the blades), priority has been on recursive and adaptive estimations schemes (where possible).

In addition to new approaches to wind power forecasting, a part of the study is devoted to problems related to evaluation of probabilistic forecasts of a very high dimension. Namely, emphasis is on the Energy score, which is one of the lead criterion for evaluating probabilistic forecasts of multivariate quantities. The work documented in Paper H illustrates that this score may be difficult to use in practice owing to its low sensitivity to changes in dependence structure and the potentially high uncertainty of the estimates.

1.2 Thesis Outline

The thesis is structured as follows. Part I introduces and summarizes the papers. Within this part, Chapter 2 introduces different aspects that constitute wind power forecasting and briefly presents different research paths that have been explored as of now. Chapter 3 comprises a summary of the main results obtained in the papers. Finally, Chapter 4 concludes Part I.

Part II is a collection of publications including the following papers:

Paper A is a journal article published in *Wind Energy*. It comprises a preliminary examination of data illustrating that wind power forecast errors do indeed have spatio-temporal characteristics.

Paper B is a peer-reviewed paper published in the *Proceedings of the European Wind Energy Conference, EWEC, 2010*. It deals with the spatio-temporal correction of wind power point forecasts.

Paper C is a journal article accepted for publication in *IEEE Transactions on Smart Grid, Special Issue on Analytics for Energy Forecasting with Applications to Smart Grid*. The paper deals with univariate probabilistic forecasts of wind power generation accounting for geographically dispersed information.

Paper D is a technical report which deals with univariate probabilistic forecasts of wind power generation. It can be viewed as a complement to Paper D,

since it describes some alternative approaches and models which were considered, but not included in the journal article, as the quality of their predictive performance was found to be not satisfactory enough.

Paper E is a technical report which deals with multivariate probabilistic forecasts, i.e. with space-time trajectories of wind power generation.

Paper F is a note submitted to *IEEE Transactions on Neural Networks and Learning Systems* which illustrates a methodological error in the CWC score which deems the score not valid for the assessment of prediction intervals.

Paper G is a discussion paper submitted to *IEEE Transactions on Sustainable Energy* which provides an additional discussion on the CWC score.

Paper H is a technical report which deals with sensitivity analysis of the Energy score.

CHAPTER 2

State-of-the-art in wind power forecasting

Before presenting the actual contribution of this thesis to the field of wind power forecasting, this chapter aims at giving a brief overview of the existing research results in the field of wind power forecasting. The objective here is not to give a thorough literature review on the subject, but rather to introduce different aspects that constitute wind power forecasting and to illustrate different research paths that have been explored as of now. The chapter, thus, comprises a short summary of the base knowledge the thesis has been built on. From this, the methods and approaches presented in the work can be better understood.

Section 2.1 presents the basic concepts of wind power forecasting. The origins of variability and predictability of wind power are discussed. The deterministic power curve model describing a conversion of wind to electric power is presented. The implications of the power curve shape on wind power variability are also explained.

In Section 2.2 we introduce different approaches to wind power forecasting and motivate the choice of the state-of-the-art prediction system used as the foundation and the principle benchmark for the methods and models developed in this study.

Section 2.3 discusses probabilistic wind power forecasts. Finally, the chapter finishes with Section 2.4 describing the important aspects of forecast verification.

2.1 Basic concepts

A wind turbine converts the kinetic energy of the wind into electric energy. The amount of power the turbine produces is directly dependent on the wind speed. The pattern of this dependency is described by a characteristic curve, also referred to as the wind turbine power curve. The shape of the power curve also depends on the generator, on the power electronics installed, as well as on the built-in control systems. The reader interested in the current status of generators and power electronics used in wind turbines is referred to [15].

Even though the turbine type affects the shape of the power curve, roughly all power curves are very similar in principle, since they are governed by the same law of physics. Figure 2.1 depicts the typical shape of a power curve.

The power curve can be split into four distinctive parts. For wind speeds below the cut-in value the turbine does not produce any power. Power production starts as the wind speed reaches the cut-in value. Further on, power generation augments sharply and reaches the nominal turbine capacity at the rated wind speed value. From the rated to the cut-off wind speed, the power production is fairly constant. For wind speeds higher than the cut-off value, the turbine stops for safety reasons and no power is produced. Some of new wind turbines have a "smooth cut-off" which means that the power does not drop abruptly when the wind speed reaches the cut-off value, but is instead reduced gradually.

The increase in power production for wind speeds between the cut-in and the rated values can be explained by the physics of the energy conversion process. It can be shown that the kinetic energy of a cylinder of air of radius R travelling with a constant wind speed v corresponds to a total wind power P_{total} within the rotor swept area of the wind turbine. This power can be expressed by:

$$P_{total} = \frac{1}{2} \rho \pi R^2 v^3 \quad (2.1)$$

Where ρ is the air density, R is the rotor radius and v is the wind speed. In reality, however, it is impossible to extract all the energy from the moving air,

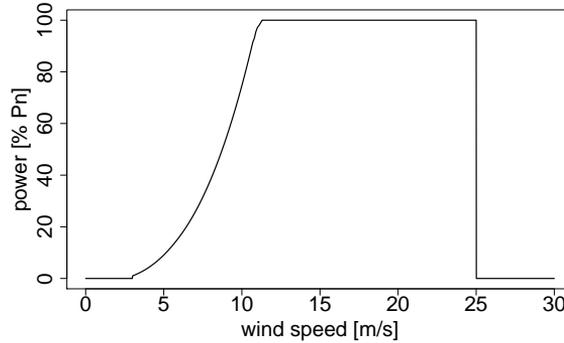


Figure 2.1: An example of an idealized power curve describing wind power generation by a single wind turbine as a function of wind speed

but rather only a fractions of it. The theoretical upper limit for this fraction is given by the Betz's limit, $C_p = 16/27$ (approximately 0.593). In practice the performance ratio reaches 0.52 – 0.55 when measured at the hub of the turbine. However, this is not taking into account the losses in the gear and the generator. If such losses are deducted, then the resulting performance ratio is in the range of 0.46-0.48 [15].

Since wind power generation is a function of wind speed, fluctuations in wind speed translate to changes in wind power generation. Fluctuations in wind speed are observed on different time scales. In this work the main interest is on short-term power forecasting, which means that the time scales of interest are in the order of hours. Fluctuations in wind speed on these time scales (from minutes to hours) fall into a part of spectrum which separate turbulent flow from the mean flow [18]. The corresponding wind speed volatility is governed by the atmospheric stability, time of the year, large cumulus clouds and rain events [18]. The fact that wind speed volatility patterns are not constant in time, but change depending on various meteorological conditions, emphasizes the fact that wind speed time series are highly non-stationary. This naturally translates to non-stationarity of wind power series.

The shape of the power curve has a very important impact on the way volatility of wind translates to the wind power variations. Fluctuations in wind speed which occur close to the cut-in or the cut-off values get dampened by the corresponding flat parts of the power curve. However, in the steep part of the power curve even a small change in wind speed leads to a large change in the power production. This is why large fluctuations in power production are normally observed when the power production is far from its natural generation bounds, while close to these bounds, the power generation is rather steady.

When looking at the power curve shown in Fig. 2.1, one could think that wind power forecasting boils down to obtaining accurate wind speed forecasts and transferring them to power generation through the manufacturer's specified curve. However, an additional challenge stems from the fact that the deterministic relationship as shown in Fig. 2.1 differs from the empirical power curve observed in practice. Partly this can be explained by the fact that the manufacturer's power curve is obtained when testing a single turbine in idealized conditions (obtained in a so called test tunnel): when the turbine is exposed to constant smooth flow of wind, with no obstacles, no turbulence, normal air pressure, etc. In reality, the behaviour of the wind is more complex than a constant flow with no turbulence, the air density also varies depending on the prevailing weather conditions – all this affects the empirical power curve. An even more important factor, however, is given by the fact that in practice wind turbines are normally gathered into wind farms. Thus, such factors as shadowing effect and, terrain particularities become very important and affect the resulting power generation. In addition, a single wind farm often aggregates turbines of different type, age, etc. All this leads to the fact that the empirical power curve differs significantly from its deterministic counterpart as discussed in more details in [19].

2.2 Point forecasts

This section does not provide a detailed overview on the history or on the state-of-the-art of wind power forecasting. Instead, the goal here is to explain which of the existing forecasting approaches has been chosen as the main benchmark and foundation for the models proposed in this thesis and why. A chronology and evolution of the short-term wind power prediction can be found in [20], while detailed reviews of the state-of-art in the field are given in [21, 13].

One way to classify wind power forecasting models is to look at the input they use - namely, do they involve Numerical Weather Predictions (NWP) or not? Typically, using NWP as input improves the quality of the resulting forecasts when considering prediction horizons larger than 3-6 hours ahead. Since the main goal of this thesis is to account for the space-time dynamics, then forecasting systems involving NWP have been considered as stronger benchmarks, since they partly account for the space-time motion of meteorological phenomena as captured by the NWP.

Methods for wind power forecasting involving NWP have been historically categorized into physical and statistical approaches. Today, however, the limit between them has become less clear, as it is commonly agreed that optimally

the two approaches should be combined.

Physical models consider numerical weather predictions and further rely on dynamical models of the atmospheric flows in order to obtain the corresponding wind speed estimate for the considered location at the hub height. Further, an idealized manufacturer's power curve is used to obtain the corresponding estimate of the resulting wind power output. Finally, model output statistics are employed to (partly) correct the residual model error.

Statistical models aim at finding a relationship between power measurements and some explanatory variables (both NWP and historical power measurements). Usually, time series, regression or artificial intelligence techniques are employed for the purpose.

Given the wealth of forecasting approaches, a natural question is which of the techniques performs the best.

Comparison of the performance of the existing prediction systems is not an easy task. The main obstacle is that, in order to compare the models, it is very important that the data used for the model estimation and validation is exactly the same. In practice this makes it almost impossible to carry out a quantitative comparison between a large number of models and methods. However, some rigorous benchmarking has been performed. Within the framework of the European Anemos project, a number of prediction models have been used to issue power predictions for a set of wind farms. The considered wind farms have been selected to cover a wide range of conditions with respect to climatology and terrain particularities.

Most of the considered predictions systems are expert-quality operational forecasting tools used by the system operators in Spain, Germany, Denmark, Ireland and Greece (at the time the study took place). Detailed results of the comparative study are given in [22, 23].

In short, the results have shown that the performance of the models depends on the terrain complexity and that none of the studied models has the best performance for every horizon and for every test case.

A general picture indicates that the considered statistical approaches tend to outperform the physical ones for short lead times (1-6 hour ahead).

Based on those results, we have considered statistical forecasting approaches (involving NWP) for forming the main benchmark in this work.

One of the statistical systems tested in the above-mentioned comparative study is the Wind Power Prediction Tool (WPPT). The results of the study have confirmed that this system provides expert-level forecasts of wind power generation. Based on these results and based on the fact that we have access to the WPPT (provided by Enfor A/S), this system has been chosen as the main benchmark in this work.

The WPPT is a statistical forecasting system originally developed at the Department of Informatics and Mathematical Modelling (IMM) at the Technical University of Denmark. The development started in 1994 with the first operational implementation at ELSAM (now DONG Energy) in 1994. For the first description of the WPPT and the experiences at power dispatch centres the reader is referred to [24]. Currently the system is a product of Enfor A/S, where it is being continuously updated.

Today the WPPT provides its users with a wide range of possibilities: e.g., generating wind power scenarios, estimating probabilities of cut-off, adaptive quantile estimation. However, here the interest is in traditional point forecasts which are given by the estimates of the expected future power generation for each location and each look-ahead time. Therefore, further in this work, when referring to the WPPT, we refer to its point forecasting module.

The corresponding model provides a point forecast following a two-step procedure.

In the first stage conditional parametric models are employed to model a statistical power curve describing wind power generation as a function of forecasted wind speed and wind direction. More information on it can be found in [25].

The second stage model (also referred to as dynamical model) uses the obtained power curve estimate as input and provides a further statistical correction, based on the recent power measurements and residual diurnal effects possibly not captured by the NWPs. More information on the dynamical part of the WPPT can be found in [26]

Following [27], the overall model writes as:

$$\begin{aligned}\hat{p}_{t+k|t}^{CP} &= \hat{f}_t(w_{t+k|t}, \theta_{t+k|t}, k) \\ \hat{p}_{t+k|t}^{PP} &= \hat{a}_t(\theta_{t+k|t}, k)p_t + \hat{b}_t(\theta_{t+k|t}, k)\hat{p}_{t+k|t}^{CP} +\end{aligned}$$

$$+ \hat{c}_t^c(\theta_{t+k|t}, k) \cos\left(\frac{2\pi d^{24}}{24}\right)_{t+k|t} + \hat{c}_t^s(\theta_{t+k|t}, k) \sin\left(\frac{2\pi d^{24}}{24}\right) \quad (2.2)$$

Where $\hat{p}_{t+k|t}^{cp}$ denotes a power curve estimate of wind generation for time $t+k$ issued at time t , $w_{t+k|t}$ is wind speed forecast for time $t+k$ available at time t , $\theta_{t+k|t}$ is wind direction forecast for time $t+k$ available at t , d_{t+k}^{24} denotes the hour of the day at time $t+k$, p_t is the power observation at time t , $\hat{p}_{t+k|t}^{pp}$ is the final power prediction for time $t+k$ issued at time t . Finally, $\hat{f}_t(\cdot)$, $\hat{a}_t(\cdot)$, $\hat{b}_t(\cdot)$, $\hat{c}_t^c(\cdot)$ and $\hat{c}_t^s(\cdot)$ are the estimates of the corresponding coefficient functions available at t .

Estimation of the model parameters can be carried out in the adaptive and recursive way, both to ease the computational load and to account for the smooth variations in the process dynamics. Detailed descriptions of various versions of the WPPT and the corresponding estimation routines can be found in [27].

Please note, that currently operational WPPT version might differ from the one given in Eq. 2.2. However, the principle has remained unchanged.

The WPPT is currently operational at a number of Danish actors in the wind power generation field (Energinet.dk, DONG, Vattenfall) as well as at a number of others outside Denmark (Nuon (Holland), AEMO (Australia), Hydro Quebec (Canada)).

2.3 Probabilistic forecasts

Point forecasts of wind power generation remain widely used by Transmission System Operators (TSOs) due to their interpretability [16], as for point forecasts, just one value is assumed to fully describe the future power generation. However, such forecasts are never perfectly accurate as there is always an element of the associated uncertainty [28].

Traditional point forecasts provide no information about the uncertainty of the predictions. Instead, it is suggested that the conditional expectation of the future outcome contains all the necessary information necessary to make an optimal decision. However, for a large class of decision-making problems the optimal solution is directly linked to other process functional than the expectation (e.g. it might be a specific quantile or some correlation measure). This is discussed when considering wind power applications in [29], while some more general theoretical derivations can be found in [30].

This renders point forecasts, only addressing the expected wind power generation, a suboptimal input to many decision-making applications, especially the ones related to stochastic optimization or risk assessment. Motivated by the above facts, primary interest is shifting from point to probabilistic wind power forecasting [31].

Examples of the decision making application requiring probabilistic wind power forecasts include wind power trading in the electricity market [32], economic load dispatch and stochastic unit commitment [33, 34, 35], optimal operation of storage [36], reserve quantification [37] and assessment of operating costs [14].

The first results on probabilistic wind power forecasting were obtained in the early work of Brown et al. [38]. In the study the authors considered a Gaussian distribution for describing wind speed data. The theoretical power curve was used to transform the Gaussian predictive intervals describing wind speed into the corresponding predictive intervals for wind power generation. The almost twenty years, scientific research mainly focused on point forecasting of wind power generation, before the probabilistic wind power forecasts attracted a new wave of attention.

In some way the two different schools for wind power point forecasting (physical and statistical) have translated to two different approaches to probabilistic forecasting.

Authors focusing on the physical approaches consider how uncertainty in wind transforms to wind power uncertainty. Predictive densities for wind speed are obtained either proposing some modelling approaches [39] or through the ensemble forecasts [40, 41] issued by the considered NWP provider. Probabilistic wind forecasts are further transformed to wind power forecasts, normally through a deterministic power curve [39, 40, 41].

The following difficulties are associated with the physical approach.

First, numerical weather predictions are obtained by solving a system of partial differential equations describing dynamics of the atmosphere. Therefore, in their essence, the numerical weather predictions are obtained from a deterministic description of the system. The ensembles are obtained by considering different initial conditions and/or considering several different models for describing the atmosphere, thus in essence such ensembles are a collection of deterministic forecasts obtained from deterministic models. Therefore, the stochastic nature of the complex meteorological phenomenon is not fully accounted for. This can be viewed as a major reason for the fact that ensemble forecasts do not provide

a reliable description of the forecast uncertainty. Forecast reliability can be improved to a certain extent if some statistical post processing is considered, e.g. model output statistics.

Second, similar as ensemble forecast is only a suboptimal probabilistic description of wind speed, a deterministic power curve is also only a suboptimal way to describe the dependency between wind speed and wind power. Ideally, a stochastic power curve should be used [19]. Recently a model for describing a stochastic power curve was presented in [42].

An alternative, following the statistical school, is to construct probabilistic forecasts for wind power generation directly, without the intermediate step of modelling the uncertainty of the wind. Advantages of this approach are (i) no need to directly account for the complexity of the stochastic power curve, (ii) owing to the geographical distribution of wind farms, the corresponding wind power data contains substantially more information than numerical weather predictions or 3-hourly data coming from the few available meteorological stations.

When interest in probabilistic forecasts re-appeared in the early 2000, there was already a wide range of high quality forecasting techniques available for issuing point predictions. Thus, a statistical approach to probabilistic power forecasting naturally took place through the probabilistic description of the point forecast errors.

One way to probabilistically describe forecast errors is to look at all the available historical forecast errors and assume that the future prediction errors will follow the same pattern. When described in such a way, the uncertainty description is constant for any considered period and is not designed to discriminate between periods of different variability. Therefore, in the literature it is common to refer to such an uncertainty estimate as “climatological” uncertainty, as opposed to the “meteorological” one.

In order to account for the fact that wind power forecast uncertainty is not constant, but depends on some explanatory variables, classification techniques have been considered. For instance, in [43] the authors considered classification according to the expected level of power generation. Even though such classification approaches are rather appealing, owing to their easy and rather intuitive interpretability, they lack continuity in uncertainty description. To cover for this, various smoothing techniques can be employed. In [44], a fuzzy-logic-based approach has been used for that purpose, while in [45] and [46] quantile regression techniques have been employed. In particular, [45] considered local

regression, which uses kernels to smooth the data, while in [46] the authors considered splines for this purpose. Both the adaptive resampling [44] and the time-adaptive quantile regression [46] approaches have been compared in [47], where it is shown that they yield similar results.

In parallel to the non-parametric approaches to probabilistic wind power forecasting, some efforts have been made to propose a parametric description of the error distributions. For example, in [48] P. Pinson considered Beta, Censored Gaussian and generalized logit-Normal densities for describing wind power generation, and subsequently compared the performance of the resulting forecasts using the test case with 10 min ahead power predictions.

Similarly, as in the case with wind power point predictions, probabilistic wind power forecasting techniques are usually optimized with respect to local information only. One of the objectives in this thesis is to introduce and evaluate a methodology enabling optimal probabilistic wind power forecasts which account for geographically dispersed information.

2.4 Scenarios

Recent methodologies for probabilistic wind power forecasting focus on providing information on prediction uncertainty for each site and each look-ahead time individually. They inform neither on the inter-dependence structure between forecast errors observed at different locations, nor on the way these errors propagate in time. However, for a number of applications such marginal predictive densities are only a suboptimal input, as the joint distribution describing wind power generation at a number of sites over a period of time might be of interest.

Multivariate predictive densities are often communicated in the form of scenarios (also referred to as trajectories or ensembles). This choice is motivated by the fact, that multivariate predictive densities often do not have an easy analytical structure. Also, trajectories are normally preferred by the end-user, since they are easy to use in the conventional deterministic optimization systems and decision tools. From the forecasters point of view, scenarios can be obtained by random draws from the associated predictive density.

Once again, the physical approach to generating such scenarios could take its way through the translation of ensemble meteorological forecasts to wind power scenarios. However, there are several difficulties related to this approach:

First, meteorological ensembles often lack calibration and need to be statistically post-processed in order to provide reliable forecasts at a given location [40]. Many post-processing techniques are local. That is, they are designed to target each site and each look-ahead time of interest individually, thus not respecting the interdependence structures.

Second, as already mentioned in this work, translation of wind to wind power is optimally described by power curves which are stochastic, site-specific and of a rather complex nature.

Third, the number of meteorological ensembles is normally in the range of 5-50 members. When interest is in power generation at many sites and over a large number of prediction horizons, the dimension of the problem becomes high, and having only 5-50 members might be not sufficient to represent the underlying multivariate predictive density.

An appealing alternative is to estimate the joint predictive densities using statistical methods. One of the techniques is given by a copula approach. It is based on decoupling the problem of finding the joint predictive distribution into two independent steps of (i) estimating the marginal densities targeting each site and each prediction horizon individually and (ii) modelling the interdependence structure between the marginals.

2.5 Forecast verification

In [49] A. H. Murphy identifies the following distinct types of forecast goodness:

1. **Consistency** is given by the correspondence between forecasts and forecaster's best judgements.
2. **Quality** is given by the correspondence between forecasts and observations
3. **Value** is given by the benefits the users gain when using forecasts

Since (i) a forecaster's judgements are internal to the forecaster and are unavailable for explicit evaluation and (ii) forecast value depends on the particular application at hand, in this work the main focus is on quality assessment.

2.5.1 Evaluating point forecasts

Evaluation of point forecasts is probably the most intuitive when compared with the assessment of probabilistic forecasts or space-time trajectories. The basic quantity used for assessment of single-value forecasts is given by the forecast error which is defined as:

$$\varepsilon_{t+k|t} = p_{t+k} - \hat{p}_{t+k|t} \quad (2.3)$$

where $\varepsilon_{t+k|t}$ denotes an error made at time $t+k$ by the corresponding forecast issued at time t , p_{t+k} is power measurement at time $t+k$ and $\hat{p}_{t+k|t}$ is power forecast issued at t for time $t+k$.

Following [50], in practical applications it is usually more convenient to introduce the normalized prediction error $\epsilon_{t+k|t}$:

$$\epsilon_{t+k|t} = \frac{\varepsilon_{t+k|t}}{P_n} \quad (2.4)$$

where P_n is the nominal capacity of the considered site. Normalizing errors permits one to compare the errors obtained at different locations, independent of their rated capacities. Since the goal of this thesis is to track the propagation of prediction errors in space and in time, normalization of the errors obtained at different locations has been performed.

In general, there exists a wealth of error measures which can be employed to evaluate the performance of point forecasts. Aiming at standardizing the procedure of point forecast verification, in [50] H. Madsen et al. present a complete protocol consisting of a set of criteria appropriate for the evaluation of wind power prediction systems. Regarding the performance measures, the authors argue that as a minimum set of error measures, the following should be used:

1. Bias, *BIAS*, which computes the mean of all errors over the validation period:

$$BIAS(k) = \frac{\sum_{t=1}^N \epsilon_{t+k}}{N} \quad (2.5)$$

This criterion informs on whether the forecasting method tends to over-predict or under-predict. If the *BIAS* = 0, this means that in the long run positive and negative errors cancel each other out, so that the resulting

predictions are unbiased. However, it does not give much information about the forecast accuracy, since it does not inform on the amplitude of the errors.

2. Mean absolute error, MAE is computed as the mean of absolute errors over the validation period:

$$MAE(k) = \frac{\sum_{t=1}^N |\epsilon_{t+k}|}{N} \quad (2.6)$$

3. Root mean squared error, $RMSE$ is computed based on squared errors over the validation period:

$$RMSE(k) = \sqrt{\frac{\sum_{t=1}^N (\epsilon_{t+k})^2}{N}} \quad (2.7)$$

Which of the scoring rules should be used as the lead criterion in practice? In general, if it is not known what precise functional of the process (e.g. the expectation, a certain quantile,...) is aimed to be described by the forecasts, then one should follow the protocol in Ref. [50] and look at the number of measures. However, if the target functional is known, there is no need to use several different scores.

This point is discussed in [51]. Here the author develops a theory for the notions of consistency and elicibility and argues that in order to make an effective point forecast, the forecaster should be told a priori what functional is of interest. Once such functional is known, the forecaster can use his best judgement to make optimal point prediction. It is then important that the scoring function is consistent with the given functional. By consistency it is implied that the expected score should be optimized when the prediction corresponds to the requested functional of the process. And a functional is elicitable if there actually exists a score which is consistent for it.

In [51] the author recalls some classical results, such as those for the mean and the median, and also derives some original results, such as, showing that scoring functions which are consistent for the value-at-risk functional do not exist.

In the case of the mean functional, the consistent scoring functions are the Bregman functions. On the basis of the work of Savage in [52], the author recalls that up to a multiplicative constant, squared error is the unique Bregman function of the prediction error form. This means that if the expectation is the functional of interest, then the $RMSE$ should be used.

Similarly, if the functional of interest is given by a quantile, then the consistent scoring functions are generalized piecewise linear. This translates to the fact that the *MAE* score should be used if the functional of interest is the median.

Since in this work, when talking about point forecasts, we refer to the estimates of the expectation of the future power generation, the *RMSE* criterion is used to evaluate predictions. For the sake of consistency, we also employ quadratic loss functions in our point forecasting models.

In this work a point forecast aims at estimating the expected power generation at a single location for a given lead time. Therefore, the evaluation is based on measuring how consistent the forecasts are with the conditional mean of the process (*RMSE*) rather than looking at other types of quality measures (such as *MAE* or *BIAS*). That is, the focus is on "how well does the point forecast represent the expected power generation" rather than on a more general question "how well does the point forecast match observations".

2.5.2 Evaluating probabilistic forecasts

2.5.2.1 Different quality aspects of probabilistic forecasts

Generally speaking, prediction quality is related to the level of correspondence between forecasts and observations. In the case of point forecasts, the measures for this correspondence are more intuitive than in the case of probabilistic forecasts. This is because assessment of probabilistic forecasts calls for a distribution-orientated approach where one has to evaluate how consistent a predictive density is with the corresponding distribution of observations.

In [49] Murphy envisages a distribution-based approach to forecast verification. Even though the author considers single-valued forecasts, he notices that both forecasts and outcomes are random variables and that their sequences have a joint distribution which contains full information on forecast quality.

Comparing predictive and observed densities is the core of probabilistic forecasts. Thus, the finding presented in the work of Murphy can be applied (with possibly some minor formulation modifications) to the case of density forecast assessment.

In [49] the author distinguishes between a number of different aspects of forecasts quality.

Here we will briefly introduce several of these aspects that are relevant for the further discussion:

- *Reliability* (also referred to as calibration) is related to statistical consistency between forecasts and observations. A probabilistic forecast is well-calibrated if it coincides with a suitable conditional distribution of observations. We follow the formal definition given in [53].

Let F_t be a predictive distribution of X at time t . The sequence $(F_t)_{t=1,2,\dots}$ is probabilistically calibrated relative to the corresponding sequence of real process generating distributions $(G_t)_{t=1,2,\dots}$ if

$$\frac{1}{T} \sum_{t=1}^T G_t \circ F_t^{-1}(p) \rightarrow p \quad (2.8)$$

for all $p \in (0, 1)$.

That is, if a forecast places a probability p on the event $X = x$, then this event must be observed with the probability p .

- *Resolution* relates to the ability of a forecasting system to issue situation-dependent predictions. High resolution means that on average different forecasts are followed by different observations.
- *Sharpness* is a property of the forecasts only and does not inform on the correspondence between forecasts and observations. That is, sharpness is given by variability of forecasts when inferred from the marginal forecast density. For example, in the case of wind power forecasting, a forecast stating that wind power generation will be 0 with probability equal to 1 is very sharp, even though it might not be consistent with the observed power generation. In the case of perfectly calibrated forecasts, sharpness and resolution are equivalent [49].
- *Uncertainty* is a property of observations only. It is related to variability of observations when looking at the marginal density of observations and thus it does not depend on forecasts at all.

2.5.2.2 Scoring probabilistic forecasts

Provided that there are many aspects of forecast quality, how should one compare several competing forecasting approaches? One way could be to check one or several measures of quality and compare the magnitudes. However, the problem is that, even if one forecast scores better in one or several quality aspects,

this does not guarantee that it will perform better in all the aspects. Moreover, it does not guarantee that this forecast is of greater value to all end-users [49].

A solution to this problem is to consider scoring rules which would address a number of quality aspects simultaneously and summarize them into a single-valued numerical score. Such a score then could be used to rank the competing forecasts. The question is then which aspects to combine, and what weights to assign to each of them? In order to answer this question first the ideal forecast should be defined, since obviously scoring rules should be constructed in such a way that the ideal forecast should be the one resulting in the optimal score.

One could argue that the real value of forecasts is given by the benefit they bring to the end user and thus a measure of forecast value should be used when ranking competing forecasting approaches. However in practice a forecaster is often not aware of the loss functions used by the users. In addition, those loss functions might also be very different in practice: for some end users forecast calibration might be of the highest importance, meanwhile others might call for forecasts of high resolution in order to optimize profit. What should the forecaster then target?

The answer is that the ultimate forecasting goal is to issue a predictive density which coincides with the real process generating density. In this idealistic case such predictive density will be preferred by all forecast users, independent of their loss function [54]. The fact that the ideal forecast is given by the real data generating density is directly related to the requirement for scoring rules to be proper.

Propriety is an essential property of a scoring rule, ensuring that the optimal score is achieved when the probabilistic forecasts coincide with the real process generating density [55, 56]. The mathematical grounds of propriety ensure that the forecaster is encouraged to be honest and issue predictions based on his best judgement, as argued in [49]. An overview of the proper scores available for verification of probabilistic forecasts is given in [56].

Without going into the details on all the available scoring rules, we will focus on the Continuous Ranked Probability Score (CRPS), which is the lead score used in this thesis for evaluation of probabilistic forecast.

2.5.2.3 Continuous ranked probability score

The scoring rule corresponding to the CRPS is defined as:

$$crps(P, x) = \int_{-\infty}^{\infty} (P(y) - \mathbf{1}(x \leq y))^2 dy \quad (2.9)$$

where P denotes predictive distribution, x denotes an observation and $\mathbf{1}$ stands for the Heavyside step function taking the value 1 if the condition inside the brackets is fulfilled and 0 otherwise.

Following this definition, the $crps$ is a negatively orientated score with the minimum value equal to 0.

One way to estimate the $crps$ is by using numerical integration techniques. Sometimes, however, the integral can be evaluated in a closed form by using the following identity:

$$crps(P, x) = E_P|X - x| - \frac{1}{2}E_P|X - X'| \quad (2.10)$$

where X and X' are independent random draws from P and $E_P(\cdot)$ denotes the expectation with respect to the probability distribution P .

For assessing a probabilistic forecast over a data set containing T observations the average of the $crps$ values for each forecast/verification pair is calculated resulting in the overall $CRPS$ value.

$$CRPS(P, x) = \frac{1}{T} \sum_{t=1}^T crps(P_t, x_t) \quad (2.11)$$

There are several features of the $CRPS$ score which make it appealing in practical applications.

First and foremost, it is a proper score [56].

Second it is a distance sensitive score, meaning that a credit is given for assigning high probabilities to the value near the one materializing.

Third, the *CRPS* for point forecasts is equivalent to the *MAE*, thus the *CRPS* provides a direct way to compare point and probabilistic forecasts.

Another particularity of the *CRPS* is that it is a robust score, compared to the likelihood-based scores (such as the logarithmic score, for instance [56]). The problem with the likelihood-based scores is that they are very sensitive to outliers which might be a problem in practical applications. For example, the logarithmic score is infinite if the vanishing probability is assigned to the value which materializes.

Following [57], the *CRPS* can be decomposed into reliability, resolution and uncertainty parts. Such decomposition can be used to obtain a better insight on the behaviour and properties of the forecasting system.

In addition, in [58] the authors proposed threshold and quantile-weighted versions of the *CRPS*. The weighted version of the *CRPS* puts more weight on the regions of interest (let it be the central part or the tails of the distribution) while retaining the crucial property of the score being proper. Threshold and quantile decompositions of the *CRPS* can also be used for evaluating the strengths and deficiencies of the forecasting system.

2.5.3 Evaluating scenarios

Essentially wind power generation scenarios (also referred to as space-time trajectories or as ensemble forecasts in meteorology) can be viewed as random draws from the joint multivariate predictive density describing wind power generation at a number of sites over a period of time. In the literature there exists rather few proposals regarding verification of multivariate probabilistic forecasts and probably the most rigorous work on this subject is given by T. Gneiting et al. in [59].

In principle verification of multivariate predictive densities is similar to that of univariate probabilistic forecasts discussed in the previous section. Most proper scoring rules available for evaluation of univariate predictive densities have the corresponding analogues available for assessment of multivariate forecasts.

For example, the *CRPS* extends to the Energy score with the related scoring rule defined as:

$$es(P, \mathbf{x}) = E_P \|\mathbf{X} - \mathbf{x}\| - \frac{1}{2} E_P \|\mathbf{X} - \mathbf{X}'\| \quad (2.12)$$

where P denotes predictive distribution, \mathbf{x} denotes an observation, \mathbf{X} and \mathbf{X}'

are independent random draws from P and $\|\cdot\|$ denotes Euclidean distance.

Then the overall Energy score over T observations is obtained from

$$ES(P, x) = \frac{1}{T} \sum_{t=1}^T es(P_t, x_t) \quad (2.13)$$

The Energy score is shown to be proper [56].

If the proper scoring rule which we considered as the lead criterion when evaluating univariate predictive densities extends to the multivariate framework, then what is the problem?

First, in case of the *CRPS* score there are decomposition techniques available which can be used for better understanding of the benefits and pitfalls of the considered forecasting system. There are no analogues of such decomposition available for the Energy score. Thus, even if we get a single-value score evaluating the forecasts, it is not clear what makes one or another forecasting system better. One way to overcome this issue could be by looking at the situation-based performance as suggested by [60].

However, the most crucial issue comes with the estimation of the score. Evaluation of the Energy score requires estimating expected values of the Euclidean distance between forecasts and observations. Most often, closed-form expressions for such expectation are unavailable and one needs to employ Monte Carlo methods in order to estimate the score [59]. When dealing with problems of a very high dimension, Monte Carlo techniques result in computational challenges. The problem of computational load translates to increased sampling uncertainty which makes it more difficult to conclude on the superiority of one forecasting system over another.

There are alternatives to the Energy score. One of them is given by the logarithmic score which is defined as

$$\text{logs}(P, \mathbf{x}) = -\ln p(\mathbf{x}) \quad (2.14)$$

where p denotes a predictive density and \mathbf{x} denotes a value which materializes.

Allowing for some affine transformations the logarithmic score is the only local proper score (see Theorem 2 in [61]). Locality means that the score depends on

the probabilistic forecast only through the value which the predictive density attains at the observation [55]. An important advantage of using local scores when dealing with multivariate predictive densities comes with the related computational benefits. When dealing with local scores, there is no need to draw random samples from the predictive density in order to evaluate the score.

The downside of the logarithmic score is its sensitivity to outliers. The score is infinite if the forecast assigns a vanishing probability to an event which occurs. In practice, when working with the real data, such sensitivity might be a problem.

Due to its sensitivity to outliers, the logarithmic score can not be used for evaluating scenarios in the same way as the Energy score. That is, the logarithmic score requires knowing the underlying predictive density. Suppose, that a forecast is given in the form of m equiprobable scenarios with the scenario-generation density being unknown. If we are to consider only the raw ensembles without dressing them with a continuous support having density, this is essentially equivalent to assuming an empirical predictive density which assigns equal probabilities of $1/m$ to each of the m available ensembles. Then, if an actual observation falls in between any two of the ensemble members rather than on one of them exactly, the logarithmic score will be infinite. That is, implementation of the logarithmic score in practice calls for the predictive density which would assign a non-zero probability to every possible outcome. Therefore, this score cannot be used for verification of ensembles directly, unless some statistical post-processing is done.

Thesis Contribution

This chapter comprises a short overview of the thesis contribution. The chapter starts with Section 3.1 describing a methodology proposed for the space-time correction of wind power point predictions. Further in Section 3.2 probabilistic forecasts are addressed. Section 3.3 gives an overview of the proposal for generating space-time trajectories. The chapter finishes with Section 3.4 presenting some remarks on the conditional discrimination ability of the Energy score.

3.1 Point forecasts

This section deals with space-time correction of point forecasts, mainly summarizing the works presented in Papers A and B.

The first part of this section gives a brief overview of wind forecasting models with off-site information as input. The overview is restricted to the approaches which were available at the time this work was initiated in order to illustrate the foundation used for building the initial proposals in this dissertation.

3.1.1 Prediction models with off-site data input

At the time this thesis was initiated, there existed several research works looking at spatio-temporal propagation of information when considering wind data. These works mainly dealt with cases where wind behaviour between the considered locations is easy to model, owing to the terrain topology or meteorological particularities of the area.

For instance, in [62] Damousis et al. have looked at the area around the Gulf of Thessaloniki which is affected by strong prevailing thermal winds. The authors showed that accounting for the up-wind information improves the quality of the resulting wind speed and wind power predictions up to 2 hours ahead.

Another example has been given in the work of Larson et al. in [63]. Here the authors have considered a potential wind farm located at the exit to the Columbia River Gorge, while meteorological observations were available at the entrance of the same Gorge. The results showed that the inclusion of the up-wind information in addition to the numerical weather predictions improves the accuracy of the resulting wind speed forecasts.

Considering the same test case of the Columbia River Gorge, Gneiting et al. in [64] have proposed a regime-switching approach which accounts for two dominant directions, thus discriminating between situations when wind enters the Gorge from two different sides. The results showed improvements in wind speed predictions up to 2 hours ahead.

In [65], the authors have proposed a generalization of the regime-switching model described in [64] by including wind direction as a circular covariate in the models.

The results of all the works mentioned above show significant improvements over the considered benchmark approaches (e.g. persistence) when testing the models on smooth terrains with known physics of the prevailing wind behaviour. However, if these methods are to be applied to other types of case studies, for which wind behaviour is more complex and where no channelling effect is present, one should not expect similar improvements.

Another important aspect is that all the studies mentioned above have focused primarily on wind speed forecasting. However, since the main interest in this work is in wind power forecasting, it is beneficial to focus on wind power data directly, without the intermediate step of considering wind speed. The direct approach is preferable for a number of reasons: (i) owing to their wide geographical spread, wind farms potentially contain more information than numerical weather predictions with their relatively coarse spatial and temporal resolutions

or few meteorological stations, (ii) it avoids modelling an empirical power curve which is known to be of a complex nature as discussed in Section 2.

Consequently, the interest in this thesis is in proposing more advanced models tailored to wind power directly and not restricted to any particular, a priori known meteorological pattern.

3.1.2 Space-time correction of wind power point predictions

The first step towards developing such models has been made by the preliminary examination of data aiming to verify whether wind power forecast errors made by a locally optimized system do indeed show any residual cross-correlation in space and in time.

For this purpose 7 months of hourly data coming from 5 groups of wind farms located in Western Denmark have been considered. Wind power forecasts for each of the groups have been made by the WPPT system. Conditional cross-correlation analysis has shown that the resulting forecast errors propagate in space and in time under the influence of meteorological conditions (mainly wind direction). The details are given in Paper A. However, it is important to stress that in this study only the power curve model of the WPPT has been considered (see eq. (2.2)). This resulted in autocorrelated forecast errors. Due to the residual autocorrelation, the magnitude of the dependencies presented in the paper is significantly higher than the one which could be observed if the full WPPT model was used. Owing to that, even though the general findings on the pattern of error propagation presented in the paper hold, the magnitude of the illustrated improvements is higher than the one which could be expected in the operational setup. Paper A is the only study where the WPPT power curve model has been considered - all the following research papers are based on the full version of the WPPT.

The work documented in Paper A has been recently elegantly and more rigorously generalized by Girard and Allard in [66] as they could base their work on a much longer data set when considering hundreds of locations spread over Western Denmark. Generally, the preliminary results in Paper A are in line with the findings uncovered in [66].

In order to account for the spatio-temporal patterns a Conditional Parametric (CP) Vector AutoRegressive (VAR) model has been proposed in Paper B. In general, CP models comprise a class of models with a linear structure, but for which the coefficients are replaced by smooth functions of other variables. In this

case, in order to account for the directional error propagation, the coefficients have been modelled as functions of forecasted wind direction.

The model coefficients have been estimated adaptively in the spirit of [67] in order to account for the long-term variations in the process dynamics.

3.1.3 Result using aggregated data

The model has been validated on 15 groups of wind farms spread throughout the territory of Western Denmark (see Fig. 3.1). One-hour-ahead predictions have been considered in this test case.

The grouping smooths out local variations and places focus on a more global phenomenon. The accuracy of the CP-VAR-corrected forecasts has been compared to that of the original WPPT forecasts based on the RMSE criterion. The reduction in the RMSE (denoted as Δ RMSE) is given as a percentage decrease in the RMSE relative to the RMSE of the WPPT forecasts for each group. The results are presented in Fig. 3.1.

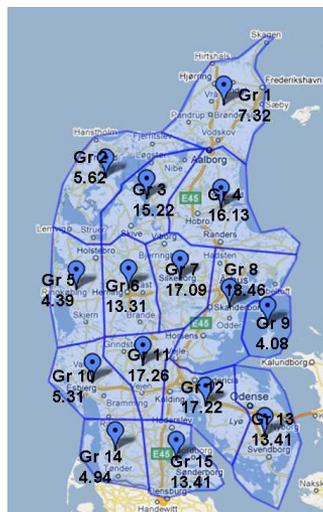


Figure 3.1: Predictive performance (improvements) given by the CP-VAR model in terms of a percentage reduction in the RMSE (Δ RMSE) of the forecast errors. (Produced using <http://maps.google.dk/>)

Note that larger improvements (17-18%) correspond to the eastern part of the region. This is in line with the fact that in Denmark the prevailing wind direction is westerly. Due to this the easterly located groups are usually situated "down-wind" and can benefit well from the information extracted from the "up-wind" territories.

An interesting point to mention is that for Group 9 the observed improvement in the RMSE (4.08%) is not as large as for the surrounding zones. This could be influenced by the fact that while other groups are formed from wind farms spread over larger territories, Group 9 covers a smaller area. This leads to more significant local variations, while making the improvement offered by the spatio-temporal model smaller. Another very likely explanation is that Group 9, in contrast to the rest of the zones, is situated off the mainland. Therefore it is very probable that the dynamics of Group 9 are different from the rest of the considered region.

3.1.4 Result using a single wind farm

The methodology presented in Paper B can be also applied to problems having a different setup. That is, instead of focusing on several locations simultaneously, one can target a single location, while using a (small) number of neighbouring territories as explanatory variables. Essentially, the estimation techniques remain the same as for the vector approach. More details can be found in Paper C.

In order to test the proposed methodology on a different test case, we have considered 20 wind farms located in Denmark. The respective locations of the considered wind farms are shown in Fig. 3.2.

Instead of targeting several locations simultaneously, focus has been on improving the quality of forecasts for the Nysted wind farm, while the surrounding sites have been used as explanatory variables.

There were two main reasons behind choosing to target Nysted. Firstly, with a rated capacity 165 MW, the Nysted wind farm accounts for about 36% of the installed capacity owned by the company operating it. Secondly, Nysted has an appealing location with many wind farms located "upwind" with respect to prevalence of westerly winds over Denmark.

Targeting an offshore wind farm is a challenging task due to the presence of large wind power fluctuations caused by local meteorological effects which are hard to model and predict [68, 69]. Such high volatility can rarely be observed onshore

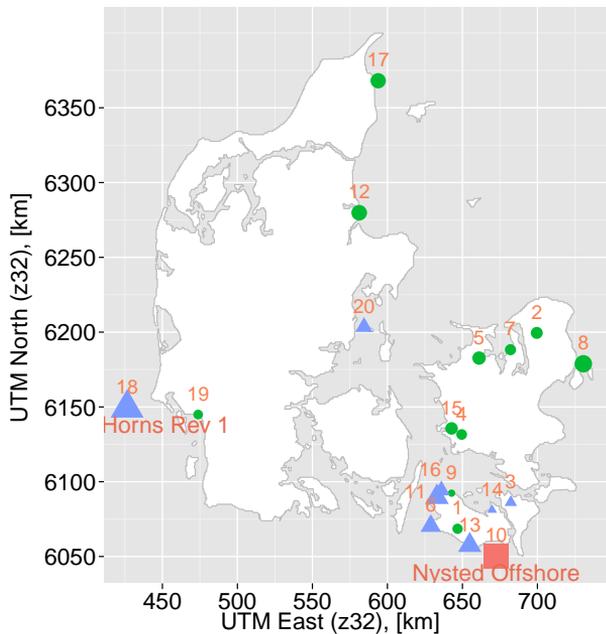


Figure 3.2: Map of the wind farms included in the analysis. The Nysted wind farm is marked as number 10 and with a square. Wind farms with triangles are used as "sensor" locations, thus improving the predictability of wind power generation at Nysted. The size of the points is proportional to the rated capacity of the wind farms, on a logarithmic scale.

where similar capacities would be spread over a much wider area, smoothing out the effects of the weather instabilities [70].

Another motivation for testing the methodology on this new test case comes with a higher temporal resolution (15 min) of the data, which could potentially also influence the magnitude of the improvements.

Finally, prediction horizons from 15 min to 8 hours ahead have been considered, thus revealing how far in the future the improvements stemming from the space-time correction reach, when looking at a small territory like Denmark.

The results depicted in Fig. 3.3 show that the peak in the improvements is observed for the lead times around 1 hours ahead. The magnitude of this improvement ($\approx 5\%$) is in line with the result observed for Group 9 in Fig. 3.1. The fact that the biggest improvements correspond to 1 hour ahead predictions

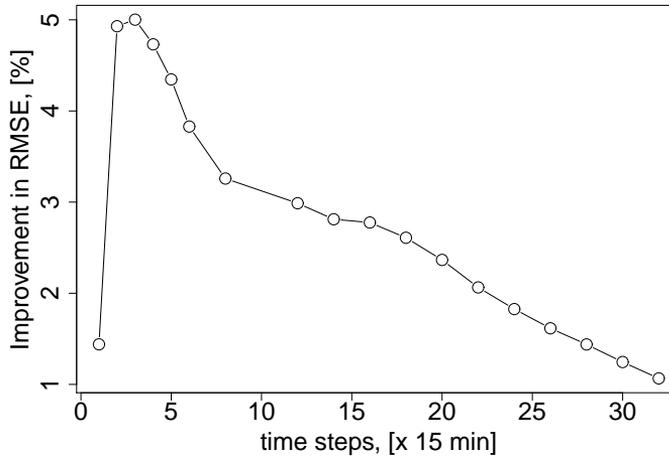


Figure 3.3: Evaluation of the point forecasts for the Nysted wind farm in terms of the relative improvements in the RMSE as a function of the considered prediction horizon.

is in line with the geographical layout of the considered wind farms and with the speed of motion of meteorological phenomenon over Denmark. Namely, it can be seen from Fig. 3.1 that almost all the reference sites are located within 50 km from Nysted. According to [66], the average speed of the error propagation over Denmark is 30-50 km/hour depending on the wind direction. This explains why the peak in improvements is observed for the lead time of 1-2 hours ahead.

Summarizing, wind power forecast errors tend to propagate following wind direction, and hence it is possible to improve predictability over a region if considering information from the neighbouring ("upwind" situated) sites. The improvements are likely to be larger if considering spatially larger areas (reduction of local variations), preferably of similar dynamics (on-shore, off-shore,...)

3.2 Probabilistic forecasts

This Section summarizes the results presented in Papers C and D. The papers deal with the extension from point to probabilistic forecasts (marginal univariate predictive densities), hence requiring description of the uncertainty associated with the point predictions previously generated.

Paper C gives a detailed description of the proposed methodology and the corresponding results. Paper D comprises additional information on the alternative approaches and models which have been considered.

The goal of the study has been to analyse the existing methodologies for predicting marginal densities of wind power generation and to propose ways to include space-time effects into the corresponding models. Since, there are currently no studies available in the literature which rigorously compare parametric and non-parametric approaches to wind power forecasting, both techniques have been considered, analysed and compared.

The proposed methodology is tailored to situations where probabilistic forecasts are to be issued for a single target location, while considering information coming from a (possibly small) number of neighbouring sites as explanatory variables. We consider a discrete formulation of the problem as opposed to a full space-time covariance model (as in [71]), which would call for a larger amount of reference sites spread throughout the considered territory.

In essence, our aim is to propose a way to optimally summarize the snapshot of forecast errors observed at time t in order to issue a predictive density describing future wind power generation at the target location at time $t + h$.

3.2.1 Considering parametric predictive densities

The parametric approach is based on the assumption that the shape of the conditional predictive densities is known and can be described by one of the known distribution functions.

In the literature Beta, generalized-Logit Normal and Censored Normal densities have been proposed as the basis for describing wind power generation. These distribution functions have also been considered in this thesis, however, since the best results have been obtained using Censored Normal distribution (see Paper D), this is the one analysed in Paper C.

Censored Normal predictive densities can be fully characterized by their location and scale parameters. These parameters can be approximated by the conditional mean and the conditional variance of power generation, respectively.

Estimating the conditional expectation of wind power generation is equivalent to point forecasting. Thus, our approach to estimating the location parameter is based on the methodology presented in Paper B. That is, we employ a Conditional Parametric model which considers point forecasts for the target location

issued by the state-of-the-art system, and corrects these by accounting for the forecast errors previously recorded at the neighbouring locations. The resulting point forecasts are then used as the estimates of the location parameter.

The scale parameter is approximated by the conditional variance of wind power generation. Thus, estimating this parameter calls for modelling wind power volatility. The volatility is not constant in time, owing to evolving wind dynamics and owing to the power curve which amplifies or dampens wind fluctuations in a non-linear manner.

We tried to account for the former aspect by employing regime-switching models, however this did not improve the accuracy of the resulting probabilistic forecasts. The corresponding models are only briefly presented in Paper D.

The effects of the power curve can be accommodated by letting the model parameters vary with the level of expected power generation, thus conditioning wind power volatility on the slope of the power curve. For this purpose a Conditional Parametric ARCH (CP-ARCH) model has been proposed, as well as its CP-ARCH-X extension involving offsite information. These models are described in more detail in Paper C.

3.2.2 Considering non-parametric predictive densities

Non-parametric predictive densities have been built using time-adaptive quantile regression. The procedure follows two main steps. First, wind power point predictions are used to determine the mean of the corresponding predictive distributions. Second, the uncertainty around the mean is shaped by building a number of quantile models. Each of the considered quantiles is modelled as a non-linear function of the expected power generation, thus accounting for the power curve effect. The offsite information can be incorporated in to the quantile models as an additional covariate, even though this has shown not to improve the quality of the resulting forecasts.

3.2.3 Results

The empirical results are obtained on the test case of a portfolio of wind farms in Denmark. The respective locations of the wind farms are shown in Fig. 3.2. The Nysted wind farm has been chosen as the target wind farm owing to its large rated capacity and then appealing location, with many wind farms located “upwind” in view of the prevailing westerly winds over Denmark.

The results have shown that the best performing parametric predictive densities are obtained when the location parameter is estimated using the CP model accounting for the directional propagation of the forecast errors, while the scale parameter is described by the CP-ARCH model accounting for the power curve effect.

The best performing non-parametric predictive densities are obtained when the mean is estimated using the CP model, thus accounting for the space-time effects, while the uncertainty around the mean (described by the quantile models) accounts for the non-linear power curve effect.

In both cases the improvements in forecast quality are achieved by the space-time correction of the conditional mean of the predictive densities. Additional inclusion of the spatio-temporal effects into the uncertainty modelling step has been shown to be superfluous, as it does not further improve the quality of the predictions.

Adaptive quantile regression with initial WPPT point forecasts as input is considered as the base benchmark. The best performing parametric and non-parametric densities are compared to the benchmark approach and the relative improvements in the CRPS (skill scores) are evaluated. The results are shown in Fig. 3.4.

The results show that accounting for the spatio-temporal effects improves the quality as measured by the CRPS of the resulting probabilistic forecasts for a range of lead times up to 5-8 hours ahead. For larger lead times, none of the proposed models outperformed the benchmark given by the locally optimized forecasts. This is in line with the scales of motion of weather systems over the region [66].

The performance of the parametric and non-parametric approaches has been compared, uncovering that they both perform similarly for lead times up to 5 hours ahead and with an advantage for non-parametric predictive densities for further lead times.

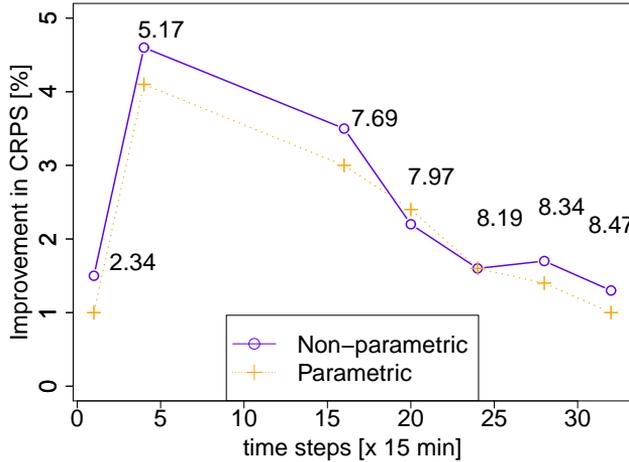


Figure 3.4: Evaluation of the density forecasts in terms of the relative improvements of the CRPS (skill scores). Point labels show the corresponding CRPS values [% of the nominal capacity] for the benchmark model.

3.3 Space-time trajectories

Paper E addresses a problem of generating space-time trajectories of wind power generation. This section gives a brief overview of the proposed methodology and highlights the main results.

The task of the study presented in E is to issue multivariate predictive densities describing wind power generation at a number of locations over a period of time. Tackling such a high dimensional problem directly is a very difficult task.

For instance, consider the spatial aspect of the problem. Power curves describing wind power generation are site-specific as they depend not only on the wind characteristics, but also on the way the turbines are positioned within a wind farm, local terrain particularities, etc. The fact that wind power dynamics are site-specific makes it more complicated to issue high quality forecasts for a large number of locations simultaneously, because the local particularities (if to be respected) keep the dimension of the problem high.

A similar situation occurs when considering the temporal aspect of the problem.

A common practice is to issue direct power forecasts for each of the prediction horizons individually rather than addressing their joint distribution. This is motivated by the fact that direct forecasts are more robust to model misspecification.

In this study we have considered a copula approach for issuing multivariate predictive densities. An important feature of copula is that it can be used to model an interdependence structure between stochastic variables independently of their marginal distribution functions. That is, multivariate predictive densities can be obtained in two-steps. First, the existing state-of-the-art forecasting tools are used to issue probabilistic forecasts for each location and each look-ahead time individually. Then, a copula is introduced in order to obtain the corresponding joint multivariate predictive densities.

A Gaussian copula has been considered in this study, suggesting that the underlying interdependence structure can be represented by the covariance structure. The main contribution of the work is given by the proposed parametrization of the covariance structure. That is, instead of tackling the covariance matrix directly, focus has been on its inverse (precision matrix). As opposed to the covariance matrix which informs on the global dependency pattern, the elements of the precision matrix represent conditional dependencies. The two matrices compare in a similar way as auto-correlation functions compare to the partial auto-correlations. We have shown that the precision matrix is very sparse, which results in several benefits.

First, working with sparse matrices results in computational benefits due to faster factorization algorithms.

Second, the precision matrix represents conditional dependences between variables as opposed to the global relations given by the covariance matrix. A zero element in the precision matrix implies that the corresponding variables are conditionally independent, given the rest. This can be used to determine the model structure.

The empirical results in this study have been obtained on the test case of 15 groups of wind farms spread throughout the territory of Western Denmark (see Fig. 3.5).

The structure of the sample precision matrix has shown that the information observed at time t at zone A depends only on local information at A at times $t+1$ and $t-1$ and on the information at four neighbouring zones of A : Northern (N), Eastern (E), Southern (S) and Western (W) neighbours at times $t-1$, t , $t+1$ (see Fig. 3.6).



Figure 3.5: Geographical locations of the 15 zones of wind farms considered in the study

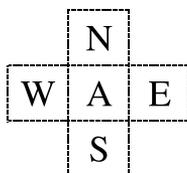


Figure 3.6: Neighbourhood specification of a single zone. The focus zone is denoted by A while N, E, S and W correspond to its Northern, Easterly, Southern and Western neighbours, respectively.

Conditional cross-correlations between the variables have been shown to be direction-dependent. That is, they are stronger in the West-East direction which is in line with the prevailing westerly winds over the territory as well as with the fact that in general, distances between the groups are closer in that direction.

Conditional precisions are also not-constant, but change with the prediction horizons.

The proposed modelling approach accounts for both the direction-dependent conditional correlations and for the horizon-dependent conditional precisions.

The results of the study have been obtained while evaluating predictive performance of the generated multivariate predictive densities. One year of data has been available for model validation. The results have shown the superiority of the proposed approach over the considered benchmarks in terms of the overall quality.

3.4 Forecast verification

This section gives a summary of the main results obtained in Papers F, G and H dealing with some methodological aspects of forecast verification.

As has been discussed in Section 2.5, in order to evaluate probabilistic forecasts one needs to employ proper scores. Propriety is seen as a basic property of a score, ensuring that forecasts coinciding with a real process generating density are given the best score value. When evaluation is done based on proper scores, the forecaster is encouraged to issue predictions based on his best judgement and aim at capturing the real process generating density rather than trying to hedge the score.

It appears that such a crucial requirement as scores being proper is not always respected. In Papers F and G emphasis has been on the recent proposal of the Coverage Width-based Criterion (CWC) as a score for evaluating prediction intervals. The papers have shown that the score is not proper and that there exist simple hedging strategies which can be used to obtain the optimal score value without providing any adequate description of the process. As a consequence the CWC score is not a valid score for evaluating predictive intervals and it is impossible to conclude on the superiority of one forecasting approach over another based on this criterion. Paper G provides a complementary discussion and examples on this issue.

Propriety, thus, can be viewed as a requirement for a score in order to be valid. It is, however, not sufficient for it to be informative. That is, by requiring scores to be proper, we essentially ask them to associate the optimal value with the real process generating density. This requirement in itself does not guarantee that such scores are able to discriminate between forecasts of different quality. For example, consider a score which always assigns a constant value to any kind of predictive density. This score is proper. However, in practice it is not useful as it is not able to rank the competing approaches.

In practice one wishes to have scores which are not only proper, but which are also able to distinguish between forecasts of different quality. Following

this, Paper H refers to the notion of discrimination. A score is said to have a high discrimination ability if differences in predictive densities translate to significant differences in the corresponding score values. In contrast, a score is said to have no discrimination ability if the same score values are assigned to various predictive densities.

Furthermore focus is on the discrimination ability of the Energy score (as defined in Eq. (2.13)). The Energy score is widely used in practice for evaluating probabilistic forecasts of multivariate quantities. This score is proper, thus in theory it can be used to assess predictive densities. However, certain concerns have been raised when trying to implement the score to evaluate the space-time trajectories generated in Paper E. The difficulties faced are explained in the following.

First, Energy score estimation calls for Monte Carlo techniques. This is computationally very expensive. More precisely, the cost of inference with a covariance matrix (if not restricted to any particular structure) is cubic in the dimension. This means that sampling from a multivariate Gaussian distribution and estimating the associated Energy score is hampered by the *big n* problem.

For example, in the particular case presented in Paper E the dimension of the problem was 645, and 8760 time steps were considered in a validation period (a year of hourly data) and for each time step 10 scenarios (which is very small given the dimension of 645) were generated in order to evaluate the score. Given this setup, it took more than 12 hours to estimate the score.

The problem, however, is not only the time it takes to estimate the score, but also in the uncertainty of the estimates stemming from the fact that rather few samples can be drawn at every time step in order to make it computationally feasible to process the whole data set.

Second, when estimating Energy scores for various predictive densities, it has been noticed that the score is not sensitive to the changes in the dependence structure. More specifically, Paper E has considered a multivariate Gaussian case with the marginal densities being known a priori. Therefore, the competing approaches have differed only in terms of the correlation structures.

Even though the data analysis has shown the variables to be highly correlated (with cross-correlation values reaching 0.8), the Energy score obtained under the assumption of independence has been very close to the one obtained when capturing the dependence structure: 1.5048 versus 1.4866, respectively. The difference is very small, even though it is still confirmed as statistically significant using Diebold-Mariano test statistics [72].

Such tiny differences in the Energy score are not exceptional. Similar results have been reported in a number of recent works focusing on multivariate probabilistic forecasts and the predictive modelling of interdependence structures, e.g. [59, 73, 74].

Owing to such small differences in Energy scores, Paper H has taken a closer look at the Energy score with the main focus being placed on its ability to discriminate between different interdependence structures.

The paper presents an analytical upper bound on the ability of the Energy score to discriminate between different correlation structures when considering multivariate Gaussian processes. This bound is derived based on the following scenario.

Suppose, that a real process generating density, G , is a multivariate Gaussian with some mean and a covariance structure

$$\Sigma_G = \sigma^2 \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \cdots & \vdots \\ 1 & 1 & \cdots & 1 \end{bmatrix} \quad (3.1)$$

This means that a single observation, $Y_t = [Y_{1t} \ Y_{2t} \ \cdots \ Y_{nt}]$ equals $[Y_1 \ Y_1 \ \cdots \ Y_1]$ with $Y_1 \sim \mathcal{N}(0, \sigma^2)$. That is, since the process is perfectly correlated, then if we know at least one element from Y_t , then we know all the rest explicitly, since $Y_{i,t} = Y_{j,t}, \forall i, j = 1, 2, \dots, n$

Suppose, a predictive density F follows a naive approach which, instead of trying to describe the correlation structure, assumes that the elements are completely independent. That is, the corresponding predictive density, F , is a n -variate Gaussian with a well-specified mean and a covariance structure

$$\Sigma_f = \sigma^2 \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} \quad (3.2)$$

That is, F suggests that information about one element in the observation vector Y_t does not give any information about any other element in Y_t .

We denote E_G and E_F the Energy scores given by the forecasts issued based on the real process generating density G and on the naive approach F , respectively.

The upper bound ΔE is the given by:

$$\Delta E = \frac{E_F - E_G}{E_G} \quad (3.3)$$

That is, ΔE describes how much the Energy score changes if instead of the perfectly correlated process generating density, one issues forecasts based on F , while totally neglecting a very strong interdependence structure between the variables.

Results in Paper H have shown that ΔE is independent of σ^2 and only depends on the process dimension n . Fig. 3.7 depicts how ΔE changes with the process dimension.

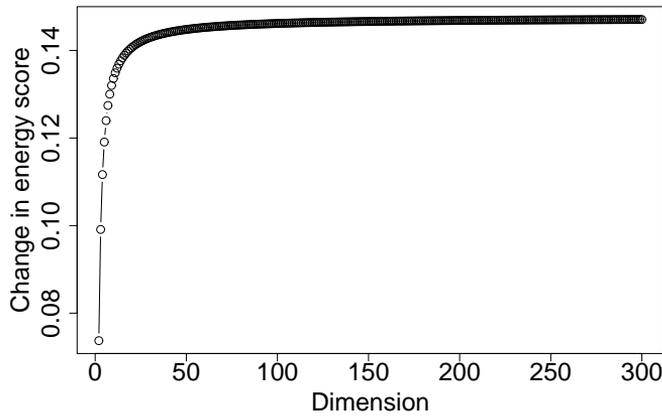


Figure 3.7: Upper bound on discrimination ability (i.e., ΔE) of the Energy score for a multivariate Gaussian process with well-predicted means and variances, as a function of the dimension n .

One can see that at the limit, the upper bound on ΔE reaches 0.14, which means that in high dimensions, even if the real process generating density is perfectly correlated, issuing forecasts which completely ignore this interdependence structure, would only result in the penalty reaching at most 14% of the score. This maximum is only achieved in this particular case of the real process generating density being perfectly correlated which is seldom (if ever at all) met in practice.

In addition Paper H has presented some simulation results which show that if the actual cross-correlation between variables is lower than 0.8, the change in the Energy scores does not exceed 3% (simulations are based on $n = 2$). This is very low and explains why in practice such minor differences in the Energy score values are reported in various scientific works which focus on modelling interdependence structures.

As a conclusion, the Energy score may be difficult to use in practice owing to its relatively low sensitivity to changes in dependence structures and potentially high uncertainty of the estimates.

Conclusions and Perspectives

4.1 Conclusions

Optimal integration of wind energy into power systems calls for high quality wind power forecasts. When initializing this research, operational wind power prediction systems were issuing forecasts for each location individually, without adequately accounting for the information coming from the neighbouring territories. However, it is intuitively expected that forecast errors made by the locally optimized approaches exhibit residual cross-correlations in space and in time owing to the inertia in meteorological systems.

In this dissertation we have shown that this intuitive dependence is indeed present, i.e. that errors made by the locally optimized prediction systems propagate in space and in time under the influence of meteorological conditions (mainly wind direction). Following this, new methods and models for capturing the residual space-time dependencies between the forecast errors have been proposed. The implementations of these models have shown that accounting for spatio-temporal effects improves the quality of the resulting wind power forecasts. Improvements have been observed when considering various types of power predictions: point forecasts, marginal predictive densities and space-time scenarios.

First, point forecasts and marginal predictive densities have been considered. The results obtained on the test case of Western Denmark have shown that the improvements in forecast quality are observed for a range of lead times up to 5-8 hours ahead. The peak improvements have been observed when considering one hour ahead predictions. This is in line with the geographical layout of the considered locations as well as with the speed of motion of weather fronts over the given territory [66]. The magnitude of the improvements depends on the positions of the “target” location and the available “sensor” sites with respect to each other, given the prevailing wind direction. More specifically, the improvements are higher when considering target locations which are situated “down-wind” with respect to the prevailing wind direction and, thus, can benefit from the information extracted from the “up-wind” territories. Also, the improvements are likely to be larger if considering aggregated territories, since data aggregation smoothens out local effects and places focus on a more global phenomenon. Furthermore, the results indicate that the improvements are larger if both the “target” and the corresponding “sensor” locations are of a similar dynamics (onshore, offshore).

It has also been shown that the obtained forecast improvements come from correcting the conditional expectation of future power generation (i.e. point forecast correction) which then leads to improvements in the probabilistic forecasts. Additional space-time correction when estimating higher order moments of the marginal predictive densities has been shown to be redundant.

Predictive performance of parametric and non-parametric probabilistic densities have been compared, uncovering that they both perform similar in terms of overall quality for lead times up to 5 hours ahead, and with an advantage for non-parametric predictive densities for further lead times. However, even though in terms of the overall quality the densities perform similarly, their conditional performance is shown to be different. Owing to the differences in conditional performance, forecasts made by the two approaches may have different values to the end-user depending on the particular application at hand.

An important part of the study has been devoted to developing new methods and models for generating space-time trajectories of wind power generation. Essentially such trajectories are given by random draws from multivariate predictive densities describing wind power generation at a number of locations over a period of time. Owing to the process complexity, estimating such joint density directly is a difficult task. Thus, we have proposed to employ a copula approach, which decouples the problem into two independent steps given by (i) modelling marginal predictive densities describing wind power generation at each site and each prediction horizon individually and by (ii) modelling the interdependence structure.

In this dissertation we have argued that a Gaussian copula is a reasonable choice for the given problem and that the best way to parametrize the dependence structure is by considering the precision matrix, i.e. the inverse of the covariance matrix.

It has been illustrated that the process precision matrix is very sparse. This has opened doors to the field of Gaussian Markov Random Fields with access to computationally efficient algorithms available for inference with sparse matrices as well as to a new way to interpret the dependence structure. That is, instead of revealing the global dependencies between the variables, a precision matrix represents the related conditional dependencies. In this study we have found these conditional dependencies are somewhat easier to model and to explain intuitively. Moreover, the proposed parametrization allows for more flexibility as one can easily obtain non-separable in space and in time dependence structures following a more complex pattern than the conventional exponential decay in time (and/or space).

Additionally, the study has revealed that the precision matrix is given by non-constant conditional precisions and conditional cross-correlations. This has put us beyond conventional approaches based on homogeneous stationary Gaussian fields. Data analysis has shown that the conditional precisions increase linearly with the prediction horizon, while the conditional cross-correlations depend on the direction. That is, when considering a test case of 15 groups of wind farms spread throughout Western Denmark, conditional dependencies between any two locations have been shown to be stronger in a West-East direction. This is in line with prevailing westerly winds over the territory as well as with the fact that in the given setup the distances between groups in that direction are generally shorter. The results have shown that the space-time trajectories generated using the proposed methodology outperform the benchmark approaches in terms of the overall quality.

Finally, the last part of the thesis has considered some methodological aspects related to verification of probabilistic forecasts of a very high dimension. Namely, focus has been on the Energy score which is a score commonly used for verification of probabilistic forecasts of multivariate quantities. We have shown (using both simulation results and some analytical derivations) that in the case of a multivariate Gaussian density, the Energy score has a rather weak ability to discriminate between different correlation patterns, provided that the marginal densities are well-specified. Such lack of sensitivity, coupled with the sampling uncertainty, means that it is difficult to use this score in practice when comparing between rival approaches which only differ from each other in terms of the dependence structure.

This is an important result, since the described Gaussian setup is not only met

when working with multivariate quantities which originally can be assumed as Normally distributed. The setup is also met when employing a Gaussian copula approach, and this is a very common choice for describing dependency between variables, independently of their marginal distributions. Thus, it is met in a wide range of practical applications.

4.2 Perspectives

The study raised a number of new questions and gave ideas for future work. In this section we do not aim to develop a general discussion on how we see or would like to see wind power forecasting in the (far) future. Rather, we follow a rather pragmatic aim to explain how we see the next steps in improving or extending the methodology presented in this thesis.

As far as point forecast corrections is concerned, the methodology proposed in this thesis is rather general as it does not require the presence of any specific meteorological or topographical patterns (e.g. strong channelling effects, strong thermal winds, etc). However, one assumption made by the model is that at a given time step a unique prevailing wind direction is sufficient to represent the weather regime over the whole territory. When moving to more complex case studies (larger areas potentially with various local wind climatologies) it could be beneficial to look at model extensions which could account for several dominant wind directions. In this case varying coefficient models or clustering techniques could be employed.

Furthermore, the results indicate that data aggregation is an important factor which helps capturing the underlying space-time dynamics. This calls for more studies on optimal aggregation techniques.

Another possibility for improving the methodology is to consider numerical weather predictions issued by different meteorological services and/or satellite images as additional explanatory variables. This data could give a better insight into the appearing fronts of imbalances between weather forecasts and observations and could help to better capture the patterns of the error propagation.

When considering marginal predictive densities, a possible improvement of the proposed methodology could be achieved by proposing better ways to quantify the uncertainty. In this work we assumed that changes in power variability can be best explained by the level of the expected power generation. However, other factors can influence the pattern of fluctuations: rain events, atmospheric stability, convective clouds, etc. One possibility is to consider radar or satellite

images in order to get new input to better explain power fluctuations.

Also, the results have shown that the proposed parametric and the non-parametric approaches provide forecasts of a similar overall quality. However, it has been shown, that the conditional performance of the resulting densities differs. Thus, it could be interesting to see how those densities compare when used as input to decision making problems involving wind power integration into energy systems. An interesting task could be to explore possibilities to combine the two approaches. That is, we could propose a regime-switching approach which could optimally choose which of the density types should be used to give the best description of power generation for the following time moment. One could also consider combining forecasts based on some meteorological conditions in the spirit of [75].

A lot of new ideas arise when considering space-time trajectories of wind power generation. The fact that we have found ourselves in the framework of Gaussian Markov Random Fields provides a lot of exciting opportunities. An interesting extension to the proposed methodology is to condition the precision matrix on meteorological conditions. This could be done by considering a regime switching approach.

Also, an interesting challenge is to move from the lattice setup considered in this study to a fully continuous approach. Based on the work of Lingren et al. [76] there is a link between stochastic partial differential equations and certain types of precision matrices. Thus, by understanding how the elements of the precision matrix depend on the distance between the zones and on the prevailing meteorological conditions, one can get a process description via stochastic partial differential equations.

In a broader context, there is a potential to generalize space-time trajectories by adding another dimension to the problem. That is, one could consider not only wind power forecasting, but, for instance, address wind and solar power simultaneously.

The interest in advancing forecasting methodologies further and further inevitably calls for more diverse and better ways to evaluate probabilistic forecasts of multivariate quantities. This calls for a lot of future research. One possibility could be to investigate whether it is possible to derive Mahalanobis distance based scores. Mahalanobis distance takes covariance structure into account and thus it is more sensitive to changes in correlation patterns than Euclidean distance (which is the core of the Energy score). However, it should be stressed that derivation of new scores is not a trivial task as one needs to ensure that the proposed scoring rules are proper. Thus, it is not certain that the idea of looking into Mahalanobis distance would lead to any new proper and useful evaluation

criterion.

Part II

Papers

PAPER A

Spatio-temporal analysis and modeling of short-term wind power forecast errors

Authors:

Julija Tastu, Pierre Pinson, Ewelina Kotwa,
Henrik Madsen, Henrik Aa. Nielsen

Published in:

Wind Energy, 36: 43–60, 2011.

Spatio-temporal analysis and modeling of short-term wind power forecast errors

Julija Tastu¹, Pinson¹, Ewelina Kotwa, Henrik Madsen¹, Henrik Aa. Nielsen²

Abstract

Forecasts of wind power production are increasingly being used in various management tasks. So far, such forecasts and related uncertainty information have usually been generated individually for a given site of interest (either a wind farm or a group of wind farms), without properly accounting for the spatio-temporal dependencies observed in the wind generation field. However, it is intuitively expected that, owing to the inertia of meteorological forecasting systems, a forecast error made at a given point in space and time will be related to forecast errors at other points in space in the following period. The existence of such underlying correlation patterns is demonstrated and analysed in this paper, considering the case-study of western Denmark. The effects of prevailing wind speed and direction on autocorrelation and cross-correlation patterns are thoroughly described. For a flat terrain region of small size like western Denmark, significant correlation between the various zones is observed for time delays up to five hours. Wind direction is shown to play a crucial role, while the effect of wind speed is more complex. Non linear models permitting capture of the interdependence structure of wind power forecast errors are proposed, and their ability to mimic this structure is discussed. The best performing model is shown to explain 54% of the variations of the forecast errors observed for the individual forecasts used today. Even though focus is on one-hour-ahead forecast errors and on western Denmark only, the methodology proposed may be similarly tested on the cases of further look-ahead times, larger areas, or more complex topographies. Such generalization may not be straightforward. While the results presented here comprise a first step only, the revealed error propagation principles may be seen as a basis for future related work.

¹DTU Informatics, Technical University of Denmark, Richard Petersens Plads, bld. 305, DK-2800 Kgs. Lyngby, Denmark

²Forecasting and Optimization for the Energy Sector A/S, Hørsholm, Denmark

A.1 Introduction

The optimal integration of wind energy into power systems requires high-quality wind power forecasts, preferably accompanied with reliable estimates of the forecast uncertainty. So far, state-of-the-art prediction systems typically provide forecasts for a single wind turbine, for a wind farm, or over a region with significant installed wind power capacities [1, 2]. Even if forecasting methodologies are developed for different spatial resolutions, see e.g. Siebert [3], the spatio-temporal interdependence structure in the wind generation field³ is seldom considered, since it is assumed to be fully captured by meteorological predictions used as input. Recently however, some research works have concentrated on wind speed prediction using spatio-temporal correlation with application to wind power prediction (for short look-ahead times, typically up to two-hours ahead), thus showing potential interest in accounting for these aspects [4, 5, 6]. These works mainly deal with cases where wind behavior between sites is easier to model, owing to terrain topology, or wind climatology. For instance, Larson & Westrick [4] consider the test case of a potential site for a wind farm located at the exit to the Columbia River Gorge, while meteorological observations are available from the entrance of this same Gorge. Another example relates to the work of Damousis *et al.* [5], for which information is available upstream of the location considered in the Thessaloniki area, and with quite steady prevailing thermal winds. In a more general setup, Gneiting *et al.* [6] have recently proposed several regime-switching models which account for two dominant wind directions while predicting wind speed up to two-hours ahead, with an interesting extension to probabilistic forecasting. The results of all the works mentioned above show significant improvements compared to benchmark prediction methods e.g. persistence. But, if these methods were to be applied to other types of case-studies, for which wind behavior is more complex and where no channeling effect is present, or for larger areas, one should then not expect similar performance of the models in terms of forecast quality. More advanced models may be needed for such cases, as discussed by Hering & Genton [7] for instance, potentially requiring significant expertise for identification of their structure or estimation of their parameters.

In operational conditions, state-of-the-art forecasting methods of wind power generation are commonly optimized with a focus on the wind farm (or aggregation of wind farms) of interest. So far, they do not account for potential information from neighboring sites, for example other wind farms or meteorological stations. Having a broader view of the forecasting problem, one could account for the possibility that, even though forecasting systems are optimized for local conditions, the inertia in meteorological systems might have the effect

³by wind generation field is meant here a complete description of the wind power generation characteristics over a domain of interest.

that a wind power forecast error at a certain point in space and time could propagate to other locations during the following period. Therefore, in view of the significant installed capacities of wind power installed all over Europe today (see current status and expected developments at www.ewea.org), analysis and understanding of the spatio-temporal characteristics of wind power forecast errors are of major importance. Indeed, errors in meteorological forecasts might translate to fronts of imbalances, i.e. taking the form of a band of forecast errors propagating across entire regions. Studies on the spatio-temporal characteristics of wind fields have already been deemed as highly informative for judging the adequacy of available generation and potential reserves in the UK for instance [8]. Regarding wind power forecasting errors, a relevant analysis of the spatial smoothing effect (thus related to the analysis of the correlation of forecast errors at the spatial level only) has been performed by Focken *et al.* [9] for the specific case of Germany. However, such an analysis does not provide information on how spatial patterns in forecast errors (or of smaller/larger forecast uncertainty) may evolve in space and time. Potential benefits of spatio-temporal analysis and associated modeling of forecast errors include global corrections of wind power forecasts, associated increased knowledge on the interdependence structure of forecast uncertainty, and correspondingly improved decision-making from the provided forecasts. This may concern both wind power producers with a geographically spread portfolio, and Transmission System Operators (TSOs) managing a grid with significant wind penetration. Better understanding of spatio-temporal dependencies may also be beneficial at the planning stage, for the optimal dispatch of wind farms in order to improve the predictability of wind generation at the regional level.

One of the main goals of this paper is to make a first step in analyzing spatio-temporal propagation of the wind power forecast errors. Therefore the first objective is to demonstrate that such a spatio-temporal interdependence structure of wind power prediction errors exists. Another objective is to show how some explanatory variables, more precisely wind speed and wind direction, may affect the nature and strength of this interdependence structure, in view of the geographical layout of the wind farms. A complementary objective is to propose models that have the ability to capture such effects. The case study considered relates to the western Denmark area, for which both hourly measurements of wind power and corresponding forecasts are available over a period of several months in 2004. Forecasts of wind speed and direction used as input to the wind power prediction method used in the analysis. A detailed description of this case study and available data is given in a first part of the paper. Subsequently, classical time-series analysis tools are employed in order to highlight the spatio-temporal characteristics of the wind power forecast errors. Based on the results of this analysis, a set of models and methods is proposed with the aim of capturing the revealed nonlinear behavior of forecast errors. Three types of statistical models are considered. Firstly a linear model based on observed

forecast errors for various groups of wind farms is presented. It is followed by a regime-switching approach permitting to switch between different linear models, depending upon the forecasted wind direction. Finally, the effect of wind speed forecasts is accounted for by generalizing the models considered, then taking the form of conditional parametric models. Some possible directions for future research are presented and discussed in a final part of the paper.

A.2 Case Study and Available Data

Owing to its already significant share of wind generation in the electricity mix as well as very ambitious objectives in the medium term, focus is given to the test case of Denmark. This country has set the goal of having 50% of the electricity demand met by wind energy in 2025 [10], which will clearly result in challenges related to the management of the grid. More precisely, the case study of this paper relates to the western Denmark area, including the Jutland peninsula and Funen island, which is connected to the UCTE (Union for the Co-ordination of Transmission of Electricity) system and has around 70% of the entire wind power capacity installed in Denmark. Another reason for the choice of this test case is that operational developments and application of wind power forecasting systems started around 1994 in Denmark [11], and it is thus common practice today to have forecasts of wind power production at different spatial resolutions and at a state-of-the-art level of accuracy. One more reason for choosing this area for the analysis is due to the orographical particularity of the territory. Denmark has a very smooth and flat terrain, while there is in general only one prevailing weather front dominating in the whole territory at any given moment. As a consequence, our analysis of spatio-temporal dependencies in forecast errors does not require for any particular orographic or vegetation particularities to be taken into account. Note that orographic effects at the very local scale are smoothed by the grouping process.

The data selected for this work comes from 22 wind farms of different nominal capacities (details can be found in [3]), and spread throughout the area considered. For all these wind farms, measurements of wind power production with an hourly resolution are available, along with wind power forecasts provided by the Wind Power Prediction Tool (WPPT). WPPT is a state-of-the-art forecasting system. Methods included in this forecasting system are described in [12, 13] (and references therein), while application results may be found in e.g. [14]. For the present case, it provides forecasts of wind power generation for each of the wind farms with a temporal resolution of one hour up to a 48-hour lead time. Forecasts are generated every hour. The inputs for WPPT are historical power measurements at the level of the wind farm con-

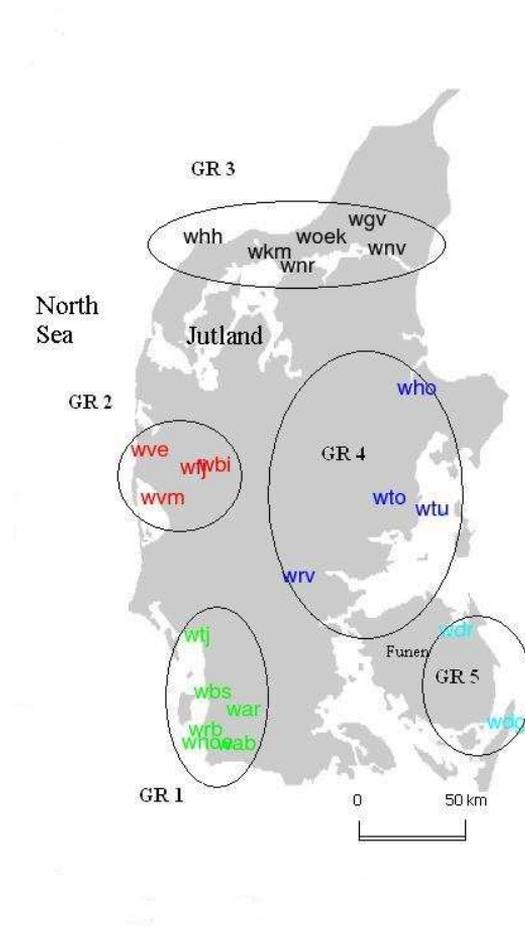


Figure A.1: Geographical locations of the 22 wind farms and selected groups of wind farms.

sidered, along with meteorological forecasts of wind speed and direction. An initial so-called power curve model permits the nonlinear conversion of wind speed and direction forecasts to power. In a second stage, a dynamical model permits recalibration of the power curve model output to correct for potential diurnal cycles not captured by the meteorological forecasts and to adapt to local conditions by accounting for the local dynamics of the wind farm considered. Adaptive estimation of the model parameters permits accommodation of long-term variations in the wind generation process characteristics because of e.g. seasonality, or ageing of the turbines. In the present case, the meteorological forecasts used as input are provided by the HIRLAM model

of the Danish Meteorological Institute (see http://www.dmi.dk/eng/index/research_and_development/dmi-hirlam-2009.htm). The forecasts are available over a 40×42 -nodes grid (horizontal resolution is 3 km) covering Denmark and surroundings, including a large part of the North Sea. To provide weather forecasts for each wind farm, data available at the HIRLAM nodes is respectively sub-sampled and interpolated (performed at DMI). Meteorological forecasts are delivered every six hours with an hourly temporal resolution up to 48 hours ahead. Wind forecasts are available at different vertical levels. Only wind forecasts at 10 meters a.g.l. (above ground level) have been considered here, though, due to the fact that WPPT also uses this particular level as input. The period for which both measurements and predictions have been made available for this study is from the fall of 2003 until July 2004. Since a new version of WPPT was installed in the fall of 2003, some time is needed for the model parameters to settle. Therefore, it was decided to disregard data originating from the last few months of 2003. The final dataset includes data from the first seven months of 2004.

Forecast errors are defined as the difference between power predictions and corresponding measurements, subsequently normalized by the installed wind power, following the framework described in [15]. Only one-hour ahead forecast errors are considered. The random variable corresponding to the forecast error at time t is denoted by x_t .

It has been chosen to study and model errors for groups of wind farms instead of concentrating on errors for each separate wind farm. This approach is preferred since spatial smoothing reduces the dependency on the local behavior and permits to focus more on global phenomena affecting the spatio-temporal movements of forecast errors. In the first step consisting of grouping the data, both the geographical layout of the wind farms and the extensive study performed by Siebert [3] have been accounted for. Based on clustering analysis, [3] proposed to form 3 groups of wind farms. We chose to further split data from 3 groups into 5 in order to have more flexibility while accounting for different error propagation directions. This splitting has been performed mainly considering the geographical layout of the wind farms. An additional correlation analysis was performed. It did not play a crucial role in our decision, as it was difficult to interpret. Since the core objective of the present paper is to check whether spatio-temporal error propagation can be modeled and used for forecast improvement, not much effort has been made to optimize the grouping of wind farms. Possibly the grouping technique could be the focus of further work, and result in additional improvement of forecast performance. The obtained groups of wind farms, along with the location of various wind farms, are depicted in Figure A.1. In the following, particular attention will be given to Group number 5 (corresponding to Funen island), as a large part of weather fronts propagation over Denmark come from the North Sea (mainly from W-

NW, see Figure A.2 demonstrating wind-rose plots for the forecasted wind at different groups). It is then expected that the most significant spatio-temporal characteristics of forecast errors will be observed if forecast errors at Group 5 are considered as the response variable to errors observed in the other groups of the Jutland peninsula.

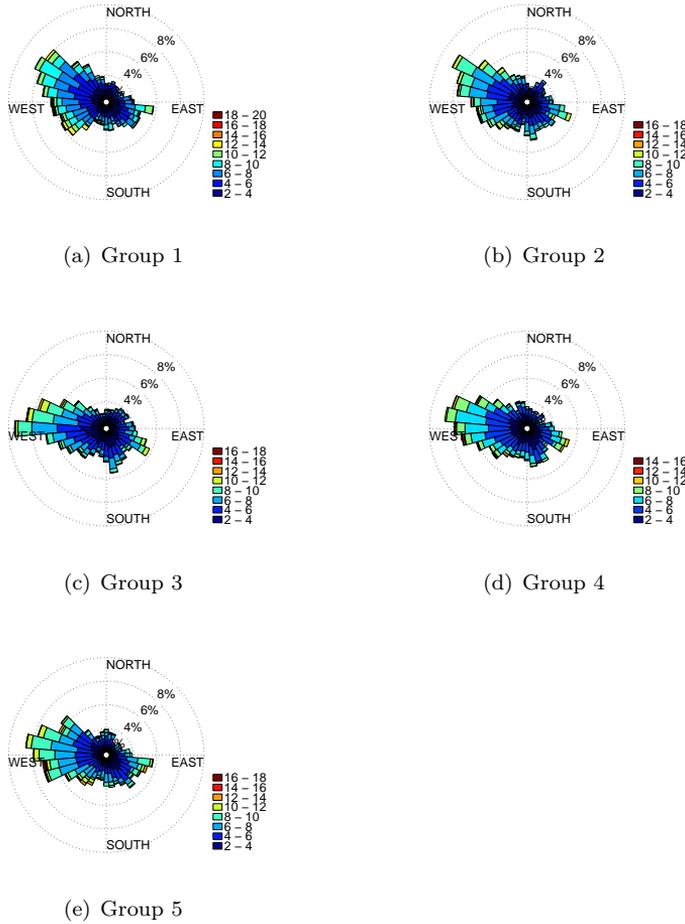


Figure A.2: Wind-rose plots showing the occurrences of forecasted wind speeds (m/s) and directions for each group.

The group errors have been calculated as an average of the errors within the groups. In parallel, since the aim is to study the effects of wind speed and direction on the spatio-temporal characteristics of forecast errors, a procedure was defined for obtaining representative wind speed and direction forecasts for each

of the groups of wind farms considered. Here, instead of employing a vectorial approach that would involve adding wind vectors, and then deriving average wind speed and direction from the norm and orientation of resulting vector, a more geometrical approach was used. The representative wind speed is given as the average of various wind speed values for the wind farms of the group. In parallel, wind direction at each of the farms defines the orientation of a set of unit vectors. The resultant vector then represents the wind direction for the group of wind farms. Note that these wind speed and direction data are forecasts, more precisely one-hour ahead forecasts, since focus is one-hour ahead forecast errors of wind power. To insist on this aspect, the notations \hat{u}_t and $\hat{\theta}_t$ will be used for wind speed and direction forecasts, respectively. Since meteorological forecasts are updated every six hours only, these one hour forecasts are obtained by using the last relevant available meteorological forecast series.

A.3 Highlighting some Spatio-temporal Characteristics of Wind Power Forecast Errors

An analysis of the available data is performed in order to reveal some of the spatio-temporal characteristics of wind power forecast errors. Such an analysis is crucial for understanding the underlying processes and for proposing a set of relevant models that would permit capture and reproduction of the various process characteristics. More precisely, the analysis performed aims at answering the following two questions:

- Is there a significant linear dependency within and between the groups, possibly with some time lag?
- Can the forecast variables, wind direction and wind speed, be used to contribute to revealing a stronger dependency?

The interest in answering the first of these two questions lies in the fact that, if linear dependency within and between groups (possibly with some time lag) is observed, it will then be straightforward to build linear models to capture such an effect. In parallel, a possible (nonlinear) relationship with some explanatory variables such as wind speed and direction forecasts could also be integrated in the proposed models with various approaches. It may appear as more relevant to study the dependency on the measured wind speed and direction, but since in real-world application such information will obviously not be available for the few following hours, forecasted values are preferred.

In the statistical literature there exists a set of standard tools that can be employed for analysing these types of linear dependencies in datasets (for more details see e.g. [16, 17]). In further works, nonlinear dependencies could be considered as well, with methods described in [18] for instance. For the questions raised in this paper the set of necessary tools includes Auto-Correlation Function (ACF) and Cross-Correlation Function (CCF). Each of these functions describes different types of dependencies and will be briefly introduced below. The analysis is structured as follows: (i) firstly the dependencies within each group of wind farms are examined; (ii) secondly the dependencies between groups are characterized; (iii) finally, the effects of wind speed and wind direction forecasts on both types of dependencies are studied.

A.3.1 Dependency within the groups

This section investigates the effects that previous values of time-series of forecast errors (for each group) have on the current state of the group. The time-series of forecast errors for the group of wind farms j is denoted by $\{x_{j,t}\}$ where t is the time index. The following analysis is based on the ACF of the time-series of forecast errors. An assumption for its use concerns the stationarity of the process considered, which in general terms translates to the idea that process characteristics do not change with time. For more information on (strictly) stationary stochastic processes see [17]. However conclusions on significant dependencies at various time lags can be formulated even though wind power forecast errors are not strictly stationary.

The ACF in lag k for the group of wind farms j , denoted by $\rho_j(k)$, is given by

$$\rho_j(k) = \rho[x_{j,t}, x_{j,t-k}] = \frac{\mathbb{E}[(x_{j,t} - \mu_j)(x_{j,t-k} - \mu_j)]}{\sigma_j^2} \quad (\text{A.1})$$

where μ_j is the mean of the time series $\{x_{j,t}\}$ and σ_j is its standard deviation. The ACF gives the correlation between the two lagged time-series $\{x_{j,t}\}$ and $\{x_{j,t-k}\}$. Therefore ρ_j takes values in $[-1, 1]$: 1 indicates a perfect positive linear dependency, -1 a perfect negative linear dependency, while 0 stands for no linear dependency at all. It is obvious that for $k = 0$ we have $\rho_j(0) = 1, \forall j$.

As focus is mainly given to Group 5, Figure A.3 illustrates the ACF of the corresponding time-series of forecast errors. Qualitatively similar results have been found for the other groups, and are not discussed here. Figure A.3 gives the value of $\rho_5(k)$ as a function of k , along with 95% confidence intervals under the assumption of independence for a Gaussian process. Please note, that the data actually is not absolutely Gaussian, even though it has some of its properties.

Therefore the 95% intervals shown are preliminary and used only for highlighting data characteristics, but can not be fully trusted for building models. In a hypothesis testing framework, one may then reject the hypothesis of independence $\{x_{5,t}\}$ and $\{x_{5,t-k}\}$ if the value of $\rho_5(k)$ lies outside of this interval. In practice, if the value of ρ for a given lag k is clearly outside this interval, one often concludes on a significant autocorrelation for that lag.

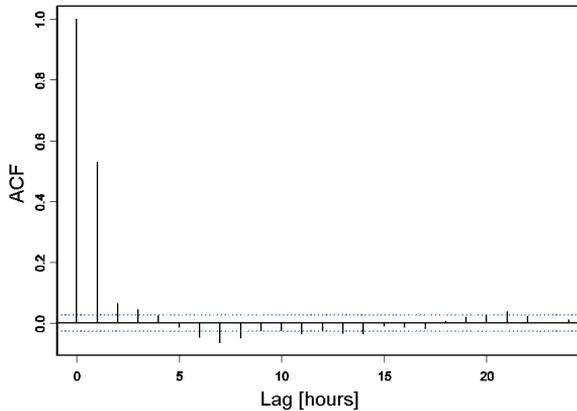


Figure A.3: ACF for Group 5, including a 95% confidence interval under the assumption of independence (dotted line). Values outside of this interval can be considered as significant correlation.

From Figure A.3, it can be seen that the ACF is a rapidly dampened exponential function, with a dominant autocorrelation in lag 1. The periodic waves for further lags are difficult to interpret. Non-negligible ACF values indicate that a fitted wind power prediction model (WPPT in this case) was not an ideal one, since the errors are not totally random. The better the fitted model is, the smaller the ACF values that would be observed. They would be 0 for all lags (starting from the lag 1) with an ideal prediction tool. Here, from looking at the ACF results, it is clear, that there is still room to improve prediction accuracy. In a general manner the $\{x_{5,t}\}$ time-series can be appropriately modeled with an Auto Regressive Moving Average (ARMA) model. This would translate to saying that there are two layers of dynamics in this time-series of forecast errors: a long-term inertia defining the MA part, and short-term dynamics making the AR part. It may be concluded from the Figure that there clearly are dependencies between forecast errors at different lags within a group of wind farms. However, some external signals (i.e. forecast errors for other groups of wind farms) may be related to such dependencies and this might better explain the observed behavior. This calls for further analysis.

A.3.2 Spatio-temporal dependencies between the groups

Once the autocorrelation pattern of forecast errors within the groups has been discussed, one can then proceed with the investigation of cross-dependencies between the groups. Information about potential cross-dependencies at certain time lags would be of great importance, since this provides crucial information for model structure identification. Demonstration of the existence of such a pattern would also translate to showing that there is spatio-temporal propagation of forecast errors between the groups.

For the case of cross-dependencies (possibly with some time-lag) between the time-series of forecast errors for the various groups of wind farms, the standard tool to consider is the CCF. The CCF between the time-series of forecast errors for Groups i and j , denoted by $\rho_{ij}(k)$, is given by

$$\rho_{ij}(k) = \rho[x_{i,t}, x_{j,t-k}] = \frac{\mathbb{E} [(x_{i,t} - \mu_i)(x_{j,t-k} - \mu_j)]}{\sigma_i \sigma_j} \quad (\text{A.2})$$

where μ_i and μ_j are the mean of the time-series $\{x_{i,t}\}$ and $\{x_{j,t-k}\}$, respectively, while σ_i and σ_j are their corresponding standard deviations.

As previously described, particular focus is given to Group 5, since its geographical location and the meteorological characteristics of western Denmark make it the most interesting group to study. Group 5 is located downwind of the other groups when the wind direction is from W-NW (which is dominant for that part of Denmark). Table A.1 summarizes the CCF evaluation (with respect to all other groups, and for lags between 0 and 5 hours) as well as the ACF evaluation performed above.

The cross-correlation values at lag 0 are significantly different from 0 for all groups, and this indicates that wind power forecasting errors for Group 5 have a tendency to be positively correlated with all the other groups. Furthermore, this correlation is typically higher for groups with a closer geographical location. Depending on the lag considered, the same group of wind farms does not always exhibit the highest correlation. For a time lag of one hour Group 4 shows the highest correlation, while for a time lag of two hours, Group 1 has the highest (as highlighted by the bold numbers in the Table). The forecast errors in the two other groups also have some correlation with forecast errors in Group 5 for the various time lags, though of minor magnitude. It is intuitively expected that this is due to the geographical layout of the various groups of wind farms and meteorological particularities of the area (prevailing W-NW wind), Groups 1 and 4 being the most strongly related to Group 5 (see Figure A.1).

Owing to the dominance of Groups 1 and 4 in the observed cross-correlation

Table A.1: Cross- and auto-correlation for Group 5.

	Group	lag					
		0	1	2	3	4	5
cross- correlation	1	0.175	0.282	0.307	0.217	0.119	0.070
	2	0.191	0.187	0.169	0.163	0.138	0.079
	3	0.1578	0.148	0.114	0.081	0.074	0.060
	4	0.2893	0.320	0.260	0.139	0.050	0.018
auto-correlation	5	1.000	0.527	0.059	0.040	0.019	-0.016

patterns, it was decided to further study their dependency on Group 5. Since a visual inspection of the CCF may be more informative, the corresponding CCFs are depicted in Figure A.4. The large cross-correlation values for the small lags denote the dependency on the lagged forecast error values for Groups 1 and 4 on the current forecast error at Group 5. More precisely, one retrieves the fact that for Groups 5 and 1 the highest cross-correlation is observed in lag 2, whereas for Groups 5 and 4 this peak is at lag 1. The reason is most likely due to the geographical layout, especially the distance between the groups. A closer look at Figure A.4 reveals periodic oscillations in the CCF for both groups for lags larger than 6-7 hours. In line with our comment about dependencies within a group, such oscillations indicate some long-term dynamics in the forecast error process.

A.3.3 Dependency on wind direction

The following analysis consists of assessing how wind direction forecasts can further characterise the spatio-temporal dependencies highlighted above. As it is known that wind direction clearly affects spatio-temporal dependencies in wind power production, similar effects are intuitively expected for the propagation of forecast errors. Group 5 is chosen here again as the group of focus, while the forecast errors from the other groups play the role of explanatory variables. Note that similar results could be obtained from considering forecast errors in any other group as the response, potentially explained by forecast errors in the remaining ones. They would not be as significant as for Group 5, as this group is ideally situated downwind from most of the other groups.

In order to examine whether wind direction has any effect on the observed spatio-temporal dependencies between forecast errors for the various groups, the available dataset of forecast errors is divided according to the forecast wind

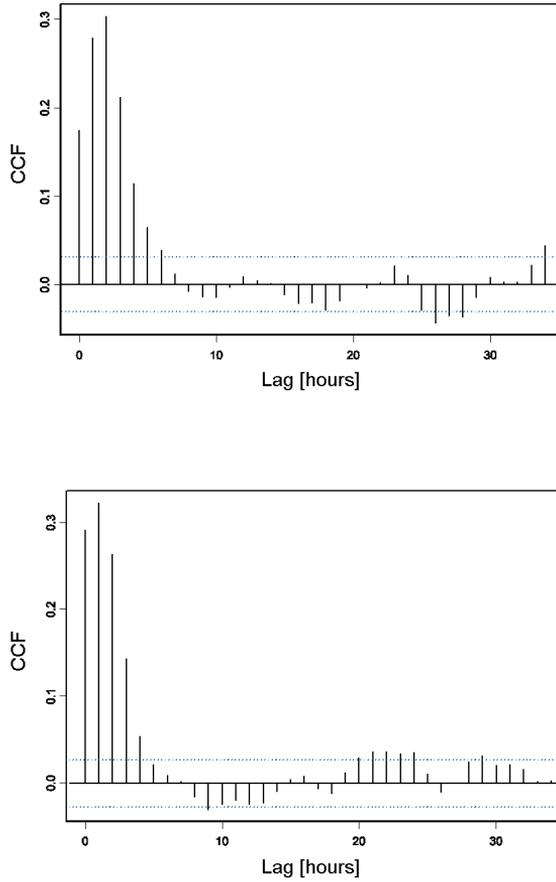


Figure A.4: CCF for the Groups 5 and 1 (top) and Groups 5 and 4 (bottom). Dotted lines show 95% confidence intervals under the assumption of independence. Values outside of such intervals can be considered as significant correlation.

direction in Group 5. The division is performed by constructing four intervals for potential wind directions: (0-90], (90-180], (180-270] and (270-360]. Therefore each interval corresponds to a specific sector, i.e. (0-90] the sector between North and East, (90-180] that between East and South, etc. For each of these sectors, a correlation analysis is performed between forecast errors at Group 5 and those of the other groups. This then translates to performing some kind of regime-based analysis of the spatio-temporal dependencies, the regime being defined by

wind direction only. Owing to the fact that the correlation structure of forecast errors is studied conditional on the wind direction, such correlation is referred to as *directional correlation* in the following. As it is shown in the previous section, that the strongest correlation structures are between Groups 5 and 1, and between Groups 5 and 4, only corresponding results are given here. The cross-correlation values for lags ranging between 0 and 5 hours are given in Tables A.2 and A.3.

Table A.2: *Directional correlation for Groups 5 and 1, for lags ranging from 0 to 5 hours.*

lag	regime			
	(0-90]	(90-180]	(180-270]	(270-360]
0	0.0457	0.1472	0.2240	0.1580
1	0.0499	0.2856	0.3597	0.2361
2	0.0672	0.3103	0.4213	0.2219
3	0.0358	0.1810	0.3218	0.1542
4	-0.0166	0.0985	0.2193	0.0519
5	0.0115	0.1130	0.1347	-0.0099

Table A.3: *Directional correlation for Groups 5 and 4, for lags ranging from 0 to 5 hours.*

lag	regime			
	(0-90]	(90-180]	(180-270]	(270-360]
0	0.1390	0.3200	0.2615	0.3460
1	0.2212	0.2691	0.2570	0.4514
2	0.1788	0.2049	0.2075	0.3762
3	0.1288	0.1555	0.0978	0.1831
4	0.1014	0.0965	0.0158	0.0485
5	0.0252	0.0735	0.0102	-0.0157

Recall that the analysis of the spatio-temporal dependencies in Table A.1 revealed that there was a maximum correlation at lag 1 between forecast errors for Groups 5 and 4, and at lag 2 between forecast errors for Groups 5 and 1. This can be seen again in Tables A.2 and A.3. Focusing on the correlation pattern between Groups 5 and 1, reveals that the correlation at lag 2 is at its maximum when the forecast wind direction is in the (180-270] sector (between South and West). Even for the other lags, the maximum correlation value between forecast errors for Groups 5 and 1 is attained in this wind direction regime. It then seems

that the wind power forecast errors have a tendency to propagate following the wind direction. Note that there is also significant correlation at various lags for the two adjacent sectors, i.e. for wind directions originating from the sectors (90-180] and (270-360], although these are of lower magnitude. This is in line with the idea that errors in wind power forecasts are directly linked to errors in weather forecasts. In case the input meteorological forecasts are wrong, they are likely to be wrong over a part of the region considered, if not the whole region, thus leading to a non-negligible correlation of wind power forecasts among the groups. For the case of Groups 5 and 1 (Table A.2) note that the values in column 2 are very similar to the ones of column 4. This could be explained by the fact that, for North-West and South-East wind, both Group 5 and Group 1 meet the weather condition at approximately the same time. None of these two groups is clearly up-wind in these two sectors. Correlation values are finally much lower for the remaining sector (wind directions in (0-90]) and this for all lags. In parallel, for the case of the correlation pattern of forecast errors for Groups 5 and 4, it is clear that in general correlation values are higher than for the case of Groups 5 and 1. This may certainly be explained by the fact that Groups 5 and 4 are geographically closer than Groups 5 and 1. Then, similar to the above, it seems that forecast errors tend to propagate following the wind direction, since the maximum correlation between Groups 5 and 4 (for a lag of one hour) is observed for wind sector (270-360], which is consistent with the geographical layout of the groups of wind farms. For this wind sector, Group 4 is located upwind of Group 5. The fact that the lag for which the maximum correlation is reached is shorter for Group 4 than Group 1 confirms the importance of distance between groups. In conclusion, it appears that the impact of wind direction on forecast errors is quite straightforward: they seem to be transported by the wind and thus propagate along the prevailing wind direction.

A.3.4 Dependency on wind speed

Since wind appears to be a driving force for the propagation of wind power forecast errors, another potential explanatory variable to be examined is the wind speed forecast. Indeed, as the distance between groups of wind farms seems to play a significant role, wind speed should also make the propagation of forecast errors slower or faster. This holds even though the speed of the error propagation is not necessarily the same than the forecasted wind speed, as the speed/direction of atmospheric features might be different with the surface wind speed/direction. In order to study the potential effect of wind speed, our strategy is to divide the dataset depending on wind speed, and to analyze the correlation pattern of forecast errors (as a function of the lag). This is done for each wind sector individually, as the impact of wind speed may be more significant for the wind sector that exhibits the clearest interdependence

of forecast errors. The propagation of forecast errors may then be seen as a finite impulse response conditioned by wind speed. As an example, focus is given here to the correlation pattern between Groups 5 and 1, and for wind sector (180-270], for which Group 1 is located directly upwind of Group 5. The dataset is divided into five smaller datasets depending on the (forecasted) wind speed in Group 5. The five wind speed intervals considered are (in m/s): 1 - [0,4), 2 - [4,6), 3 - [6,8), 4 - [8,10), 5 - [10,25). The CCF is then calculated for each of these wind speed intervals, and for lags between one and seven hours. Figure A.5 then illustrates how the correlation pattern varies depending on the wind speed. Note that this analysis has some restrictions, as the number of observations is not the same among the intervals. Each of them contains 200-400 data points, which makes the results significant, but not straightforward to compare, since the level of significance for the estimates differs from interval to interval. However, it may still allow us to observe some general features that would be explained by the wind speed level.

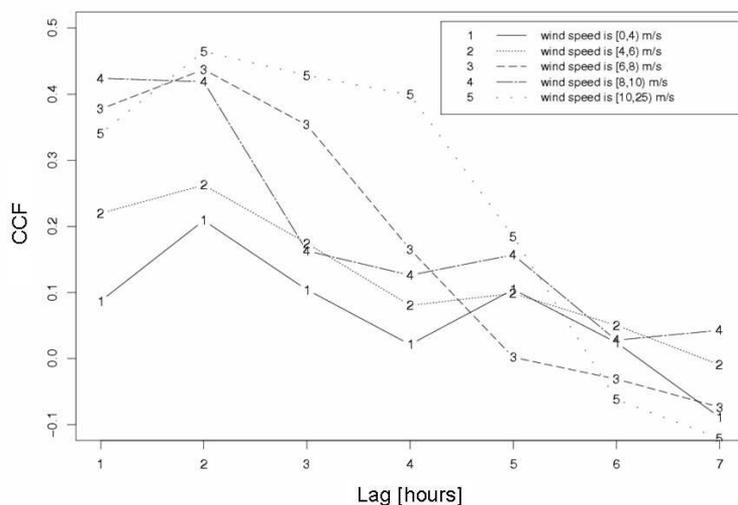


Figure A.5: Cross-correlation between forecast errors for Groups 5 and 1 and for the wind sector (180,270]. Cross correlation is given for different wind speed levels, and as a function of the lag.

Note from Figure A.5 that there is a general trend such that the average correlation of forecast errors between Groups 5 and 1 (and for the wind sector considered) increases as the wind speed gets larger. Indeed for low wind speeds,

forecast errors may be mainly due to local phenomena, and thus do not propagate at all to the neighboring groups of wind farms. For higher wind speeds, one retrieves the finite impulse behavior mentioned earlier, with wind speed directly influencing the magnitude of the correlation between forecast errors, as well as the lag for which this correlation is maximum. This particular lag is of 1-2 hours. Note that the high correlation for lags up to 5 hours in the case of high wind speeds (i.e. here between 10 and 25 m/s) may be due to very large and long-lasting discrepancies e.g. phase shifts between forecasts and measurements. Such phase shifts correspond to timing (or phase) errors in the forecasts and directly translate to clusters of errors of significant magnitude with the same sign, thus increasing their observed autocorrelation. This phenomenon is more common for higher levels of wind speed, in relation to meteorological fronts crossing the area, and to significant ramping in wind power generation. In a general the effect of wind speed on the propagation of wind power forecast errors appears to be more difficult to perceive than that of wind direction and is clearly nonlinear (Figure A.5). If the dependence was linear, the shape of the CCF for different wind speed intervals would be the same with a possible shift in the dominant lag or potential linear deformation of the CCF. This nonlinear effect can also be seen for other wind sectors and other groups (though not shown and commented on here), even though this effect is also conditioned by the geographical layout of the groups of wind farms, mainly their respective positioning and the distance between them.

A.4 Proposal of Relevant Models

In the above analysis, it has been demonstrated that wind power forecast errors indeed have some spatio-temporal characteristics at the level of western Denmark, and that this propagation of forecast errors is also affected by wind speed and direction. Our objective in this section is then to propose a set of relevant models that may be used to capture and reproduce the observed behavior of forecast errors. Remember that focus is here is on one-hour ahead forecast errors here but that an analogous methodology could be applied for the modeling of forecast errors related to further look-ahead times. Since the most significant correlation patterns observed over the whole data analysis are those obtained when concentrating on forecast errors at Group 5, it is decided to concentrate on this case. The overall methodology and set of models are thus introduced for this specific case, though they could be similarly derived if considering other groups of wind farms. The one-hour ahead forecast errors at Group 5 are seen as the response variable and denoted by y_t (instead of $x_{5,t}$), while the explanatory variables, which consist of the one-step ahead forecast errors in Groups 1 to 4, are denoted by $x_{1,t}, \dots, x_{4,t}$, t being the time index. The notation for the

errors in Group 5 is here changed in order to make it easier to see the difference between response and explanatory variables.

Three types of models appear to be relevant for modeling the observed spatio-temporal characteristics of the wind power forecast errors. Firstly since there is some significant linear correlation between forecast errors at different time lags and for different groups, a straightforward starting point is to use Auto-Regressive models with eXogenous input (referred to ARX models in the following). Such models also comprise a natural benchmark against which more complex models should be evaluated. Indeed, for capturing the effect of wind direction on the spatio-temporal characteristics of forecast errors, it is proposed in a second stage to use a regime-switching approach. Such a regime-switching approach will permit switching between different ARX models, depending on the forecast wind direction. Finally, the more complex effect of wind speed on these spatio-temporal characteristics is accounted for by upgrading ARX models to conditional parametric models in each wind direction regime, thus making the coefficients of the model a nonparametric function of wind speed. In all cases, it is assumed that the time-series considered have stationary properties. This assumption may be relaxed in the future, and model coefficients may be adaptively estimated in an estimation framework including exponential forgetting.

A.4.1 Linear models

As the most simple linear model to be employed for the modeling of one-hour ahead forecast errors, one may think of a simple AutoRegressive (AR) model. However, since our aim here is to consider the spatio-temporal effects highlighted above, it appeared more relevant to also account for some explanatory variables, namely the one-hour ahead forecast errors observed in the other groups and for different points in time in the past. This then led to the building of an ARX model. For more information related to the theory behind the building of ARX models, we refer to [17, 16, 19]. The general structure of an ARX model is given by

$$y_t = \beta_0 + \sum_{l=1}^p \beta_l y_{t-l} + \sum_{i=1}^n \sum_{j=1}^{k_i} \beta_{i,j} x_{i,t-j} + \epsilon_t \quad (\text{A.3})$$

where the response variable y_t is linearly explained by its p previous values in the auto-regressive part, and by n external input variables, each up to lag k_i ($i = 1, \dots, n$). ϵ_t is a purely random variable with zero mean and finite variance, which represents the noise that cannot be explained by the model.

The estimation of ARX model parameters can be straightforward performed

with Least Squares (LS) estimation methods. Again, extensive details and discussion on this topic can be found in [17, 16, 19]. The procedure employed for selecting the input explanatory variables and their lags is detailed in [20]. Finally, the model structure obtained is the following

$$y_t = \beta_0 + \sum_{l=1}^7 \beta_l y_{t-l} + \sum_{i=1}^3 \beta_{1,i} x_{1,t-i} + \sum_{j=1}^2 \beta_{4,j} x_{4,t-j} + \epsilon_t \quad (\text{A.4})$$

meaning that the current one-hour ahead forecast error in Group 5 can be explained by a linear combination of its last 7 values, in addition to last 3 forecast errors made for Group 1 and the last 2 forecast errors made for Group 4. This is consistent with the results from the analysis of the spatio-temporal characteristics of forecast errors performed in the previous section. Note that number of lags used in the model is different from the number suggested by Figures A.3 and A.4. Information from those figures is only used as a first step towards understanding and highlighting data characteristics. Final model structure is decided on the basis of Akaike's Information Criterion (AIC) and optimization of determination coefficients (see [20] for exact details) in order to achieve the best possible performance of the model.

A.4.2 Regime-switching models based on wind direction

Here we recall the idea of what we defined above as directional correlation, which was used in the data analysis performed above. The main purpose of defining such directional correlation is to analyze and model the effect of wind direction on the spatio-temporal dependencies of forecast errors. We claim that if the wind direction is compatible with the direction of the vector having its beginning in a given group of wind farms and ending in another group of wind farms, then the dependency between errors for these two groups (possibly with some lag) should be higher than in case of different directions.

Regime-switching models extend the idea of linear models by having a set of linear models, each of them being active in a certain regime. The switch between regimes can be governed by previous values of the response variable, external signals or unobservable stochastic processes. Here, focus is on the second type of regime-switching models as the regime switches will be governed by the wind direction forecast in Group 5. Regimes are defined by threshold values for the wind direction variable $\hat{\theta}_t$. These thresholds correspond to the upper bounds of the intervals in which the given 'sub-model' is active. The corresponding models employed may then be referred to as Threshold AutoRegressive with eXternal input (TARX) models. This type of regime-switching model has initially been introduced in [21], and extensively described in [22]. For the specific case of the

wind power application, basic concepts of regime-switching modeling may be found in [23].

The potential range of values for the wind direction variable $\hat{\theta}_t$ is $\mathcal{I} = (0, 360]$. Define intervals $R_1 \cup \dots \cup R_k = \mathcal{I}$ such that $R_i \cap R_j = \emptyset, i \neq j$. Each interval is given by $R_i = (r_{i-1}, r_i]$. The values r_0, \dots, r_k are the so-called threshold values which define switches between regimes. The threshold values are in general to be estimated from the data. However here, we consider the case when the values are known in advance, since they have been derived from an analysis of the data similar to that performed above. The motivation for such an assumption is that we analyze the case for which the regimes are governed by wind direction. From physical knowledge and intuition about the process characteristics, the choice of regimes may be fairly straightforward. The general form of the models examined further is

$$y_t = \beta_0^{(s_t)} + \sum_{l \in L_y^{(s_t)}} \beta_l^{(s_t)} y_{t-l} + \sum_{i=1}^4 \sum_{j \in L_{x_i}^{(s_t)}} \beta_{i,j}^{(s_t)} x_{i,t-j} + \epsilon_t \quad (\text{A.5})$$

where

$$s_t = \begin{cases} 1, & \text{if } \hat{\theta}_t \in R_1 \\ 2, & \text{if } \hat{\theta}_t \in R_2 \\ \vdots & \\ k, & \text{if } \hat{\theta}_t \in R_k \end{cases} \quad (\text{A.6})$$

In the above, $\hat{\theta}_t$ serves as the external signal which determines regime switching, t being the time index. In parallel, y_t is the response variable i.e. the one-hour ahead forecast errors at Group 5, the $x_{i,t-j}$ are the forecast errors for Group i and at lag j , and $\{\epsilon_t\}$ is zero mean white noise. $L_y^{(s_t)}$ and $L_{x_i}^{(s_t)}$ are sets of non-negative integers defining the auto-regressive and input lags (for Group i) of the model. The superscript (s_t) indicates that these sets of integers may be different for each of the regimes, i.e. along for different model structures depending on wind direction. The $\beta_{j,i}^{(s_t)}$ coefficients are the linear coefficients to be estimated in each regime s_t . Since the thresholds are known, the estimation problem for TARX models is solved by fitting different linear models to the data in each of the regimes. The estimation method to be employed is described in detail in [23].

For the test considered in the present paper, after analysis of the data in order to split it into various wind direction regimes, and then in each regime in order to identify the structure of the linear models (for more details, see [20]), the following general structure of the TARX model was obtained. First of all, the

regimes are defined as follows:

$$s_t = \begin{cases} 1, & \text{if } \hat{\theta}_t \in (0, 90] & \text{(North-East sector)} \\ 2, & \text{if } \hat{\theta}_t \in (90, 180] & \text{(East-South sector)} \\ 3, & \text{if } \hat{\theta}_t \in (180, 270] & \text{(South-West sector)} \\ 4, & \text{if } \hat{\theta}_t \in (270, 360] & \text{(West-North sector)} \end{cases} \quad (\text{A.7})$$

Originally, the choice for these regimes was dictated by easiness of interpreting the effect of wind direction forecasts which in this case is compatible with geographical cardinal directions. In fact, other divisions of the range of wind direction values were studied, and the improvement in model fit was considered insignificant or none. In a second stage, focus is on the linear models to be fitted in each of the regimes. The optimal number of lags is selected separately for each of the regimes. Table A.4 describes the structure of the resulting TARX model. While building the models, AIC was used to decide on the final number of lags used. The choice of the variables seems to be reasonable if the position of the groups of wind farms is taken into account (see Figure A.1). For example, the sector $(270, 360]$ corresponds to situations with the wind direction forecast from the North-West sector and in this case the effect on Groups 1 and 4 is seen to be most significant. Also, the maximum lags taken for Groups 1 and 4 conform with the directional distance from Group 5 in this regime. By the directional distance in this case we consider a projection of the distance between the corresponding groups on the axis following the middle wind direction of the current regime (e.g. equal to 315 for regime 4).

Table A.4: Threshold model structure: number of lags in the autoregressive part of the model, and selected lags for each of the other groups.

s_t	AR	Group 1	Group 2	Group 3	Group 4
1	10	-	-	4th	1st
2	5	1st	-	-	1st
3	6	1st, 2nd, 4th	-	-	1st
4	6	1st and 3rd	-	-	1st

A.4.3 Conditional parametric models with regime-switching

It is now aimed at upgrading the previous regime-switching model by integrating the complex nonlinear effect of wind speed on the spatio-temporal characteristics of forecast errors. The underlying idea is that the time delay for the propagation

of errors is directly linked to wind speed. For the purpose of accounting for such an influence, it is proposed here to transform the linear models in each of the regimes of the TARX model described above by conditional parametric models. Conditional parametric models comprise a class of models with a linear structure (like an ARX model), but for which the linear coefficients are replaced by smooth functions of other variables. For an extensive description of conditional parametric models, we refer to [24, 25].

More specifically, it is chosen to employ conditional parametric ARX models in each of the regimes in order to obtain a conditional parametric regime-switching approach. A conditional parametric ARX model with the model coefficient being smooth functions of wind speed, and with regime switches based on wind direction, can be written as

$$y_t = \beta_0^{(s_t)}(\tilde{u}_t) + \sum_{l \in L_y^{(s_t)}} \beta_l^{(s_t)}(\tilde{u}_t) y_{t-l} + \sum_{i=1}^4 \sum_{j \in L_{x_i}^{(s_t)}} \beta_{i,j}^{(s_t)}(\tilde{u}_t) x_{i,t-j} + \epsilon_t \quad (\text{A.8})$$

where the regime switches with respect to wind direction forecast are governed by (A.6). $\{\epsilon_t\}$ is a white noise sequence, i.e. a sequence of independent and identically distributed random variables with zero mean and finite variance. In addition, as in the case for the simpler TARX models introduced above, $L_y^{(s_t)}$ and $L_{x_i}^{(s_t)}$ define the model structure (i.e. the lags to be considered), with the superscript (s_t) indicating that these sets of integers may be different for each of the regimes. Note that for simplification and for direct comparison with the results that will be obtained with TARX models, the structure of the conditional parametric regime-switching model is defined similarly to the TARX model described above, that is, by (A.7) for the regime switches, and by Table A.4 for the model structure. In the following, conditional parametric models with regime-switching will be abbreviated as CP-TARX models.

Then, in contrast to the TARX models, the $\beta_{j,i}^{(s_t)}$ coefficients are smooth functions of a representative wind speed \tilde{u}_t (discussed below). Since the thresholds on wind direction are known, the estimation problem simplifies to the independent estimation of a conditional parametric model in each of the regimes. For this purpose, the LFLM (Local Fitting of Linear Model) software developed at the Technical University of Denmark [26] is employed. For an extensive description of the estimation methods involved, we refer to [13]. Coefficient functions have been locally approximated with first-order polynomials, for a number of 150 fitting points uniformly spread over the range of potential wind speed values. Tricube kernels have been chosen, with a nearest-neighbor bandwidth covering the 40% wind speed data closest to each fitting point, allowing smooth local estimates of the coefficient functions.

The variable \tilde{u}_t in equation (A.8) is a filtered wind speed at time t , which is

representative of the wind field potentially affecting forecast errors at the group of wind farms considered. Firstly, it was decided to take wind speed in Group 5 ($\hat{u}_{5,t}$) at time t as the representative wind speed \tilde{u}_t . Of course, in this case the information about wind speeds in other groups was lost and not accounted for by the model. It was therefore decided that \tilde{u}_t should be a summary of wind speed information at all the groups included in the model, based on a filter employing weighted linear regression. The weights were selected according to the corresponding coefficients of a linear regression of y_t on the errors from the other groups included in the model. For instance, assume that we want to explain y_t using $x_{1,t-2}$ and $x_{2,t}$. Then, in order to obtain the representative wind speed, the linear model $\tilde{u}_t = a\hat{u}_{1,t-2} + b\hat{u}_{2,t}$ is employed, where $\hat{u}_{1,t-2}$ and $\hat{u}_{2,t}$ denote wind speeds in Group 1 at time $t-2$ and in Group 2 at time t , respectively. The coefficients a and b in this model are the weight coefficients estimated for the model $y_t = ax_{1,t-2} + bx_{2,t}$. Such representation of a forecasted wind speed showed a better model performance in terms of R^2 , therefore was chosen for the further analysis. Note that filtered wind speed values certainly are different from the forecasted wind speeds. The values for the filtered wind speed range between 0 and 5 m/s.

A.5 Application Results

The objective of this section is to illustrate and analyze the ability of the various models presented above to capture the spatio-temporal characteristics of wind power forecast errors, as well as the effects of both wind speed and direction on those characteristics. The modeled errors are subtracted from the original forecasts issued by WPPT in order to get the forecasts adjusted after consideration of spatio-temporal dependencies. The accuracy of these adjusted forecasts is compared to that of original WPPT forecasts based on two different criteria. Here, models are fitted on the dataset considered, which has a limited size (seven months). Ideally, one year or more of data would be preferred. In addition, since regime-switching uses different models for each regime, this further reduces the amount of data used for estimation of model parameters. In order to optimally use this limited dataset, the approach employed is firstly to fit the various models to the whole dataset (seven months from the year 2004), with the aim of evaluating their ability to capture the effects highlighted in the previous section. In a second stage, a cross-validation exercise allows us to comment on the generalization ability of the models, i.e. on their potential ability to reproduce observed and modeled effects if trained and used on different data.

A.5.1 Model fitting

Comparison is made between the linear ARX model of equation (A.4), the regime-switching TARX model of equation (A.5), and the regime-switching conditional parametric CP-TARX model of equation (A.8). The structure of the last two models is detailed in Table A.4. Remember that the linear ARX model only accounts for autoregressive effects and linear effects from neighboring groups of wind farms, while the TARX model additionally accounts for the dependency on wind direction and the CP-TARX model aims at capturing the dependency on both wind speed and direction.

For evaluation of the fit of the various models, two criteria are employed. On the one hand, the coefficient of determination R^2 tells how much of the variations in the wind power forecast errors at Group 5 are explained by the models. Its value is between 0 and 1, 1 being a perfect power of explanation. It may then conveniently be expressed in percentage units. On the other hand, it is chosen to employ the Root Mean Square Error (RMSE) criterion. The RMSE is a quadratic error measure, thus giving more weight to large residuals and being in line with idea of LS fitting of the models. It is given here as a percentage of the installed capacity of Group 5. For more details on evaluation of statistical model fitting, we refer to [27, 19], and also to [15] for the specific case of the wind power application. Table A.5 gathers the corresponding results. Note that RMSE values have been calculated for the same data set on which the model parameters have been estimated, thus informing about the quality of the fit of the models in a LS sense. It may therefore be that the higher ability of some of the models to better explain the errors come from some form of overfitting. This will be discussed in more details and accounted for in the following subsection, when performing a cross-validation exercise.

The linear ARX model already has a certain ability to explain variations in wind power forecast errors at Group 5, since it has an R^2 of 47.8%. However in a general manner, this ability is increased by accounting for the effects of wind speed and direction. Indeed, TARX and CP-TARX exhibit higher values for the coefficient of determination, reaching 49.9% and 54.2%, respectively. The overall RMSE values for these two models are also lower than for the linear ARX model, with a non-negligible advantage for the more complex CP-TARX model. For comparison, the RMSE for one-hour ahead forecasts for this group of wind farms is 11.67% of nominal capacity before application of the various models studied here. Note that the fairly high level of original prediction error may be explained by the fact the nominal capacity for Group 5 is small. Such reduction in the RMSE criterion means that whatever the type of model chosen, the most reduction in forecast errors actually comes from the initial idea of accounting for spatio-temporal effects, while going for complex models, including wind speed

Table A.5: Evaluation of the fitting of the various models over the whole dataset. This evaluation is based on the coefficient of determination R^2 and on the error criterion RMSE. The regime is determined by the wind direction forecast.

Regime	ARX model		TARX model		CP-TARX model	
	R^2 [%]	RMSE [%]	R^2 [%]	RMSE [%]	R^2 [%]	RMSE [%]
1	-	-	38.4	7.5	48.5	5.3
2	-	-	46.1	4.4	48.3	4.3
3	-	-	49.3	6.8	55.6	6.4
4	-	-	54.9	5.6	57.7	5.4
Overall	47.8	5.8	49.9	5.7	54.2	5.4

and direction, mainly allows for better performance in certain meteorological conditions. Indeed, going into more detail, one notices differences in the values of evaluation criteria among the various regimes. These differences may be due to the more or less appropriate structures of the (sub)models in each regime, or due to different amounts of data used for model fitting, as well as different inherent predictability levels in various meteorological conditions. Since wind primarily blows from western directions over this region, a large share of the data available corresponds to regimes 3 and 4. Differences among regimes are of higher magnitude for the TARX model, with the CP-TARX model always having higher R^2 values as well as lower RMSE. The most significant improvements are observed for regimes 1 and 3, corresponding to the North-East and South-West direction, and for which the model structure is quite different. In the former case, the model mainly has an autoregressive pattern, while for the latter case the model has a lighter autoregressive pattern and relies more on past forecast errors at Group 1. This confirms a general interest of having the model coefficients as a function of wind speed.

After verifying that model residuals are not correlated, a bootstrapping technique (following the framework introduced in [28] and more specifically the functions described in [26]) is applied to check the level of uncertainty associated with the estimates of model coefficients. As an example, the results for regime 3 are shown in Figure A.6. Results for other regimes are qualitatively similar, and not discussed here. They are extensively commented on in [20]. A first interesting point with Figure A.6 is the noticeable evolution of the model coefficients as a function of the wind speed level. One sees for instance that as the wind speed level increases, there is a general trend that the autoregressive coefficients get closer to zero, while the coefficients values for the different

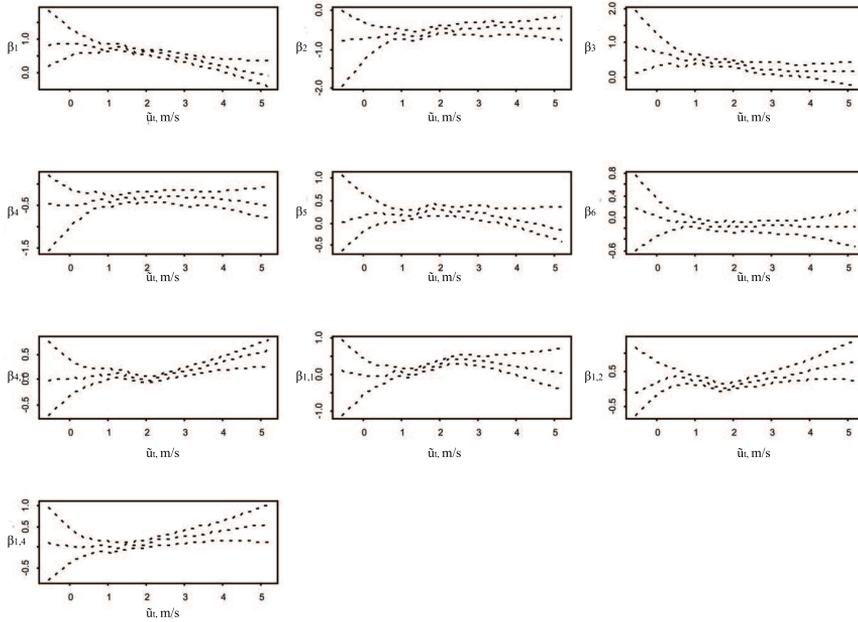


Figure A.6: Coefficients of the CP-TARX model (see equation (A.8)) fitted in regime 3, along with 95% confidence intervals based on 200 bootstrap replicates.

lags in forecast errors at Groups 1 and 4 globally increase. This observation is mainly based on the results from the area where the bootstrap confidence intervals are narrow enough to make it possible to conclude on the behavior of coefficients, i.e. where the wind speed level is between 1 and 3 m/s. In the areas corresponding to very low or high wind speed levels, due to the lack of data, the confidence intervals are broad, prohibiting determination of coefficient behavior. This translates to saying that the higher the wind speed, the larger the effect of upstream information (from Groups 1 and 4 in regime 3, see map in Figure A.1) and the less significant is the autoregressive pattern. Such a behavior may actually be fairly intuitive: as a wind front is stronger and moves faster, it possibly could transport forecast errors and dominate over local effects, which in contrast may be the main source of forecast errors for calm periods (thus corresponding to low wind speeds). In parallel, the impact of the distribution of representative wind speed values on the uncertainty of model coefficients is visible: as filtered wind speed values are more concentrated between 1 and 4 m.s⁻¹ the 95% bootstrap confidence intervals are fairly tight, while they get wider for representative wind speed values outside of this range. This uncertainty in the value of the model coefficients directly relates to an insufficient

amount of data available for those intervals. As may be noticed from the Figure, coefficient functions actually prolong for representative wind speed values below zero. This is due to the estimation method employed, and does not mean that representative wind speed values below zero may be encountered.

A.5.2 Cross-Validation and generalization ability

In the model-fitting exercise carried out above, both R^2 and RMSE measures have been calculated for the same data for which the model parameters have been estimated. Therefore, the higher ability of some of the models to better explain the variations of prediction errors may come from a numerical artifact, namely the so-called over-fitting. As a consequence, in order to verify if the models would perform similarly if applied to new (unseen) data, a cross validation procedure is employed. The idea of cross-validation is to use a subset of data for estimation of the model parameters, while the other subset is employed for model evaluation. More precisely, 3-fold cross validation is applied (as described in [19, 16]). The data in each regime is divided into three equal subsets. Two of the constructed subsets are used for parameter estimation and the third subset is used for checking the model performance. By repeating the procedure three times, one obtains three different estimations of the model parameters, with corresponding evaluation on independent subsets. The results are presented in Table A.6, with a focus on regimes 3 and 4 only, since these are deemed as more interesting in the above analysis, owing to the higher amount of data available, and better performance of the fitted models. Also, emphasis is on the TARX and CP-TARX models, since the effect of wind speed on the spatio-temporal characteristics of forecast errors is more complex, and the way conditional parametric models permit (or not) to capture them should be verified.

Table A.6: 3-fold cross validation results for both TARX and CP-TARX models in regimes 3 and 4.

Model (regime n°)	subset1		subset2		subset3	
	R^2 [%]	RMSE [%]	R^2 [%]	RMSE [%]	R^2 [%]	RMSE [%]
CP-TARX (regime 3)	47.8	5.89	40.4	7.78	49.2	8.03
TARX (regime 3)	48.1	5.93	47.8	7.20	47.5	7.96
CP-TARX (regime 4)	52.5	5.58	51.6	5.81	57.0	6.89
TARX (regime 4)	53.4	5.43	52.9	5.56	55.3	6.60

Cross validation results show that TARX models seem to have better generaliza-

tion ability than the more complex CP-TARX models. R^2 values for the three subsets for TARX models are fairly stable and at an almost similar level as for fitting performed on the whole dataset in the previous section (see Table A.5). RMSE values exhibit higher differences though. In parallel, the cross validation exercise for CP-TARX models yields more significant differences in both RMSE and R^2 from one evaluation subset to the other, with a significant decrease in R^2 if compared to the model-fitting results of Table A.5. And, in a general manner forecast accuracy for the TARX models is slightly better than that of the CP-TARX models, while this was not the case for using the whole data set above. Such results may be interpreted as a higher generalization ability of TARX models in comparison to the CP-TARX models. However, it is important to note the limited amount of data used in the present study. As already mentioned, the available data covers a period of seven months only. The number of observations in regimes 3 and 4 for this period is 1536 and 1640, respectively. When it comes to the cross validation exercise, each of the constructed subsets includes data from 2.33 months period only, making the number of observations available for the estimation step drop to around 1000 (which is 2 subsets or 4.66 months) for the specified regime. Taking into account that each set also has to be divided according to wind direction and that different wind speed levels have to be considered, it is likely that this seven-month period is actually not sufficient to draw final conclusions on the ability of CP-TARX models to capture the spatio-temporal characteristics of forecast errors accounting for the effects of wind speed and direction. Results obtained for the TARX models may appear as more trustworthy as they are based on more data for each wind regime considered (since no division according a wind speed is needed), but this may not be still the case if extending this study to longer periods.

Considering only the data in regimes 3 and 4, the RMSE averaged between the cross-validation subsets is 6.71% and 6.49% for the CP-TARX and TARX models, respectively. For comparison, the RMSE estimated on the same data subset before applying any of the studied spatio-temporal models is 8.99%. This evidences that both of the presented models can significantly reduce forecasting errors of the state-of-the-art prediction tool.

A.6 Conclusions and Perspectives

The present paper can be seen as the first step towards understanding and capturing the complex nature of spatio-temporal propagation of wind power forecast errors. The test case of the western Denmark area is of particular relevance, in view of the significant installed wind power capacities spread over this region, and of the resulting management challenges for the TSO or for power

producers with a geographically spread wind portfolio. A thorough analysis of the available forecast and measurement data has permitted formulation of a set of important conclusions. Such conclusions go along the line of our main objective, which is to show that there clearly exists some spatio-temporal patterns in the characteristics of wind power prediction errors. First of all, there exists in general a significant cross-correlation between forecast errors for neighboring areas with lags of a few hours. For the present case study, lags with significant dependency are up to five hours, while the lags with most effect are the one and two-hour lags. This cross-correlation pattern is clearly conditioned by the prevailing weather situation, mainly characterized by wind speed and direction. Wind direction is shown to play a crucial role, while the effect of wind speed is more complex. Prevailing wind speed affects the dependency in the following way: the higher the wind speed the stronger the dependency on more remote places; while in case of lower wind speeds, more influence comes from a local origin (thus exhibiting an autoregressive pattern).

In terms of modeling, this means that the dependency on wind direction may be easily accounted for by state-of-the-art regime-switching approaches, while dependency on wind speed should be captured by more complex models. This has been performed here by embedding conditional parametric models in the regime-switching approach. The superiority of such a proposal for capturing the complex effect of wind speed has not been demonstrated, possibly because of the limited size of the available dataset (only data from a seven-month period was available). The best spatio-temporal model proposed has been shown to explain up to 54% of one-hour ahead wind power forecast errors in terms of R^2 . When applied to new, "unseen" data, the regime-switching model has shown the ability to reduce the forecast errors from the initial 8.99% to 6.49% in terms of the RMSE criterion.

Note that owing to the choice of such a short look-ahead (1 hour ahead), forecast errors may be due to large ramps in wind power generation, which are difficult to predict when a strong weight is given to the past few power measurements (as is done by a state-of-the-art model like WPPT for forecasts up to ca. six-hours ahead). The various potential origins of the forecast errors do not alter the interest of the proposed approach, since they involve statistically characterizing spatio-temporal patterns in forecast errors, and subsequently taking advantage of this knowledge for forecast correction. The proposed analysis and methodology could also be extended to the case of errors for further look-ahead times (up to several hours ahead) if working on the same terrain as Denmark. In order to make models valid for data coming from a larger region or from a region with a more complex terrain than Denmark, some adjustments would have to be done in the modeling approach due to the fact that it is not always possible to use one prevailing wind speed or direction as a representative of the situation in the whole region. The methodology presented in this paper could

be evaluated for such more complicated cases. Possibly if working with a larger region and small time-lags, the region could be divided into sub-regions and each sub-region could be analyzed separately. If considering further look-ahead times (more than several hours ahead), then data from a larger region should be considered along with a thorough examination of the weather forecasts, in order to evaluate how weather fronts normally move along the region and which parts of the entire region may affect each other at the time scales considered. Such a generalization of the proposed methodology might not be straightforward. We believe, however, that the principles introduced for highlighting spatio-temporal characteristics of forecast errors, model building and estimation, can be seen as generic in future related work.

For a small area like western Denmark, which is the first to be touched by fronts coming from North-West, the use of online measurements from the United Kingdom, or from measurement devices in the North Sea, might lead to highly significant improvements on a longer time horizon. In parallel, considering the number of turbines spread over western Denmark, it appears crucial to propose a modeling approach that would allow for dynamic evolution of the overall wind installations. Indeed new wind farms should be easily accounted for in the model, without having to re-estimate all coefficients and/or change the structure of the existing models. A potential solution could be to employ a lattice approach, for which a data assimilation step would permit accommodation of all online measurements before modeling the spatio-temporal dynamical process. Then, in order to make the general approach more generic, and potentially applicable to larger regions (potentially with various local wind climatologies), methodology adjustments should account for the fact that it may not be possible to consider a unique prevailing wind speed and direction as being representative of the weather regime over the whole area. For such more complicated cases it may be needed to switch from conditional parametric models to varying-coefficient models.

A.7 Acknowledgments

The authors would like to thank Energi E2 A/S (now part of DONG Energy) for providing the wind power data, the Danish Meteorological Institute (DMI) for the meteorological forecasts, as well as Torben S. Nielsen and the WPPT team for the wind power forecasts. This work has been partly supported by the Danish Public Service Obligation (PSO) fund, through the ‘Improved Wind Power Prediction’ project (under contract PSO-5766), which is also acknowledged. Acknowledgements are finally due to two reviewers whose comments allowed to enhance the paper.

References A

- [1] G. Giebel, R. Brownsword, G. Kariniotakis, M. Denhard, and C. Draxl, “The state-of-the-art in short-term prediction of wind power—A literature overview.,” tech. rep., Technical University of Denmark, 2003.
- [2] 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*, vol. 12, no. 6, pp. 1725–1744, 2008.
- [3] N. Siebert, “Development of methods for regional wind power forecasting,” *Ph.D. Dissertation*, 2008.
- [4] K. A. Larson and K. Westrick, “Short-term wind forecasting using off-site observations,” *Wind energy*, vol. 9, no. 1-2, pp. 55–62, 2006.
- [5] I. G. Damousis, M. C. Alexiadis, J. B. Theocharis, and P. S. Dokopoulos, “A fuzzy model for wind speed prediction and power generation in wind parks using spatial correlation,” *Energy Conversion, IEEE Transactions on*, vol. 19, no. 2, pp. 352–361, 2004.
- [6] T. Gneiting, K. Larson, K. Westrick, M. G. Genton, and E. Aldrich, “Calibrated probabilistic forecasting at the stateline wind energy center: The regime-switching space–time method,” *Journal of the American Statistical Association*, vol. 101, no. 475, pp. 968–979, 2006.
- [7] A. S. Hering and M. G. Genton, “Powering up with space-time wind forecasting,” *Journal of the American Statistical Association*, vol. 105, no. 489, pp. 92–104, 2010.

-
- [8] M. S. Miranda and R. W. Dunn, "Spatially correlated wind speed modelling for generation adequacy studies in the UK," in *Power Engineering Society General Meeting, 2007. IEEE*, pp. 1–6, IEEE, 2007.
- [9] U. Focken, M. Lange, K. Mönnich, H.-P. Waldl, H. G. Beyer, and A. Luig, "Short-term prediction of the aggregated power output of wind farms—a statistical analysis of the reduction of the prediction error by spatial smoothing effects," *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 90, no. 3, pp. 231–246, 2002.
- [10] E. E. Analyses, "50% Wind Power in Denmark in 2025-English Summary," 2007.
- [11] H. Madsen, "Models and methods for wind power forecasting. Elsam/IMM, Denmark," tech. rep., ISBN 87-87090-29-5, 1996.
- [12] T. S. Nielsen, H. A. Nielsen, and H. Madsen, "Prediction of wind power using time-varying coefficient functions," in *Proceedings of World Congress on Automatic Control, Barcelona, Spain*, 2002.
- [13] H. A. Nielsen, T. S. Nielsen, A. K. Joensen, H. Madsen, and J. Holst, "Tracking time-varying-coefficient functions," *International Journal of Adaptive Control and Signal Processing*, vol. 14, no. 8, pp. 813–828, 2000.
- [14] N. Cutler, M. Kay, K. Jacka, and T. S. Nielsen, "Detecting, categorizing and forecasting large ramps in wind farm power output using meteorological observations and WPPT," *Wind Energy*, vol. 10, no. 5, pp. 453–470, 2007.
- [15] H. Madsen, P. Pinson, G. Kariniotakis, H. A. Nielsen, and T. S. Nielsen, "Standardizing the performance evaluation of shortterm wind power prediction models," *Wind Engineering*, vol. 29, no. 6, pp. 475–489, 2005.
- [16] C. Chatfield, *The analysis of time series: An introduction, 6th edition*. CRC press, New York, 2003.
- [17] H. Madsen, *Time Series Analysis*. Chapman & Hall/CRC: London, 2007.
- [18] H. A. Nielsen and H. Madsen, "A generalization of some classical time series tools," *Computational Statistics & Data Analysis*, vol. 37, no. 1, pp. 13–31, 2001.
- [19] H. Madsen and J. Holst, "Modelling non-linear and non-stationary time series," *Lecture Notes, Technical University of Denmark, Dpt. of Informatics and Mathematical Modeling, Kgs. Lyngby, Denmark*, 2000.
- [20] E. Kotwa and J. Vlasova, "Spatio-temporal modeling of short-term wind power prediction errors," *M.Sc. Thesis Dissertation*, 2007.

-
- [21] H. Tong, *Threshold models in non-linear time series analysis. Lecture notes in statistics, No. 21*. Springer-Verlag, 1983.
- [22] H. Tong, *Non-linear time series: a dynamical system approach*. Oxford University Press, 1990.
- [23] P. Pinson, L. Christensen, H. Madsen, P. E. Sørensen, M. H. Donovan, and L. E. Jensen, “Regime-switching modelling of the fluctuations of offshore wind generation,” *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 96, no. 12, pp. 2327–2347, 2008.
- [24] J. M. Chambers, T. Hastie, *et al.*, *Statistical models in S*. Chapman & Hall London, 1992.
- [25] W. S. Cleveland and S. J. Devlin, “Locally weighted regression: an approach to regression analysis by local fitting,” *Journal of the American Statistical Association*, vol. 83, no. 403, pp. 596–610, 1988.
- [26] H. A. Nielsen, “An S-PLUS/ R library for locally weighted fitting of linear models,” *Technical report, 1997-22*, 1992.
- [27] D. A. Pierce, “ R^2 measures for time series,” *Journal of the American Statistical Association*, vol. 74, no. 368, pp. 901–910, 1979.
- [28] B. Efron and R. J. Tibshirani, “An Introduction to the Bootstrap (Chapman & Hall/CRC Monographs on Statistics & Applied Probability),” 1994.

PAPER B

Multivariate Conditional Parametric models for a spatio-temporal analysis of short-term wind power forecast errors

Authors:

Julija Tastu, Pierre Pinson, Henrik Madsen

Published in:

Scientific Proceedings of the European Wind Energy Conference, Warsaw (PL),
77–81, 2010.

Multivariate Conditional Parametric models for a spatio-temporal analysis of short-term wind power forecast errors

Julija Tastu¹, Pinson¹, Henrik Madsen¹

Abstract

Optimal integration of wind energy into power systems requires high quality wind power forecasts, preferably accompanied by reliable estimates of the forecast uncertainty. So far, state-of-the-art wind power prediction systems generate forecasts for each site of interest individually, without properly accounting for information from the neighbouring territories. However, due to the inertia in meteorological systems, there exists a spatio-temporal inter-dependence between the sites, i.e. the errors in wind power forecasts propagate in space and time under the influence of meteorological conditions. In this work multivariate (vector) conditional parametric models are proposed to capture this phenomenon. It is shown that the adjusted wind power point forecasts result in a reduction in prediction errors. An uncertainty level associated with the new, adjusted forecasts is evaluated by providing a probabilistic density function based on a truncated multivariate normal distribution. The models are validated on the test case of western Denmark by considering one-hour-ahead wind power predictions. However the proposed methodology could be similarly tested on the basis of other areas with spatially sparse data and on the basis of cases with further look-ahead times.

B.1 Introduction

In operational conditions, state-of-the-art forecasting methods of wind power generation are commonly optimized with focus on the wind farm (or aggregation of wind farms) of interest. So far, they do not account for potential information from neighbouring sites, for example other wind farms or meteo-

¹DTU Informatics, Technical University of Denmark, Richard Petersens Plads, bld. 305, DK-2800 Kgs. Lyngby, Denmark

rological stations. With a broader view of the forecasting problem, one could account for the possibility that, even though forecasting systems are optimized for local conditions, the inertia in meteorological systems might have the effect that a wind power forecast error at a certain point in space and time could propagate to other locations during the following period. Therefore, in view of the significant installed capacities of wind power installed all over Europe today (see current status and expected developments at www.ewea.org), analysis and understanding of the spatio-temporal characteristics of wind power forecast errors are of major importance. Indeed, errors in meteorological forecasts might translate to fronts of imbalances, taking the form of a band of forecast errors propagating across entire regions. Studies on the spatio-temporal characteristics of wind fields have already been deemed as highly informative for judging the adequacy of available generation and potential reserves in the UK for instance [1]. Regarding wind power forecasting errors, a relevant analysis of the spatial smoothing effect (related to the analysis of the correlation of forecast errors at the spatial level only) has been performed by Focken *et al.* [2] for the specific case of Germany. However, such an analysis does not provide information on how spatial patterns in forecast errors (or of smaller/larger forecast uncertainty) may evolve in space and time. Potential benefits of spatio-temporal analysis and associated modelling of forecast errors include global corrections of wind power forecasts, associated increased knowledge of the interdependence structure of forecast uncertainty, and correspondingly improved decision-making from the forecasts available. This may concern both wind power producers with a geographically spread portfolio, and Transmission System Operators (TSOs) managing a grid with significant wind penetration. Better understanding of spatio-temporal dependencies may also be beneficial at the planning stage, for the optimal dispatch of wind farms in order to improve the predictability of wind generation at regional level.

The first step towards checking the existence and possibility of capturing spatio-temporal patterns in wind power forecast errors has been done in previous work by Tastu *et al.* [3]. It was demonstrated that wind power forecast errors do indeed have some spatio-temporal characteristics at the level of western Denmark, and that the propagation of the forecast errors is also affected by forecasted wind speed and direction. To capture the dependence on the wind direction, a regime-switching approach was suggested, i.e. the data was divided into four intervals according to the wind direction and in each regime a linear model was constructed. As a result, a difference in the model structure and performance observed in each interval showed clearly that forecasted wind direction plays an important role in propagation of wind power forecast errors. None of the models proposed in [3] is a ready-to-use-tool, but rather a demonstration of the existence of such spatio-temporal patterns. The main objective of this paper is to improve and extend the methodology described in [3] by presenting a ready-to-use-tool for correcting wind power forecasts based on spatio-temporal effects and

to evaluate the uncertainty level associated with the adjusted predictions. The proposed improvements concern both changes in the point forecasting approach and a general extension of the methodology from point to probabilistic forecasting. The latter is in line with recent publications showing that a transition from point to full probability forecasting can be very beneficial for an optimal decision-making process (see [4] and [5] among others). As far as changes in point forecasts are concerned, the following steps are taken:

- Instead of an abrupt division of data into intervals (regime-switching method) a smoother approach is proposed. It permits tracking the changes in coefficients within the whole range of possible values, not only within the limited number of intervals.
- As the previous work was mainly a preliminary data analysis and an examination of whether spatio-temporal models could be beneficial, it only concentrated on making predictions for one chosen group. In this work a multivariate approach is applied in order to model an entire region at the same time.
- The estimation method has been changed from offline to online when coefficients are being estimated recursively. Such an approach also allows for an exponential forgetting of old observations, which leads to the model being adaptive with respect to the long-term variations in the process characteristics.

In order to track uncertainty level associated with new, adjusted point forecasts, a parametric approach is employed: predictive densities are modelled as truncated multivariate normal distributions (following the idea of Gneiting *et al.* [6]).

B.2 Case study and Available Data

Owing to its already significant share of wind generation in the electricity mix as well as very ambitious objectives in the medium term, focus is given to the test case of Denmark. Denmark has set the goal to meet 50% of electricity demand with wind energy in 2025 [7], and this will clearly result in challenges related to the management of the grid. More precisely, the case study of this paper relates to western Denmark, including the Jutland peninsula and the island of Funen, which is connected to the UCTE (Union for the Co-ordination of Transmission of Electricity) system and stands for around 70% of the entire

wind power capacity installed in Denmark. Another reason for the choice of this test case is that operational developments and application of wind power forecasting systems started here in around 1994 [8], and it is common practice today to have forecasts of wind power production at different spatial resolutions and at a state-of-the-art level of accuracy.

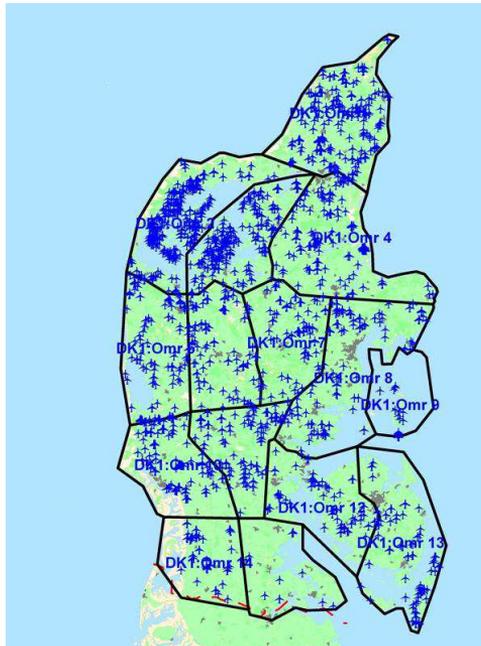


Figure B.1: Geographical locations of the groups of wind farms.

The data selected for this work comes from 15 groups of wind farms spread throughout the considered area. The chosen grouping corresponds to the actual resolution map used by the Danish TSO. For all the 15 groups, measurements of wind power production with an hourly resolution are available, along with wind power forecasts provided by the Wind Power Prediction Tool (WPPT). WPPT is a state-of-the-art forecasting system. Methods included in this forecasting system are described in [9, 10] (and references therein), while application results may be found in e.g. [11]. In the present case, forecasts of wind power generation are provided for each of the wind farms with a temporal resolution of one hour and up to a 48-hour lead time. Forecasts are generated every hour. Meteorological forecasts used as input are provided by the HIRLAM model of the Danish Meteorological Institute. Forecasts are available over a 40×42 -node grid (horizontal resolution is 3 km) covering Denmark and surroundings, including a large part of the North Sea. To provide weather forecasts for each group of wind farms, data available at the HIRLAM nodes is respectively sub-sampled

and interpolated (performed at DMI). Meteorological forecasts are delivered every six hours with an hourly temporal resolution up to 48 hours ahead. Wind forecasts are available at different vertical levels. Only wind forecasts at 10 meters a.g.l. (above ground level) have been considered in this work due to the fact that WPPT also uses this particular level as input. The period for which both measurements and predictions have been made available for this study is from the 1st of January, 2006 to the 24th of October, 2007.

Forecast errors are defined as the difference between power predictions and corresponding measurements, subsequently normalized by the installed wind power, following the framework described in [12]. Only one-hour-ahead forecast errors are considered.

B.3 Point forecasts

In this section focus is put on the point forecasting of the WPPT prediction errors. The goal is to build adequate multivariate models accounting for spatio-temporal inter-dependencies among the groups and to evaluate if the implementation of such models would result in a reduction of wind power prediction errors. Two types of models are considered. First, a vector autoregressive model allowing capture of a linear spatio-temporal inter-dependence structure, is fitted to data. This model also comprises a benchmark against which more complex model should be evaluated. For capturing the effect of meteorological forecasts on spatio-temporal characteristics, it is proposed in a second stage to upgrade the vector autoregressive model to a conditional parametric one, making the coefficients of the model smooth non-parametric functions of meteorological forecast data.

B.3.1 The VAR model

For capturing spatio-temporal patterns in wind power forecast errors, a Vector AutoRegressive (VAR) model (B.1) is fitted to the data:

$$\mathbf{w}_t = \sum_{i=1}^p \mathbf{A}_i \mathbf{w}_{t-i} + \boldsymbol{\epsilon}_t \quad (\text{B.1})$$

where \mathbf{w}_t is a vector of the dimension $[m \times 1]$ showing wind power forecast errors at m groups of wind farms ($m = 15$) obtained for time t , $t = 1 \dots N$. p indicates

the order of the model. In this work p is set to 1, as predictions made by similar models of higher orders did not show significant improvement in terms of the Root Mean Squared Error (RMSE) criterion. $\boldsymbol{\epsilon}_t$ term has a dimension $[m \times 1]$ and is assumed to be distributed multivariate with zero mean, \mathbf{A} is a coefficient matrix to be estimated from the data. The estimation of VAR model parameters can be straightforwardly performed with the recursive Least Squares (LS) estimation method. Details and discussion on this topic can be found in [13, 14, 15].

The results from this model corrected forecasts are only used in this work as a benchmark for evaluating the performance of more advanced models (Conditional Parametric): accounting also for the effects of meteorological forecasts on the spatio-temporal characteristics of the wind power forecast errors.

B.3.2 The CP-VAR model

Conditional parametric models comprise a class of models with a linear structure (like a Vector AutoRegressive model (VAR), for instance), but for which the coefficients are replaced by smooth functions of other variables, i.e.

$$\mathbf{w}_t = \sum_{i=1}^p \mathbf{A}_i(\mathbf{z}_t) \mathbf{w}_{t-i} + \boldsymbol{\epsilon}_t \quad (\text{B.2})$$

which translates to replacing the \mathbf{A}_i coefficients in the VAR model (B.1) with coefficient functions. The model can be fitted to data, with \mathbf{w}_t being a vector of the dimension $[m \times 1]$ showing wind power forecast errors at m groups of wind farms ($m = 15$) obtained for time t , $t = 1 \dots N$, \mathbf{z}_t is a vector $[1 \times l]$ representing a signal obtained at time t which conditions model coefficients \mathbf{A} . If assumed that \mathbf{A}_i depends on average forecasted wind direction for time t (wd_t), then $l = 1$ and $\mathbf{z}_t = wd_t$. If \mathbf{A}_i is conditioned on both average wind speed and direction, then $l = 2$ and $\mathbf{z}_t = [wd_t, ws_t]$, where ws_t denotes an average forecasted wind speed for m groups of wind farms. The results from CP-VAR models with respect to both wind speed and direction forecasts will not be presented and discussed in this work, as they did not show any improvement in terms of forecast RMSE compared to a model conditioned on the forecasted wind direction only. For similar reasons it was decided to present only the results for the CP-VAR model with $p = 1$. This model will be referred to as CP-VAR in the following. If using model (B.2) for a forecasting application, the one-step-ahead forecast at time t (denoted as $\hat{\mathbf{w}}_{t+1|t}$) will be given by

$$\hat{\mathbf{w}}_{t+1|t} = \sum_{i=0}^{p-1} \hat{\mathbf{A}}_{i,t}(\mathbf{z}_t) \mathbf{w}_{t-i} \quad (\text{B.3})$$

where $\hat{\mathbf{A}}_{i,t}$ is the estimate of \mathbf{A}_i evaluated at time t .

B.3.3 Estimation in the CP-VAR model

The estimation process combines a general framework used for estimation in univariate conditional parametric models (CP models) with the difference in the least squares (LS) algorithm applied: instead of a univariate LS here a multivariate recursive LS algorithm is used. Extensive information on estimation in CP-models can be found in [16], multivariate recursive LS is described in [17]. Below a summary of the estimation process is given.

For simplicity, the estimation process is described for a generic CP-VAR model, expressed as:

$$\mathbf{w}_t = \mathbf{A}(\mathbf{z}_t)\mathbf{x}_t + \boldsymbol{\epsilon}_t \quad (\text{B.4})$$

Estimation in (B.4) aims at estimating the functions $\mathbf{A}(\cdot)$ with the space spanned by the observations of \mathbf{z} . The functions are only estimated for discrete values of the argument \mathbf{z} . Below $\mathbf{z}_{(j)}$ denotes a single of these points and $\hat{\mathbf{A}}(j)$ denotes the estimate of coefficient functions when evaluated at $\mathbf{z}_{(j)}$. After the local coefficients $\mathbf{A}_{(j)}$ are estimated at a number of fitting points, the coefficient value for any value \mathbf{z}_t can be obtained by linear interpolation.

One solution to the estimation problem is to assume, that the coefficient function is locally constant at $\mathbf{z}_{(j)}$.

In this work online setting of the estimation process is used, i.e. when one aims at tracking the local coefficients by using a recursive estimation. This approach also allows for an exponential forgetting of old observations, which leads to the model being adaptive with respect to the long-term variations in the process characteristics. From here on it is considered that at time n a set of n past observations is available, and thus the dataset grows as time increases.

First, let us introduce the objective function to be minimized at each time n :

$$S_n(\mathbf{A}_{(j)}) = \sum_{t=1}^n \Lambda_{(j),n}(t) c_{(j),t} \rho(\mathbf{w}_t - \mathbf{A}_{(j)}\mathbf{x}_t) \quad (\text{B.5})$$

where

- $\Lambda_{(j),n}$ is the function that permits exponential forgetting of past observations, i.e.

$$\Lambda_{(j),n}(t) = \begin{cases} \lambda_{(j),n}^{\text{eff}} \Lambda_{(j),n-1}(t-1), & 1 \leq t \leq n-1 \\ 1 & , \quad i = n \end{cases} \quad (\text{B.6})$$

In the above definition, $\lambda_{(j),n-1}^{\text{eff}}$ is the effective forgetting factor for the fitting point $\mathbf{z}_{(j)}$ which makes it possible to account for the weighting in the formulation of (B.5). The effective forgetting factor ensures that old observations are downweighted only when new information is available. Following the definition given by Nielsen *et al.* [10], $\lambda_{(j),n}^{\text{eff}}$ is a function of $c_{(j),n}$, so that

$$\lambda_{(j),n}^{\text{eff}} = 1 - (1 - \lambda)c_{(j),n} \quad (\text{B.7})$$

where λ is the classical user-defined forgetting factor, $\lambda \in (0, 1)$. In this work $\lambda = 0.999$ is used as this is the value that empirically minimizes the RMSE of the model-based one-hour-ahead predictions.

- ρ is a quadratic criterion, i.e. such that $\rho(\boldsymbol{\epsilon}) = \boldsymbol{\epsilon}\boldsymbol{\epsilon}^T/2$.
- the weights $c_{(j),t}$ are assigned by a Kernel function of the following form

$$c_{(j),t} = T \left(\prod_{k=1}^l \frac{|z_{t-1,k} - z_{(j),k}|_k}{\hat{h}_{(j),k}(\alpha_k)} \right) \quad (\text{B.8})$$

In the above, $|\cdot|_k$ denotes the chosen distance on the k^{th} dimension of \mathbf{z} . For the CP-VAR models considered, one would for instance certainly choose a polar distance if considering \mathbf{z} as a wind direction informing on the global wind regime.

In (B.8), $\hat{h}_{(j),k}$ is the bandwidth for that particular fitting point $\mathbf{z}_{(j)}$ and for the k^{th} dimension of $\mathbf{z}_{(j)}$. In this work a bandwidth for a forecasted wind direction is set to 45 degrees for all the fitting points. This value was chosen empirically as the one minimizing the RMSE of the model-based predictions.

Finally in (B.8), T is defined as the tricube function, i.e.

$$\begin{aligned} T &: v \in \mathbb{R}^+ \rightarrow T(v) \in [0, 1], \\ T(v) &= \begin{cases} (1 - v^3)^3, & v \in [0, 1] \\ 0 & , \quad v > 1 \end{cases} \end{aligned} \quad (\text{B.9})$$

as introduced and discussed by e.g. Cleveland and Devlin in [18].

The local coefficients $\hat{\mathbf{A}}_{(j),n}$ at time n for model (B.5) are then given by

$$\hat{\mathbf{A}}_{(j),n} = \arg \min_{\mathbf{A}_{(j)}} S_n(\mathbf{A}_{(j)}) = \quad (\text{B.10})$$

$$= \arg \min_{\mathbf{A}_{(j)}} \sum_{t=1}^n \Lambda_{(j),n}(t) c_{(j),t} \rho(\mathbf{w}_t - \mathbf{A}_{(j)} \mathbf{x}_t) \quad (\text{B.11})$$

$$(\text{B.12})$$

The recursive formulation for an adaptive estimation of the local coefficients $\hat{\mathbf{A}}_{(j),n}$ leads to the following three-step updating procedure at time n :

$$\epsilon_{(j),n} = \mathbf{w}_n - \hat{\mathbf{A}}_{(j),n-1} \mathbf{x}_{n-1} \quad (\text{B.13})$$

$$\mathbf{R}_{(j),n} = \lambda_{(i),n}^{\text{eff}} \mathbf{R}_{(j),n-1} + c_{(j),n} \mathbf{x}_n \mathbf{x}_n^\top \quad (\text{B.14})$$

$$\hat{\mathbf{A}}_{(j),n}^\top = \hat{\mathbf{A}}_{(j),n-1}^\top + c_{(j),n} (\mathbf{R}_{(j)})^{-1} \mathbf{x}_n \epsilon_{(j),n}^\top \quad (\text{B.15})$$

where $\lambda_{(j),n}^{\text{eff}}$ is again the effective forgetting factor. One sees that when the weight $c_{(j),n}$ equals 0 (thus meaning that the local estimates should not be affected by the new information), then one has $\hat{\mathbf{A}}_{(j),n} = \hat{\mathbf{A}}_{(j),n-1}$ and $\mathbf{R}_{(j),n} = \mathbf{R}_{(j),n-1}$. This confirms the role of the effective forgetting factor, i.e. down-weight old observations, but only when new information is available.

For initializing the recursive process, the matrices $\mathbf{R}_{(j),0}$, $j = 1, \dots, m$, can be chosen as

$$\mathbf{R}_{(j),0} = \delta \mathbf{I}, \quad \forall j \quad (\text{B.16})$$

where δ is a small positive number and \mathbf{I} is an identity matrix of appropriate size. In parallel, the local coefficients $\hat{\mathbf{A}}_{(j),0}$ are initialized, i.e. as a matrix of zeros. Note that one may not want to apply (B.15) as long as $\mathbf{R}_{(j),n}$ is not invertible. This can simply be checked for the first time steps, the updating formula (B.15) being skipped as long as the condition number of $\mathbf{R}_{(j),n}$ is not judged good enough.

B.3.4 Assessment of the point forecasts

Since after installation of the models some time is needed for parameter values to settle, in this work it was decided to disregard the first 5000 data points (approximately 30% of the data) in the evaluation step for both point forecast and probabilistic forecast assessments. The objective of this section is to illustrate and analyse the ability of the presented models to capture the spatio-temporal characteristics of wind power forecast errors, as well as the effects of forecasted

wind direction on those characteristics. The forecasted errors are added to the original WPPT predictions to get the adjusted forecast based on spatio-temporal dependencies. The accuracy of such corrected forecasts is compared to that of original WPPT forecasts based on the RMSE criterion.

Comparison is made between:

1. The RMSE of the original WPPT forecast, i.e. estimate of the errors from the state-of-the-art wind power forecasting system, without accounting for the spatio-temporal characteristics.
2. The RMSE of the VAR model (B.1) based predictions, i.e. estimate of the errors resulting from the corrected forecast based on linear spatio-temporal patterns without considering the effects of meteorological conditions.
3. The RMSE of the CP-VAR model (B.2) based predictions, i.e. estimate of the errors resulting from the corrected forecast based on spatio-temporal patterns with consideration of the effects of the average forecasted wind direction.

Results are presented in table B.1. For each of the models, RMSE estimates are given for all the 15 groups together with the estimates of the error reduction in terms of the RMSE. The reduction in RMSE (denoted as Δ RMSE) is given as a percentage decrease in RMSE in comparison to the RMSE of the WPPT forecast for each group. It is seen that accounting for the spatio-temporal characteristics (VAR model) results in a reduction in RMSE for all the groups. The CP-VAR model outperforms VAR and this proves that the forecasted wind direction influences the spatio-temporal patterns and taking it into consideration in the model permits more accurate predictions.

Figure B.2 shows the distribution of Δ RMSE resulting from the CP-VAR model through the considered geographical area. One can note that the larger improvements correspond to the eastern part of the region. This is in line with the fact that in Denmark the prevailing wind direction is westerly, so the easterly located groups are usually situated "down-wind" from the rest of the region. Therefore the spatio-temporal models show better predictive performance on the eastern part of the region as the information propagates following the wind direction. An interesting point to mention is that for Group 9 the observed improvement in the RMSE (4.08%) is not as large as for the other surrounding zones. This could be explained by the fact that Group 9, in contrast to the rest of the groups, is situated off the mainland. Therefore it is very probable that the dynamics of Group 9 are different from the rest of the groups.

Group	WPPT	VAR		CP-VAR	
	RMSE [%]	RMSE [%]	Δ RMSE [%]	RMSE [%]	Δ RMSE [%]
1	3.32	3.11	6.40	3.08	7.32
2	2.98	2.88	3.13	2.81	5.62
3	3.39	2.99	11.77	2.87	15.22
4	3.29	2.83	13.98	2.76	16.13
5	3.15	3.06	2.65	3.01	4.39
6	3.27	2.92	10.82	2.83	13.31
7	3.53	3.01	14.69	2.92	17.09
8	2.93	2.47	15.45	2.39	18.46
9	3.34	3.22	3.46	3.20	4.08
10	3.58	3.45	3.60	3.39	5.31
11	3.29	2.83	14.18	2.72	17.26
12	3.21	2.75	14.38	2.66	17.22
13	2.97	2.63	11.63	2.57	13.41
14	3.77	3.64	3.50	3.58	4.94
15	3.49	3.11	10.76	3.01	13.67

Table B.1: Evaluation of the forecast performance of the various models in terms of the RMSE.

B.4 Probabilistic Forecasts

Focus in this section is on probabilistic forecasts. The main objective is to estimate the uncertainty associated with the previously presented point forecasts by providing the probability density function of the corresponding random variable. In order to build such probabilistic forecasts for the prediction errors, a parametric approach employing a truncated MultiVariate Normal distribution (MVN) is used. This work will not go into detail on the properties of MVN distribution and would rather refer readers to [19] for the detailed information. Briefly, the assumption is that:

$$\mathbf{w}_t \sim N_{\mathbf{a}_t}^{\mathbf{b}_t}(\hat{\mathbf{w}}_{t|t-1}, \Sigma(\mathbf{z}_t)) \quad (\text{B.17})$$

where $\hat{\mathbf{w}}_{t|t-1}$ denotes the mean of the distribution, which is assumed to be equal to the point forecast of the wind power forecast errors obtained from (B.3). b_t and a_t are vectors of the dimension $[m \times 1]$ denoting upper and lower truncation limits of the distribution. The need to truncate the distribution arises from the fact that standardized wind power predictions issued by WPPT for time t (denoted by \mathbf{pp}_t) lie between 0 and 1, since one cannot obtain negative power as well as a quantity larger than a nominal capacity. Therefore the errors of the

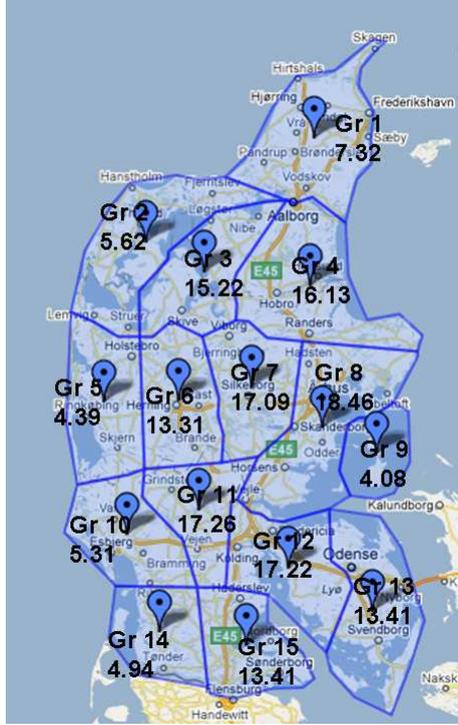


Figure B.2: Predictive performance of the CP-VAR model in terms of a percentage reduction in the RMSE (ΔRMSE) of the forecast errors. (Produced using <http://maps.google.dk/>)

power predictions are also bounded:

$$\mathbf{b}_t = 1 - \mathbf{pp}_t \mathbf{a}_t = -\mathbf{pp}_t \quad (\text{B.18})$$

$\Sigma(\mathbf{z}_t)$ is a covariance matrix of the distribution which is conditional on the external signal \mathbf{z}_t which in this case equals an average forecasted wind direction at time t . Estimation is performed in a recursive adaptive way, similar to the framework of estimation in CP-VAR models.

Analogously to the estimation in CP-VAR models, Σ is only estimated for discrete values of \mathbf{z} . Again, if $\mathbf{z}_{(j)}$ denotes a single one of these points and $\hat{\Sigma}_{(j),t}$ denotes the estimate of Σ at $\mathbf{z}_{(j)}$ evaluated at time t , then

$$\hat{\Sigma}_{(j),t} = \lambda_{(j),t}^{\text{eff}} \hat{\Sigma}_{(j),t-1} + (1 - \lambda_{(j),t}^{\text{eff}}) \boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t^T \quad (\text{B.19})$$

where $\epsilon_t = \mathbf{w}_t - \hat{\mathbf{w}}_{t|t-1}$ and $\lambda_{(j),t}^{\text{eff}}$ is defined as in (B.7). For initializing the recursive process $\hat{\Sigma}_{(j),0}$ has to be chosen. It can be a zero matrix.

B.4.1 Assessment of the probabilistic forecasts

A primary requirement for probabilistic forecasts relates to their calibration, which corresponds to the statistical consistency between the probabilistic forecasts and the observations [20]. In the univariate case, calibration can be verified using the Probability Integral Transform (PIT). In the ideal situation, i.e. if the observations were drawn from the predictive distribution, the PIT would have a uniform distribution on the unit interval $[0, 1]$ [20]. Therefore, in order to assess calibration for a univariate case, one can plot a PIT histogram and check for its uniformity. Assessing probabilistic forecasts of multivariate quantities is more complex. Some of the tools are presented in [21]. Since in this work 15-dimensional data having a truncated distribution with varying parameters is analysed, the implementation and evaluation of genuinely multivariate approaches presented in [21] become troublesome. Instead, as the first step, it was decided to check for the adequacy in the behaviour of the individual marginals of the estimated multivariate density. In what follows univariate PITs are applied to individual marginal distributions for each group and checked for uniformity. This also permits to scan for non-uniform directions. After checking PIT histograms for each of the 15 groups, it was observed that the results for all groups look very similar. Figure B.3 shows as an example PITs obtained for Groups 5, 8 and 9. These groups are chosen for the demonstration purposes due to the differences in their geographical positions. Group 5 is chosen as a representative of the western part of the region, Group 8 as a representative of the eastern part. As already stated before, Group 9 is situated off the mainland, and thus it might have some off-shore dynamics differing from the rest of the groups. One can see that the histograms shown are all close to uniform.

Figure B.4 depicts episodes with forecasts and measurements for Groups 5, 8 and 9 for the one-week period beginning on the 3rd of August, 2007 at 4 a.m. The demonstrated forecast intervals relate to the mean of the predictive distributions, therefore they are usually symmetric around the point forecasts. Exceptions from this rule are the cases when the mean is too close to the limits of truncation of the predictive density. In such cases the predictive intervals are correspondingly shifted to remain of the same nominal proportions. One can see that the width of the prediction intervals changes in time. Once the observed values do not vary too much, the probabilistic forecasts get sharper and the corresponding forecast intervals more narrow. When there is a larger dispersion in the observed WPPT errors, the corresponding uncertainty estimates react

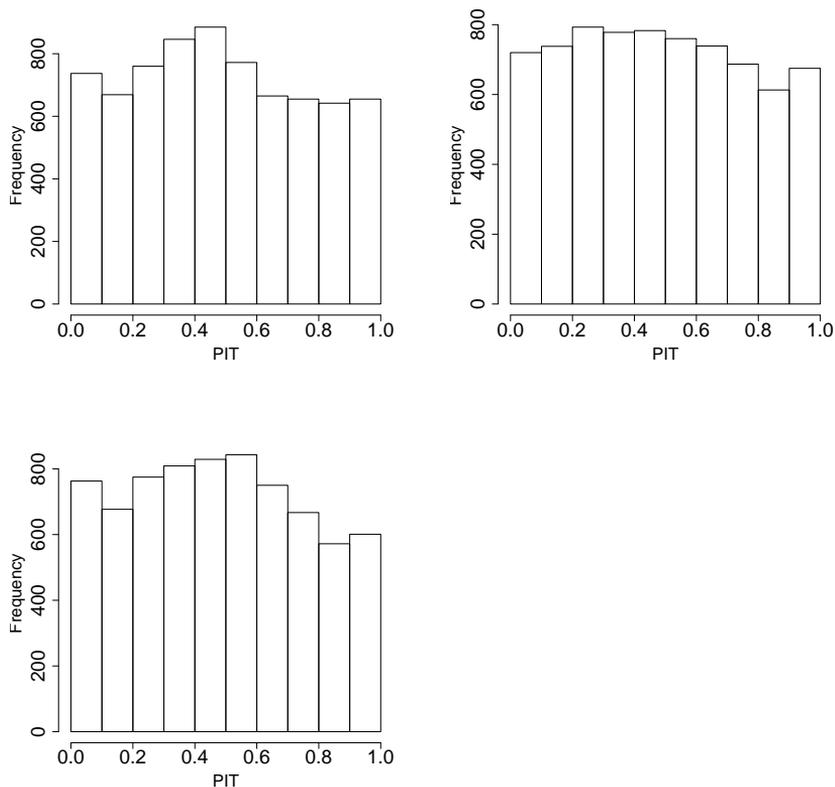


Figure B.3: PIT histograms for the reliability assessment of the univariate probabilistic forecasts for Groups 5 (left), 8 (center) and 9 (right)

adequately and the prediction intervals become wider.

The performed assessment indeed evidences that the proposed MVN predictive distribution is reliable if making probabilistic forecasts for each group individually, based on the corresponding univariate marginals of the estimated MVN. However, the performed evaluation is not sufficient to claim whether, if making a multi-dimensional forecast, a calibration with MVN probabilistic density would be achieved. This requires additional research and is our point of interest for the future work.

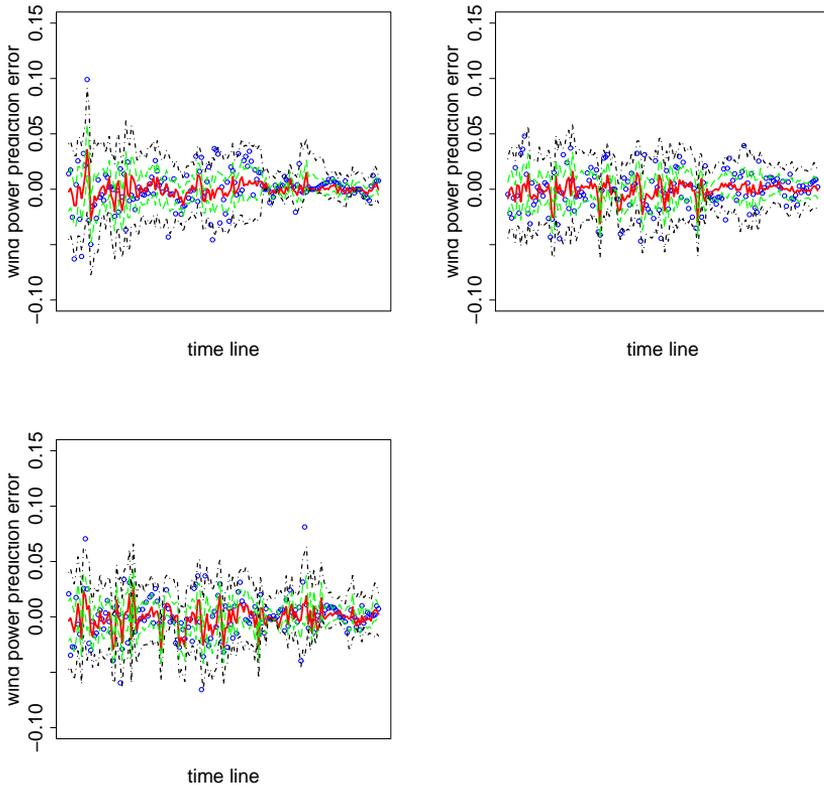


Figure B.4: Episodes with forecasts and measurements corresponding to a one-week period from 4 a.m. 2007-08-03 to 4 a.m. 2007-08-10 for Groups 5 (left), 8 (center) and 9 (right). Blue circles denote the observed WPPT errors, red solid lines show the corresponding point forecasts produced by the CP-VAR model. Green and black broken lines correspond to upper and lower limits of the 50% and 90% forecast intervals (based on the quantiles of marginal MVN distributions).

B.5 Conclusions

Due to the fact that state-of-the-art wind power prediction tools generate forecasts for individual sites only, without properly accounting for the information coming from the neighbouring territories, there is potential for improving the forecast quality by including the spatio-temporal aspects into the models. In

this paper a CP-VAR model is proposed for this purpose. The model coefficients being recursively estimated smooth functions of the forecasted average wind direction permit to consider the influence of meteorological data on the spatio-temporal inter-dependence among the sites and also to account for the long-term variations in the process characteristics. It is shown that correspondingly corrected forecasts when evaluated on the test case of western Denmark result in a reduction of prediction errors up to 18.46% in terms of RMSE. The presented adjusted forecasts are accompanied with the estimates of the associated uncertainty. Predictive densities are modelled as truncated multivariate normal distribution. The performed assessment evidences that the proposed method results in reliable univariate probabilistic forecasts for each individual group. However, additional research is needed before concluding if the multidimensional forecast is calibrated with the estimated multivariate density. This is one of the points of interest for the future work.

In this paper all the models are validated on the test case of western Denmark by considering one-hour ahead predictions. The proposed methodology could be extended to the case of further look ahead times (up to several hours ahead) if working on the same terrain as Denmark. Having in mind that western Denmark is the first to be touched by fronts coming from North-West, the use of on-line measurements from the United Kingdom or from the measurement devices in the North Sea, might lead to significant improvements in making predictions for longer time horizons. In order to make models valid for data coming from a region with a more complex wind climatology than Denmark, some adjustments would have to be done in the modelling approach due to the fact that it is not always possible to use average wind direction as a representative of the situation in the whole region. For such more complicated cases it may be needed to switch from conditional parametric models to varying-coefficient models. Since CP-VAR model structure and complexity are highly dependent on the number of sites considered, in the future it appears crucial to propose a modelling approach which would permit to easily include or exclude a new wind installation into the model without having to change the whole structure and to re-estimate the coefficients. A potential solution for this could be to employ a lattice approach.

B.6 Acknowledgement

The work presented has been partly supported by the European Commission under the SafeWind project (ENK7-CT2008-213740), which is hereby acknowledged. The authors would like to additionally thank Energinet.dk, the Transmission System Operator in Denmark, for providing the data used in the paper.

References B

- [1] M. S. Miranda and R. W. Dunn, “Spatially correlated wind speed modelling for generation adequacy studies in the UK,” in *Power Engineering Society General Meeting, 2007. IEEE*, pp. 1–6, IEEE, 2007.
- [2] U. Focken, M. Lange, K. Mönnich, H.-P. Waldl, H. G. Beyer, and A. Luig, “Short-term prediction of the aggregated power output of wind farms—a statistical analysis of the reduction of the prediction error by spatial smoothing effects,” *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 90, no. 3, pp. 231–246, 2002.
- [3] J. Tastu, P. Pinson, E. Kotwa, H. Madsen, and H. A. Nielsen, “Spatio-temporal analysis and modeling of short-term wind power forecast errors,” *Wind Energy*, vol. 14, no. 1, pp. 43–60, 2011.
- [4] T. Gneiting, “Editorial: probabilistic forecasting,” *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, vol. 171, no. 2, pp. 319–321, 2008.
- [5] P. Pinson, C. Chevallier, and G. N. Kariniotakis, “Trading wind generation from short-term probabilistic forecasts of wind power,” *Power Systems, IEEE Transactions on*, vol. 22, no. 3, pp. 1148–1156, 2007.
- [6] T. Gneiting, K. Larson, K. Westrick, M. G. Genton, and E. Aldrich, “Calibrated probabilistic forecasting at the stateline wind energy center: The regime-switching space–time method,” *Journal of the American Statistical Association*, vol. 101, no. 475, pp. 968–979, 2006.
- [7] E. E. Analyses, “50% Wind Power in Denmark in 2025-English Summary,” 2007.

-
- [8] H. Madsen, “Models and methods for wind power forecasting. Elsam/IMM, Denmark,” tech. rep., ISBN 87-87090-29-5, 1996.
- [9] T. S. Nielsen, H. A. Nielsen, and H. Madsen, “Prediction of wind power using time-varying coefficient functions,” in *Proceedings of World Congress on Automatic Control, Barcelona, Spain, 2002*.
- [10] H. A. Nielsen, T. S. Nielsen, A. K. Joensen, H. Madsen, and J. Holst, “Tracking time-varying-coefficient functions,” *International Journal of Adaptive Control and Signal Processing*, vol. 14, no. 8, pp. 813–828, 2000.
- [11] N. Cutler, M. Kay, K. Jacka, and T. S. Nielsen, “Detecting, categorizing and forecasting large ramps in wind farm power output using meteorological observations and WPPT,” *Wind Energy*, vol. 10, no. 5, pp. 453–470, 2007.
- [12] H. Madsen, P. Pinson, G. Kariniotakis, H. A. Nielsen, and T. S. Nielsen, “Standardizing the performance evaluation of shortterm wind power prediction models,” *Wind Engineering*, vol. 29, no. 6, pp. 475–489, 2005.
- [13] H. Madsen, *Time Series Analysis*. Chapman & Hall/CRC: London, 2007.
- [14] C. Chatfield, *The analysis of time series: An introduction, 6th edition*. CRC press, New York, 2003.
- [15] H. Madsen and J. Holst, “Modelling non-linear and non-stationary time series,” *Lecture Notes, Technical University of Denmark, Dpt. of Informatics and Mathematical Modeling, Kgs. Lyngby, Denmark, 2000*.
- [16] H. A. Nielsen, “An S-PLUS/ R library for locally weighted fitting of linear models,” *Technical report, 1997-22, 1992*.
- [17] B. Lindoff and J. Holst, “Bias and covariance of the recursive least squares estimator with exponential forgetting in vector autoregressions,” *Journal of Time Series Analysis*, vol. 17, no. 6, pp. 553–570, 1996.
- [18] W. S. Cleveland and S. J. Devlin, “Locally weighted regression: an approach to regression analysis by local fitting,” *Journal of the American Statistical Association*, vol. 83, no. 403, pp. 596–610, 1988.
- [19] W. C. Horrace, “Some results on the multivariate truncated normal distribution,” *Journal of Multivariate Analysis*, vol. 94, no. 1, pp. 209–221, 2005.
- [20] T. Gneiting, F. Balabdaoui, and A. E. Raftery, “Probabilistic forecasts, calibration and sharpness,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 69, no. 2, pp. 243–268, 2007.

-
- [21] T. Gneiting, L. I. Stanberry, E. P. Gritmit, L. Held, and N. A. Johnson, “Assessing probabilistic forecasts of multivariate quantities, with an application to ensemble predictions of surface winds,” *Test*, vol. 17, no. 2, pp. 211–235, 2008.

PAPER C

Probabilistic forecasts of wind power generation accounting for geographically dispersed information

Authors:

Julija Tastu, Pierre Pinson, Pierre-Julien Trombe, Henrik Madsen

Accepted for publication in :

IEEE TRANSACTIONS ON SMART GRID, 2013

Probabilistic forecasts of wind power generation accounting for geographically dispersed information

Julija Tastu¹, Pierre Pinson¹, Pierre-Julien Trombe¹, Henrik Madsen¹

Abstract

Forecasts of wind power generation in their probabilistic form are a necessary input to decision-making problems for reliable and economic power systems operations in a smart grid context. Thanks to the wealth of spatially distributed data, also of high temporal resolution, such forecasts may be optimized by accounting for spatio-temporal effects that are so far merely considered. The way these effects may be included in relevant models is described for the case of both parametric and non-parametric approaches to generating probabilistic forecasts. The resulting predictions are evaluated on the real-world test case of a large offshore wind farm in Denmark (Nysted, 165 MW), where a portfolio of 19 other wind farms is seen as a set of geographically distributed sensors, for lead times between 15 minutes and 8 hours. Forecast improvements are shown to mainly come from the spatio-temporal correction of the first order moments of predictive densities. The best performing approach, based on adaptive quantile regression, using spatially corrected point forecasts as input, consistently outperforms the state-of-the-art benchmark based on local information only, by 1.5%-4.6%, depending upon the lead time.

C.1 Introduction

One of the underlying challenges in implementing smart grid concepts is the efficient integration of renewable energy sources, especially wind energy [1]. Indeed, the stochastic nature of wind power, with its variability and limited predictability, induces difficulties in operating and managing power grids, particularly for

¹DTU Informatics, Technical University of Denmark, Richard Petersens Plads, bld. 305, DK-2800 Kgs. Lyngby, Denmark

balancing electricity consumption and production [2]. Today, the development of advanced wind power prediction systems is considered as one of the most cost-effective solutions for mitigating the impact of the uncertainty stemming from the integration of wind power into power grids. In a recent survey of grid operators' views on wind power integration, 94% of the respondents indicated that the integration of a significant amount of wind power will ultimately depend on the accuracy of wind power forecasts [3]. A history of the short term wind power forecasting and an overview of the state-of-the-art methodology are given in [4] and [5], respectively.

Today the main interest turns from point to probabilistic forecasting [6]. This is driven by the complexity of the related decision making tasks which calls for the forecast uncertainty quantification. For example, when trading wind power on the Danish electricity market, one deals with a non-symmetric penalty function as overproduction and underproduction are not penalized in the same way, when settled through the balancing market. Therefore, in order to bid optimally it is not sufficient to know the expected power generation only. One should also be informed about the possibilities of actual energy production to exceed or to be less than the expected value, hence allowing minimizing expected balancing costs [7]. Other applications of probabilistic forecasts to power grid operations include economic load dispatch and stochastic unit commitment [8, 9, 10], optimal operation of storage [11], reserve quantification [12] and assessment of operating costs [13].

More generally speaking, the benefits of probabilistic forecasts can be justified by the fact that for a large class of decision making problems the optimal solution is directly linked to a specific quantile rather than the expectation of the future outcome [14].

For continuous variables like wind power generation, probabilistic forecasts take the form of predictive density functions, fully describing that random variable for a set of lead times.

Up to now the number of studies on probabilistic wind power forecasting is relatively small compared with point forecasting. A part of the available studies focus on indirect wind power forecasts, i.e. when firstly wind speed predictive densities are obtained and secondly they are transformed to wind power forecasts. Most often idealized deterministic power curve is used for such transformation [15, 16]. However in practice power curves are stochastic and of a rather complex nature [17]. To account for it, stochastic power curve models can be built, as for example in [18]. Another possible approach (which is also considered in this work) is to construct predictive densities for wind power directly, without the intermediate step of modelling the uncertainty of the wind. Advantages of this approach are (i) no need to directly account for the com-

plexity of the stochastic power curve, (ii) owing to the geographical distribution of wind farms, the corresponding wind power data contains substantially more information than numerical weather predictions or 3 hourly data coming from the few available meteorological stations.

For constructing direct wind power density forecasts, one could follow two different families of approaches: a *parametric* or a *non-parametric* one. By the parametric approach we refer to a distribution-based methodology, which requires an assumption on the shape of the predictive densities. An example can be found in Ref. [19]. By the non-parametric one we refer to the distribution-free techniques, i.e. to the ones that are based on estimating the predictive densities directly from the data, without any constraints on the shape of the resulting distribution. As an example, adaptive resampling [20], time-adaptive quantile regression [21] and time-adaptive quantile-copula [22] techniques were recently described, their evaluation suggesting that they have a similar performance level [22, 23].

Our objective here is to introduce and evaluate a methodology allowing to issue probabilistic wind power forecasts optimally accounting for geographically dispersed information. The methods are tailored to targeting a single site of interest while using a number of neighbouring sites as explanatory variables. Focus is placed on time-adaptivity in order to reduce the computational load, and also to allow for smooth variations in the process dynamics, as induced by seasonal effects for instance [24]. The forecasting methodology is evaluated on the test case of a portfolio with 20 wind farms in Denmark, where the offshore wind farm at Nysted (165MW) is the target one, while the others are used as sensors. Predictions have a temporal resolution of 15 minutes with lead times up to 8 hours ahead.

The remainder of the paper is structured as follows. Section C.2 introduces the new challenges and opportunities related to wind power forecasting in a smart grid context. Section C.3 describes the two-step procedures for generating probabilistic forecasts with high temporal resolutions and optimally accounting for geographically distributed information, in both parametric and non-parametric frameworks. The data used for the empirical study is presented in Section C.4, while the results obtained are subsequently discussed in Section C.5. The paper ends with concluding remarks in Section C.6.

C.2 Wind power forecasting in a smart grid context

The evolution from traditional power systems operation to smart grid concepts has two major implications for wind power forecasting applications.

First, it is expected to enhance the way information collected by utilities is used in operational practice [1], with a transition towards higher frequencies for power generation scheduling, from hours to minutes, potentially reducing reserve requirements [25]. This translates to new challenges for the prediction of wind power generation at high temporal resolutions (in the order of few minutes), as recently addressed by [19] (and references therein). These challenges stem from *(i)* the concentration of wind turbines within relatively small geographical areas, hence magnifying power fluctuations, and *(ii)* the lack of intra-hour information in traditional forecasts based on Numerical Weather Prediction (NWP) models. This has for instance led to the consideration of new meteorological observations from remote sensing (e.g., satellite or weather radar images), available at high spatio-temporal resolutions and to be integrated into prediction systems [26].

Second, the evolution towards smart grids is likely to result in an increased volume of available information in space and in time, on both generation and consumption sides [1]. Owing to their geographical dispersion, wind farms comprise a dense network of atmospheric sensors capable of capturing valuable information on the spatio-temporal propagation of meteorological systems, and thereby on the propagation of wind power forecast errors [27, 28]. Historically, most state-of-the-art wind power prediction systems are optimized and run locally, for a single region, site or wind farm of interest, using on-site information only (e.g., meteorological forecasts, historical measurements of wind power) [24]. Yet, a new trend in wind power forecasting consists in exploiting spatio-temporal correlations in wind (power) data collected from neighbouring sites and integrating off-site information into prediction systems. The potential gains in terms of forecast accuracy have been underlined and quantified in a number of recent studies.

These studies can be divided into two groups. The first group considers cases for which dominant meteorological conditions are known a priori and the models are designed accordingly. To this group one can assign Ref. [29, 30] where the authors consider situations with one dominant wind direction, and Ref. [31, 15] designed for the situations with a strong channelling effect (considering two dominant directions). The second group does not rely on any dominant, known in advance weather patterns. Instead, the considered models are designed to capture the corresponding effects directly from the data. To this group one can

assign Ref.[32, 27, 33, 34, 35].

In view of this evolving context and of the limitations with existing approaches and input data in wind power forecasting, a path towards improving short-term predictability in wind power generation, and with high temporal resolution, may come from combining information from different sources, different locations and potentially different temporal resolutions. The methodology described in the following can be seen as part of this ongoing effort for making better use of available data and information in a smart grid context.

C.3 Methodology

The objective of the methodology introduced here is to generate probabilistic forecasts of wind power generation accounting for geographically dispersed information, which are to be of higher quality than forecasts produced based on local information only.

The main idea is based on answering the following question: if one has a snapshot of forecast errors currently (or previously) observed at the number of reference sites, then how does this information translate to the situation at the target location at time $t + k$. Thus, the approach is tailored to a situation with one target location and a number (possibly very small) of reference sites. We do not intend to describe a full space-time covariance structure for the error propagation. Instead, we focus on a rather pragmatic approach, which is focused on capturing as much of information available at present time t as possible and translating it to the situation at time $t + k$.

Depending on the layout of considered wind farms and specificities of the motion of weather systems over the considered territory, the optimal amount of information for explaining a situation at the target location at future time $t + k$ can be obtained from the errors observed at present time t or some past time $t - h$ at the reference sites. In other words, if the reference sites are rather remote and it takes longer than k for the information to propagate from the reference to the target point, then a snapshot of the past errors ($t - h$) should be used as explanatory variables. If on average it takes less than k for the information to propagate, one should use the corresponding snapshot taken at time t . Preliminary data analysis (for example, cross-correlation analysis of the forecast errors) can be used to get a hint on the average speed of error propagation over the territory [28]. Further in this work we focus on the case with $h = 0$, i.e we use the latest available information as explanatory variables. This is motivated by our will to ease the notation and the fact that this setup was optimal for the

considered test case.

The proposed procedure follows two main steps. First, the original single-valued predictions (also referred to as point forecasts) are corrected by integrating off-site information. Subsequently, these are upgraded to full probabilistic forecasts in the form of predictive densities, also allowing for off-site information to shape these predictive densities. Both parametric and non-parametric approaches are described: the former one is based on censored Gaussian distributions, while latter one relies on time-adaptive quantile regression.

C.3.1 Parametric predictive densities

Some initial considerations are to direct our choice for relevant predictive densities. Indeed when normalized by the nominal capacity of the turbine, farm or portfolio of interest, wind power generation is double-bounded between 0 and 1. Also, the non-linear and sigmoid-shaped conversion from wind speed to power results in conditional heteroskedasticity, i.e., in a non-constant variability of wind power generation [36]. Finally, when considering single wind farms rather than aggregated territories, wind power generation may equal 0 and 1 with a non-zero probability. This results in a non-negligible concentration of probability mass at the bounds. These aspects can be also seen in Fig. C.1: when the expected power is far from the natural generation bounds (Fig. C.1(b)), the conditional histogram resembles that of a Gaussian distribution (a characteristic bell-shape around the expected value can be seen). The closer to the bounds, the less dispersed distributions become and the higher the probability concentration at the closest bound can be noted. Predictive densities must be able to account for these specificities.

Various proposals for density functions were made in the literature, including the generalized Logit-Normal, Censored Normal and Beta distributions compared in [19], also considered in the present work. Since the best results were obtained with Censored Normal distributions, only these are introduced and discussed in the following.

C.3.1.1 Censored Normal distribution

Wind power generation as a Censored Normal (CN) variable follows an ordinary Normal law within the open unit interval. However, since the values outside $[0, 1]$ cannot be taken, the tails of the Normal distribution are cut and converted to probability masses at the corresponding bounds (0 and 1, respectively). For-

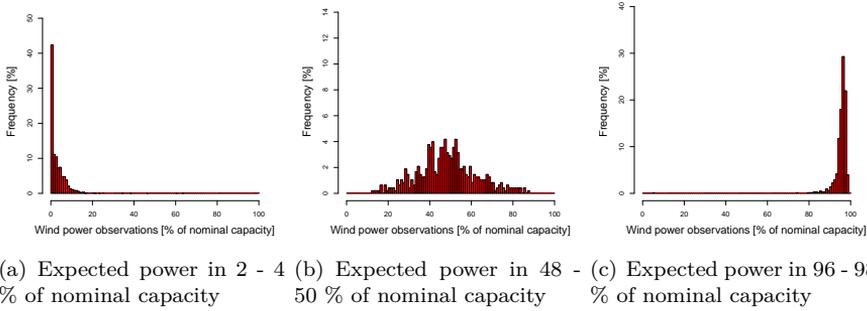


Figure C.1: Distribution of the observed power conditional to different levels of the expected power generation given by the state-of-the-art forecasting system, WPPT. 1 hour ahead forecasts are considered. Note that the range of frequencies on the y-axis varies from one plot to the next.

mally, a CN predictive density for wind power generation p_t at time t is defined as

$$f_p(y; \mu_t, \sigma_t^2) = w_t^0 \delta_0(y) + f^{(0,1)}(y; \mu_t, \sigma_t^2) + w_t^1 \delta_1(y), \quad (\text{C.1})$$

where $y \in [0, 1]$, δ_0 and δ_1 are Dirac functions at 0 and 1, respectively, while w_t^0 and w_t^1 are the weights representing probability mass concentration at the bounds. These are given by

$$w_t^0 = \Phi\left(\frac{-\mu_t}{\sqrt{\sigma_t^2}}\right); \quad w_t^1 = 1 - \Phi\left(\frac{1 - \mu_t}{\sqrt{\sigma_t^2}}\right), \quad (\text{C.2})$$

with $\Phi(\cdot)$ denoting the standard Normal distribution function. In parallel, $f^{(0,1)}(x; \mu_t, \sigma_t^2)$ follows a Gaussian density function within the open unit interval $(0, 1)$ and equals 0 outside this interval,

$$f^{(0,1)}(y; \mu_t, \sigma_t^2) = \begin{cases} (2\pi\sigma_t^2)^{-\frac{1}{2}} \exp\left\{-\frac{(y-\mu_t)^2}{2\sigma_t^2}\right\}, & y \in (0, 1) \\ 0, & \text{otherwise} \end{cases}. \quad (\text{C.3})$$

CN predictive densities as in Eq. (C.1) can be fully characterized by their location, μ , and scale, σ^2 , parameters which correspond to the mean and variance of the latent Gaussian process. These may be well approximated by the mean and variance of the censored process in practice, since the corresponding bias is found to be of a limited magnitude [19]. Both parameters are here predicted employing conditional parametric models. Therefore, a short presentation of a generic conditional parametric models is given below, followed by the specifics of the models for μ and σ^2 .

C.3.1.2 Generic conditional parametric models

A generic conditional parametric model reads

$$y_t = \boldsymbol{\theta}^\top(z_t)\mathbf{x}_t + \epsilon_t, \quad (\text{C.4})$$

where y_t is the value for the response variable at time t , $\mathbf{x}_t = [x_{1,t}, x_{2,t}, \dots, x_{l,t}]^\top$ and z_t are two groups of explanatory variables. $\boldsymbol{\theta}(\cdot) = [\theta_1(\cdot), \theta_2(\cdot), \dots, \theta_l(\cdot)]^\top$ is a vector of coefficient functions to be estimated and ϵ_t is a noise term. The estimation of $\boldsymbol{\theta}(\cdot)$ can be performed in an adaptive recursive manner as presented in Ref. [37], to which the reader is referred to for more details. Adaptivity in parameter estimation reduces computational costs significantly, hence comprising an essential element for operational implementation.

Briefly, the estimation of $\boldsymbol{\theta}(\cdot)$ is carried out in a semi-parametric way, i.e., without imposing any particular shape for the coefficient functions. The only assumption is that these are smooth enough to be locally approximated by constants (or polynomials in a more general setting). The estimation problem then boils down to estimating those local constants (polynomials) at a number m of fitting points $z_{(j)}, j = 1, \dots, m$. This is done by estimating linear models at each of these fitting points, hence yielding local estimates of $\boldsymbol{\theta}(z_{(j)})$. After the local coefficients are estimated, the values of coefficient functions at any given point z_t can be obtained by interpolation techniques. In this work we considered linear interpolation. From our experience if a sufficient number of fitting points is considered, linear interpolation is sufficient and the increase of model complexity by considering splines is unnecessary.

The number of fitting points, the forgetting factor and the bandwidth can be chosen empirically as the values optimizing the performance of the resulting model.

C.3.1.3 Models for the location parameter

The location parameter of CN predictive densities is given by wind power point forecasts $\hat{p}_{t|t-k}$ issued at time $t - k$ for time t , k being the lead time,

$$\hat{\mu}_t = \hat{p}_{t|t-k}. \quad (\text{C.5})$$

Two alternatives for obtaining $\hat{p}_{t|t-k}$ are considered, which are:

1. *Local* forecasts, $\hat{p}_{t|t-k} = \tilde{p}_{t|t-k}$, are obtained using a traditional wind power forecasting tool. Thus, they are optimized with respect to local

information, only, and do not account for information available at neighbouring sites;

2. *Spatio-temporal* forecasts, $\hat{p}_{t|t-k} = \check{p}_{t|t-k}$, are obtained by adjusting the local forecasts based on the geographically distributed information from other wind farms. This is here carried out with the method of Ref. [27], using a conditional parametric model for tracking spatio-temporal dependencies, as rapidly described hereafter.

We suppose that the local forecast error made at time t at the target location, $\xi_t = p_t - \tilde{p}_{t|t-k}$, depends on the errors previously recorded at a set I of neighbouring sites, $\xi_{t-k}^{(i)}$, $i \in I$. This dependency is assumed to be governed by the forecasted wind direction as

$$\xi_t = \boldsymbol{\theta}^\top (\hat{w}_{t|t-k}) \mathbf{x}_t + \epsilon_t, \quad (\text{C.6})$$

with $\hat{w}_{t|t-k}$ denoting the wind direction predicted at time $t-k$ for time t . It is in practice given by the global average of the wind field forecast over the considered territory. In parallel, \mathbf{x}_t is a column vector of the lagged local forecast errors at the set of neighbouring locations, $\xi_{t-k}^{(i)}$, $i \in I$.

Estimation of coefficient functions $\boldsymbol{\theta}$ in model (C.6) is performed as for the generic model (C.4). We have used 9 fitting points in the estimation procedure: from 0° to 320° with increments of 40° . Once the estimates of $\boldsymbol{\theta}$ are obtained, one can then correct the local wind power forecasts $\tilde{p}_{t|t-k}$ with

$$\check{p}_{t|t-k} = \tilde{p}_{t|t-k} + \hat{\boldsymbol{\theta}}_{t-k}^\top (\hat{w}_{t|t-k}) \mathbf{x}_t. \quad (\text{C.7})$$

C.3.1.4 Models for the scale parameter

The scale parameter σ_t^2 is approximated by the conditional variance of wind power generation. It is estimated by modelling squared residuals ε_t^2 :

$$\varepsilon_t^2 = (p_t - \hat{p}_{t|t-k})^2, \quad (\text{C.8})$$

where, as explained in the above, $\hat{p}_{t|t-k} = \tilde{p}_{t|t-k}$ or $\hat{p}_{t|t-k} = \check{p}_{t|t-k}$, for the case of local and spatio-temporal forecasts, respectively. The variance is consequently given by the expectation of these squared residuals, for which relevant models are to be proposed.

The volatility of wind power generation is not constant in time, owing to evolving dynamics in the wind itself, but also owing to the power curve that amplifies or

dampens wind fluctuations in a nonlinear manner. The former aspect could for instance be accounted for with regime-switching models [33]. This was not done here since they were not found to improve the skill of the resulting probabilistic forecasts. In parallel, the effect of the power curve can be accommodated by letting model parameters vary with the level of expected power generation, translating to conditioning wind power volatility on the slope of the power curve. For that purpose, the so-called Conditional Parametric ARCH (CP-ARCH) model is introduced, as well as its CP-ARCHX extension when adding geographically distributed information. An extensive comparison of these various approaches and models (ARCH, GARCH and regime-switching models) can be found in Ref. [38].

CP-ARCH – using local information only The CP-ARCH model can be formulated as

$$\begin{aligned}\varepsilon_t &= \sigma_t r_t, \\ \sigma_t^2 &= \alpha_0(\hat{p}_{t|t-k}) + \alpha_1(\hat{p}_{t|t-k})\varepsilon_{t-k}^2,\end{aligned}\tag{C.9}$$

where both point forecasts $\hat{p}_{t|t-k}$ and squared residuals ε_{t-k}^2 can again relate to local or spatio-temporal forecasts, r_t is a noise term and α_0 , α_1 are coefficient functions to be estimated. This model somewhat states that the conditional variance of wind power generation σ_t^2 , at the target location and at time t , is a function of the previously observed forecast errors at that location, ε_{t-k}^2 , only. The model coefficients α_0 and α_1 are made a function of the expected power generation, as given by the point forecasts. This is where the impact of the non-linear shape of the underlying power curve is accounted for.

The model of Eq. (C.9) is essentially a conditional parametric AR model for ε_t^2 , which can be rewritten as

$$\varepsilon_t^2 = \alpha_0(\hat{p}_{t|t-k}) + \alpha_1(\hat{p}_{t|t-k})\varepsilon_{t-k}^2 + r_t.\tag{C.10}$$

Consequently, the estimation of α_0 and α_1 is performed as for the generic conditional parametric model (C.4), more precisely by setting $y_t = \varepsilon_t^2$, $\boldsymbol{\theta} = [\alpha_0 \ \alpha_1]^\top$, $\mathbf{x}_t = [1 \ \varepsilon_{t-k}^2]^\top$ and $z_t = \hat{p}_{t|t-k}$. We have used 20 fitting points in the estimation procedure: from 0% to 100% quantile of wind power generation with increments of 5% (excluding median). The initial selection of fitting points was done arbitrary. Further attempts to increase the number of discretization points did not result in model improvements. Thus, provided that the computational cost was acceptable (see Section C.5.2), we used the initial setup.

A forecast $\hat{\sigma}_t^2$ of the scale parameter σ_t^2 , issued at time $t-k$ for time t , is finally given by

$$\hat{\sigma}_t^2 = \hat{\alpha}_{0,t-k}(\hat{p}_{t|t-k}) + \hat{\alpha}_{1,t-k}(\hat{p}_{t|t-k})\varepsilon_{t-k}^2.\tag{C.11}$$

CP-ARCHX – accounting for geographically distributed information

Additional spatio-temporal effects can be included into the CP-ARCH model by extending it to a CP-ARCHX one, yielding

$$\begin{aligned}\varepsilon_t &= \sigma_t r_t, \\ \sigma_t^2 &= \alpha_0(\hat{p}_{t|t-k}) + \alpha_1(\hat{p}_{t|t-k})\varepsilon_{t-k}^2 + \gamma(\hat{p}_{t|t-k})\xi_{t-k}^{(j)2},\end{aligned}\quad (\text{C.12})$$

where, in addition to the variables and parameters of the CP-ARCH model (C.9), $\xi_{t-k}^{(j)2}$ denotes squared errors from the local point forecasts at the j^{th} neighbouring site. The estimation of α_0 , α_1 and γ is similar to the case of generic conditional parametric models, by setting $y_t = \varepsilon_t^2$, $\boldsymbol{\theta} = [\alpha_0 \ \alpha_1 \ \gamma]^\top$, $\mathbf{x}_t = [1 \ \varepsilon_{t-k}^2 \ \xi_{t-k}^{(j)2}]^\top$ and $z_t = \hat{p}_{t|t-k}$.

Finally, a forecast for the scale parameter of CN predictive densities, issued at time $t - k$ for time t , is given by

$$\begin{aligned}\hat{\sigma}_t^2 &= \hat{\alpha}_{0,t-k}(\hat{p}_{t|t-k}) + \hat{\alpha}_{1,t-k}(\hat{p}_{t|t-k})\varepsilon_{t-k}^2 \\ &\quad + \hat{\gamma}_{t-k}(\hat{p}_{t|t-k})\xi_{t-k}^{(j)2}.\end{aligned}\quad (\text{C.13})$$

C.3.2 Non-parametric predictive densities

A non-parametric approach to issuing predictive densities of wind power generation does not rely on any assumption of a known distribution for the data. Instead, it suggests predicting a set of m quantiles $q_t^{(\tau)}$ and reconstructing full cumulative distribution function based on interpolation techniques. This gives more flexibility, though at a cost, since requiring to setup and estimate m models, resulting in a larger number of parameters, while still only partly describing densities. The quantiles defining the predictive densities are quantile forecasts for pre-defined nominal proportions, here obtained in a quantile regression framework.

C.3.2.1 Generic quantile regression models

The most basic form of a quantile regression model for a response y_t and a nominal proportion τ , as introduced in Ref. [39], is

$$q_t^{(\tau)} = \mathcal{F}_{y_t}^{-1}(\tau|\mathbf{x}_t) = \boldsymbol{\beta}^\top \mathbf{x}_t + r_t, \quad (\text{C.14})$$

where $\mathcal{F}_{y_t}^{-1}$ is the inverse distribution function of y_t , \mathbf{x}_t is a column vector of explanatory variables, $\boldsymbol{\beta}$ is a vector of parameters to be estimated and r_t is a

noise term. Given a set of N observations on which the model is to be fitted, estimates of β , $\hat{\beta}$ are obtained by solving the following linear programming problem:

$$\hat{\beta} = \operatorname{argmin}_{\beta} \sum_{t=1}^N (y_t - \beta^\top \mathbf{x}_t) (\tau - \mathbf{1}(y_t < \beta^\top \mathbf{x}_t)), \quad (\text{C.15})$$

where $\mathbf{1}(\cdot)$ is an indicator function, equal to 1 if the condition between brackets realizes, and to 0 otherwise.

The optimization problem (C.15) can be solved in a time-adaptive fashion by applying the method described in Ref. [21]. Briefly, it consists in updating the dataset used for estimation in a sensible way each time new data points become available, so as to minimize computational costs while allowing for smooth time variations in the model parameters and still covering the whole range of variations for the various explanatory variables.

Finally based on the parameters for the quantile regression model estimated based on past observations, a forecast for $q_t^{(\tau)}$, issued at time $t - k$, can be obtained with

$$\hat{q}_{t|t-k}^{(\tau)} = \hat{\beta}^\top \mathbf{x}_t \quad (\text{C.16})$$

By having a bank of m quantile regression models with, say, $\tau = j/(m + 1)$, $j = 1, \dots, m$, and then issuing quantile forecasts for these various nominal proportions, full predictive distributions are constructed.

C.3.2.2 Quantile regression models for wind power generation

Building non-parametric predictive densities for wind power generation by using quantile regression is performed in two steps. First, wind power point predictions, $\hat{p}_{t|t-k}$ are used to determine the mean of the corresponding predictive distributions (see Section C.3.1.3). Second, uncertainty around the mean is shaped, thus upgrading point forecasts to full predictive densities. This is done by estimating a conditional distribution of the point forecast errors at the target location (ε_t) and adding it to the estimate of the expected power generation, i.e.,

$$\mathcal{F}_{p_t} = \hat{p}_{t|t-k} + \mathcal{F}_{\varepsilon_t}, \quad (\text{C.17})$$

where \mathcal{F}_{p_t} and $\mathcal{F}_{\varepsilon_t}$ stand for the cumulative density functions of p_t and ε_t , respectively.

When defining quantile models for the distribution of the forecast errors, it is essential to account for some of the important characteristics of the process. Here namely, the uncertainty is known to be shaped by the power curve [36].

This dependence is known to be non-linear and thus, the quantile models we study here are given by

$$q_t^{(\tau)} = \mathcal{F}_{\varepsilon_t}^{-1}(\tau|\hat{p}_{t|t-k}) = g(\hat{p}_{t|t-k}) + r_t. \quad (\text{C.18})$$

No particular shape is imposed for the function g . Instead, we estimate it in a non-parametric way, using a spline representation. In other words, it is assumed that g can be viewed as a piecewise cubic function of $\hat{p}_{t|t-k}$ and, thus, can be expressed as a linear combination of the known basis functions b_j , resulting in the following quantile models

$$q_t^{(\tau)} = \beta_0 + \sum_{j=1}^{K-1} b_j(\hat{p}_{t|t-k})\beta_j + r_t, \quad (\text{C.19})$$

where b_j are natural cubic B-spline basis functions, K is the number of knots used for the spline construction and $\beta_i(\tau)$ are coefficients to be estimated. Such spline representation of g permits to use the estimation techniques valid for linear models such as that in Eq. (C.15), by setting

$$y_t = \varepsilon_t, \mathbf{x}_t = [1 \ b_1(\hat{p}_{t|t-k}) \ \dots \ b_{K-1}(\hat{p}_{t|t-k})]^\top \text{ and } \boldsymbol{\beta} = [\beta_0 \ \beta_1 \ \dots \ \beta_{K-1}]^\top. \text{ Model (C.19) is referred to as QR in the following.}$$

In that setup, additional spatio-temporal effects captured by point forecast errors previously recorded at the j^{th} sensor location may be accounted for. This translates to having $\xi_{t-k}^{(j)}$ as a supplementary explanatory variable in the QR model, as a linear or nonlinear term (represented by splines). Since the spline-based representation did not result in any improvement, we focus on the simpler linear case. The resulting model is denoted by QR-X and is written as

$$\mathcal{F}_{\varepsilon_t}^{-1}(\tau|\hat{p}_{t|t-k}) = g(\hat{p}_{t|t-k}) + \gamma\xi_{t-k}^{(j)} + r_t. \quad (\text{C.20})$$

Estimation is similar to the case of model (C.15), by setting $y_t = \varepsilon_t$, $\mathbf{x}_t = [1 \ b_1(\hat{p}_{t|t-k}) \ \dots \ b_{K-1}(\hat{p}_{t|t-k}) \ \xi_{t-k}^{(j)}]^\top$ and $\boldsymbol{\beta} = [\beta_0 \ \beta_1 \ \dots \ \beta_{K-1} \ \gamma]^\top$.

Quantile regression is used to provide 18 quantile forecasts with nominal proportions going from 5% to 95% by 5% increments, except for the median. 0% and 100% quantiles are set to 0 and 1, respectively. The setup corresponds to the one used in [23]. Linear interpolation is used to reconstruct full distribution functions from the set of quantiles. From our experience if a sufficient number of quantiles is considered, the linear interpolation is sufficient and the increase of complexity by considering splines is unnecessary. Empirical study could be performed in order to check whether an increase in the number of fitting points improves the performance of the models.

C.4 Data

The data used in this study were provided for 20 wind farms located in Denmark. All wind farms are owned and operated by the same power company. The respective locations of these wind farms are shown in Fig. C.2. For each wind farm, the following information is available:

- Wind power measurements at a temporal resolution of 15 minutes. They are normalized by the respective nominal capacities P_n of the various wind farms;
- Point forecasts of wind power generation, with lead times from 0 to 48 hours, and temporal resolution of 15 minutes. These predictions were generated with the Wind Power Prediction Tool (WPPT) [24], which is one of the state-of-the-art prediction models for the short-term wind power forecasting as discussed in Ref. [5];
- Meteorological forecasts of wind speed and wind direction at 10 meters above ground level, with lead times from 1 to 48 hours, and temporal resolution of 1 hour. These forecasts were generated by the HIRLAM model operated by the Danish Meteorological Institute (DMI) [40].

The data covers a period from May 1, 2008 to December 31, 2009. A first part of the data from May 1, 2008 to December 31, 2008 was used as a burn-in period in order to allow time-adaptive parameters not to be influenced anymore by their initial values. Forecast evaluation was carried out over the remainder of the dataset, from January 1, 2009 to December 31, 2009. Due to the large number of missing values, the effective evaluation period was eventually consisting of approximately 8.5 months (more than 25.000 forecast series for each of the lead times considered).

The Nysted wind farm was chosen as the target wind farm in this study. Nysted is located offshore on the Rødsand sand bank, near Lolland, Denmark, and is the southernmost of all wind farms shown in Fig. C.2. There are two main reasons behind this choice. Firstly, Nysted was the largest wind farm in Eastern Denmark until 2010, with a rated capacity of 165MW, and therefore was one of the main contributors to the aggregated amplitude of wind power fluctuations in that region. It also accounted for about 36% of the installed capacity owned by the company operating it. And secondly, Nysted has an appealing location with many wind farms located “upwind” in view of the prevalence of westerly flows over Denmark [41]. Indeed, improvements in forecast accuracy resulting from the use of off-site information are expected to be larger for wind farms located “downwind”, as shown by Ref. [27] for instance.

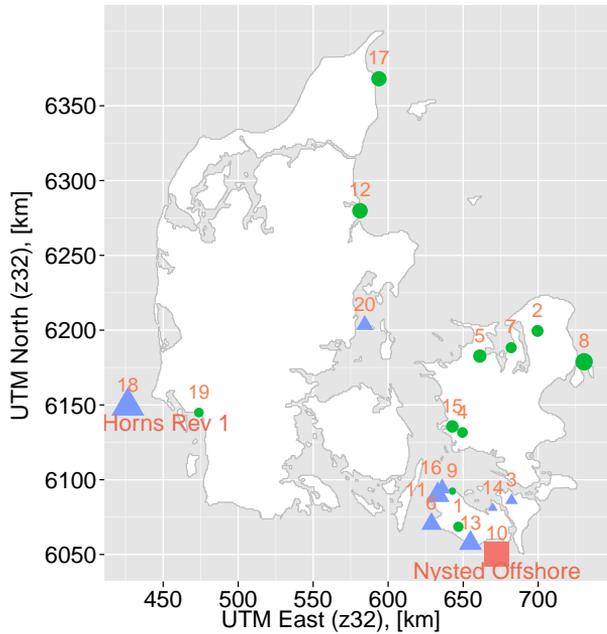


Figure C.2: Map of the 20 wind farms included in the analysis. The Nysted wind farm is marked as number 10 and with a square. Information from the wind farms with triangles contributed to improve the predictability of wind power generation at Nysted. Information from the wind farms with bullet points was found not to improve the predictability of wind power generation at Nysted. The size of the points is proportional to the rated capacity of the wind farms, on a logarithmic scale.

Out of the 19 nearby wind farms, only 8 wind farms (number 3, 6, 11, 13, 14, 16, 18 and 20) were used as explanatory variables. The selection was performed empirically, based on the stepwise selection.

In Ref. [28] it was found that correlations in wind power forecast errors between two wind farms were very small for distances larger than 50 km, over the Western Denmark area. Our findings are consistent with these results since most wind farms that contribute to improve wind power predictability at Nysted are located within such range (see Fig. C.2). More surprisingly, despite the large distances separating Nysted from wind farms 18 and 20 (approximately 265 km and 176 km, respectively), integrating information from these last two wind farms also led to substantial gains in forecast accuracy. Both wind farms 18 and 20 are located out in the open sea as Nysted, while the remaining wind farms in this

analysis are located onshore or near-shore. This difference with the results in Ref. [28] could be explained by a higher spatial persistence and homogeneity of wind field dynamics over waters than over lands, where the terrain roughness is known to be a very influential factor.

C.5 Empirical results

C.5.1 Model notation

The following notations are used for the model names:

- CN:CP/CP-ARCH refers to CN predictive densities with μ given by the spatio-temporal point forecasts (C.7) and σ^2 estimated using the CP-ARCH model (C.9);
- CN:CP/CP-ARCHX refers to CN predictive densities with μ given by the spatio-temporal point forecasts (C.7) and σ^2 estimated using the CP-ARCHX model (C.12);
- WPPT/QR stands for the non-parametric predictive densities based on the time-adaptive quantile regression (QR) (C.19). Local forecasts (WPPT) are used as input;
- CP/QR stands for the non-parametric predictive densities based on the time-adaptive quantile regression (QR) (C.19). Spatio-temporal forecasts given by (C.7) are used as input;
- CP/QR-X stands for non-parametric predictive densities with additional consideration of the spatio-temporal effects in the uncertainty modelling step as in (C.20). Spatio-temporal forecasts given by (C.7) are used as input;

C.5.2 Computational details

The parameters for the parametric densities were updated every 15 min. A single update step took less than a second of computing time. The quantile regression models were updated daily. A single update step took approximately 3 seconds when evaluating 20 different quantile models. The computations were performed on a laptop, having a processor Intel i 7-2620M CPU 2.70 GHz and the installed RAM of 8 GB

C.5.3 Overall evaluation

The evaluation and comparison of probabilistic forecasting approaches follows the guidelines, scores and diagnostic tools described in Ref. [42]. The lead score is the Continuous Ranked Probability Score (CRPS), which is a proper score for density forecasts. This score is negatively oriented: the smaller it is, the better the forecasts are.

Adaptive quantile regression with original WPPT point forecasts as input (denoted by WPPT/QR) is considered the base benchmark. Other predictive densities are compared to the benchmark approach and the relative improvements in CRPS (skill scores) are calculated. The summary of the results is given in Fig. C.3.

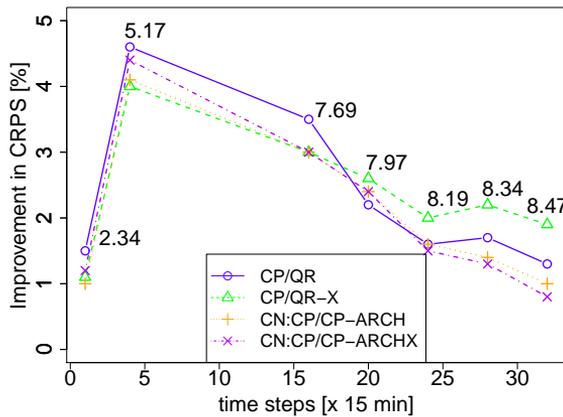


Figure C.3: Evaluation of predictive densities in terms of relative CRPS improvement, as a function of lead time. Point labels indicate the CRPS values [% of nominal capacity] for the benchmark model WPPT/QR.

The models have similar levels of CRPS improvement for all the lead times considered. This was also confirmed by using the Diebold-Mariano test statistics t_n [43], which may be more generally used to test for equal performance of probabilistic forecasts. t_n is asymptotically standard Normal under the null hypothesis of vanishing score differentials. This hints at the fact that forecast improvements brought in by the proposed methodologies mainly come from the space-time correction of the point forecasts defining the conditional mean of

Table C.1: Testing for equal forecast performance with the Diebold-Mariano test statistics t_n . p is that the probability that the corresponding value of t_n is achieved with the null hypothesis being true. ‘*’ symbols mark statistically significant differences in CRPS.

Horizon		15 min	1 h	4 h	5 h	6 h	7 h	8 h
CP/QR vs WPPT/QR	t_n	-9.29	-12.32	-5.63	-3.43	-2.34	-2.60	-3.19
	p	0.00*	0.00*	0.00*	0.00*	0.02*	0.01*	0.00*
CN:CP/CP-ARCH vs WPPT/QR	t_n	-6.06	-10.45	-3.22	-2.28	-1.34	-1.07	-0.65
	p	0.00*	0.00*	0.00*	0.02*	0.18	0.28	0.51
CP/QR vs CN:CP/CP-ARCH	t_n	-1.60	-0.90	-0.41	-0.25	-0.00	-0.39	-1.08
	p	0.11	0.37	0.68	0.80	1.00	0.70	0.28

predictive densities. Further consideration of space-time dynamics in the modelling of uncertainty around the mean does not seem to bring additional benefit. This is confirmed by the fact that CP/QRX does not outperform CP/QR, and similarly CN:CP/CP-ARCHX does not outperform CN:CP/CP-ARCH.

Consequently, we further focus on the evaluation of the CP/QR and CN:CP/CP-ARCH predictive densities, since comprising the best performing non-parametric and parametric probabilistic forecasts, respectively. Both types of probabilistic forecasts outperform the benchmark approach given by WPPT/QR (see Fig. C.3). The statistical significance of those improvements has been verified using the Diebold-Mariano test statistics t_n . The corresponding results are given in Table C.1.

The non-parametric densities accounting for the space-time dynamics (CP/QR) show statistically significant improvements for lead time up to 8 hours ahead. Such an observation is consistent with the spatio-temporal scales for the inertia of weather systems passing over Denmark. The improvements for the parametric alternative based on CN predictive densities can be considered statistically significant up to 5 hours ahead. At the same time, it is not possible to reject the null hypothesis such that CP/QR and CN:CP/CP-ARCH perform similarly for all lead times up to 8 hours ahead. From the 5-hour lead time, the benefits from space-time considerations start fading away. Mainly, this is caused by the fact that the benchmark approach is chosen non-parametric. Thus, CP/QR and WPPT/QR are of the same model family, while CN:CP/CP-ARCH differs from WPPT/QR in its nature, since being parametric. This results in higher variance of score differentials once comparing CN:CP/CP-ARCH and WPPT/QR, and subsequently lower t_n values.

Another interesting point to mention is that the peak in the improvements is observed for the prediction horizons of 1 hour ahead. This is in line with the

layout of the considered wind farms. From the map in Fig. C.2 one can see that almost all the considered reference sites are within 50 km from Nysted. According to [28], an average speed of error propagation over Denmark is 30-50 km/h (depending on the prevailing wind direction). This result is consistent with the fact that the peak of cross-correlations between Nysted and almost all the reference sites comes at lags of approximately 1 hour ahead and correspondingly results in the highest improvements.

C.5.4 Conditional evaluation

Emphasis is then placed on the situation-dependent performance of probabilistic forecasts, through a conditional forecast evaluation exercise. Results are shown and discussed for 1-hour ahead forecasts only, since these are qualitatively similar for the other lead times from 15 min to 8 hours ahead.

Firstly, since predictive densities evolve with the level of expected power generation, the overall skill of these densities is assessed based on the CRPS score, as a function of the point forecast values. The corresponding results are depicted in Fig. C.4.

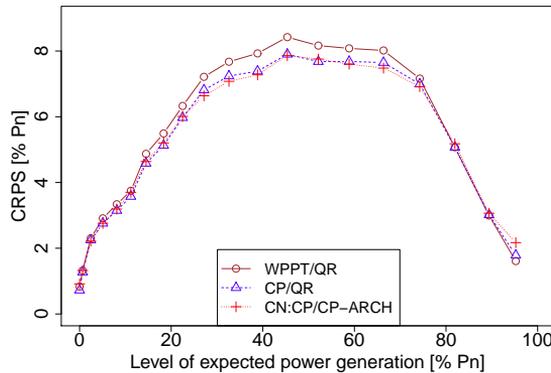


Figure C.4: CRPS conditional to the expected wind power generation (given by point forecasts) for the 1-hour lead time, with WPPT/QR considered as the benchmark. Power levels are given by 20 equally populated classes deduced from the distribution of point forecasts.

The CRPS for all the predictive densities considered increases for expected power levels in the medium range. This can be explained by the fact that

higher uncertainties in the wind power generation are faced in periods with medium power generation, i.e., in the steep slope part of the power curve. The performances of the parametric and the non-parametric densities are similar in this predicted power range. Some differences can be noted, however, close to the generation bounds. Closer to the generation bounds, the censoring effect in the parametric densities is more present (see Figure C.1), hence leading to a higher bias in parameter estimates. For the particular case of the the upper bound, that is, when expected power is close to nominal capacity, it is also that power down-regulation actions were not always flagged and discarded from the dataset, as much as they should be, then affecting the parameter estimation and evaluation of the forecasts. Better results are therefore expected to be seen in case down-regulation actions are better dealt with when gathering wind power generation datasets in the future.

Similar differences in conditional forecast skill can be observed when assessing the skill of predictive densities conditional to actual power measurements, see Fig. C.5. Parametric densities perform better during periods when observed power is not close to the generation bounds. In contrast during periods with low and high power generation, the non-parametric densities show superior results. This suggests that CN:CP/CP-ARCH forecasts have a better ability to discriminate among the observations when the power is in its medium range and a worse ability to discriminate as the power generation gets closer to its natural bounds.

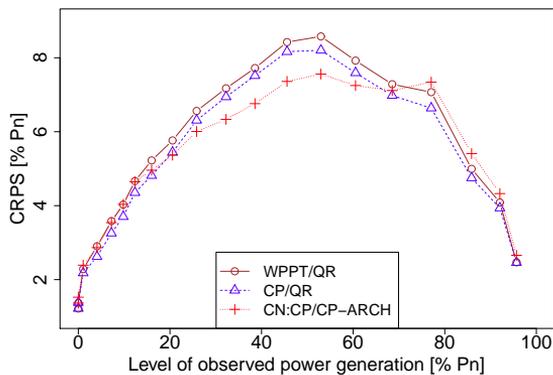


Figure C.5: CRPS conditional to the observed power generation level for the 1-hour lead time, with WPPT/QR being the benchmark. Power levels are given by 20 equally populated classes deduced from the distribution of power measurements.

As an illustration of the type of probabilistic forecasts finally obtained with the

various approaches proposed here, Figure C.6 gives the example of predictive densities issued by the CP/QR model, issued on the 24th of November, 2009, at 17:15, their shape evolving with the level of power and with the lead time.

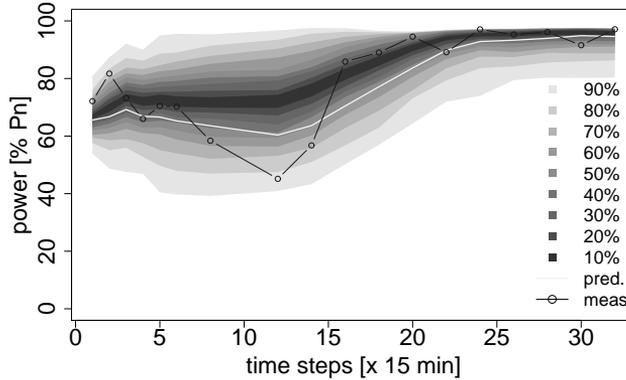


Figure C.6: Predictive densities of wind power generation from the CP/QR model, depicted by a set of prediction intervals with increasing nominal coverage rates. These forecasts were issued on the 24th of November, 2009, at 17:15, for lead times between 15 mins and 8 hours ahead.

C.6 Conclusions

Focus was given to probabilistic wind power forecasting with consideration of geographically distributed information, hence permitting to capture additional space-time dynamics. The proposed methodology can be used for issuing density forecasts for a single site of interest, while using information from the other wind farms as explanatory variables.

The methodology is based on discrete formulation of the problem as opposed to proposing a full space-time covariance model which normally would call for a larger amount of reference sites spread throughout the territory. The approach chosen here is based on proposing ways to best summarize a snapshot of forecast errors observed at time t when issuing probabilistic forecasts for time $t + k$.

Two ways of constructing predictive densities were described and analysed. The parametric approach relied on CN distributions while the non-parametric one employed quantile regression techniques. All estimation methods were introduced in a time-adaptive framework in order to reduce computational costs,

while allowing for long-term variations in the process dynamics, as induced by meteorological systems for instance.

The empirical results obtained on the test case of a portfolio of wind farms in Denmark show that accounting for spatio-temporal effects improves the quality of probabilistic forecasts for a range of lead times, here up to 8 hours. It was shown to be sufficient to focus on correcting the conditional expectation of wind power generation. Additional inclusion of spatio-temporal effects into the uncertainty modelling step did not significantly further improve the skill of the predictive densities. None of the proposed approaches outperformed the benchmark for lead times further than 8 hours ahead, in line with the scales of motion of weather systems over a small region such as that covered by Denmark. Thus, likely, for a different test case, depending on the geographical layout of the considered wind farms and some meteorological particularities of the area, this maximum lead time would differ.

The performance of parametric and non-parametric approaches were also compared, uncovering that they both performed similar for lead times up to 5 hours ahead, and with an advantage for non-parametric predictive densities for further lead times. Based on overall skill, the highest-quality forecasts were obtained by adaptive quantile regression with spatially corrected point predictions, with CRPS improvements between 1.5% to 4.6% depending upon the lead time. Further research on both parametric and non-parametric approaches to wind power probabilistic forecasting may challenge the comparison carried out here.

C.7 Acknowledgement

The authors gratefully acknowledge DONG Energy for the data and financial support provided through the project “Spatio-temporal correction of wind power forecasts”. Additional acknowledgements are due to Tryggvi Jónsson for his help with quantile regression.

References C

- [1] H. Farhangi, “The path of the smart grid,” *Power and Energy Magazine, IEEE*, vol. 8, pp. 18–28, 2010.
- [2] T. Ackermann *et al.*, *Wind power in power systems*, vol. 140. Wiley Online Library, 2005.
- [3] L. Jones and C. Clark, “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*, 2011.
- [4] 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*, vol. 12, no. 6, pp. 1725–1744, 2008.
- [5] G. Giebel, R. Brownsword, G. Kariniotakis, M. Denhard, and C. Draxl, “The state-of-the-art in short-term prediction of wind power—A literature overview. ANEMOS.plus,” tech. rep., Technical University of Denmark, 2011. available at: <http://orbit.dtu.dk>.
- [6] T. Gneiting, “Editorial: probabilistic forecasting,” *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, vol. 171, no. 2, pp. 319–321, 2008.
- [7] A. Botterud, Z. Zhou, J. Wang, R. Bessa, H. Keko, J. Sumaili, and V. Miranda, “Wind power trading under uncertainty in LMP markets,” *Power Systems, IEEE Transactions on*, vol. 27, no. 2, pp. 894–903, 2012.

-
- [8] X. Liu, "Economic load dispatch constrained by wind power availability: A wait-and-see approach," *Smart Grid, IEEE Transactions on*, vol. 1, no. 3, pp. 347–355, 2010.
- [9] A. Botterud, Z. Zhou, J. Wang, J. Sumaili, H. Keko, J. Mendes, R. J. Bessa, and V. Miranda, "Demand dispatch and probabilistic wind power forecasting in unit commitment and economic dispatch: A case study of Illinois," *Sustainable Energy, IEEE Transactions on*, vol. 4, no. 1, pp. 250–261, 2013.
- [10] C. Liu, J. Wang, A. Botterud, Y. Zhou, and A. Vyas, "Assessment of impacts of PHEV charging patterns on wind-thermal scheduling by stochastic unit commitment," *Smart Grid, IEEE Transactions on*, vol. 3, no. 2, pp. 675–683, 2012.
- [11] Á. Duque, E. Castronuovo, I. Sánchez, and J. Usaola, "Optimal operation of a pumped-storage hydro plant that compensates the imbalances of a wind power producer," *Electric Power Systems Research*, vol. 81, no. 9, pp. 1767–1777, 2011.
- [12] R. Bessa, M. Matos, I. Costa, L. Bremermann, I. Franchin, R. Pestana, N. Machado, H. Waldl, and C. Wichmann, "Reserve setting and steady-state security assessment using wind power uncertainty forecast: a case study," *IEEE Transactions on Sustainable Energy*, 2012.
- [13] M. Ortega-Vazquez and D. Kirschen, "Assessing the impact of wind power generation on operating costs," *Smart Grid, IEEE Transactions on*, vol. 1, no. 3, pp. 295–301, 2010.
- [14] T. Gneiting, "Quantiles as optimal point forecasts," *International Journal of Forecasting*, vol. 27, no. 2, pp. 197–207, 2011.
- [15] T. Gneiting, K. Larson, K. Westrick, M. Genton, and E. Aldrich, "Calibrated probabilistic forecasting at the stateline wind energy center," *Journal of the American Statistical Association*, vol. 101, no. 475, pp. 968–979, 2006.
- [16] J. W. Taylor, P. E. McSharry, and R. Buizza, "Wind power density forecasting using ensemble predictions and time series models," *Energy Conversion, IEEE Transactions on*, vol. 24, no. 3, pp. 775–782, 2009.
- [17] I. Sanchez, "Short-term prediction of wind energy production," *International Journal of Forecasting*, vol. 22, no. 1, pp. 43–56, 2006.
- [18] J. Jeon and J. W. Taylor, "Using conditional kernel density estimation for wind power density forecasting," *Journal of the American Statistical Association*, vol. 107, no. 497, pp. 66–79, 2012.

- [19] 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)*, 2012.
- [20] P. Pinson and G. Kariniotakis, “Conditional prediction intervals of wind power generation,” *Power Systems, IEEE Transactions on*, vol. 25, no. 4, pp. 1845–1856, 2010.
- [21] J. Møller, H. Nielsen, and H. Madsen, “Time-adaptive quantile regression,” *Computational Statistics & Data Analysis*, vol. 52, no. 3, pp. 1292–1303, 2008.
- [22] R. Bessa, V. Miranda, A. Botterud, Z. Zhou, and J. Wang, “Time-adaptive quantile-copula for wind power probabilistic forecasting,” *Renewable Energy*, vol. 40, no. 1, pp. 29–39, 2012.
- [23] P. Pinson, H. Nielsen, J. Møller, H. Madsen, and G. Kariniotakis, “Non-parametric probabilistic forecasts of wind power: required properties and evaluation,” *Wind Energy*, vol. 10, no. 6, pp. 497–516, 2007.
- [24] H. Nielsen, T. Nielsen, and H. Madsen, “An overview of wind power forecasts types and their use in large-scale integration of wind power,” in *Proceedings of the 10th International Workshop on Large-Scale Integration of Wind Power into Power Systems*, 2011.
- [25] D. Lew and R. Piwko, “Western wind and solar integration study,” tech. rep., National Renewable Energy Laboratories, 2010.
- [26] P. Trombe, P. Pinson, T. Bøvith, N. Cutululis, C. Draxl, G. Giebel, A. Hahmann, N. Jensen, B. Jensen, N. Le, H. Madsen, L. Pedersen, A. Sommer, and C. Vincent, “Weather radars - The new eyes for offshore wind farms?,” 2012. preprint, under review.
- [27] J. Tastu, P. Pinson, and H. Madsen, “Multivariate conditional parametric models for a spatiotemporal analysis of short-term wind power forecast errors,” in *Scientific Proceedings of the European Wind Energy Conference, Warsaw (PL)*, pp. 77–81, 2010.
- [28] R. Girard and D. Allard, “Spatio-temporal propagation of wind power prediction errors,” *Wind Energy*, 2012.
- [29] I. G. Damousis, M. C. Alexiadis, J. B. Theocharis, and P. S. Dokopoulos, “A fuzzy model for wind speed prediction and power generation in wind parks using spatial correlation,” *Energy Conversion, IEEE Transactions on*, vol. 19, no. 2, pp. 352–361, 2004.
- [30] M. Alexiadis, P. Dokopoulos, and H. Sahsamanoglou, “Wind speed and power forecasting based on spatial correlation models,” *Energy Conversion, IEEE Transactions on*, vol. 14, no. 3, pp. 836–842, 1999.

- [31] K. A. Larson and K. Westrick, "Short-term wind forecasting using off-site observations," *Wind energy*, vol. 9, no. 1-2, pp. 55–62, 2006.
- [32] A. Hering and M. Genton, "Powering up with space-time wind forecasting," *Journal of the American Statistical Association*, vol. 105, no. 489, pp. 92–104, 2010.
- [33] P.-J. Trombe and P. Pinson, "High-resolution forecasting of wind power generation with regime-switching models and off-site observations," tech. rep., Technical University of Denmark, Dpt. of Applied Mathematics and Computer Science, 2012. available at: <http://orbit.dtu.dk>.
- [34] A. Lau, *Probabilistic Wind Power Forecasts: From Aggregated Approach to Spatio-temporal Models*. PhD thesis, Mathematical Institute, University of Oxford, 2011.
- [35] R. Jursa and K. Rohrig, "Short-term wind power forecasting using evolutionary algorithms for the automated specification of artificial intelligence models," *International Journal of Forecasting*, vol. 24, no. 4, pp. 694–709, 2008.
- [36] M. Lange, "On the uncertainty of wind power predictions—analysis of the forecast accuracy and statistical distribution of errors," *Journal of Solar Energy Engineering*, vol. 127, p. 177, 2005.
- [37] H. Nielsen, T. Nielsen, A. Joensen, H. Madsen, and J. Holst, "Tracking time-varying-coefficient functions," *International Journal of Adaptive Control and Signal Processing*, vol. 14, no. 8, pp. 813–828, 2000.
- [38] J. Tastu, P. Pinson, P.-J. Trombe, and H. Madsen, "Spatio-temporal correction targeting Nysted offshore – probabilistic forecasts," tech. rep., Technical University of Denmark, 2012.
- [39] R. Koenker and G. Bassett Jr, "Regression quantiles," *Econometrica: journal of the Econometric Society*, pp. 33–50, 1978.
- [40] B. Sass, N. Nielsen, J. Jørgensen, B. Amstrup, M. Kmit, and K. Mogensen, "The operational DMI-HIRLAM system 2002-version," tech. rep., DMI - Danmarks Meteorologiske Institut, 2002.
- [41] J. Cappelen and B. Jørgensen, "Observed wind speed and direction in Denmark - with climatological standards normals, 1961-90," tech. rep., DMI - Danmarks Meteorologiske Institut, 1999.
- [42] T. Gneiting and A. Raftery, "Strictly proper scoring rules, prediction, and estimation," *Journal of the American Statistical Association*, vol. 102, no. 477, pp. 359–378, 2007.

-
- [43] F. X. Diebold and R. S. R. S. Mariano, "Comparing predictive accuracy," *Journal of Business & Economic Statistics*, vol. 13, no. 3, pp. 253–263, 1995.

PAPER D

Spatio-temporal correction targeting Nysted Offshore. Probabilistic forecasts

Authors:

Julija Tastu, Pierre Pinson, Pierre-Julien Trombe, Henrik Madsen

Published as :

Technical report

Spatio-temporal correction targeting Nysted Offshore. Probabilistic forecasts

Julija Tastu¹, Pierre Pinson¹, Pierre-Julien Trombe¹, Henrik Madsen¹

Abstract

The report is a part of the project *DONG Spatio-temporal* which is a project between DONG Energy, DTU Informatics and ENFOR A/S financed by DONG Energy A/S. The study investigates whether accounting for the geographically distributed information improves the quality of the resulting probabilistic forecasts for the Nysted offshore wind farm. Different approaches for issuing predictive densities are studied and compared.

The results show that the spatio-temporal correction of the conditional expectations of the predictive densities improves the quality of the corresponding forecasts for a range of lead times up to 5-8 hours ahead. Similar correction of the higher order moments is shown to be unnecessary as does not improve the skill of the resulting predictions.

The best performing of the studied models is based on the adaptive quantile regression using the spatially corrected point predictions as input. This model is shown to outperform the benchmark approach given by the locally optimized forecasts in terms of the CRPS score by 1.5%-8.29% depending on the considered prediction horizon.

D.1 Introduction

Previous studies within *DONG Spatio-temporal* concern improvements in the forecasts of the expected power at Nysted Offshore using spatial information. Such predictions are never 100% accurate as they are always associated with some uncertainty due to incomplete knowledge one has about the future events. The uncertainty of the point forecasts is highly variable and depends on the considered look-ahead time, the quality of the related meteorological forecasts

¹DTU Informatics, Technical University of Denmark, Richard Petersens Plads, bld. 305, DK-2800 Kgs. Lyngby, Denmark

and many more factors. As the verity of the point-forecasts is time-varying, they are not to be trusted in the same way at every step. Providing information on how much those forecasts can actually be trusted at a certain moment, plays a very important role in many decision making tasks. Such information is directly related to knowing how much risk is being taken by relying on the prediction. Complementing point predictions with the estimates of the associated uncertainty is a task of probabilistic forecasting.

The main objective of this work is to extend the point forecasts for Nysted Offshore to probabilistic forecasts. The aim is to "dress" the point predictions with the uncertainty estimates in order to obtain full predictive densities of the generated wind power instead of giving only the expected value (which is the case for the point forecasts). There are two ways of getting such predictive densities. One way is to assume that the wind power generation follows a known distribution. Then in order to build a predictive density one needs to estimate the parameters of the assumed distribution using the past observations. This approach is called a parametric, or distribution-based method. Another way is to assume that the exact distribution is unknown. In that case one needs to consider a wind power expectation given by the point prediction and assume that the accuracy of it will be directly related to the previously observed performance of the corresponding point forecast model. The quantiles of the past point forecast errors are then estimated. A 0.95 quantile, for example, gives a value which will not be exceeded by the observation with a probability (nominal proportion) of 0.95. After estimating the quantiles, the point forecasts are being "dressed" with a set of prediction intervals of different nominal coverage rates in order to obtain a full density. This approach is called a non-parametric, or distribution-free approach.

When building probabilistic forecasts for a wind power variable which account for the spatial information, some knowledge could be taken from the earlier works concerning spatial modelling of a wind speed data with a consideration of probability forecasts. Gneiting et al. [1] propose regime switching models which account for two dominant wind directions while predicting wind speed up to two hours ahead. The corresponding probabilistic forecasts for the wind speed are based on a truncated Normal distribution. The truncation is carried out in order to ensure that the final forecasts are within the valid range, i.e. no negative wind speed values occur. Hering and Genton [2] extend the work of Gneiting by considering a circular wind direction and applying both truncated Normal distribution and a skew-T distribution for describing the model related uncertainties. In their work the authors also link the predictive wind speed densities to the corresponding wind power variable using a power curve model. Lau in [3] looked at the spatio-temporal models for a wind power generation while considering multi-step ahead forecasts described by the censored Normal distribution with spatial covariance structures.

In parallel, without considering spatial effects, but rather concentrating on a single site, Pinson [4] compares several types of parametric predictive densities for describing very short-term wind power generation at Horns Rev wind farm. A censored Normal, a Beta and a generalized logit-Normal distributions are considered in the work. The author shows that the generalized logit-Normal distribution is a better candidate for describing the wind power generation than the censored Normal or Beta distributions.

In addition some research was carried out for proposing non-parametric approaches for probabilistic forecasting of the generated wind power. Møller et al. [5] introduced an adaptive estimation scheme for a quantile regression and used the method for issuing probabilistic wind power forecasts. In parallel, in [6] Pinson and Kariniotakis proposed another non-parametric approach called adaptive re-sampling. Similarly to the adaptive quantile regression this approach does not assume any particular distribution for the wind power generation. Instead it builds prediction intervals of different nominal coverage rates based on the past deviations from the expected power. Both mentioned non-parametric approaches have been compared on the same test case - hourly wind power predictions at Klimt wind farm. The results are documented in [7] where it is shown that the two methods perform similarly.

To our knowledge, no studies comparing parametric and non-parametric approaches to probabilistic wind power forecasting exist. As far as the spatial models for the wind power data are concerned, nothing but the censored Normal distribution has been considered (see [3]). The goal of this study is therefore to apply both parametric and non-parametric methods for issuing probabilistic forecasts for Nysted Offshore. The comparison of the forecasts performance will be carried out considering different prediction horizons. The most important questions which are aimed to be answered are:

- which modelling approach gives better results: parametric or non-parametric one? Is it consistent for all the considered horizons?
- does accounting for the spatial information improve the performance of the probabilistic forecasts?

The outline of the report is as follows:

A short description of the data used in this work is given in Section D.2. It is followed by the methodological aspects on the parametric probabilistic forecasting (Section D.3). The distributions considered in this work are presented in Sections D.3.1 and D.3.2. The parameter estimation for the considered densities

is discussed in Section D.3.3 and D.3.4. A non parametric approach for building the predictive densities is presented in Section D.4. Section D.5 discusses the methods for probabilistic forecast assessment. The results are presented in Section D.6. The paper ends with the concluding remarks in Section D.7.

D.2 Data and previous work

All analysis within *Dong Spatio-temporal* are primary based on the WPPT power (point) forecasts with a resolution of 15 min and a horizon of 48 hours available at a number of wind farms listed in Table D.1. The location of the farms are shown in Figure D.1. The WPPT forecasts are based on the power measurements with a resolution of 5 min and meteorological forecasts with a time resolution of 1 hour and a horizon of 48 hours. The meteorological forecasts include wind direction and wind speed at an altitude of 10 meters. Meteorological forecasts are provided by the HIRLAM model operated by the Danish Meteorological Institute (DMI). The WPPT forecasts used in the study are historical runs where the delay of the meteorological forecasts has been set to 4 hours. This implies that the meteorological forecasts originally issued, for instance, at midnight become available as an input to the wind power prediction models only at 4 a.m. This means that the wind power forecasts issued at 4 a.m. and targeting 5 a.m. (one hour ahead predictions) use the meteorological forecast issued at midnight as input. That implies that the one-hour ahead wind power predictions issued at 4 a.m. are actually using five-hours ahead meteorological forecasts as input. The data covers the period from the 1st of May, 2008 to the 1st of February, 2010. All the models considered in the project have been run for the period from the 1st of May, 2008 to the 31st of December, 2009. The actual evaluation of the models' performance is based on the period from the 1st of January, 2009 to the 31st of December, 2009.

Spatio-temporal correction of the WPPT power forecasts at Nysted Offshore was carried out in [8, 9]. Different configurations of the correction models have been tried in [8, 9]. The best performance was observed when the spatial correction models (Conditional Parametric models) for Nysted Offshore have been run using a reduced set of explanatory variables proposed in [9] (the corresponding model is denoted as a CP-model) and additionally to that incorporated data transformations described in [8] (the corresponding model is denoted as a Logit-CP model). Detailed analysis of those two models is given in [8]. Complementary to the initial WPPT forecasts in this work the corrected predictions made by the CP and the Logit-CP models are considered.

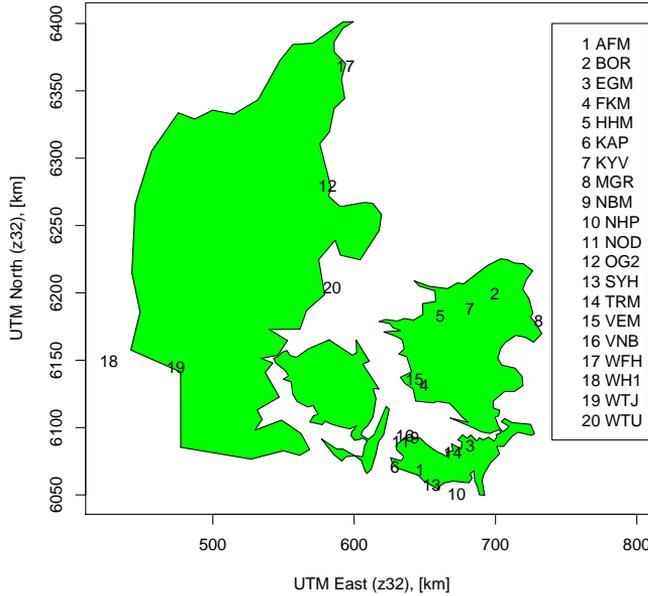


Figure D.1: Map of wind farms included in the analysis. Nysted (target wind farm) is marked by number 10

D.3 Parametric approaches to probabilistic forecasting

Parametric approach is based on the assumption that a distribution function of the wind power generation is known. In order to choose a specific type of density, the crucial features of the wind power variable are to be taken into the account. Firstly, when normalized by the nominal capacity, it is double bounded between a minimum production of 0 and a maximum one of 1. It is in parallel a non-linear function of the forecasted wind speed in the shape of a sigmoid, which means that higher variations in the wind power are observed when the predicted wind speed is in its medium range. For the wind speeds close to the cut-in or cut-off values the fluctuations in the observed power are lower. This means that a distribution of the observed power changes with the level of the predicted wind speed and as the latter is directly related to the expected power, one can say that the density of the observed power depends on

Name	Abbrev.	Cap. (MW)	Number of turbines	UTM E	UTM N	Hub h.
Askoe Faeland	AFM	2.40	4	646777.00	6068525	40.00
Borup (sum)	BOR	3.96	6	699583.00	6199488	44.00
Egelev	EGM	1.80	4	682256.00	6086204	35.00
Frankerup/Faarstrup	FKM	2.70	4	649515.50	6131569	40.50
Hagesholm	HHM	6.00	3	661093.00	6182659	68.00
Kappel	KAP	10.20	24	628911.00	6070409	32.00
Kyndby	KYV	3.06	19	682049.00	6188266	30.00
Middelgrunden	MGR	20.00	10	730597.30	6178847	64.00
Nyboelle	NBM	1.00	2	642797.00	6092329	35.00
Nysted Offshore	NHP	165.60	72	672504.00	6050365	68.00
Noejsohmheds Odde	NOD	21.00	21	633010.00	6089671	50.00
Overgaard 2	OG2	11.50	5	581199.00	6279823	80.00
Syltholm	SYH	18.75	25	654876.00	6057536	44.00
Taars	TRM	1.00	2	669788.00	6081120	40.00
Vemmelev (sum)	VEM	4.65	8	642791.67	6135634	38.33
Vindeby ("offshore")	VNB	4.95	11	635956.00	6094187	38.00
Frederikshavn	WFH	10.60	4	593839.00	6368078	80.00
Horns Rev 1	WH1	160.00	80	426487.00	6149224	80.00
Tjaereborg Enge	WTJ	2.00	1	473833.00	6144928	57.00
Tunoe Knob	WTU	5.00	10	584391.00	6203478	45.00

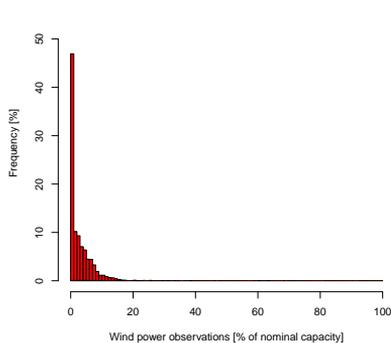
Table D.1: Wind farms included in the analysis. UTM coordinates corresponds to zone 32.

the level of the expected power. This is further demonstrated in Figure D.2. Another point clearly seen in Figure D.2 is a non-negligible concentration of probability mass at the bounds. When the expected power is far from the natural generation bounds (see Figure D.2(c)), the conditional histogram resembles that of a Gaussian distribution (a characteristic bell-shape around the expected value can be seen). The closer to the bounds, the less dispersed distributions become and the higher the probability concentration at the closest bound can be noted.

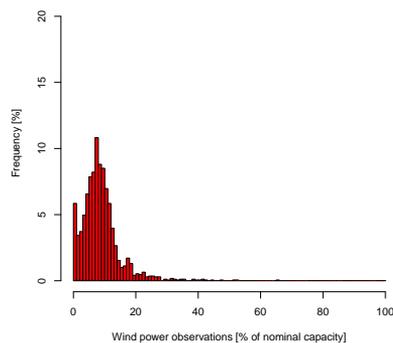
D.3.1 Censored Normal distribution

For modelling those effects a Censored Normal distribution with varying parameters has been considered in [4, 3]. The censoring is needed due to the fact that a wind power variable is bounded. A Censored Normal wind power variable follows an ordinary Gaussian law within the open unit interval. Since the values outside $[0,1]$ cannot be taken, the corresponding tails of the ordinary Gaussian distribution are cut and deposited to the bounds of the defined interval (to 0 and 1, respectively).

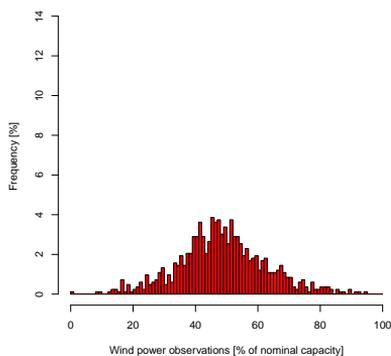
Formally the Censored Normal predictive density for the wind power generation



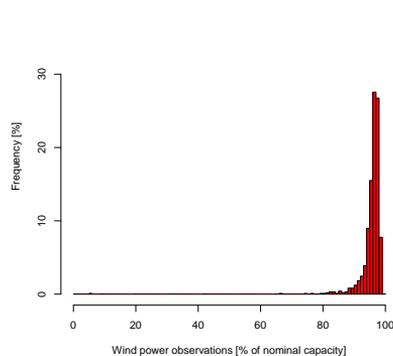
(a) Expected power in 2 - 4 % of nominal capacity



(b) Expected power in 8 - 10 % of nominal capacity



(c) Expected power in 48 - 50 % of nominal capacity



(d) Expected power in 96 - 98 % of nominal capacity

Figure D.2: Distribution of the observed power conditional to different levels of the expected power given by the WPPT forecasts. Note that the range of frequencies on the y-axis varies from one plot to the next.

X_{t+k} at time $t+k$ can be written as:

$$X_{t+k} \sim w_{t+k}^0 \delta_0(x) + f^{(0,1)}(x; \mu_{t+k}, \sigma_{t+k}^2) + w_{t+k}^1 \delta_1(x), x \in [0, 1] \quad (\text{D.1})$$

where δ_0 and δ_1 are Dirac delta functions at 0 and 1, respectively, representing the location of the potential concentration of probability mass:

$$\delta_j(x) = \begin{cases} \infty & \text{if } x = j \\ 0 & \text{otherwise} \end{cases}$$

w_{t+k}^0 and w_{t+k}^1 are the weights representing the levels of probability mass concentration at the corresponding bounds of the unit interval.

$$w_{t+k}^0 = F(0; \mu_{t+k}, \sigma_{t+k}^2) \quad (\text{D.2})$$

$$w_{t+k}^1 = 1 - F(1; \mu_{t+k}, \sigma_{t+k}^2) \quad (\text{D.3})$$

with $F(x; \mu, \sigma^2)$ being a cumulative Gaussian distribution function with parameters μ and σ^2 . $f^{(0,1)}(x; \mu, \sigma^2)$ follows a Gaussian density function (with a location parameter μ and a scale parameter σ^2) within the open unit interval $(0,1)$ and equals 0 outside this interval:

$$f^{(0,1)}(x; \mu, \sigma^2) = \begin{cases} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} & \text{if } x \in (0, 1) \\ 0 & \text{otherwise} \end{cases}$$

D.3.2 Generalized Logit-Normal distribution

The aforementioned characteristics of the wind power variable (bounds, non-constant variance) indicate that some data transformations could be employed in order to stabilize the variance, reduce the influence of the bounds and as a result could make the assumption of a Gaussian distribution more appropriate. For that purpose a generalized logit (GL) transformation (see equation (D.4)) has been proposed in [4]. In [8] this transformation has been incorporated into the models for spatially correcting the expected wind power generation at Nysted Offshore and resulted in the improved model performance.

For an original time series x_t the generalized logit transform is given by

$$y_t = \gamma(x_t; v) = \ln \left(\frac{x_t^v}{1 - x_t^v} \right), v > 0, |x_t| < 1 \quad (\text{D.4})$$

while the inverse transformation is defined as

$$x_t = \gamma^{-1}(y_t; v) = \left(1 + \frac{1}{\exp(y_t)} \right)^{-1/v}, v > 0 \quad (\text{D.5})$$

A shape parameter v has been set to 0.01 (see [8] for details).

Based on (D.1), but now considering Y_{t+k} , the GL-transform of the corresponding wind power generation X_{t+k} , the form of its predictive density is given by

$$Y_{t+k} \sim w_{t+k}^0 \delta_{-\infty}(y) + f^{(0,1)}(y; \tilde{\mu}_{t+k}, \tilde{\sigma}_{t+k}^2) + w_{t+k}^1 \delta_{\infty}(y), y \in \mathcal{R} \quad (\text{D.6})$$

In other words, instead of assuming that the wind power generation X_{t+k} follows a censored Normal distribution, this distribution is assumed for the GL-transform of the original variable. Otherwise stated - the original wind power generation is assumed to follow a generalized logit Normal distribution.

D.3.3 Estimation of the location parameter

A parallel between parameter estimation in the censored Normal and the generalized logit-Normal densities. Both the censored Normal and the generalized logit-Normal densities are characterized by the location and the scale parameters. The methodology for estimating the parameters is analogical for both densities. The only difference comes from the input used in the estimation routines. Suppose μ (the location parameter of the censored Normal density) is described as a function f of the available wind power measurements p and the point predictions \hat{p} , i.e. $\mu = f(p, \hat{p})$. Then the location parameter of the generalized logit-Normal density $\tilde{\mu}$ can be found from $f(\gamma(p; v), \gamma(\hat{p}; v))$.

The same parity holds for the scale parameters σ^2 and $\tilde{\sigma}^2$. Thus in the following we focus in details on the estimation of μ and σ^2 , keeping in mind that the equivalent techniques with a corresponding input adjustment can be used for finding $\tilde{\mu}$ and $\tilde{\sigma}^2$.

From Normal to censored-Normal distribution: approximation used in the parameter estimation routines As has been mentioned before, the censored Normal predictive density can be fully characterised by the location (μ) and the scale (σ^2) parameters. In case of an ordinary (non-Censored) Gaussian distribution the location parameter is described by the mean (expectation) of the distribution and the scale parameter - by its variance. It is not exactly the case for the Censored Normal distribution. However, as previously demonstrated in Figure D.2 censoring effect is only significant close to the bounds. But even close to them it is assumed that the location and the scale parameters can be well approximated by the expectation and the variance, respectively. Such approximation introduces a certain bias (lack of accuracy) as the observed power is close to the generation bounds, but nevertheless this estimation method is being used in the state-of-the art research works (see [4] as an example) where it is argued to show a satisfactory performance.

Estimating the location parameter Thus the location parameter of the censored Normal distribution is approximated by the mean (expected value) of the corresponding density. When modelling wind power, the expectation is given by the point forecast. Therefore an estimate of the location parameter which describes the density of the power generation at time $t + k$ (when predicting k -steps ahead) is given by:

$$\widehat{\mu}_{t+k|t} = \widehat{p}_{t+k|t} \quad (\text{D.7})$$

where $\widehat{\mu}_{t+k|t}$ denotes the estimate (forecasted value) of the location parameter $\mu_{t+k|t}$ and $\widehat{p}_{t+k|t}$ stands for the point prediction of the wind power generation issued at time t for time $t + k$.

Recall, the location parameter of the generalized logit-Normal density could be then found from:

$$\widehat{\mu}_{t+k|t} = \gamma(\widehat{p}_{t+k|t}; v) \quad (\text{D.8})$$

where $\widehat{\mu}_{t+k|t}$ denotes an estimate of $\widetilde{\mu}_{t+k|t}$ and $\widehat{p}_{t+k|t}$ stands for the expected wind power generation at time $t + k$ forecasted at time t .

In this work several different point predictions ($\widehat{p}_{t+k|t}$) are considered: initial

WPPT forecasts and two spatially corrected forecasts which showed the best improvements over the WPPT in the previous studies ([8]). More precisely, the following power expectations are considered:

1. *WPPT forecasts*
2. *CP forecasts.* Conditional parametric models are used for the spatial correction of the WPPT forecasts. The correction model is run using a reduced set of explanatory variables proposed by ENFOR in [9]. The model itself and the resulting predictions are analysed in details in [8].
3. *Logit-CP forecasts.* WPPT forecast correction is performed using the conditional parametric model which is run on the reduced set of explanatory variables and incorporates the GL data transformation (D.4). A detailed description of the model is given in [10].

D.3.4 Estimation of the scale parameter

The scale parameter in case of the Censored Normal density is approximated by the variance of the wind power distribution. Variance is estimated by modelling squared residuals (ϵ^2) made by the point forecasting model:

$$\epsilon_{t+k}^2 = (p_{t+k} - \hat{p}_{t+k|t})^2 \quad (\text{D.9})$$

where p_{t+k} denotes wind power observation at time $t+k$ and $\hat{p}_{t+k|t}$ is the corresponding wind power point forecast issued for time $t+k$ at time t (k -step ahead point prediction).

An estimate ($\hat{\sigma}^2$) of the scale parameter σ^2 is given by the expectation (prediction) of the future squared deviation, i.e.:

$$\hat{\sigma}_{t+k|t}^2 = \hat{\epsilon}_{t+k|t}^2 \quad (\text{D.10})$$

where $\hat{\epsilon}_{t+k|t}^2$ denotes forecasted value of the squared deviation issued at time t for time $t+k$.

As discussed before, the scale parameter of the generalized logit-Normal density is then found from

$\widehat{\sigma}^2$ can be obtained by modelling the squared deviations ($\widetilde{\epsilon}_t^2$) of the transformed wind power variable i.e. :

$$\widehat{\sigma}_{t+k|t}^2 = \widetilde{\epsilon}_{t+k|t}^2 \quad (\text{D.11})$$

where

$$\widetilde{\epsilon}_{t+k}^2 = \left(\gamma(p_{t+k}; v) - \gamma(\hat{p}_{t+k|t}; v) \right)^2 \quad (\text{D.12})$$

The same techniques are used for modelling $\widetilde{\epsilon}_{t+k}^2$ as for ϵ_{t+k}^2 :

1. Exponential smoothing
2. GARCH / GARCHX model
3. Conditional Parametric ARCH model without/ with eXogenous input (CP-ARCH/ CP-ARCHX)
4. Markov switching models

D.3.4.1 Exponential smoothing

One of the easiest to implement techniques for the adaptive estimation of the scale parameter σ^2 can be performed using an exponential smoothing scheme. This method can be viewed as a simple weighted average between the previous smoothed statistic and the information given by the recently available observation. The smoothing scheme writes as:

$$\hat{\sigma}_{t+k|t}^2 = (1 - \alpha)\hat{\sigma}_{t+k-1|t-1}^2 + \alpha(p_t - \hat{p}_{t|t-k})^2 \quad (\text{D.13})$$

where α is a smoothing parameter. It is arbitrary set to 0.9997 in this work. The chosen value assumes a locally constant value of the variance, allowing for the long-term variations only.

D.3.4.2 GARCH and GARCHX models

In order to suggest a more advanced model for the squared deviations, the crucial features of the ϵ_t^2 series have to be revealed. Figure D.3 depicts squared residuals of the 15-min ahead WPPT point forecasts for the period of time spanning the 20th of March, 2009 to the 20th of June, 2009. Volatility of the prediction errors exhibits clustering in time: large errors tend to be followed by large errors and similarly small errors tend to be followed by the small ones.

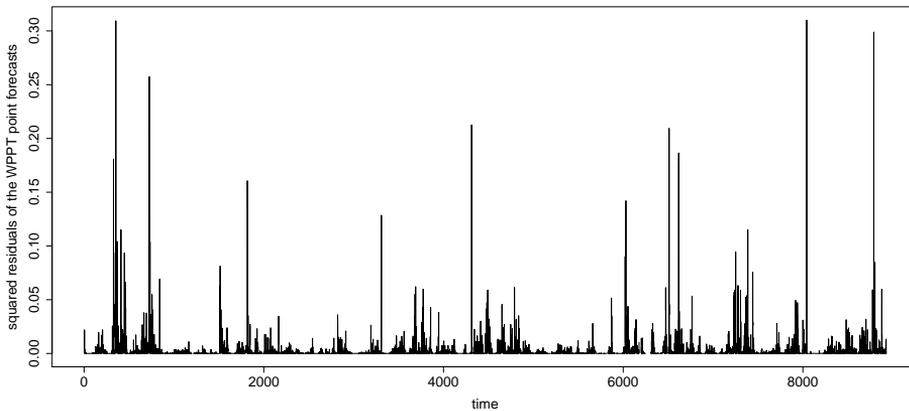
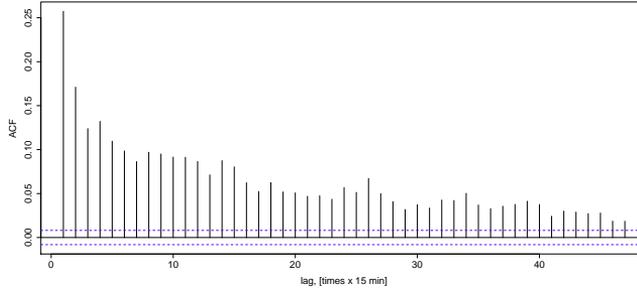


Figure D.3: Squared residuals of the WPPT forecasts (prediction horizon is 15 min) in the period of the 20th of March, 2009 to the 20th of June, 2009

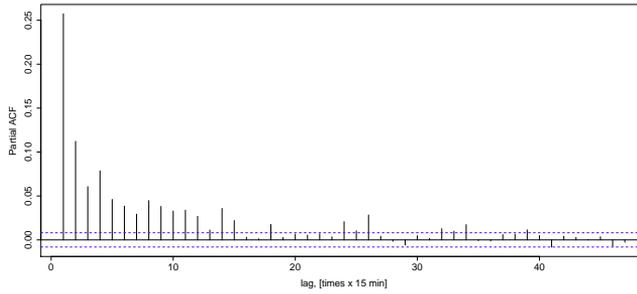
The autocorrelation function (ACF) and the partial ACF of the squared errors made by the WPPT 15-min ahead forecasts (see Figures D.4) indicate that the conditional variance should be modelled as a weighted sum of approximately the last 18 squared errors (suggesting a high order AutoRegressive Conditional Heteroscedasticity (ARCH) model).

However, for the sake of parsimony, an ARCH model of a high order can be substituted by a Generalized ARCH (GARCH) specification [11].

Following [11] a noise process ϵ_t is called a GARCH process of order (p, q)



(a) Autocorrelation of the squared errors



(b) Partial autocorrelation of the squared errors

Figure D.4: Autocorrelation and partial autocorrelation of the squared errors made by the WPPT 15-min ahead forecasts. Estimated on the data spanning the 1st of January, 2009 to the 31st of December, 2009

if it satisfies:

$$\begin{aligned} \epsilon_t &= \sigma_t v_t, & v_t &\stackrel{i.i.d.}{\sim} D(0,1) \\ \sigma_t^2 &= \alpha_0 + \sum_{j=1}^q \alpha_j \epsilon_{t-j}^2 + \sum_{k=1}^p \beta_k \sigma_{t-k}^2 \end{aligned} \quad (\text{D.14})$$

Where $D(0, 1)$ denotes a distribution with a zero mean and a unit variance. For the specific case of the wind power variable analysed in this work (considering

various prediction horizons k), the model (D.14) is applied to data as:

$$\begin{aligned} \epsilon_{t|t-k} &= \sigma_{t|t-k} v_t, & v_t &\stackrel{i.i.d.}{\sim} D(0, 1) \\ \sigma_{t|t-k}^2 &= \alpha_0 + \sum_{j=k}^t \alpha_j \epsilon_{t-j|t-j-k}^2 + \sum_{l=k}^t \beta_l \sigma_{t-l|t-l-k}^2 \end{aligned} \quad (\text{D.15})$$

where $\epsilon_{t|t-k}$ stands for the errors made by a wind power (point) forecasting model (see (D.9)), $\sigma_{t|t-k}^2$ denotes a conditional variance of the point forecast residuals $\epsilon_{t|t-k}$. In other words, when building a GARCH model with targeting a k steps prediction horizon, GARCH(k,k) model is applied on data with some parameters fixed to zero: $\alpha_j = 0$ and $\beta_j = 0$, where $j = 1, \dots, k-1$. For the sake of simplicity further on in the report we will skip the notation of the k steps ahead prediction horizon and stay with the simpler notation of the models given by ϵ_t rather than $\epsilon_{t|t-k}$.

It is known that if ϵ_t follows a GARCH(p,q) process, then ϵ_t^2 follows an ARMA(r,q) process where $r=\max(p,q)$ [11]. Adaptive estimation of the model parameters can be carried out using a recursive least squares algorithm [12].

Based on the previous research works documented in [8, 9], the WPPT forecasts for Nysted Offshore can be improved if considering a spatial information in the prediction scheme. As previously highlighted, a point forecast corresponds to the mean (the first order moment) of the wind power variable. When moving to probabilistic framework an important question is whether spatial information can be further used for improving higher order moments, not only the mean.

Censored Gaussian distribution is fully characterised by the location and the scale parameters. They are approximated by the the mean and the variance of the predictive density, respectively. Therefore, the question boils down to wondering whether the spatial information can be incorporated into the models for describing the variance.

Figure D.5 shows an empirical cross correlation function between the pre-whitened series of squared deviations at Kappel wind farm and the filtered series of the squared prediction errors observed at Nysted Offshore. The considered wind farms are numbered as 6 and 10 on the map, respectively (see Figure D.1). The plot reveals that the strongest correlation is between the squared deviations at time t at Nysted Offshore and the squared WPPT forecast errors at Kappel observed one hour prior. In other words, the situation at Nysted Offshore "now" depends on the situation at Kappel "one hour ago". Such pre-

liminary diagnostics suggests that the GARCH model (D.14) can be further extended by including an eXogenous input given by the squared errors observed at the distant location. The corresponding model is denoted by GARCHX and is defined as:

$$\begin{aligned} \epsilon_t &= \sigma_t v_t, & v_t &\stackrel{i.i.d.}{\sim} D(0,1) \\ \sigma_t^2 &= \alpha_0 + \sum_{j=1}^q \alpha_j \epsilon_{t-j}^2 + \sum_{k=1}^p \beta_k \sigma_{t-k}^2 + \sum_{j=1}^l \gamma_j \xi_{t-j}^2 \end{aligned} \quad (\text{D.16})$$

where additionally to the notations of the GARCH model ξ_t^2 denotes squared errors made by WPPT at the remote location. For the prediction horizons up to 1 hour squared residuals at Kappel wind farm are considered. For the larger horizons ξ_t^2 is given by the squared errors at Horns Rev I (farm number 18). The choice is based on the analysis of the cross-correlation functions. The results are in line with the geographical layout of the farms (see Figure D.1) and the considered temporal resolution.

If ϵ_t follows a GARCHX(p,q,l) process, then ϵ_t^2 follows an ARMAX(r,q,l) process where $r=\max(p,q)$. Adaptive estimation of the model parameters can be carried out using a recursive least squares algorithm [12].

D.3.4.3 CP-ARCH and CP-ARCHX models

Recall the results demonstrated in Figure D.2: conditional densities of the observed power change with the level of the predicted power. This effect can be accounted for if letting the coefficients of the model be dependent on the level of the predicted power. This calls for considering Conditional Parametric ARCH (CP-ARCH) models for estimating σ_t^2 .

Analogously to ARCH, but with non-constant parameter values, one can define a CP-ARCH process of order (q) as a noise process ϵ_t satisfying:

$$\begin{aligned} \epsilon_t &= \sigma_t v_t, & v_t &\stackrel{i.i.d.}{\sim} D(0,1) \\ \sigma_t^2 &= \alpha_0(\hat{p}_t) + \sum_{j=1}^q \alpha_j(\hat{p}_t) \epsilon_{t-j}^2 \end{aligned} \quad (\text{D.17})$$

Where $D(0,1)$ denotes a distribution with a zero mean and a unit variance. For the specific case of the wind power variable analysed in this work, ϵ_t stands for the errors made by a wind power (point) forecasting model, σ_t^2 denotes a

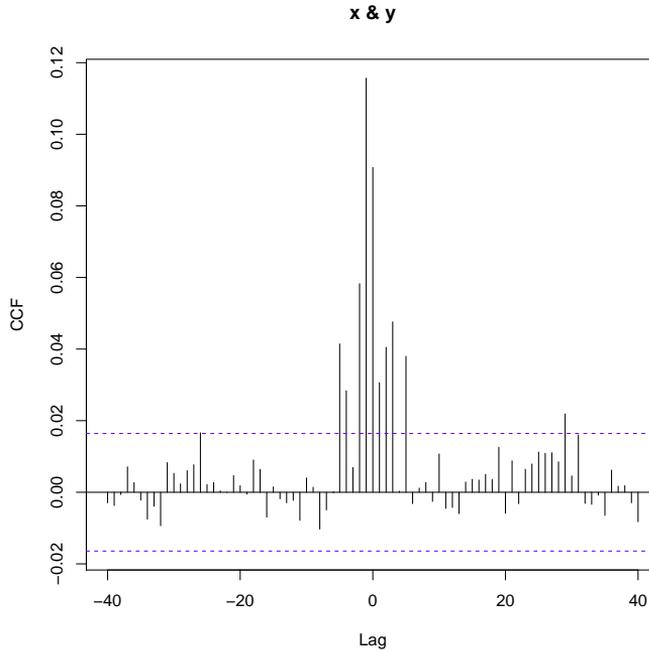


Figure D.5: The empirical cross correlation function between the pre-whitened input series (x) of the squared errors made by the WPPT forecast at Kappel wind farm and the filtered output series (y) of the squared errors made by the CP model for Nysted Offshore. Lags are given in hours. 1-hour ahead predictions are considered.

conditional variance of the residuals ϵ_t . The main difference between the ARCH and the CP-ARCH models comes from the nature of the model parameters. The parameters of the CP-ARCH model α_j , $j = 0, \dots, q$ instead of considered constant (as in case of the ARCH model) are assumed to be smooth functions of \hat{p}_t . In this particular case \hat{p}_t stands for the wind power (point) prediction at Nysted Offshore issued for time t .

A spatial information (an eXogenous signal) can be included into the CP-ARCH

model resulting in the CP-ARCHX specification:

$$\begin{aligned} \epsilon_t &= \sigma_t v_t, \\ \sigma_t^2 &= \alpha_0(\hat{p}_t) + \sum_{j=1}^q \alpha_j(\hat{p}_t) \epsilon_{t-j}^2 + \sum_{j=1}^l \gamma_j(\hat{p}_t) \xi_{t-j}^2 \end{aligned} \quad v_t \stackrel{i.i.d.}{\sim} D(0, 1) \quad (\text{D.18})$$

where additionally to the notations of the CP-ARCH model ξ_t^2 denotes squared errors of the WPPT forecast at the remote location. For the prediction horizons up to 1 hour squared deviations at Kappel wind farm are considered. For the larger horizons ξ_t^2 is given by the squared errors at Horns Rev I (farm number 18).

The CP-ARCH process for ϵ_t can be written as a Conditional Parametric AutoRegressive (CP-AR) model for the ϵ^2 . Similarly CP-ARCHX translates to a Conditional Parametric AutoRegressive model with an eXogenous input (CP-ARX) for the squared residuals. The parameters can then be estimated using an adaptive recursive algorithm used for the parameter estimation in the conditional parametric models. For the estimation details see [10].

D.3.4.4 Markov switching models

Figure D.3 indicates that the variance of the wind power forecast errors is not stationary. The succession of periods with volatility of higher and lower magnitudes suggests a regime switching approach. Factors conditioning the changes in the volatility dynamics are not known to the full extent. Motivated by the results shown in Figure D.2, the CP-ARCH and CP-ARCHX models assume that the changes in the wind power densities depend on the level of the predicted power. However, this assumption is just a simplification of the underlying complex process. As an alternative Markov Switching models allow a sequence of the regimes to be governed by a hidden, unknown signal, instead of being directly explained by some observable process. The estimation of the transitions between the regimes is carried out in a probabilistic way. Such types of models have been successfully implemented for describing the volatile behaviour of the offshore wind power generation in [13, 14].

Following [15] this can be written as following . Let ϵ_t^2 , $t = 1, \dots, n$ be the time series of the squared deviations from a wind power point forecast. In parallel, consider c_t a regime sequence taking a finite number of discrete values,

$c_t \in \{1, \dots, m\}, \forall t$. In this work $m = 2$. It is assumed that ϵ_t^2 is an AutoRegressive process (The corresponding Markov Switching model is then denoted by MS-ARCH) governed by the regime sequence in the following way:

$$\epsilon_t^2 = \alpha_0^{(c_t)} + \sum_{j=1}^p \alpha_j^{(c_t)} \epsilon_{t-j}^2 + e_t^{(c_t)} \quad (\text{D.19})$$

or an AutoRegressive process with an eXogenous input (The corresponding Markov Switching model is then denoted by MS-ARCHX) described by:

$$\epsilon_t^2 = \alpha_0^{(c_t)} + \sum_{j=1}^p \alpha_j^{(c_t)} \epsilon_{t-j}^2 + \sum_{j=1}^l \gamma_j^{(c_t)} \xi_{t-j}^2 + e_t^{(c_t)} \quad (\text{D.20})$$

where analogically to model (D.16), ξ_t^2 denotes squared errors of the WPPT forecast at the remote location. For the prediction horizons up to 1 hour squared deviations at Kappel wind farm are considered. For the larger horizons ξ_t^2 is given by the squared errors at Horns Rev I (farm number 18). $e_t^{(c)}$ is a Gaussian white noise process in regime c , i.e. a sequence of independent random variables with the density function η :

$$\eta^{(c)}(e) = \frac{1}{\sigma_t^{(c)} \sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(\frac{e}{\sigma_t^{(c)}}\right)^2\right) \quad (\text{D.21})$$

In addition it is assumed that the regime sequence c_t follows a first order Markov chain on the finite space $1, \dots, m$:

$$p_{ij} = P(c_t = j | c_{t-1} = i, c_{t-2}, \dots, c_0) = P(c_t = j | c_{t-1} = i) \quad (\text{D.22})$$

The probabilities governing the switches between the regimes are gathered in the transition probability matrix Γ :

$$\mathbf{\Gamma} = \begin{pmatrix} p_{11} & p_{12} & \dots & p_{1m} \\ p_{21} & p_{22} & \dots & p_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ p_{m1} & p_{m2} & \dots & p_{mm} \end{pmatrix} \quad (\text{D.23})$$

where $p_{ij} \geq 0, \forall i, j$ and $\sum_{j=1}^m p_{ij} = 1, \forall i$.

A likelihood L of the Markov Switching model (D.19) is then given by:

$$L_T = \boldsymbol{\delta} \mathbf{P}(e_1) \boldsymbol{\Gamma} \mathbf{P}(e_2) \dots \boldsymbol{\Gamma} \mathbf{P}(e_T) \mathbf{1}' \quad (\text{D.24})$$

where $\mathbf{P}(e_j) = \begin{pmatrix} \eta^{(1)}(e_j) & & 0 \\ & \ddots & \\ 0 & & \eta^{(m)}(e_j) \end{pmatrix}$

and $\boldsymbol{\delta}$ is an initial distribution which is assumed to correspond to the stationary process distribution found from:

$$\boldsymbol{\delta} (\mathbf{I}_m - \boldsymbol{\Gamma} + \mathbf{U}) = \mathbf{1} \quad (\text{D.25})$$

where $\mathbf{1}$ is a row vector of ones, \mathbf{I}_m is an identity matrix of order m and \mathbf{U} is an $m \times m$ matrix of ones.

The set of coefficients allowing to fully characterize the Markov switching model (D.19) is given by $\boldsymbol{\Theta} = [\boldsymbol{\Gamma}, \alpha_0^{(1)}, \dots, \alpha_p^{(1)}, \dots, \alpha_0^{(m)}, \dots, \alpha_p^{(m)}, \sigma^{(1)}, \dots, \sigma^{(m)}]$. Estimation of the model parameters is performed by numerically maximizing the likelihood function with respect to $\boldsymbol{\Theta}$. In order to avoid dealing with the constraints in the optimization routine, a re-parametrization of $\boldsymbol{\Gamma}$ and $\sigma^{(1)}, \dots, \sigma^{(m)}$ is carried out as proposed in [15].

In this work the parameters of the MS-AR model are estimated in a non-adaptive way. We have tried to employ an adaptive estimation scheme (using a sliding window of 5000 observations) and the results were shown not to improve. It was therefore decided to stick to the non-adaptive estimation procedure for the sake of simplicity.

D.4 Non-parametric approach to probabilistic forecasting

A non-parametric approach for estimating the predictive densities of the wind power generation does not assume a specific, known distribution for the data. Instead, it suggests estimating quantiles of the data directly from the previous observations. Adaptive quantile regression offers a way of estimating quantiles in the slowly varying non-stationary systems using linear regression techniques.

Quantile regression as presented in [16] is based on a linear model:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{r} \quad (\text{D.26})$$

where \mathbf{y} denotes a vector of realizations of the random variable Y , \mathbf{X} is a design matrix, i.e. a matrix containing explanatory variables, \mathbf{r} denotes a vector of the error terms and $\boldsymbol{\beta}$ is a parameter vector to be estimated from the data. Commonly, in case of an ordinary linear regression, a quadratic loss function is applied for obtaining the parameter estimate $\hat{\boldsymbol{\beta}}$. The resulting estimate $\hat{y}_t = \mathbf{x}_t \hat{\boldsymbol{\beta}}$ corresponds to the conditional mean, i.e. the expected value of the random variable Y . Instead of being quadratic, a loss function $Q_\tau(r)$ for the quantile regression is based on a weighting of the absolute values of the residuals:

$$Q_\tau(r) = \begin{cases} \tau r & \text{if } r \geq 0 \\ (\tau - 1)r & \text{if } r < 0 \end{cases}$$

Given N observations, the best estimate of $\boldsymbol{\beta}$ is:

$$\hat{\boldsymbol{\beta}}(\tau) = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \sum_{t=1}^N Q_\tau(r_t) \quad (\text{D.27})$$

The resulting estimate of y_t given by $\hat{y}_t = \mathbf{x}_t \hat{\boldsymbol{\beta}}(\tau)$ is then equal to the conditional τ^{th} quantile of the random variable Y , i.e. the value which is not exceeded by an observation with a nominal proportion (probability) of τ . Solutions for (D.27) are obtained using linear programming techniques.

A time-adaptive parameter estimation method allows to update the estimate of the parameters as the new observations become available. The method is introduced and described in details in [5]. The main idea behind it is to update the design matrix \mathbf{X} by leaving out the oldest observations as a new data becomes available. The parameters are then re-estimated based on the updated version of the design matrix and using the previously estimated parameter value for initialization of the new optimization procedure (D.27). In this work we follow the routine as described in [5], but skip the idea of using the

old parameter value as an initial step in the updating procedure (i.e. we rather follow the exact approach of [17]). This is done for the sake of implementation simplicity and due to the fact that the quality of the results is not affected by omitting the aforementioned step.

Application to the case study The quantile models studied here are given by

$$\epsilon_t = g(\hat{p}_t, \tau) + r_t = \beta_{0,t}(\tau) + f_t(\hat{p}_t, \tau) + r_t = \beta_{0,t}(\tau) + \sum_{j=1}^{K-1} b_j(\hat{p}_t) \beta_{j,t} + r_t \quad (\text{D.28})$$

where ϵ_t is an error made by the point forecast, \hat{p}_t is the forecasted power, τ is the required quantile, r_t is a noise term. The b_j are natural cubic B-spline basis functions, K is the number of knots used for the spline construction. In other words, the model assumes that the τ^{th} quantile of the forecast error ϵ_t is a non-linear function g of the predicted power. Function g is viewed as a piecewise cubic function of \hat{p}_t and can therefore be expressed as a linear combination of the known basis functions b_j . This representation permits to use the estimation techniques valid for the linear model given by (D.26).

Some technical details: the model examined in this work has the knots for the spline basis functions placed in steps of 25% quantiles of the forecasted wind power. The design matrix contains 2100 recent observation. It is updated and the model parameters are re-estimated daily at 00:00. Considered quantiles $\tau \in T = \{0.05, 0.10, \dots, 0.95\}$

The intuitive explanation of model (D.28) is the following. Preliminary data analysis given in Figure D.2 showed that the density of the wind power generation depends on the level of the predicted power. Model (D.28) permits to estimate in an optimal way the set of quantiles T of the point forecast prediction error as a non-linear function of the expected wind power. Knowing the quantiles, it is possible to "dress" the point forecasts of the wind power with a set of prediction intervals of different nominal coverage rates and obtain a full predictive density. For instance, knowing 0.05 and 0.95 quantiles of the ϵ_t (denoted as $\epsilon_{0.05,t}$ and $\epsilon_{0.95,t}$, respectively) it is possible to say that an actually observed power generation at time t will with 90% certainty be covered by the interval $[\hat{p}_t + \epsilon_{0.05,t}, \hat{p}_t + \epsilon_{0.95,t}]$.

D.5 Methods for probabilistic forecasts assessment

Forecast quality relates to the degree of correspondence between forecasts and observations. When evaluating such similarity there are many aspects to be considered. The most common approach is to compute a measure of the overall correspondence between predictions and outcomes. For instance, the accuracy of the forecasts is given by a measure of the average correspondence between individual pairs of forecasts and observations. In this work the accuracy is evaluated using a Conditional Ranked Probability Score (it is defined and discussed below). Computation of measures of such overall correspondence is very helpful and useful when several probabilistic forecasts are to be compared. However, the drawback is that it shuffles different aspects of forecast quality, weights each of them in a certain way and provides a corresponding summarizing value. In order to truly understand how the forecast operates in different situations and how consistent are the issued predictive densities with the observed proportions the overall score has to be decomposed and different aspects have to be assessed separately. In this work additionally to the overall accuracy and skill of the forecasts we will focus on such aspects as calibration, sharpness and conditional evaluation of the predictive densities.

Reliability (calibration) refers to the statistical consistency between the distributional forecasts and the observations: it reflects how close the nominal and the empirical proportions are. Over an evaluation set of a significant size, the observed and the predicted coverage rates should be as close as possible. Calibration is viewed as a very important quality for the probabilistic forecasts. However, alone it does not make for a useful forecast [18]. For example a climatology forecast (forecast based on the overall histogram of the historical data, not based upon the dynamic implications of the current state of the system) constitutes a reliable forecast. This forecast though is constant, not flexible enough and simply unable to satisfactorily represent possible outcomes in the current situation. Due to that, besides the wish for the forecasts to be calibrated, they are also desired to provide forecast users with a situation-dependent informative predictions. This is closely related to the quality of sharpness.

Sharpness refers to the degree of concentration of the distribution of the probabilistic forecast. If the density forecast takes the form of a delta function, this would correspond to a maximum possible sharpness and would equate with the idealized concept of the perfect point forecast. In contrast, the climatology forecast is generally not sharp as there is a probability that the future observa-

tion can take on any value that has been observed in the past.

Conditional evaluation concerns assessment of the conditional distributions. As has been previously discussed the predictive densities of the wind power generation are not constant, - they change with the level of expected power. Thus is it interesting to check how well the issued probabilistic forecasts perform in the moments when the power generation is high, low or medium.

All three aforementioned characteristics (reliability, sharpness and quality of the conditional densities) are very important for making a good probabilistic forecast and it is a combination of them which plays the most important role. One forecast might be reliable, but lacking sharpness. Another one might be sharp, but lacking calibration. Which one is better? The answer is commonly given by some proper scoring rules.

Scoring rules assess the quality of the predictive density by assigning a numerical score based on the forecast and on the event or value which materializes. They assess different quality aspects simultaneously providing a single numerical value summarizing the quality of the forecast performance. Following [19] let $S(F, x)$ denote a score assigned when a forecaster issues the predictive density F and x is the value which actually materializes. We consider scores as penalties which the forecaster wishes to minimize on average. A scoring rule is proper if the expected value of the penalty $S(F, x)$ for an observation x drawn from G is minimized if $F = G$, i.e. if the forecast is perfect. In other words, the smaller the obtained score is, the better the issued forecast is (the closer it is to the perfect forecast). The score is called strictly proper if the minimum is unique. In estimation problems proper scoring rules encourage a forecaster to do careful assessments and to be honest. Every attempt to speculate on, for instance, sharpness with the price of losing in calibration (or vice versa) is being penalized and is correspondingly reflected in the value of the obtained score.

Conditional Ranked Probability Score (crps) has gained popularity as a means of evaluating wind power probabilistic forecasts. It is defined as:

$$crps(F, x) = - \int_{-\infty}^{\infty} (F(y) - \mathbf{1}\{y \geq x\})^2 dy \quad (\text{D.29})$$

where $F(y)$ is a predictive cumulative distribution function, x is a value which materialized (an actual observation), $\mathbf{1}\{y \geq x\}$ denotes a function that gives the value 1 if $y \geq x$ and 0 otherwise.

For assessing a probabilistic forecast over a data set containing T observations the average of the crps values for each forecast/verification pair is calculated.

$$CRPS = \frac{1}{T} \sum_{t=1}^T crps(F_t, x_t) \quad (\text{D.30})$$

The choice of using CRPS is motivated by the facts, that firstly, it is a proper scoring rule [20]. Secondly, it is a distance sensitive rule, meaning that a credit is given for assigning high probabilities to values near, but not identical to the one materializing. Another useful property of the CRPS score arises from the fact that for point forecasts it reduces to the absolute error. Thus CRPS provides a direct way to compare point and probabilistic forecasts.

Skill score For comparing forecast models it is convenient to introduce a measure of the relative improvement in CRPS with respect to the considered reference forecast model. Such improvements are given by the corresponding skill score δ which is defines as:

$$\delta = \frac{CRPS_{ref} - CRPS_m}{CRPS_{ref}} 100\% \quad (\text{D.31})$$

where $CRPS_m$ corresponds to the CRPS score of the considered forecast model and $CRPS_{ref}$ is the CRPS score of the reference model.

D.6 Results

D.6.1 Notation

Before discussing the results the following notation is introduced: all the considered models are noted as "Expectation Model/Volatility Model" where in front of "/" stands a type of the point predictions used as input and after the "/" the used uncertainty estimation method is introduced. Following that:

Non-parametric models ("/QR"):

- "Model"/QR stands for the non-parametric predictive densities based on the adaptive quantile regression. "Model" specifies the type of point predictions used as input. (WPPT/QR, CP/QR, Logit-CP/QR).

Parametric models under the assumption of the censored Normal distribution ("CN:") :

- CN:WPPT/Exp.smooth - A parametric model under the assumption of the censored Normal distribution. The WPPT point predictions are used as input. The scale parameter of the suggested distribution is estimated using the exponential smoothing technique (D.13).
- CN:CP/Exp.smooth - A parametric model under the assumption of the censored Normal distribution. The CP point predictions are used as input. The scale parameter of the suggested distribution is estimated using the exponential smoothing technique (D.13).
- CN:CP/GARCH - A parametric model under the assumption of the censored Normal distribution. The CP point predictions are used as input. The scale parameter of the suggested error distribution is estimated using the GARCH approach (D.14).
- CN:CP/GARCHX - A parametric model under the assumption of the censored Normal distribution. The CP point predictions are used as input. The scale parameter of the suggested error distribution is estimated using the GARCH approach (D.16).
- CN:CP/CP-ARCHX - A parametric model under the assumption of the censored Normal distribution. The CP point predictions are used as input.

The scale parameter of the suggested error distribution is estimated using the CP-ARCHX approach (D.18).

- CN:CP/CP-ARCH - A parametric model under the assumption of the censored Normal distribution. The CP point predictions are used as input. The scale parameter of the suggested error distribution is estimated using the CP-ARCH approach (D.17).
- CN:Logit-CP/CP-ARCH - A parametric model under the assumption of the censored Normal distribution. The Logit-CP point predictions are used as input. The scale parameter of the suggested error distribution is estimated using the CP-ARCH approach (D.17).
- CN:CP/MS-AR - A parametric model under the assumption of the censored Normal distribution. The CP point predictions are used as input. The scale parameter of the suggested error distribution is estimated using the MS-AR approach (D.19).
- CN:CP/MS-ARX - A parametric model under the assumption of the censored Normal distribution. The CP point predictions are used as input. The scale parameter of the suggested error distribution is estimated using the MS-ARX approach (D.20).

Parametric models under the assumption of the generalized logit-Normal distribution ("GLN:") :

- GLN:Logit-CP/Exp.smooth - A parametric model under the assumption of the generalized logit-Normal distribution. The Logit-CP point predictions are used as input. The scale parameter of the suggested error distribution is estimated using the exponential smoothing technique (D.13).
- GLN:Logit-CP/GARCHX - A parametric model under the assumption of the generalized logit-Normal distribution. The Logit-CP point predictions are used as input. The scale parameter of the suggested error distribution is estimated using the GARCHX approach (D.14).
- GLN:Logit-CP/CP-ARCHX - A parametric model under the assumption of the generalized logit-Normal distribution. The Logit-CP point predictions are used as input. The scale parameter of the suggested error distribution is estimated using the CP-ARCHX approach (D.18).
- GLN:Logit-CP/MS-AR - A parametric model under the assumption of the generalized logit-Normal distribution. The Logit-CP point predictions are used as input. The scale parameter of the suggested error distribution is estimated using the MS-AR approach (D.19).

- GLN:Logit-CP/MS-ARX - A parametric model under the assumption of the generalized logit-Normal distribution. The Logit-CP point predictions are used as input. The scale parameter of the suggested error distribution is estimated using the MS-ARX approach (D.20).

D.6.2 Accuracy and skill assessment

Different parametric and non-parametric models have been used for issuing probabilistic forecasts for Nysted Offshore. The CRPS scores for all the models have been evaluated over a period from the 1st of January, 2009 to the 31st of December, 2009. Only the data points where none of the considered models have a missing value were used in the evaluation set. This resulted in approximately 25 000 (≈ 8.5 months) active data points for every of the considered prediction horizons. The CRPS results for all the considered models run with the different prediction horizons are presented in Table D.2.

As a benchmark it is chosen to use the adaptive quantile regression using the WPPT forecast as input. This is a very strong benchmark as both the WPPT and the adaptive quantile regression are the state-of-the-art approaches. Other models are compared to the benchmark model and the relative improvements δ are calculated. The results as they are presented in Table D.2 may look a bit overwhelming - too many of them to be able to note all the details. In order to make it easier to spot the main characteristic, in the following paragraphs several snapshots of the overall results are taken and discussed in more details.

D.6.2.1 Parametric predictive densities

Estimating the first order moments

Spatial correction of the first order moments One way for describing a random variable in a probabilistic framework is to estimate all order moments (mean, variance, skewness, kurtosis, higher order moments). In case of the parametric approach, i.e. when the data is assumed to follow a known distribution, a finite number of moments will fully characterize the variable. The parametric densities considered in this work can be fully characterized by the first two order moments: the mean and the variance (see Section D.3.1 for details).

This implies that using those two moments as input we can fully recreate the predictive density. The mean is given by the point predictions of the wind

Horizon	15 min		1 hour		4 hours		5 hours		6 hours		7 hours		8 hours	
Model name	CRPS, %	δ , %												
Non-parametric densities														
<i>WPPT/QR</i>	<i>2.34</i>		<i>5.17</i>		<i>7.69</i>		<i>7.97</i>		<i>8.19</i>		<i>8.34</i>		<i>8.47</i>	
CP/QR	2.31	1.5	4.94	4.6	7.43	3.5	7.73	2.9	8.00	2.3	8.13	2.5	8.29	2.1
Logit-CP/QR	2.31	1.5	4.92	4.9	7.42	3.5	7.74	2.8	8.02	2.1	8.15	2.3	8.31	1.9
Censored Normal predictive densities														
CN:WPPT/Exp. smooth.	2.49		5.40		7.89		8.17		8.39		8.54		8.67	
CN:CP/Exp. smooth.	2.46	-5.2	5.14	0.6	7.63	0.8	7.94	0.3	8.21	-0.2	8.37	-0.4	8.53	-0.7
CN:CP/GARCH	2.37	-1.4	5.06	2.2	7.60	1.2	7.92	0.6	8.06	1.6	8.36	-0.2	8.52	-0.5
CN:CP/GARCHX	2.35	-0.5	5.03	2.7	7.60	1.3	7.92	0.6	8.20	-0.1	8.35	-0.1	8.52	-0.5
CN:CP/CP-ARCH	2.32	1.0	4.96	4.1	7.46	3.0	7.77	2.4	8.06	1.6	8.23	1.4	8.39	1.0
CN:CP/CP-ARCHX	2.31	1.2	4.95	4.4	7.46	3.0	7.78	2.4	8.07	1.5	8.24	1.3	8.40	0.8
CN:Logit-CP/CP-ARCH	2.31	1.2	4.93	4.7	7.42	3.5	7.75	2.7	8.03	1.9	8.19	1.8	8.36	1.4
CN:CP/MS-AR	2.37	-1.3	5.06	2.2	7.62	0.9	8.31	-4.4	8.21	-0.2	8.41	-0.8	8.57	-1.1
CN:CP/MS-ARX	2.37	-1.1	5.05	2.4	7.61	1.1	7.97	0.0	8.21	-0.3	8.37	-0.4	8.55	-0.9
Generalized logit-Normal predictive densities														
GLN:Logit-CP/Exp. smooth	2.43	-3.6	5.09	1.7	7.64	0.7	7.96	0.1	8.23	-0.5	8.38	-0.5	8.55	-0.9
GLN:Logit-CP/GARCH	2.33	0.6	4.95	4.2	7.59	1.4	7.93	0.4	8.21	-0.2	8.36	-0.3	8.54	-0.7
GLN:Logit-CP/CP-ARCH	2.31	1.5	4.94	4.6	7.54	1.9	7.87	1.3	8.15	0.5	8.32	0.3	8.50	-0.3
GLN:Logit-CP/MS-AR	2.35	-0.2	4.98	3.8	7.65	0.5	8.01	-0.5	8.26	-0.8	8.42	-1.0	8.63	-1.8
GLN:Logit-CP/MS-ARX	2.34	-0.2	4.97	3.9	7.63	0.9	7.99	-0.3	8.23	-0.5	8.39	-0.6	8.58	-1.3

Table D.2: Evaluation of the density forecasts with a CRPS criterion. An adaptive quantile regression over the initial WPPT point forecasts is considered a benchmark(given in italic). Other models are compared to the benchmark model and the corresponding relative improvements are given by δ .

Horizon	15 min		1 hour		4 hours		5 hours		6 hours		7 hours		8 hours	
	CRPS, %	δ , %												
Censored Normal predictive densities														
CN:WPPT/Exp. smooth.	2.49		5.40	0.6	7.89	0.8	8.17	0.3	8.39	-0.2	8.54	-0.4	8.67	-0.7
CN:CP/Exp. smooth.	2.46	-5.2	5.14	0.6	7.63	0.8	7.94	0.3	8.21	-0.2	8.37	-0.4	8.53	-0.7

Table D.3: Evaluation of the density forecasts with a CRPS criterion. Comparison of the Censored Normal predictive densities when using the WPPT and the spatially corrected CP forecasts as input. δ shows the corresponding relative improvements over the WPPT/QR

power generation. The results given in Table D.3 show that the quality of the parametric predictive densities is improved if instead of the original WPPT forecasts the spatially corrected point predictions given by the CP model are considered. This indicates that accounting for the spatio-temporal effects while estimating the first order moment of the predictive density improves the quality of the corresponding probabilistic forecasts.

Incorporating the generalized logit transformation in to the input correction model In [8] it is shown that the quality of the point predictions is further improved if the generalized logit data transformation is incorporated into the spatial correction models (resulting in the Logit-CP point predictions). The results given in Table D.4 show that using the Logit-CP predictions instead of the CP ones as input improves the quality of the corresponding predictive densities. In other words, considering the data transformation when estimating the expectation of the Censored Normal distribution brings slight improvements in all the considered horizons.

Horizon	15 min		1 hour		4 hours		5 hours		6 hours		7 hours		8 hours	
	CRPS, %	δ , %												
Non-parametric densities														
WPPT/QR	2.34		5.17		7.69		7.97		8.19		8.34		8.47	
Censored Normal predictive densities														
CN:CP/CP-ARCH	2.32	1.0	4.96	4.1	7.46	3.0	7.77	2.4	8.06	1.6	8.23	1.4	8.39	1.0
CN:Logit-CP/CP-ARCH	2.31	1.2	4.93	4.7	7.42	3.5	7.75	2.7	8.03	1.9	8.19	1.8	8.36	1.4

Table D.4: Evaluation of the density forecasts with a CRPS criterion. Comparison of the Censored Normal predictive densities using the CP and the Logit-CP forecasts as input. δ shows the corresponding relative improvements over the WPPT/QR

Estimating the second order moments

Accounting for the dynamics in variance of the wind power generation

As has been discussed in Section D.3 the variance of the wind power generation is not constant. Accounting for this improves the results significantly. This is demonstrated in Table D.5 where several different approaches for estimating the variance are compared.

Horizon	15 min		1 hour		4 hours		5 hours		6 hours		7 hours		8 hours	
Model name	CRPS, %	δ , %												
Non-parametric densities														
<i>WPPT/QR</i>	<i>2.34</i>		<i>5.17</i>		<i>7.69</i>		<i>7.97</i>		<i>8.19</i>		<i>8.34</i>		<i>8.47</i>	
Censored Normal predictive densities														
CN:CP/Exp. smooth.	2.46	-5.2	5.14	0.6	7.63	0.8	7.94	0.3	8.21	-0.2	8.37	-0.4	8.53	-0.7
CN:CP/GARCH	2.37	-1.4	5.06	2.2	7.60	1.2	7.92	0.6	8.06	1.6	8.36	-0.2	8.52	-0.5
CN:CP/CP-ARCH	2.32	1.0	4.96	4.1	7.46	3.0	7.77	2.4	8.06	1.6	8.23	1.4	8.39	1.0

Table D.5: Evaluation of the density forecasts with a CRPS criterion. Comparison of the Censored Normal predictive densities based on different techniques to model the dynamics of the variance. δ shows the corresponding relative improvements over the WPPT/QR

Exponential smoothing technique assumes a locally constant variance, allowing for the long-term variations, only. The GARCH model accounts for the changes in the dynamics of the variance based on the previous (most recent) observations. This technique accounts for the changing dynamics. The main drawback of the GARCH method is that it is based purely on the past observations, meaning that the predictive performance of the model has a certain delay - i.e. the model is not capable to predict when exactly the change will occur, but once the dynamics has actually changed the model is able to notice this and adapt correspondingly. That is why the model performs best on the data exhibiting the clustering effect. The CP-ARCH model (as discussed in Section D.3.4.3) captures the changes in the dynamics driven by variations in the level of the expected power generation. This approach is motivated by the results depicted in Figure D.2 where it is shown that the variance of the wind power generation depends on the level of the expected power. Differently from the GARCH approach, the conditional parametric method allows to describe the dynamics based not only on the previously observed fluctuations, but takes into the consideration the power forecast as well. Therefore this model is better adapted to foreseeing the changes in the dynamics of the wind power generation. The results given in Table D.5 show that the CP-ARCH model provides the best input for the Censored Normal predictive density. The fact that the exponential smoothing shows the poorest results, proves that the variance of the wind power generation

is not constant and accounting for the changing dynamics is important. As the GARCH model is outperformed by the CP-ARCH model, it is possible to conclude that the changes in the dynamics of the variance can be (partly) explained by the level of the expected power.

Modelling the dynamics of the variance: explicit modelling conditional on the level of the expected power versus the implicit Markov switching approach As has been previously discussed, conditioning the density of the wind power generation on any observable input would always be just an approximation of the actual complex meteorological phenomenon. Markov switching models propose an alternative approach by modelling the dynamics driven by the unknown signal. As can be seen from the results shown in Table D.6 the explicit approach suggesting that the dynamics is driven by the level of the expected power outperforms the implicit one given by the Markov switching model.

Horizon	15 min		1 hour		4 hours		5 hours		6 hours		7 hours		8 hours	
	CRPS, %	δ , %												
Non-parametric densities														
<i>WPPT/QR</i>	<i>2.34</i>		<i>5.17</i>		<i>7.69</i>		<i>7.97</i>		<i>8.19</i>		<i>8.34</i>		<i>8.47</i>	
Censored Normal predictive densities														
CN:CP/CP-ARCH	2.32	1.0	4.96	4.1	7.46	3.0	7.77	2.4	8.06	1.6	8.23	1.4	8.39	1.0
CN:CP/MS-AR	2.37	-1.3	5.06	2.2	7.62	0.9	8.31	-4.4	8.21	-0.2	8.41	-0.8	8.57	-1.1
Generalized logit-Normal predictive densities														
GLN:Logit-CP/CP-ARCH	2.31	1.5	4.94	4.6	7.54	1.9	7.87	1.3	8.15	0.5	8.32	0.3	8.50	-0.3
GLN:Logit-CP/MS-AR	2.35	-0.2	4.98	3.8	7.65	0.5	8.01	-0.5	8.26	-0.8	8.42	-1.0	8.63	-1.8

Table D.6: Evaluation of the density forecasts with the CRPS criterion. Two methods for capturing density dynamics are compared. The explicit method suggests that it is the level of the power generation which causes a variability in the density dynamics. The implicit method assumes that the driving force is unknown and tries to implicitly capture it by the Markov switching approach. An adaptive quantile regression over the initial WPPT point forecasts is considered as benchmark(given in italic). Other models are compared to the benchmark model and the corresponding relative improvements are given by δ .

This supports the idea that the dynamics of the variance of the wind power generation are indeed strongly affected by the level of the expected power. Therefore accounting for it directly in the modelling procedure is more beneficial than trying to capture the underlying complex phenomenon by applying Markov switching methods. The latter does not seem to capture any additional information than is extracted directly from the changes in the power expectation.

Accounting for the spatial information As has been discussed, the spatial correction of the first order moments of the predictive densities improves the performance of the corresponding probabilistic forecast. As the considered parametric densities are fully described by both the mean and the variance, an interesting question is whether the spatial correction of the variance can bring additional ameliorations. The answer is given by the results shown in Table D.7. No significant improvements are achieved when considering the CP/GARCHX (CP/CP-ARCHX) model instead of the CP/GARCH (CP/CP-ARCH). This shows that no additional improvements in the model performance are achieved when including the spatial information into the variance estimation.

Horizon	15 min		1 hour		4 hours		5 hours		6 hours		7 hours		8 hours	
	CRPS, %	δ , %												
Non-parametric densities														
<i>WPPT/QR</i>	<i>2.34</i>		<i>5.17</i>		<i>7.69</i>		<i>7.97</i>		<i>8.19</i>		<i>8.34</i>		<i>8.47</i>	
Censored Normal predictive densities														
CN:CP/GARCH	2.37	-1.4	5.06	2.2	7.60	1.2	7.92	0.6	8.06	1.6	8.36	-0.2	8.52	-0.5
CN:CP/GARCHX	2.35	-0.5	5.03	2.7	7.60	1.3	7.92	0.6	8.20	-0.1	8.35	-0.1	8.52	-0.5
CN:CP/CP-ARCH	2.32	1.0	4.96	4.1	7.46	3.0	7.77	2.4	8.06	1.6	8.23	1.4	8.39	1.0
CN:CP/CP-ARCHX	2.31	1.2	4.95	4.4	7.46	3.0	7.78	2.4	8.07	1.5	8.24	1.3	8.40	0.8

Table D.7: Evaluation of the density forecasts with a CRPS criterion. Comparison of the models performance with and without accounting for the spatial effects when estimating the variance. δ shows the corresponding relative improvements over the WPPT/QR.

Comparing two type of parametric densities: Censored Normal versus the Generalized logit-Normal Previously it has been discussed that applying the generalized logit transformation on the data permits to estimate the expectation of the wind power generation in a more accurate way (see [8] for more details). This suggests that the proposed transformation is suitable for stabilizing the variance of the wind power generation (making it less dependent on the power expectation) allowing for more robust point predictions. The following question is whether the transformation also makes the data look more Gaussian, ie whether the assumption that the transformed data is Normally distributed surpasses the suggestion that the non-transformed data is Gaussian. This is analogical to wondering which of the studied parametric densities describes the data better - the Censored Normal distribution or the generalized logit-Normal one. The results given in table D.8 indicate that the censored Normal distribution shows better results. This indicates that even though the generalized logit transformation helps to stabilize the variance, makes it less dependent on the bounds, it does not make the assumption of Gaussianity more appropriate.

Horizon	15 min		1 hour		4 hours		5 hours		6 hours		7 hours		8 hours	
	CRPS, %	δ , %												
Censored Normal predictive densities														
CN:Logit-CP/CP-ARCH	2.31	1.2	4.93	4.7	7.42	3.5	7.75	2.7	8.03	1.9	8.19	1.8	8.36	1.4
Generalized logit-Normal predictive densities														
GLN:Logit-CP/CP-ARCH	2.31	1.5	4.94	4.6	7.54	1.9	7.87	1.3	8.15	0.5	8.32	0.3	8.50	-0.3

Table D.8: Evaluation of the density forecasts with a CRPS criterion. Comparison of the Censored Normal and the generalized logit-Normal distributions. δ shows the corresponding relative improvements over the WPPT/QR

D.6.2.2 Non-parametric predictive densities

spatial correction of the input Similarly to the parametric predictive densities, the performance of the non-parametric predictive densities improves if instead of the WPPT forecasts the spatially corrected point predictions are considered as input. This once again proves that the quality of the probabilistic forecasts can be improved if the spatial effects are taken into consideration.

Horizon	15 min		1 hour		4 hours		5 hours		6 hours		7 hours		8 hours	
	CRPS, %	δ , %												
Non-parametric densities														
WPPT/QR	2.34		5.17		7.69		7.97		8.19		8.34		8.47	
CP/QR	2.31	1.5	4.94	4.6	7.43	3.5	7.73	2.9	8.00	2.3	8.13	2.5	8.29	2.1
Logit-CP/QR	2.31	1.5	4.92	4.9	7.42	3.5	7.74	2.8	8.02	2.1	8.15	2.3	8.31	1.9

Table D.9: Evaluation of the non-parametric densities forecasts with a CRPS criterion.

Considering data transformation for the input correction One can see that in case of the non-parametric predictive densities, the implementation of the generalized logit transformation does not help improving the corresponding probabilistic densities.

D.6.2.3 Parametric versus the non-parametric densities

In order to compare the parametric and the non-parametric approaches, we consider the CN:Logit-CP/CP-ARCH and the CP/QR which are the best per-

forming parametric and non-parametric models, respectively. The results given in Table D.10 show that the non-parametric densities outperform the parametric ones when considering larger prediction horizons. For shorter prediction horizons both approaches show similar results.

Horizon	15 min		1 hour		4 hours		5 hours		6 hours		7 hours		8 hours	
Model name	CRPS, %	δ , %												
Non-parametric densities														
<i>WPPT/QR</i>	2.34		5.17		7.69		7.97		8.19		8.34		8.47	
CP/QR	2.31	1.5	4.94	4.6	7.43	3.5	7.73	2.9	8.00	2.3	8.13	2.5	8.29	2.1
Censored Normal predictive densities														
CN:Logit-CP/CP-ARCH	2.31	1.2	4.93	4.7	7.42	3.5	7.75	2.7	8.03	1.9	8.19	1.8	8.36	1.4

Table D.10: Evaluation of the density forecasts with a CRPS criterion. Comparison of parametric and non-parametric approaches

D.6.3 Reliability assessment

As has been previously discussed in Section D.5 one of the most important qualities for the distributional forecast is reliability. In the earlier scientific works reliability is even viewed as a requirement rather than a desired property. The recent development, however, indicates that, as has been argued in Section D.5 reliability by itself does not guarantee a useful forecast and should be only evaluated in relation with the sharpness and resolution (the joint performance evaluation is given by the score functions). However in this section the focus is on reliability, due to the particular importance given to it.

In the following analysis it is chosen to focus only on the benchmark model (WPPT/QR), the best performing non-parametric model (CP/QR) and two best performing parametric models (CN:Logit-CP/CP-ARCH) and (GLN:Logit-CP/CP-ARCH).

Figure D.6 depicts reliability diagrams for the considered models when working on 1 hour, 4 hours and 8 hours ahead predictions. One can see that the non-parametric approach (based on the adaptive quantile regression) provides calibrated forecasts. On another hand, the models based on the parametric assumptions, deviate more from the nominal proportions and therefore one can claim that the parametric approaches result in probabilistically biased forecasts. Why are those forecasts biased? Partly the lack of calibration could be explained by the fact that parametric approaches imposes certain theoretical shapes for the distribution of the wind power generation. An important part of the assumption is based on the suggestion that with a non-zero probability a

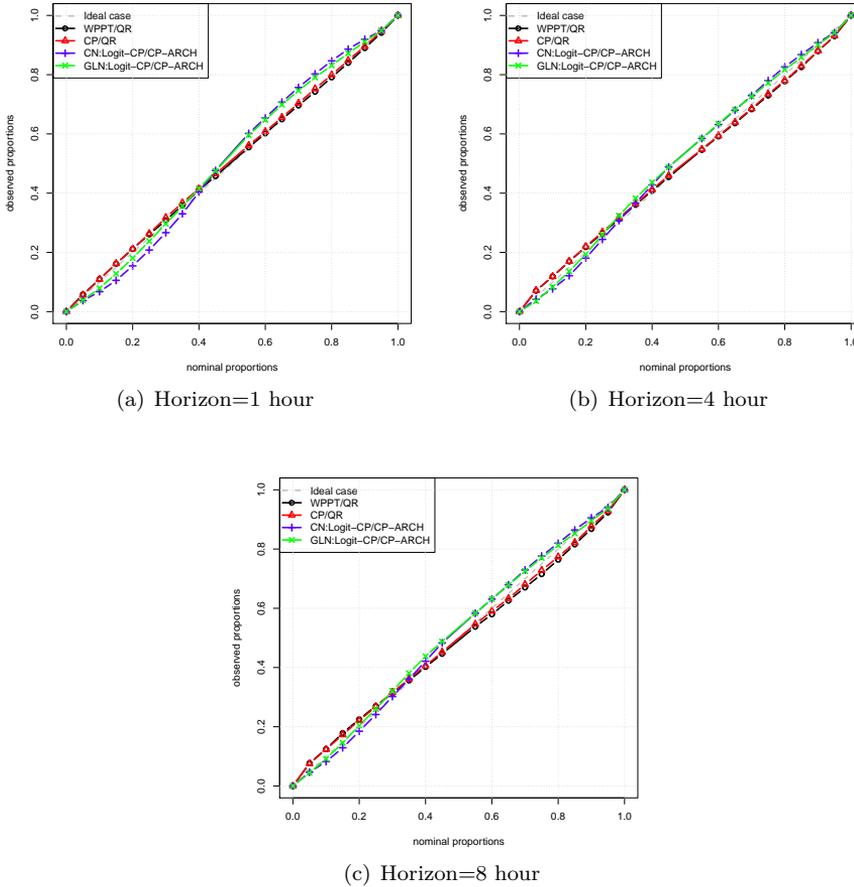


Figure D.6: Reliability diagram

measured wind power will reach 0 and 100% of the nominal capacity. In fact, in the considered data set there were no observations reaching 100%. What was actually happening when the measured wind power was approaching the nominal capacity is shown in Figure D.7. There is a clear indication that some human factor has been involved, i.e. most probably the power was down-regulated at those periods. Since such regulations are not of a constant level through the considered data set, it is difficult to replace the theoretical maximum of the nominal capacity by the adjusted value of the down-regulated power. Such deviations from the theoretical settings, where the power is allowed to reach the nominal capacity create a certain bias which is also reflected in the reliability

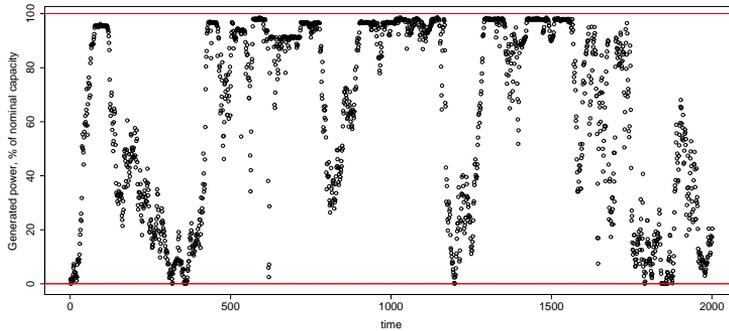


Figure D.7: A period of the observed power generation at Nysted Offshore recorded in the period from 19:45 on the 4th of November, 2008 to 15:45 on the 25th of November, 2008

diagrams of the parametric models. This could be corrected if the corresponding information on the wind farm regulation policy was available.

D.6.4 Conditional operation

In this section the focus is on the conditional evaluation of the performance of the WPPT/QR, CP/QR, CN:logit-CP/CP-ARCH and GLN:Logit-CP/CP-ARCH models. In Section D.6.2 it is shown that when evaluated on all the data available in the validation set, the CP/QR shows the best results in terms of the CRPS score. The goal of this section is to check whether this conclusion holds for various subsets of the validation set.

D.6.4.1 Accuracy as a function of time

. Firstly the validation period is divided into 5 equally populated subsets and the CRPS scores of the considered models are calculated. The results are given in Figure D.8.

One can see that the CP/QR model consistently (in all the considered subsets with different prediction horizons) outperforms the benchmark approach

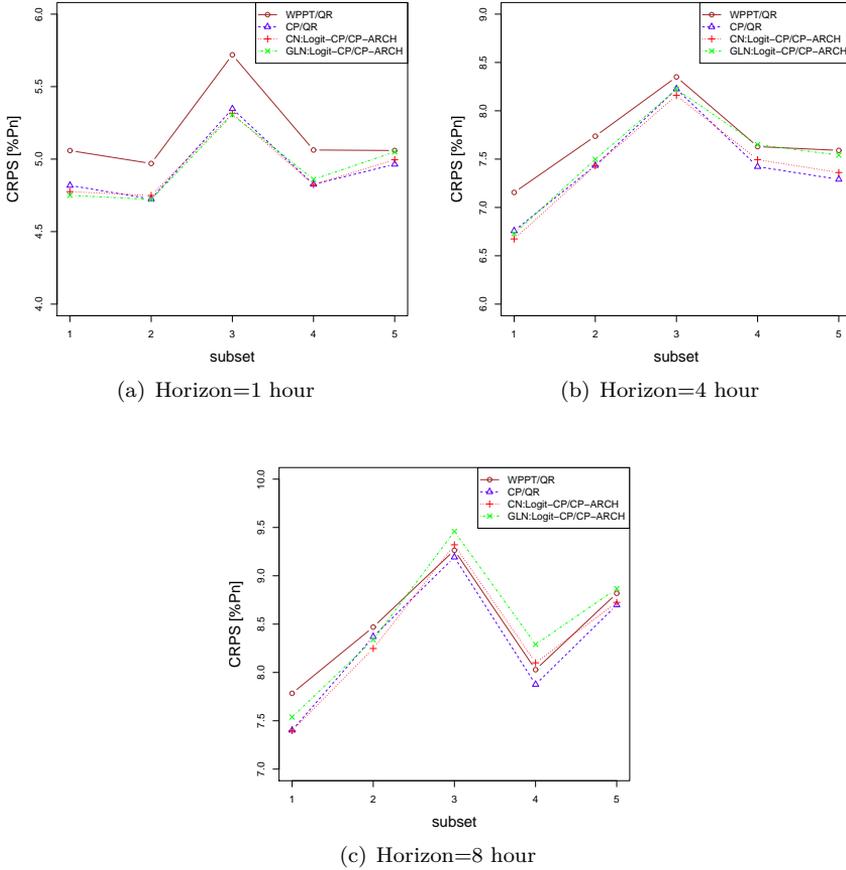


Figure D.8: CRPS evaluated over 5 equally sized subsets of the validation data set. P_n denotes the nominal capacity.

(WPPT/QR). The performance of the parametric densities with respect to the WPPT/QR is less steady. When considering 1 hour ahead forecasts, the parametric predictive densities, similarly to CP/QR outperform the WPPT/QR in all the subsets of the validation set. However, as the prediction horizon increases, the parametric models fail to outperform the benchmark approach in major part of the subsets of the validation set.

D.6.4.2 Sharpness as a function of the level of expected power generation

Recall, sharpness corresponds to the ability of probabilistic forecasts to concentrate the probabilistic information about future outcomes. In this work, following the approach by Pinson et al. [7] the sharpness is assessed by the mean widths of the central predictive intervals with a nominal coverage rate of 50%, i.e. if writing

$$\beta_{t,h} = \hat{q}(0.75)_{t|t-h} - \hat{q}(0.25)_{t|t-h} \quad (\text{D.32})$$

the size of the central interval forecast estimated at time $t - h$ for lead time t . Here $\hat{q}^{0.75}$ and $\hat{q}^{0.25}$ define the corresponding quantiles of the predictive densities. Then a measure of sharpness for horizon h is given by $\bar{\beta}_h$, the mean size of the intervals:

$$\bar{\beta}_h = \frac{1}{N} \sum_{t=1}^N \beta_{t,h} \quad (\text{D.33})$$

The results given in Figure D.9 show that with all the considered prediction horizons, the non-parametric predictive densities provide sharper forecasts than the parametric ones. This holds for all the levels of the expected power.

D.6.4.3 Accuracy and skill as functions of the level of expected power generation

As has been previously discussed in this work, the densities of the wind power generation depend on the level of the expected power. It is thus interesting to see how the models perform conditional to the level of the forecasted power generation. The results in terms of conditional accuracy given by the CRPS scores are depicted in Figure D.10 and the corresponding skill scores are shown in Figure D.11.

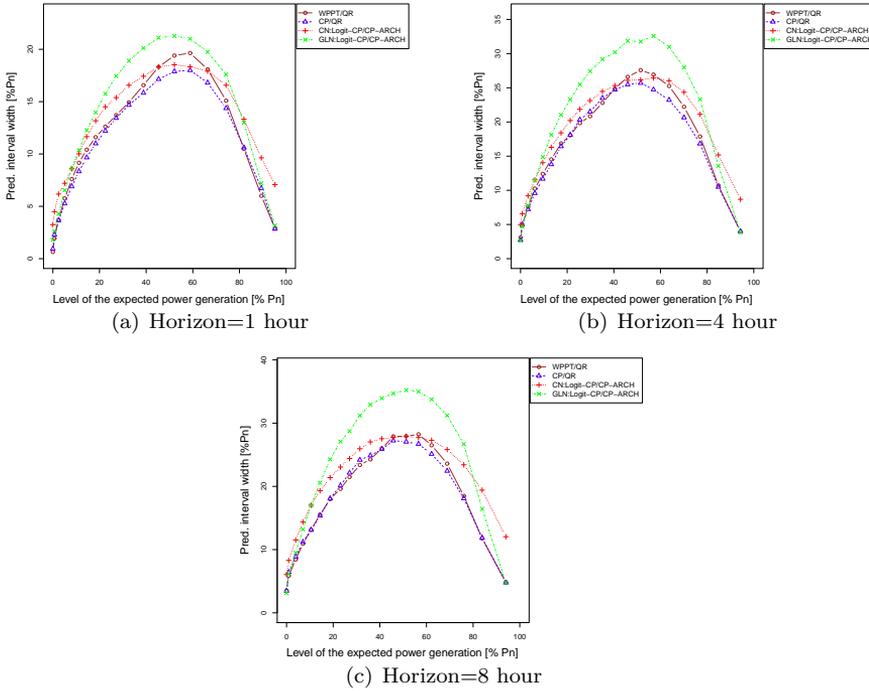


Figure D.9: Sharpness is given by the width of the prediction intervals with the nominal coverage of 50 %. It is evaluated as a function of the expected power generation level. Levels of the point predictions represent 20 equally populated classes based on the quantiles of the CP point forecasts. P_n denotes the nominal capacity.

From Figure D.10 one can see that the CRPS scores increase when the expected power is not close to the generation bounds. This leads to higher uncertainty associated with the corresponding predictive densities. As the result the increase in the CRPS scores is observed.

The analysis of the conditional skill score depicted in Figure D.11 indicates that the parametric models perform similarly to the non-parametric ones when the expected power is its medium range. However, close to the generation bounds the parametric densities perform worse than the non-parametric ones. This can be explained by the fact that the closer to the bounds, the more significant the censoring effect in the parametric densities become. This leads to higher bias in parameter estimates. The drop in operation quality of the parametric predictive densities is especially evident when the level of the expected power is close to

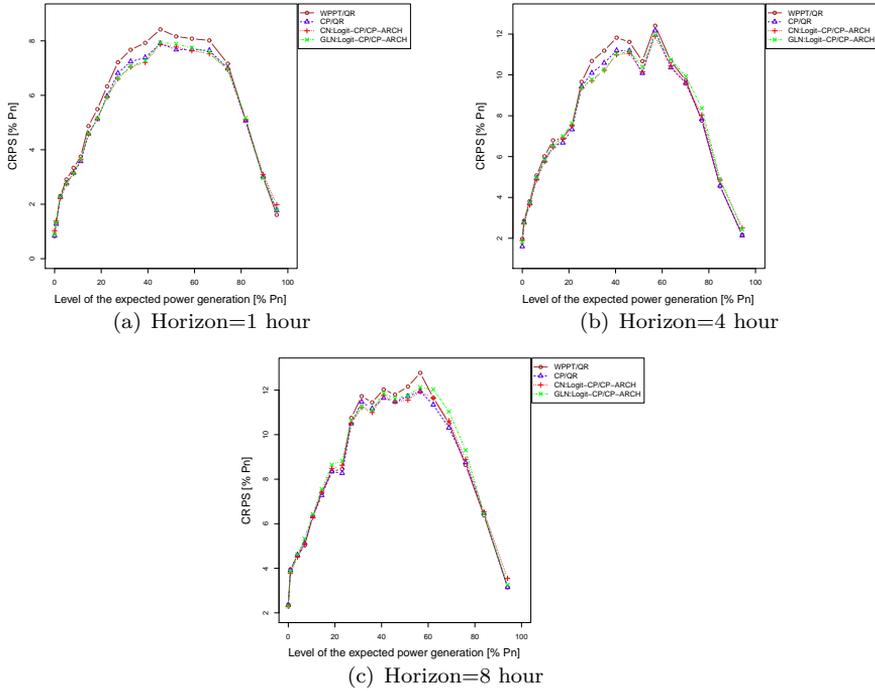


Figure D.10: CRPS evaluated as a function of the expected power generation level. Levels of the point predictions represent 20 equally populated classes based on the quantiles of the CP point forecasts. P_n denotes the nominal capacity.

the nominal capacity. This is in line with the discussion given in Section D.6.3 on the down-regulation policy.

D.6.4.4 Accuracy and skill as functions of the measured power generation

Figures D.12 and D.13 depict the accuracy and skill scores as functions of the observed power generation. From those figures one can conclude that the parametric densities managed to describe the situations when the observed power was in its medium range better than the non-parametric ones. In the situations when the power measurements fell close to the generation bounds, the non-

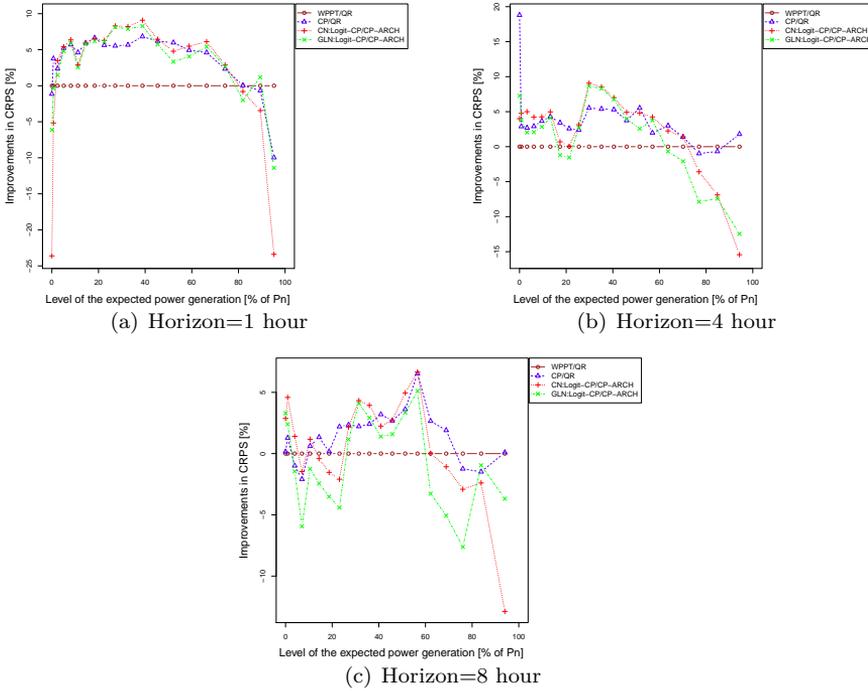


Figure D.11: Relative improvements in CRPS when compared to the benchmark model WPPT/QR. Evaluation is conditional to the expected power generation level. Levels of the point predictions represent 20 equally populated classes based on the quantiles of the CP point forecasts. P_n denotes the nominal capacity.

parametric models performed much better than the parametric densities. This once again indicates that the non-parametric densities outperform the parametric ones mainly in the situations when the observed power approaches 0 or the nominal capacity. This is the region where the parametric densities suffer the most from the approximations used in parameter estimation methods and the power regulation policies.

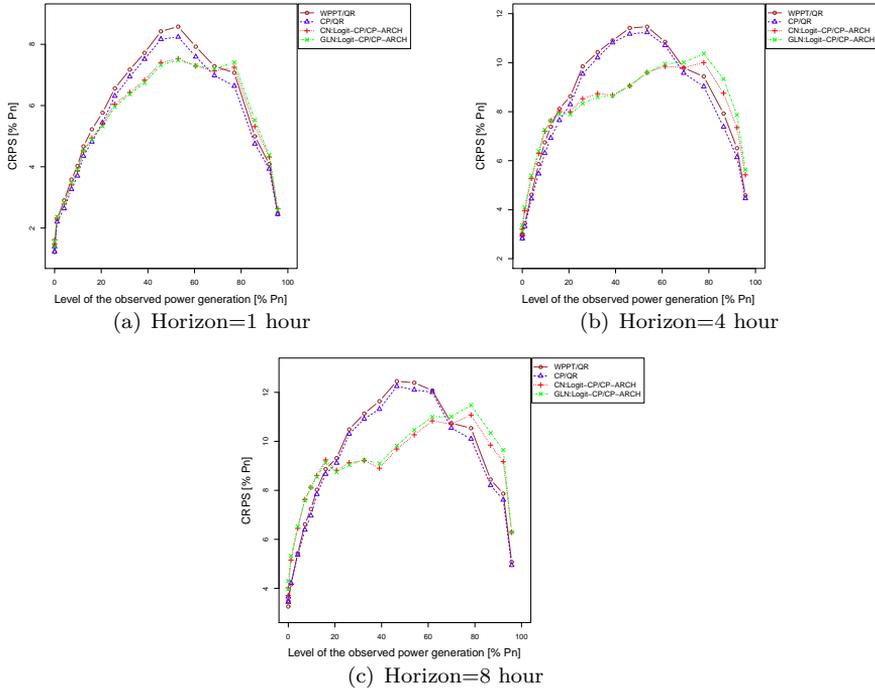
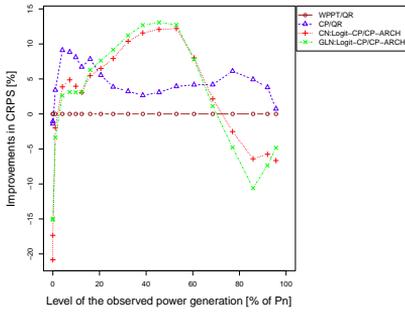


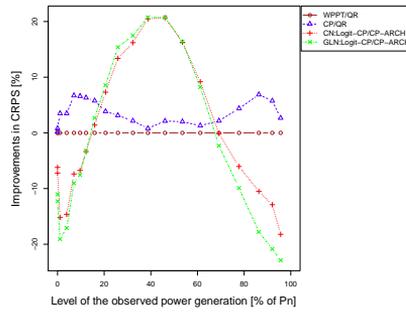
Figure D.12: CRPS evaluated as a function of the power measurements. Levels of the power measurements represent 20 equally populated classes based on the quantiles of the CP point forecasts. P_n denotes the nominal capacity.

D.6.5 Demonstration of the operation of the best-performing model

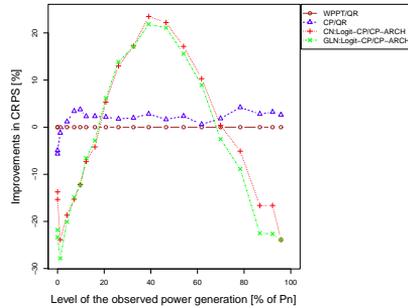
Figure D.14 gives an example of probabilistic predictions given by the CP/QR and the benchmark approach. In general one can see that the difference in the distributions seems rather subtle and mainly related to the expectation (mass center) of the predictive densities. Figure D.15 demonstrates an example of how the CP/QR model could be used in practice- at a given time probabilistic predictions could be issued for up to agreed amount of hours ahead. In this work we consider the predictions up to 8 hours ahead.



(a) Horizon=1 hour



(b) Horizon=4 hour



(c) Horizon=8 hour

Figure D.13: Relative improvements in CRPS when compared to the benchmark model WPPT/QR. Evaluation is conditional to the expected power generation level. Levels of the power measurements represent 20 equally populated classes based on the quantiles of power observations. P_n denotes the nominal capacity.

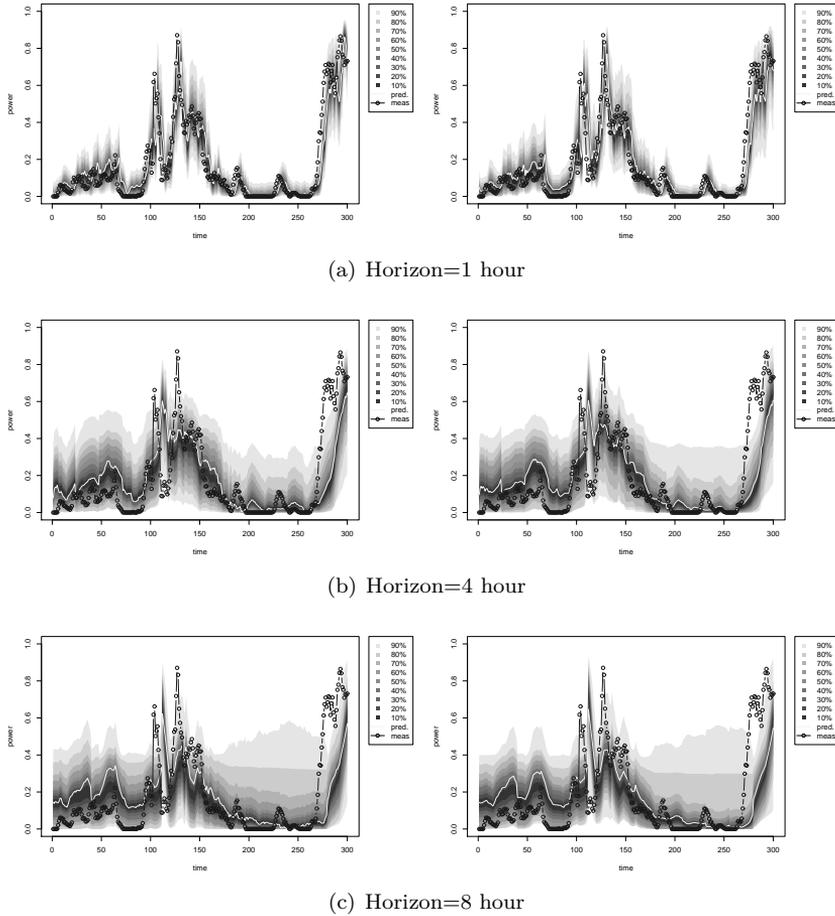


Figure D.14: Example of the predictive densities for the wind power generation at Nysted Offshore given by the CP/QR (left) and the WPPT/QR (right) models. Considered period is from 10:15 on the 27th of July, 2009 to 13:45 on the 30th of July, 2009 (corresponding to 300 time steps of 15 min). Graphically the difference in predictive densities seems subtle and mainly related to the difference in the means of the predictive densities.

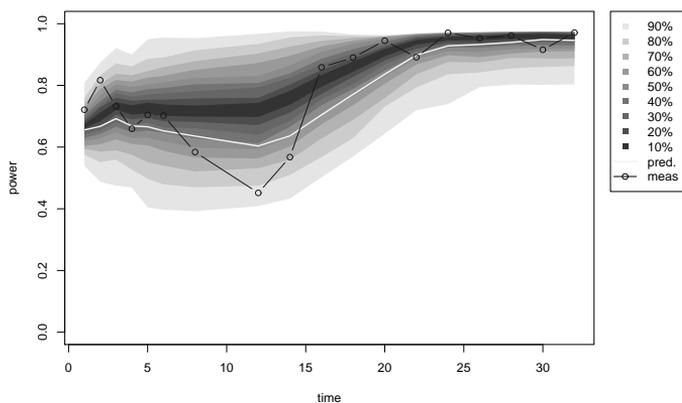


Figure D.15: Example of probabilistic forecasts of wind power generation obtained with the CP/QR model. Predictive density is given by the prediction intervals of the different nominal coverage rates. Power values are standardized by the nominal capacity of the wind farm. Predictions for different lead times (from 15 min to 8 hours ahead) are issued at 17:15 on the 24th of November, 2009. A solid line shows the point predictions.

D.7 Conclusions

Focus has been given to probabilistic forecasts for Nysted Offshore. Both parametric and non-parametric methods for building the predictive densities have been considered. Following the parametric approach two different densities have been proposed for modelling the wind power generation: the Censored Normal and the generalized logit-Normal. The non-parametric approach was based on the time-adaptive quantile regression. The results show that using the spatially corrected point forecasts instead of the original WPPT predictions as input to building the predictive densities improves the performance of the models in both parametric and non-parametric approaches. It translates to saying that the spatial correction of the first order moments improves the quality of the corresponding predictive densities. The spatial correction of the higher order moments was shown not to ameliorate the quality of the predictive densities any further.

It has been shown in the work that the densities of the wind power generation are not constant. Two different methods for modelling the dynamics of the predictive densities have been proposed. Firstly, it has been noted that the distribution of the wind power generation is related to the level of the expected power. Capturing this effect was shown to improve the performance of the corresponding forecasts. In the second place, it has been decided to model the changes in the predictive densities implicitly, i.e. assume that the dynamics of the system is governed by some unobservable process rather than by any particular explanatory variable. This was implemented using the Markov-switching models. It has been shown that the direct approach with conditioning the density dynamics of the level of the expected power outperforms the Markov switching models. This indicates that the dynamics of the predictive densities can be rather well explained by the changes in the expected level of the power generation. Thus accounting for it directly in the modelling procedure is more beneficial than trying to capture the underlying complex phenomenon in a probabilistic framework.

In this work two types of the proposed parametric densities have been compared. It is shown that the censored Normal distribution describes the data better than the generalized logit-Normal. This holds for all the considered prediction horizons.

The parametric and the non-parametric probabilistic forecasts have also been compared. It is shown that both approaches perform similarly (in terms of the average accuracy) in the short prediction horizons (up to 5-6 hours ahead). In the longer horizons the difference in performance becomes more significant with the quantile-regression based models taking the leading position. Even in the

shorter horizons, even though the two approaches show similar results in terms of accuracy, the corresponding densities are different. The parametric densities are shown to be less sharp than the non-parametric ones. The analysis of the situation-based performance of the two approaches has been carried out. It is shown that the parametric densities outperform the non-parametric ones in the periods when the power generation is in its medium range. Closer to the bounds, the operation quality of the parametric models is poor compared to the quantile regression based approach.

Summarizing, based on the overall operation quality, the best performing model is the CP/QR. This is a probabilistic forecast based on the adaptive quantile regression using the spatially corrected CP point predictions as input. The model consistently outperforms the benchmark approach in all the considered horizons. The relative improvements in overall quality (compared to the benchmark approach WPPT/QR) are ranging from 1.5% to 8.29% depending on the prediction horizon.

References D

- [1] T. Gneiting, K. Larson, K. Westrick, M. G. Genton, and E. Aldrich, “Calibrated probabilistic forecasting at the stateline wind energy center: The regime-switching space–time method,” *Journal of the American Statistical Association*, vol. 101, no. 475, pp. 968–979, 2006.
- [2] A. S. Hering and M. G. Genton, “Powering up with space-time wind forecasting,” *Journal of the American Statistical Association*, vol. 105, no. 489, pp. 92–104, 2010.
- [3] A. Lau, *Probabilistic Wind Power Forecasts: From Aggregated Approach to Spatiotemporal Models*. PhD thesis, University of Oxford, 2011.
- [4] 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)*, vol. 61, no. 4, pp. 555–576, 2012.
- [5] J. K. Møller, H. A. Nielsen, and H. Madsen, “Time-adaptive quantile regression,” *Computational Statistics & Data Analysis*, vol. 52, no. 3, pp. 1292–1303, 2008.
- [6] P. Pinson and G. Kariniotakis, “Conditional prediction intervals of wind power generation,” *Power Systems, IEEE Transactions on*, vol. 25, no. 4, pp. 1845–1856, 2010.
- [7] P. Pinson, H. A. Nielsen, J. K. Møller, H. Madsen, and G. N. Kariniotakis, “Non-parametric probabilistic forecasts of wind power: required properties and evaluation,” *Wind Energy*, vol. 10, no. 6, pp. 497–516, 2007.

-
- [8] J. Tastu, P. Pinson, and H. Madsen, “Spatio-temporal analysis and correction targeting Nysted Offshore- 2,” tech. rep., Technical University of Denmark, 2011.
- [9] “Spatio-temporal analysis and correction targeting Nysted Offshore,” Tech. Rep. ENFOR/04EKS0017A002-A, Enfor A/S, 2011.
- [10] S. C. Thomsen, S. Otterson, J. Tastu, and H. Madsen, “Spatio-temporal correction of DONG forecast errors,” tech. rep., Technical University of Denmark, 2011.
- [11] T. Bollerslev, “Generalized autoregressive conditional heteroskedasticity,” *Journal of econometrics*, vol. 31, no. 3, pp. 307–327, 1986.
- [12] H. Madsen, *Time series analysis*, vol. 72. CRC Press, 2008.
- [13] P.-J. Trombe, P. Pinson, and H. Madsen, “A general probabilistic forecasting framework for offshore wind power fluctuations,” *Energies*, vol. 5, no. 3, pp. 621–657, 2012.
- [14] P. Pinson and H. Madsen, “Adaptive modelling and forecasting of offshore wind power fluctuations with Markov-switching autoregressive models,” *Journal of Forecasting*, vol. 31, no. 4, pp. 281–313, 2012.
- [15] W. Zucchini and I. L. MacDonald, *Hidden Markov models for time series: an introduction using R*. CRC Press, 2009.
- [16] R. Koenker and G. Bassett J., “Regression quantiles,” *Econometrica: journal of the Econometric Society*, pp. 33–50, 1978.
- [17] T. Jónsson, P. Pinson, and H. Madsen, “Predictive densities for day-ahead electricity prices using time-adaptive quantile regression,” *working paper*.
- [18] T. M. Hamill, “Interpretation of rank histograms for verifying ensemble forecasts,” *Monthly Weather Review*, vol. 129, no. 3, pp. 550–560, 2001.
- [19] T. Gneiting, F. Balabdaoui, and A. E. Raftery, “Probabilistic forecasts, calibration and sharpness,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 69, no. 2, pp. 243–268, 2007.
- [20] T. Gneiting and A. E. Raftery, “Strictly proper scoring rules, prediction, and estimation,” *Journal of the American Statistical Association*, vol. 102, no. 477, pp. 359–378, 2007.

PAPER E

Space-time scenarios of wind power generation produced using a Gaussian copula with parametrized precision matrix.

Authors:

Julija Tastu, Pierre Pinson and Henrik Madsen

Technical report

Space-time scenarios of wind power generation produced using a Gaussian copula with parametrized precision matrix.

Julija Tastu¹, Pierre Pinson², Henrik Madsen¹

Abstract

The emphasis in this work is placed on generating space-time trajectories (also referred to as scenarios) of wind power generation. This calls for prediction of multivariate densities describing wind power generation at a number of distributed locations and for a number of successive lead times. A modelling approach taking advantage of sparsity of precision matrices is introduced for the description of the underlying space-time dependence structure. The proposed parametrization of the dependence structure accounts for such important process characteristics as non-constant conditional precisions and direction-dependent cross-correlations. Accounting for the space-time effects is shown to be crucial for generating high quality scenarios.

E.1 Introduction

Large scale integration of wind energy into power grids induces difficulties in operation and management of power systems due to the stochastic nature of wind, with its variability and limited predictability [1]. For optimal integration of wind energy into power systems high quality wind power forecasts are required [2]. A history of short-term wind power forecasting and an overview of the state-of-the-art methodology are given in [3] and [4], respectively.

Owing to the complexity of the related decision making tasks, it is preferable that the forecasts provide the user not only with the expected value of the future

¹Technical University of Denmark, Applied Mathematics and Computer Science (DTU Compute), Kgs. Lyngby, Denmark

²Technical University of Denmark, Electrical Engineering (DTU Elektro), Kgs. Lyngby, Denmark

power generation, but also with the associated uncertainty estimates. This calls for probabilistic, rather than point forecasting [5]. Applications of probabilistic forecasts to power grid operations include trading wind energy [6], economic load dispatch and stochastic unit commitment [7, 8, 9], optimal operation of storage [10], reserve quantification [11] and assessment of operating costs [12].

Usually probabilistic wind power forecasts are generated on a per-site and per-look-ahead time basis. As a result, they do not inform about the interdependence structure between forecast errors obtained at different times and/or at different sites.

Addressing each site of interest individually is motivated by the fact that power curves describing the conversion of meteorological variables to power are often given by complex non-linear functions of meteorological conditions, number and type of the considered wind turbines, their interposition within the wind farm, some topographical particularities of the area, etc. The fact that wind power dynamics is so site-specific makes it more complicated to issue high quality forecasts for a large number of sites simultaneously, because the local particularities (if to be respected) keep the dimension of the problem high.

Similarly, a common practice is to issue direct power forecasts for each of the time horizons of interest individually, rather than addressing the joint distribution. This can be explained by the fact that such direct forecasts are more robust to model misspecification. Iterated multistep-ahead predictions as a rule are more efficient if the model is correctly specified. Given the complexity of the underlying process, a correct specification is hard to achieve in practice, therefore direct forecasts are often preferred.

As a result, what is often available in practice for the decision maker is a set of marginal predictive distributions for N sites of interest and T lead times. For some decision tasks marginal densities are a suboptimal input, since the joint behaviour of power generation at all sites and the considered lead times might be of interest.

Having a set of marginal distributions, the joint density can be restored using a copula approach. One important feature of copulas is that they can be used to model dependency between stochastic variables independently of the marginal distribution functions. This is important because, as mentioned previously, modelling wind power generation at individual sites while targeting a specific lead time is already a complex task. Therefore, it is an advantage to decouple the problem of estimating marginal densities from the estimation of the space-time dependence structure.

Copulas have been widely used in many fields for modelling the dependence

between stochastic variables, including a number of problems related to wind power. As an example in [13], predictive densities for wind power generation were built by modelling the relation between wind speed and wind power using copulas. In [14], copulas have been used to estimate system net load distribution when accounting for the dependence structure between wind activity at different locations and its relation to the system load. In [15], a copula has been used to model the dependence between wind speed at a number of sites.

In [16], the authors focused on a single wind farm. A Gaussian copula, fully characterized by an empirical covariance structure, has been used to derive joint predictive distributions (multivariate in time) from the set of marginal densities. Furthermore, in [17], the author placed emphasis on wind power generation at a pair of sites and, considered different types of copulas for modelling the dependence between wind power generation at these sites for a given lead time. The present study generalizes these works by looking at the interdependence of wind power generation in time and in space. It is aimed at issuing joint predictive density of wind power generation from a set of marginal predictive distribution. The problem then boils down to specifying and estimating a suitable dependence structure.

In this work a modelling approach taking an advantage of sparsity of precision matrix is introduced for the description of the underlying dependence structure. In order to make the methodology mathematically tractable in high dimensions, a parametrization of the precision matrix is proposed. This proposal goes beyond the conventional assumptions of homogeneous stationary Gaussian Random fields, since the presented parametrization accounts for the boundary points and considers non-constant conditional variances and direction-dependent conditional correlations.

The paper has the following outline. Section E.2 introduces the data set used in the study. The methodology is described in Section E.3. It consists of some preliminaries and definitions, introduction to copula modelling and explanation on how precision matrices relate to the Gaussian copula approach. Further, Section E.4 presents the proposed parametrization of the dependence structure. The estimation process is discussed in Section E.5, while the empirical results are given in Section E.6. The paper finishes with the conclusions and perspectives presented in Section E.7.

E.2 Data

The case study relates to western Denmark, including the Jutland peninsula and the island of Funen, which produces approximately 2.5 GW, or 70% of the entire wind power capacity installed in Denmark.

Besides the significant share of wind generation, a reason for placing the focus on Denmark is given by its climate and terrain characteristics. That is, the territory is small enough for the incoming weather fronts to affect all its parts. In addition, the terrain is smooth, therefore passing weather fronts do not meet any obstacles when propagating over the country. These aspects make the test case an ideal candidate for understanding space-time effects before moving to more complex cases.

The data selected for this work comes from 15 groups of wind farms spread throughout the territory of Western Denmark. The chosen grouping corresponds to the resolution map used by the Danish Transmission System Operator. For all 15 groups measurements of wind power production with an hourly resolution are available, along with the related marginal predictive densities, derived using the adaptive re-sampling method as described in [18]. This forecasting method is one of the state-of-the-art approaches which yields reliable probabilistic forecasts with high skill. [19].



Figure E.1: Geographical locations of the 15 zones of wind farms

The available data covers a period from the 1st of January, 2006 to the 24th of October, 2007. The data set has been divided into two subsets. The first of them covering a period from the 1st of January, 2006 to the 30th of November, 2006 has been used for data analysis, the model building and the estimation. The second subset covering a period from the 30th of November, 2006 to the 24th of October, 2007 has been used for evaluating the predictive performance of the models.

E.3 Methodology

The objective of the methodology introduced here is to generate multivariate probabilistic forecasts describing wind power generation at a number of distributed locations and for a number of successive lead times.

The proposed approach follows two main steps. First, a state-of-the-art forecasting system is used to issue probabilistic forecasts for each location and each lead time individually. Subsequently, these are upgraded to full multivariate predictive densities using a copula function.

The focus in this work is on parametrization of the copula function.

E.3.1 Preliminaries and definitions

In general, the problem has the following setup. At every time step t the interest is in predicting wind power generation for times $t + 1, t + 2, \dots, t + T$ at N distributed locations. That is, there are in total $n = NT$ quantities of interest which are denoted in the following by $Y_{t,1}, Y_{t,2}, \dots, Y_{t,n}$. The enumeration is done so that $Y_{t,1}, \dots, Y_{t,T}$ represent wind power generation at the first location for the lead times $1, \dots, T$, then $Y_{t,T+1}, \dots, Y_{t,2T}$ represent wind power generation at the second location for the lead times $1, \dots, T$, and so on.

Uppercase letters represent stochastic variables, while lowercase letters denote the corresponding observations. Bold font is used to emphasize vectors and matrices. For example, $\mathbf{y}_t = [y_{t,1}, y_{t,2}, \dots, y_{t,n}]^\top$ stands for the realization of \mathbf{Y}_t .

The aim of the forecaster is to issue a multivariate predictive distribution F_t , describing a random vector $\mathbf{Y}_t = [Y_{t,1}, Y_{t,2}, \dots, Y_{t,n}]^\top$

$$F_t(y_1, y_2, \dots, y_n) = P(Y_{t,1} < y_1, Y_{t,2} < y_2, \dots, Y_{t,n} < y_n) \quad (\text{E.1})$$

There are two different families of approaches to probabilistic forecasting: parametric and non-parametric ones. The parametric approach refers to a distribution-based methodology, which requires an assumption on the shape of predictive densities. The non-parametric one refers to the distribution-free techniques, i.e. to the ones that are based on estimating the predictive densities directly from the data, without any constraints on the shape of the resulting distribution. An advantage of the non-parametric approach is given by the fact that it is fully data driven and, thus, can account for any level of asymmetry, any dependence structure, etc. The drawback, however, is that in high dimensions a fully non-parametric approach becomes intractable, even if only a climatological distribution is considered. If one wishes to issue conditional predictive densities, the curse of dimensionality becomes even more evident. Therefore, some parametrization ought to be proposed in order to make the estimation of predictive densities mathematically tractable.

Parametrizing F_t directly implies a simultaneous description of both marginal densities as well as the space-time interdependence structure. Considered distributions should account for the non-Gaussian, bounded nature of wind power generation as well for non-constant wind power variability. Unfortunately, there is no obvious distribution function which could address all the required aspects together. Copulas propose a solution by decomposing the problem of estimating F_t into two parts.

First, the focus is on marginal predictive densities, $F_{t,i} = P(Y_{t,i} < y_i)$, $i = 1, 2, \dots, n$, describing wind power generation at each location and for each lead time individually. As opposed to multivariate predictive densities, for which not many proposals exist in the literature, marginal predictive densities for wind power generation have been considered more. Thus, at this point the forecaster might take advantage of the state-of-the-art methods for probabilistic wind power forecasting. In this thesis an adaptive resampling has been used for obtaining marginal predictive densities $F_{t,i}$. The method was first described in [18]. The results documented both in [18] and in [19] confirm that it yields reliable wind power forecasts with high skill.

Subsequently, the marginal predictive densities are upgraded to F_t using a copula function. Mathematically the foundation of copulas is given by Sklar's theorem in [20]. The theorem states that: For any multivariate cumulative distribution function F_t with marginals $F_{t,1}, F_{t,2}, \dots, F_{t,n}$ there exists a copula C such that

$$F_t(y_1, y_2, \dots, y_n) = C(F_{t,1}(y_1), F_{t,2}(y_2), \dots, F_{t,n}(y_n)) \quad (\text{E.2})$$

This means that, given a set of marginal distributions, the task of getting the joint distribution boils down to finding a suitable copula function.

E.3.2 Copulas for wind power data

In general, copulas can be classified into parametric and non-parametric. In this work focus is on the former ones, since the latter become intractable in very high dimensions.

Several parametric copula types have been considered for wind power data. Namely, in [16] the authors advocate that a Gaussian copula is an adequate choice when generating multivariate in time predictive densities when describing wind power generation at a single location.

In parallel, in [17] the author has considered different copula types for modelling the dependence between wind power generation at two sites when focusing on a single lead time. The results have shown that a Gumbel copula performs best, however Gaussian and Frank copulas also fit the data adequately.

When moving to higher dimensions, the construction of Archimedean copulas (e.g. Gumbel) becomes complex. For instance, a traditional approach for constructing the n -variate Gumbel copula requires the n^{th} order derivative of the inverse of the process generating function. Even considering explicit formulas for those derivatives given in [21], the complexity remains high compared to the Gaussian copula approach. Moreover, in Ref. [22] Guzman shows that in higher dimensions Gaussian copulas outperform their Gumbel's counterparts. However, the results should be interpreted with care as they depend on the site characteristics as well as on the type of the marginal predictive densities considered.

The works mentioned above indicate that the Gaussian copula is an adequate choice for describing spatial and temporal dependencies which are present in wind power data. However, these works have not considered spatio-temporal dependencies. Thus, the first step in this study involved a preliminary data examination to verify whether the Gaussian copula was consistent with the observed space-time dependence structure.

For example, consider $Y_{t,5*43+5}$ and $Y_{t,4*43+4}$ which represent wind power generation at zone 6 at time $t + 5$ and wind power generation at zone 5 at $t + 4$, respectively. The dependence between random variables $Y_{t,5*43+5}$ and $Y_{t,4*43+4}$ can be graphically represented looking at the ranks of the uniform variables $F_{t,5*43+5}(y_{t,5*43+5})$ and $F_{t,4*43+4}(y_{t,4*43+4})$.

The scatterplot of the corresponding ranks characterizes the dependence structure between $Y_{t,5*43+5}$ and $Y_{t,4*43+4}$, while the overlaying contour plot represents the so called empirical copula [23]. The empirical copula is then compared to the corresponding Gaussian copula and the results are illustrated in Fig. E.2. Both patterns are very similar, and this is an indication that the Gaussian copula is appropriate for describing the spatio-temporal dependence structure. The results obtained while considering different pairs of variables have been qualitatively similar.

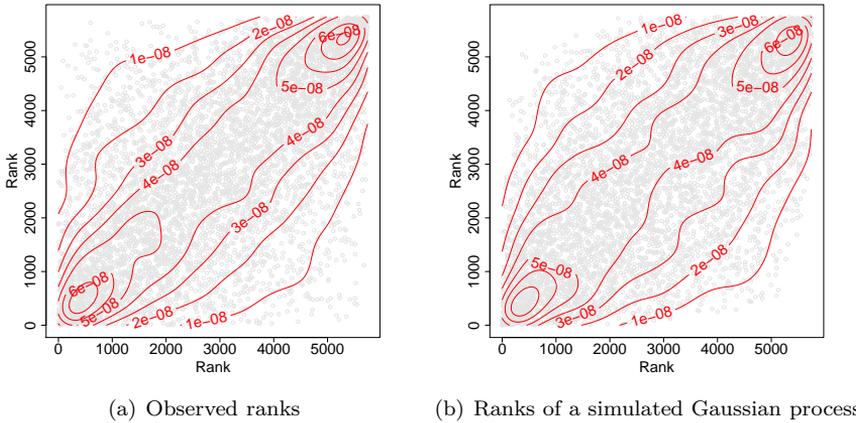


Figure E.2: Left: Scatterplot with contour overlay given by the ranks of $F_{t,5*43+5}(y_{t,5*43+5})$ and $F_{t,4*43+4}(y_{t,4*43+4})$. Right: Scatterplot with contour overlay of the simulated bivariate Gaussian process having the same rank correlation as the observed data illustrated on the left.

One should note, that the considered verification scheme does not guarantee that the Gaussian copula is the best choice for modelling the dependence structure. It should be only seen as an indication that there are no obvious inconsistencies between the Gaussian copula and the data. The reason not to consider other copula types has been given by a strong preference to use Gaussian copulas, since they have an advantage of being simple to use in high dimensions, widely used and having a strong theoretical linkage to a large class of mathematical theories.

E.3.3 Gaussian Copula

Gaussian copula is given by

$$C(F_{t,1}(y_1), \dots, F_{t,n}(y_n)) = \Phi_{\Sigma}(\Phi^{-1}(F_{t,1}(y_1)), \dots, \Phi^{-1}(F_{t,n}(y_n))) \quad (\text{E.3})$$

where Φ^{-1} denotes the inverse of the univariate standard Gaussian distribution function and $\Phi_{\Sigma}(\cdot)$ is the n -variate Gaussian distribution function with zero mean, unit marginal variances and correlation matrix Σ .

That is, the copula is built by transforming wind power generation $y_{t,i}$ to the latent standard Gaussian variable $x_{t,i}$ by applying the following:

$$x_{t,i} = \Phi^{-1}(F_{t,i}(y_{t,i})) \quad (\text{E.4})$$

The resulting $\mathbf{x}_t = [x_{t,1}, \dots, x_{t,n}]^{\top}$ are realization of the corresponding random process $\mathbf{X} = [X_1, \dots, X_n]^{\top}$ which is distributed as multivariate Gaussian with zero mean, unit marginal variances and a correlation matrix Σ , i.e.

$$\mathbf{X} \sim \mathcal{N}(0, \Sigma) \quad (\text{E.5})$$

In other words, it is assumed that a joint multivariate predictive density for \mathbf{Y}_t can be represented by the multivariate Gaussian density in the transformed domain given by \mathbf{X} :

$$F_t(y_1, \dots, y_n) = \Phi_{\Sigma}(\Phi^{-1}(F_{t,1}(y_1)), \dots, \Phi^{-1}(F_{t,n}(y_n))) \quad (\text{E.6})$$

Note, that in this setup, even though the marginal distributions $F_{t,i}$ as well as the joint distributions F_t are time-dependent, the underlying dependence structure is fully represented by the time-invariant correlation matrix Σ , thus there is no time index in the notation of the random variable \mathbf{X} .

The goal is to propose a sensible parametrization for Σ . This is done by focusing on \mathbf{X} .

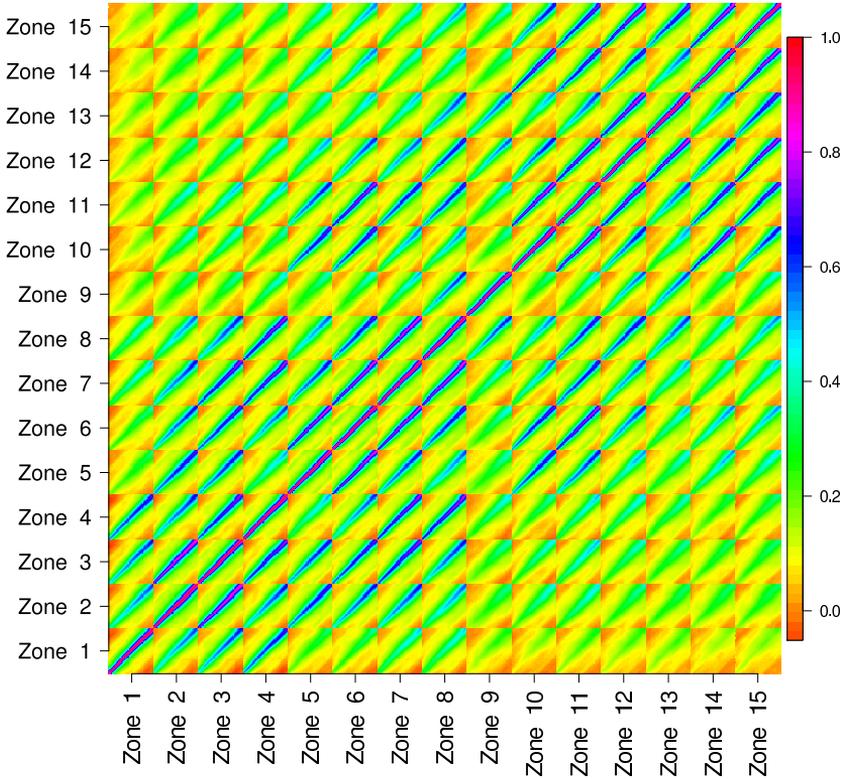


Figure E.3: Sample correlation matrix

E.3.4 Modelling as a conditional autoregression

Consider, a set of available wind power observations corresponding to \mathbf{y}_t , $t = 1, \dots, T$. The observations are transformed to the latent Gaussian variables \mathbf{x}_t and the covariance structure of the latter ones is studied. As can be seen from Fig. E.3, the sample covariance matrix, Σ is dense. This implies that inference with such a matrix has a computational complexity of $\mathcal{O}(n^3)$. In order to make the proposed methodology applicable for problems of high dimension, instead of modelling the covariance matrix directly, we focus on its inverse, denoted by \mathbf{Q} [24]. The inverse of a covariance matrix is called a precision matrix.

The sample precision matrix (see Fig. E.4) is very sparse. This opens the doors to the framework of Gaussian Markov Random Fields (GMRF), allowing us to benefit from computationally efficient algorithms derived for the inference with sparse matrices. More specifically, by switching from a dense covariance matrix

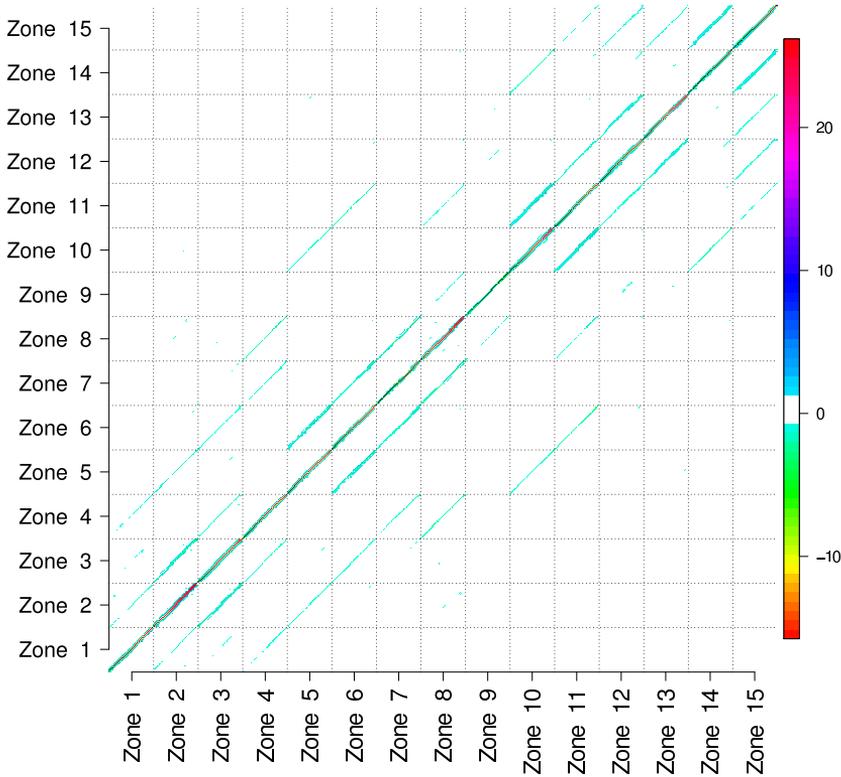


Figure E.4: Sample precision matrix

to its sparse inverse, we reduce the computational complexity from $\mathcal{O}(n^3)$ to the range from $\mathcal{O}(n)$ to $\mathcal{O}(n^{3/2})$, depending on the process characteristics [24].

In contrast to covariance structure which informs of marginal dependence between variables, the precision matrix represents conditional interdependencies. The elements of the precision matrix have a useful conditional interpretation.

The diagonal elements of \mathbf{Q} are the conditional precisions of X_i given $\mathbf{X}_{-i} = [X_1, X_2, \dots, X_{i-1}, X_{i+1}, \dots, X_n]^\top$ while the off-diagonal elements, with a proper scaling, provide information about the conditional correlations between variables. For a zero mean process, the following holds:

$$E(X_i | \mathbf{X}_{-i}) = -\frac{1}{Q_{ii}} \sum_{j \neq i} Q_{ij} X_j \quad (\text{E.7})$$

$$\text{Var}(X_i|\mathbf{X}_{-i}) = 1/Q_{ii} \quad (\text{E.8})$$

A very important relation is that $Q_{ij} = 0$ if and only if elements X_i and X_j are independent given the rest, $\mathbf{X}_{-\{i,j\}}$. This means that the non-zero pattern of \mathbf{Q} determines the neighbourhood of the conditional dependence between variables and can be used to provide parametrization of the precision matrix. Of course, one still has to keep in mind that \mathbf{Q} is required to be symmetric positive-definite (SPD).

The relationship given by eq. (E.8) is sometimes used for an alternative specification of Gaussian Markov Random Field through full conditionals. This approach was pioneered by Besag in [25] and the resulting models are also known as conditional autoregressions, abbreviated as CAR. When specifying GMRF through CAR, instead of considering the entries of \mathbf{Q} , Q_{ij} , directly, focus is on modelling terms $\kappa_i = Q_{ii}$ and $\beta_{ij} = Q_{ij}/Q_{ii}$, $i, j = 1, \dots, n$.

From eq. (E.8) it is seen that elements β_{ij} are given by the coefficients of the corresponding conditional autoregression models, while κ_i informs on the related variances.

This translates to the following equality:

$$\mathbf{Q} = \boldsymbol{\kappa} * \mathbf{B}. \quad (\text{E.9})$$

where $\boldsymbol{\kappa}$ denotes a diagonal matrix of dimension $n \times n$, the diagonal elements of which are given by κ_i , $i = 1, \dots, n$. \mathbf{B} is a coefficient matrix consisting of coefficients β_{ij} . In other words, \mathbf{B} equals the precision matrix standardized by its diagonal.

CAR specification is sometimes easier to interpret and we will use it to propose a parametrization for \mathbf{Q} in this work.

E.4 Parametrization of the precision matrix

The CAR specification (see eq. E.8) decouples the problem of describing \mathbf{Q} into two parts. First, parametrization of conditional precisions, $\boldsymbol{\kappa}$ is discussed. Then, parametrization of the coefficient matrix \mathbf{B} is presented.

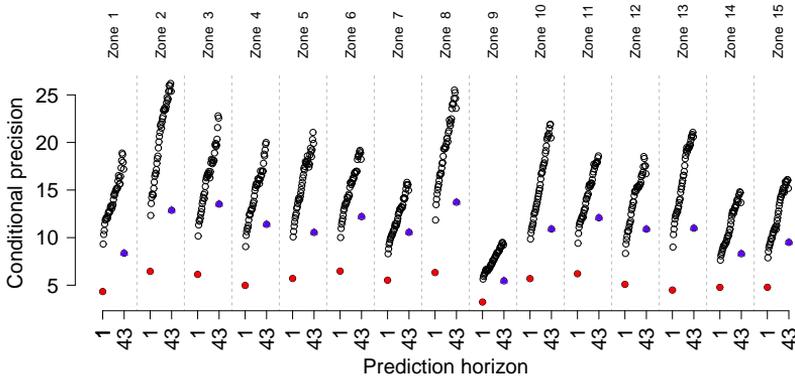


Figure E.5: Diagonal elements of the sample precision matrix, \mathbf{Q} . Boundary points given by the conditional precisions related to the horizons of 1 and 43 hours ahead are marked with red and blue circles, respectively

E.4.1 Structure of the diagonal elements

Conventionally, CAR models are given by stationary GMRF. Stationarity implies rather strong assumptions on both the neighbourhood structure and the elements of \mathbf{Q} . Firstly, the structure of the neighbourhood allows for no special arrangement for the boundary points. Secondly, the full conditionals have constant parameters not depending on i . In other words, the conditional precisions given by κ_i , $i = 1, \dots, n$ are assumed to be constant. However, the data analysis has shown this assumption would be very restrictive in this case.

The diagonal of the sample precision matrix \mathbf{Q} is depicted in Fig. E.5. One can note, that its elements are not constant. Their variation has some structure, which is captured in the following.

E.4.1.1 Conditional precisions for different zones

First, it can be seen that the conditional precisions describing different zones are rather similar. The most significant deviation from the global picture is observed for zone 9. This is also in line with the results shown in [26] and could be explained by two main factors. On the one hand, it could be caused by the fact that group 9 covers a smaller territory compared to the other zones. This

leads to more significant local variations, which results in the lower conditional precisions.

Another possible explanation is that zone 9, in contrast with the rest of the territories, is situated off the mainland. Therefore, it is very likely that the dynamics in zone 9 are different from the rest of the considered region.

If looking at the rest of the zones, then the observed pattern of conditional precisions is very similar. The differences are present, however, as of now, we have not been able to explain them by any of the available explanatory variables. An assumption that the precision pattern could depend on whether a zone is located in the centre of the considered territory or on the boundary has not been supported by the data. It has been also considered that patterns of conditional precisions could depend on the overall level of power variability. This, however, has not found a support in the data, either. Further investigation of this matter is left for future work. In this study it is considered that the pattern is the same for all zones. That is, any potential differences are disregarded and the following parametrization is proposed:

$$\begin{aligned} \text{diag}(\mathbf{Q}) &= [\kappa_1, \kappa_2, \dots, \kappa_{645}]^\top = & \text{(E.10)} \\ &= [\kappa_1, \kappa_2, \dots, \kappa_{43}, \kappa_1, \kappa_2, \dots, \kappa_{43}, \dots, \dots, \kappa_1, \kappa_2, \dots, \kappa_{43}]^\top & \text{(E.11)} \end{aligned}$$

where $\mathbf{K} = [\kappa_1, \kappa_2, \dots, \kappa_{43}]^\top$ is a vector of conditional precisions corresponding to a single zone.

E.4.1.2 Conditional precisions for different lead times

Since it has been assumed that, at zone level, the conditional precisions follow the same dynamics, focus is on a single zone. The corresponding conditional precisions are given by $\mathbf{K} = [\kappa_1, \kappa_2, \dots, \kappa_{43}]^\top$. κ_1 and κ_{43} correspond to the temporal boundary and this explains why they stand out from the general pattern as show in Fig. E.5. The temporal boundary for κ_1 and κ_{43} is given by the fact that we do not consider lead times of less than 1 hour ahead and more than 43 hours ahead.

Further from the temporal boundaries, i.e. for the lead times from 2 to 42 hours ahead, Fig. E.5 suggests that the conditional precisions increase with the lead time. For accounting for this effect, the following parametrization is proposed:

$$\kappa_i = \kappa_{i-1} * \rho \quad \text{(E.12)}$$

for $i = 2, \dots, 42$. Here ρ is a ratio parameter.

E.4.1.3 Final parametrization of the conditional precisions

Summarizing the reasoning presented in the previous sections, the following parametrization for the diagonal elements of the precision matrix is proposed:

$$\kappa = \begin{array}{c} \text{zone} \\ 1 \\ 2 \\ \vdots \\ 15 \end{array} \begin{array}{cccc} 1 & 2 & \cdots & 15 \\ \left(\begin{array}{cccc} \mathbf{K} & & & \\ & \mathbf{K} & & \\ & & \ddots & \\ & & & \mathbf{K} \end{array} \right) \end{array} \quad (\text{E.13})$$

where

$$\mathbf{K} = \begin{array}{c} \text{lead time} \\ 1 \\ 2 \\ 3 \\ \vdots \\ 42 \\ 43 \end{array} \begin{array}{cccccc} 1 & 2 & 3 & \cdots & 42 & 43 \\ \left(\begin{array}{cccccc} q_1 & & & & & \\ & \rho & & & & \\ & & \rho^2 & & & \\ & & & \ddots & & \\ & & & & \rho^{41} & \\ & & & & & q_{43} \end{array} \right) \frac{1}{\sigma^2} \end{array} \quad (\text{E.14})$$

Thus, the diagonal of \mathbf{Q} can be described with four parameters. q_1 and q_{43} describe the temporal boundary conditions, ρ describes a proportional increase in conditional precisions and σ^2 represents a base level of variation.

E.4.2 Structure of the standardized precision matrix

Next step is to propose a parametrization for \mathbf{B} . This requires understanding the neighbourhood structure of \mathbf{Q} , i.e. identifying which elements are non-zero.

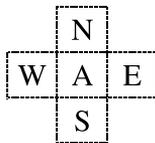


Figure E.6: Neighbourhood specification of a single zone. The focus zone is marked A, while W, N, E and S denote its Western, Northern, Eastern and Southern neighbours, respectively.

E.4.2.1 Spatial neighbourhood

Consider a single zone, further denoted by A. A careful look at Fig. E.4 reveals that information at zone A is only dependent on local information at A and on the four closest neighbouring zones: Northern (N), Eastern (E), Southern (S) and Western (W) neighbours of A (see Fig. E.6).

E.4.2.2 Temporal neighbourhood

Fig. E.4 shows that information observed at zone A at time t is only dependent on a very small amount of elements at zones A, N, E, S, and W.

Since precision matrices ought to be symmetric, it is sufficient to focus on the dependency between A and its Western and Southern neighbours, without direct consideration of the Eastern and Northern neighbours. Let us zoom-in to some relevant blocks of the sample coefficient matrix \mathbf{B} when focusing on zone 6.

From the results depicted in Fig. E.7 one can note that the corresponding conditional correlations of zone A with its North and the West side neighbours differ. Information at zone A observed at time t is conditionally dependent only on the simultaneous information at zone N. Meanwhile, the conditional correlation with zone W is significant at times $t - j$, $j = -2, \dots, 2$. This difference in the dependency pattern can be partly explained by the fact that in Denmark prevailing winds are westerly. Thus, forecast errors most often propagate from West to East, as discussed in [27]. This means that usually zones A and N are influenced by the upcoming weather front at similar time, while zone W is exposed to it earlier. Of course, one should also keep in mind, that in our test case distances between zones A and N are larger than those between A and W, and this can be another factor influencing different patterns of the related dependencies.

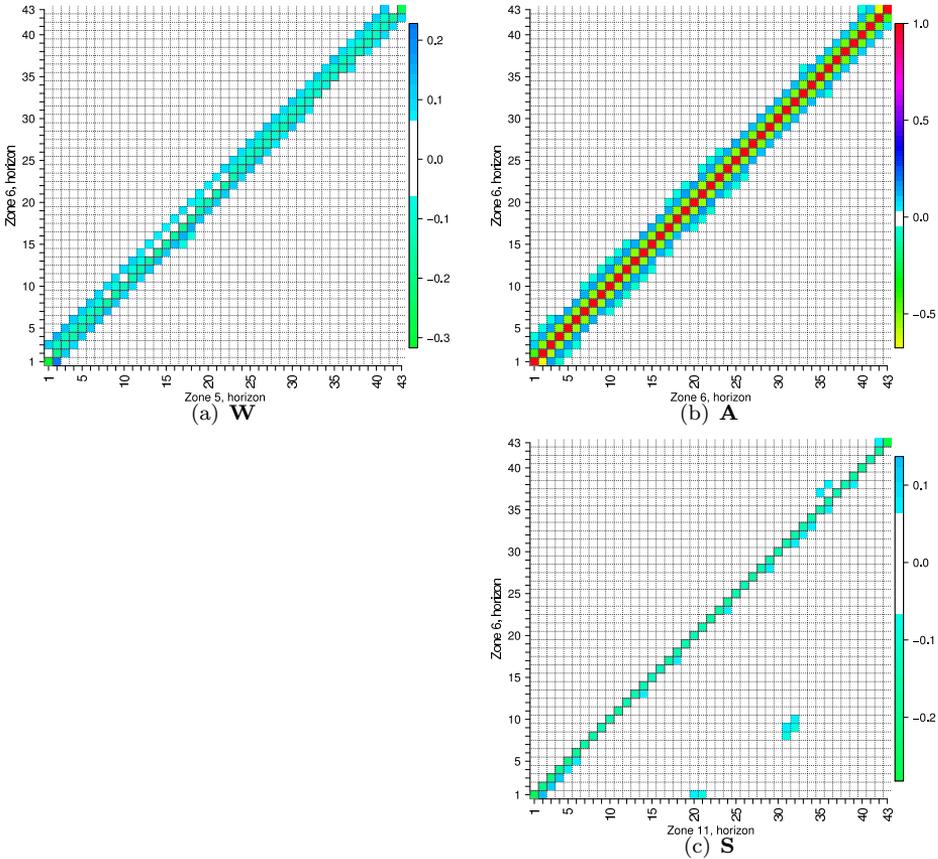


Figure E.7: Zoomed in blocks of the standardized (by its diagonal) sample precision matrix

In general, the results depicted in Fig. E.7 show that information corresponding to lead time h for zone A is dependent on the variables at the neighbouring zones corresponding to lead times $h - j$, where $j = -2, \dots, 2$. Thus, visually the data suggests a second order (temporal) process. In this work both the second ($j = -2, \dots, 2$) and the first order ($j = -1, 0, 1$) models have been considered. Since the corresponding difference in the performance of the resulting predictive densities was rather minor, in this study the focus is on the first order model ($j = 1$). Extension to higher order models is rather straight-forward and all the discussed parametrization and estimation procedures apply.

In this work a directional non-stationary CAR model, abbreviated as DCAR, is considered. That is, the conditional correlations are made direction-dependent.

In this respect the work is inspired by [28] where the authors consider a directional (in space) CAR model. We refer the reader to that work for a clear description of the modelling approach. The current proposal can be viewed as a generalization of the work presented in [28] since space-time neighbourhoods are considered along with the non-constant precisions.

When considering DCAR models, directional neighbourhoods should be chosen carefully so that each of them forms a (directional) clique. That is, consider two elements from the full random vector \mathbf{X} : X_i and X_j . Then, given that X_i is a "west-side" and "one-hour-ago" neighbour of X_j , X_j should be assigned as the "east-side" and "one-hour-ahead" neighbour of X_i . This is essential for ensuring the symmetry of the precision matrix.

E.4.2.3 Final parametrization of the standardized matrix

Summarizing, data analysis has suggested that information coming from zone A at lead time h conditionally depends only on information coming from zones N, E, S, and W with lead times $h - 1$, h , $h + 1$ and on the local situation at zone A for lead times $h - 1$ and $h + 1$. In terms of the CAR specification given in eq. (E.8) this translates to:

$$E(x_h^{(A)}) = - \sum_{j=\{-1,1\}} a_j x_{h+j}^{(A)} - \sum_{j=\{-1,0,1\}} (b_j x_{h+j}^{(W)} + b_j^* x_{h+j}^{(E)} + c_j x_{h+j}^{(N)} + c_j^* x_{h+j}^{(S)}) \quad (\text{E.15})$$

Here $x_h^{(\cdot)}$ refers to a single element from the latent Gaussian vector \mathbf{x} corresponding to the information obtained at zone "." when considering marginal forecasts for lead times h . a_j , b_j , b_j^* , c_j and c_j^* denote the corresponding coefficients which are the building blocks for \mathbf{B} .

Data analysis has shown that a_j , b_j , b_j^* , c_j and c_j^* do not depend on the considered lead time h . It can be also seen in Fig. E.7 that there is no indication of any increase/decrease of coefficient values with the lead time. The only values which drop out from the constant picture are the ones corresponding to the temporal boundaries and this will be accounted for when scaling by the corresponding conditional precisions.

In this work it is assumed that the corresponding coefficients are constant for all zones. Further work could be done in order to explain spatial variation in the coefficient values.

Some restrictions have to be imposed on the parameters in \mathbf{B} to ensure that the resulting \mathbf{Q} is symmetric positive definite. Imposing symmetry reduces the parameter space significantly, since coefficients a_1 can be derived from a_{-1} , b_j^* from b_{-j} and c_j^* from c_{-j} , $j = -1, 0, 1$. This will be formulated below in eq. (E.23).

E.4.3 Final parametrization of the precision matrix

The precision matrix is given by:

$$\mathbf{Q} = \kappa \mathbf{B} \tag{E.16}$$

where κ represents the diagonal elements which are assumed to be independent on the considered zone, but dependent on the lead time:

$$\kappa = \begin{matrix} & \text{zone} & 1 & 2 & \cdots & 15 \\ \begin{matrix} 1 \\ 2 \\ \vdots \\ 15 \end{matrix} & \left(\begin{matrix} \mathbf{K} & & & \\ & \mathbf{K} & & \\ & & \ddots & \\ & & & \mathbf{K} \end{matrix} \right) \end{matrix} \tag{E.17}$$

where \mathbf{K} describes how conditional precisions change with the lead time. It is given by:

$$\mathbf{K} = \begin{matrix} & \text{lead time} & 1 & 2 & 3 & \cdots & 42 & 43 \\ \begin{matrix} 1 \\ 2 \\ 3 \\ \vdots \\ 42 \\ 43 \end{matrix} & \left(\begin{matrix} q_1 & & & & & & & \\ & \rho & & & & & & \\ & & \rho^2 & & & & & \\ & & & \ddots & & & & \\ & & & & \rho^{41} & & & \\ & & & & & & q_{43} & \end{matrix} \right) \frac{1}{\sigma^2} \end{matrix} \tag{E.18}$$

Standardized by the diagona) precision matrix is given by **B**:

$$\mathbf{B} = \begin{matrix} & \begin{matrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 15 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \\ 10 \\ 11 \\ 12 \\ 13 \\ 14 \\ 15 \end{matrix} & \left(\begin{matrix} \mathbf{A} & & & \mathbf{S} & & & & & & & & & & & \\ & \mathbf{A} & \mathbf{E} & & \mathbf{S} & & & & & & & & & & \\ & \mathbf{W} & \mathbf{A} & \mathbf{E} & & \mathbf{S} & & & & & & & & & \\ \mathbf{N} & & \mathbf{W} & \mathbf{A} & & & & \mathbf{S} & & & & & & & \\ & \mathbf{N} & & & \mathbf{A} & \mathbf{E} & & & & \mathbf{S} & & & & & \\ & & \mathbf{N} & & \mathbf{W} & \mathbf{A} & \mathbf{E} & & & & \mathbf{S} & & & & \\ & & & \mathbf{N} & & \mathbf{W} & \mathbf{A} & \mathbf{E} & & & & \mathbf{S} & & & \\ & & & & \mathbf{N} & & & \mathbf{W} & \mathbf{A} & & & & \mathbf{S} & & \\ & & & & & \mathbf{N} & & & & \mathbf{A} & \mathbf{E} & & & \mathbf{S} & \\ & & & & & & \mathbf{N} & & & \mathbf{W} & \mathbf{A} & \mathbf{E} & & & \mathbf{S} \\ & & & & & & & \mathbf{N} & & & \mathbf{W} & \mathbf{A} & \mathbf{E} & & \\ & & & & & & & & \mathbf{N} & & & \mathbf{W} & \mathbf{A} & \mathbf{E} & \\ & & & & & & & & & \mathbf{N} & & & \mathbf{W} & \mathbf{A} & \mathbf{E} \end{matrix} \right) \end{matrix} \tag{E.19}$$

where **W**, **N** represent the blocks of conditional dependencies between the focus zone and its Western and Northern neighbours, respectively, while **A** represent local dependencies at zone A itself. The blocks are parametrized in the following way:

$$\mathbf{A} = \begin{matrix} \text{lead time} & \begin{matrix} 1 & 2 & 3 & 4 & \dots & 41 & 42 & 43 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ \vdots \\ 41 \\ 42 \\ 43 \end{matrix} & \left(\begin{matrix} 1 & & & & & & & \\ & \frac{\rho}{q_1} a_{-1} & & & & & & \\ a_{-1} & 1 & \rho a_{-1} & & & & & \\ & a_{-1} & 1 & \rho a_{-1} & & & & \\ & & a_{-1} & 1 & & & & \\ & & & & \ddots & & & \\ & & & & & 1 & \rho a_{-1} & \\ & & & & & a_{-1} & 1 & \rho a_{-1} \\ & & & & & & \frac{\rho^{42}}{q_{43}} a_{-1} & 1 \end{matrix} \right) \end{matrix} \tag{E.20}$$

$$\mathbf{W} = \begin{array}{c} \text{lead time} \\ 1 \\ 2 \\ 3 \\ 4 \\ \vdots \\ 41 \\ 42 \\ 43 \end{array} \begin{array}{ccccccc} 1 & 2 & 3 & 4 & \cdots & 41 & 42 & 43 \\ \left(\begin{array}{ccccccc} b_0 & \frac{b_1}{q_1} & & & & & & \\ b_{-1} & b_0 & b_1 & & & & & \\ & b_{-1} & b_0 & b_1 & & & & \\ & & b_{-1} & b_0 & & & & \\ & & & & \ddots & & & \\ & & & & & b_0 & b_1 & \\ & & & & & b_{-1} & b_0 & b_1 \\ & & & & & & \frac{\rho^{42}}{q_{43}} b_{-1} & b_0 \end{array} \right) \end{array} \quad (\text{E.21})$$

$$\mathbf{N} = \begin{array}{c} \text{lead time} \\ 1 \\ 2 \\ 3 \\ 4 \\ \vdots \\ 41 \\ 42 \\ 43 \end{array} \begin{array}{ccccccc} 1 & 2 & 3 & 4 & \cdots & 41 & 42 & 43 \\ \left(\begin{array}{ccccccc} c_0 & \frac{c_1}{q_1} & & & & & & \\ c_{-1} & c_0 & c_1 & & & & & \\ & c_{-1} & c_0 & c_1 & & & & \\ & & c_{-1} & c_0 & & & & \\ & & & & \ddots & & & \\ & & & & & c_0 & c_1 & \\ & & & & & c_{-1} & c_0 & c_1 \\ & & & & & & \frac{\rho^{42}}{q_{43}} c_{-1} & c_0 \end{array} \right) \end{array} \quad (\text{E.22})$$

with

$$\begin{aligned} \mathbf{E} &= \mathbf{K}^{-1} \mathbf{W}^\top \mathbf{K} \\ \mathbf{S} &= \mathbf{K}^{-1} \mathbf{N}^\top \mathbf{K} \end{aligned} \quad (\text{E.23})$$

to ensure symmetry of \mathbf{Q}

Thus, we can model \mathbf{Q} as a function of a parameter vector $\boldsymbol{\theta}$, where:

$$\boldsymbol{\theta} = [q_1, \rho, q_{43}, \sigma^2, a_{-1}, b_0, b_{-1}, b_1, c_0, c_{-1}, c_1]^\top \quad (\text{E.24})$$

E.5 Estimation

This section discusses how to fit the GMRF defined by $\mathbf{Q}(\boldsymbol{\theta})$ to the observations. This task can be divided into two parts. Firstly, one needs to decide on the discrepancy measure between the observed data and the suggested GMRF. Secondly, one needs to propose a way to ensure that the parameter estimates belong to the valid parameter space Θ^+ which would ensure that the resulting precision matrix is symmetric positive definite (SPD).

E.5.1 The valid parameter space

In section E.4 the precision matrix \mathbf{Q} is described as a function of the parameter vector $\boldsymbol{\theta}$. In this section we discuss how to ensure that parameter estimates $\hat{\boldsymbol{\theta}}$ belong to the valid parameter space Θ^+ which would ensure that the resulting precision matrix $\mathbf{Q}(\hat{\boldsymbol{\theta}})$ is SPD.

Symmetry of \mathbf{Q} is imposed by its construction (see Section E.4). Thus, we are left with the concerns of whether the matrix is positive definite.

Unfortunately, in general it is hard to determine Θ^+ . There are some analytical results available for precision matrices that are Toeplitz [29]. This could be used when working with homogeneous stationary GMRF, but this is not the case in this study. When there is no knowledge on Θ^+ available, the common practice is to consider a subset of Θ^+ which is given by the sufficient condition of \mathbf{Q} being diagonal dominant.

Diagonal dominance is most often easy to treat analytically. On the downside, this approach becomes more and more restrictive for an increasing number of parameters. This issue is discussed in more detail in [29]. For instance, for our particular test case we could see that the assumption of diagonal dominance was too restrictive, as the estimated parameters (if no such restriction was imposed) far from fulfilled the criterion of diagonal dominant precision matrix.

If the full valid parameter space Θ^+ is unknown and its diagonal dominant subset is deemed as too restrictive, it is always possible to use a "brute force" approach (following terminology of [29]). This entails checking if $\hat{\boldsymbol{\theta}} \in \Theta^+$ by direct verification of whether the resulting $\mathbf{Q}(\hat{\boldsymbol{\theta}})$ is SPD or not. This is most easily done by trying to compute the Cholesky factorization which will be successful if and only if \mathbf{Q} is positive definite. The "brute force" method was the one used in this work.

However, it is worth mentioning some advantages given by the diagonal dominance approach over the "brute force" method. An important one is that if one estimates parameters while requiring the diagonal dominance, then one can be sure that if a new territory is to be included to the considered setup, there is no strict necessity (other than aiming for optimality) to re-estimate the parameters. In other words, one can be sure that the "old" parameter vector would guarantee a valid covariance structure for the enlarged lattice. This is not exactly the case for the "brute force" approach. If we want to take an additional zone into consideration, we cannot be guaranteed that the previously estimated parameters would result in a valid covariance structure. That is, we might need to re-estimate. However, the experiments have shown a "new" set of parameters being very close to the "old" one. Thus, if we use previously estimated parameters as the initial condition for the optimization routine, then we can expect to get fast estimates of the "new" parameter vector.

E.5.2 Choosing an appropriate optimization criterion

When estimating θ from the real data, one needs to decide on some discrepancy measure between the imposed GMRF and the observations.

In this work we focused on parameter estimation using maximum likelihood theory. In [30] the authors argue that maximum likelihood estimators for GMRF are not robust with respect to model errors and might result in coefficient estimates which do not describe well the global properties of the data. See [30] for more details. The authors propose a new optimization criterion which resolves this difficulty. The criterion is based on a norm distance between the estimated and the observed correlation structures. In this work we considered both the norm-based discrepancy optimization and the likelihood approach. Since estimates obtained with both approaches were consistent, further focus is on the likelihood based inference. This choice is made, since, following [30], if a GMRF describes the data adequately, then maximum likelihood-based inference is more efficient than the norm-optimization. The reader is referred to [30] for a broader discussion on the existing alternatives.

E.5.3 Parameter estimation using maximum likelihood

Let us focus on a single time t and recall some of the notation introduced in Section E.3. The corresponding observation of the latent Gaussian field $\mathbf{x}_t = [x_{t,1}, x_{t,2}, \dots, x_{t,n}]^\top$ is then given by the corresponding transformations of the related power measurements $y_{t,1}, y_{t,2}, \dots, y_{t,n}$. That is,

$$\mathbf{x}_t = [\Phi^{-1}(F_{t,1}(y_{t,1})), \Phi^{-1}(F_{t,2}(y_{t,2})), \dots, \Phi^{-1}(F_{t,n}(y_{t,n}))]^\top$$

The essence of the presented methodology is based on the assumption that \mathbf{x}_t follows a multivariate Gaussian distribution with zero mean and correlation matrix given by \mathbf{Q}^{-1} .

Then the log likelihood contribution given by \mathbf{x}_t writes as:

$$l_t = -\frac{n}{2} \ln(2\pi) + \frac{1}{2} \ln |\mathbf{Q}| - \frac{1}{2} (\mathbf{x}_t)^\top \mathbf{Q} \mathbf{x}_t \tag{E.25}$$

Given H realizations of the random process \mathbf{X} , the overall likelihood is given by

$$l(\boldsymbol{\theta}) = \sum_{t=1}^H l_t = -\frac{nH}{2} \ln(2\pi) + \frac{H}{2} \ln |\mathbf{Q}(\boldsymbol{\theta})| - \frac{1}{2} \sum_{t=1}^H (\mathbf{x}_t)^\top \mathbf{Q}(\boldsymbol{\theta}) \mathbf{x}_t \tag{E.26}$$

By solving $\frac{\partial l(\boldsymbol{\theta})}{\partial \sigma^2} = 0$ with respect to σ^2 yields the following profile maximum likelihood estimator for σ^2

$$\widehat{\sigma^2} = \frac{\sum_{t=1}^H \mathbf{x}_t^\top \mathcal{P} \mathbf{x}_t}{Hn} \tag{E.27}$$

with

$$\mathcal{P} = \begin{matrix} & \text{lead time} & 1 & 2 & 3 & \dots & 42 & 43 \\ \begin{matrix} 1 \\ 2 \\ 3 \\ \vdots \\ 42 \\ 43 \end{matrix} & & \begin{pmatrix} q_1 & & & & & & & \\ & \rho & & & & & & \\ & & \rho^2 & & & & & \\ & & & \ddots & & & & \\ & & & & & & \rho^{41} & \\ & & & & & & & q_{43} \end{pmatrix} & \end{matrix} \tag{E.28}$$

Having the profile likelihood estimate for σ^2 , we view \mathbf{Q} as a function of the parameter vector $\boldsymbol{\theta}^-$:

$$\boldsymbol{\theta}^- = [q_1, \rho, q_{43}, a_{-1}, b_0, b_{-1}, b_1, c_0, c_{-1}, c_1]^\top \tag{E.29}$$

Estimate for θ^- is obtained by a numerical optimization of the likelihood function given in eq. (E.26) with respect to the parameter vector θ^- .

The requirement for the resulting $\hat{\mathbf{Q}}$ to be symmetric positive definite is equivalent to requiring all eigenvalues to be positive. Similarly to [30], we approach the constrained optimization problem as an unconstrained one, adding an infinite penalty if some of the eigenvalues are negative. This approach works well in practice.

Also, $\Sigma = \mathbf{Q}^{-1}$ is required to have a unit diagonal. In practice this is achieved by the corresponding scaling of the estimate $\hat{\mathbf{Q}}$ as suggested in [29].

E.6 Results

E.6.1 Assessing global model fit

Verification starts with examination of the global properties of the estimated dependence structure. This is done in the spirit of [30], i.e. by visually comparing the estimated covariance structure with the sample one. The estimated correlation matrix is illustrated in Fig. E.8, while the sample one is shown in Fig. E.3. The fit seems adequate.

The motivation for checking the global resemblance between the dependence structures in addition to the overall likelihood evaluation is given by the following. When optimizing the likelihood, the optimal fit is given by fitting the covariances within the neighbourhood exactly, while the remaining ones are determined by the inversion of the fitted precision matrix [30]. This may result in the estimates, which instead of capturing dependencies between all the variable pairs in some reasonable way, capture just some of them with a very high precision, while ignoring the others.

The fact, that the the estimated covariance matrix is in line with the the sample one, indicates that the model describes the data adequately and that the resulting joint density can be used for inferring on the global properties of the process.

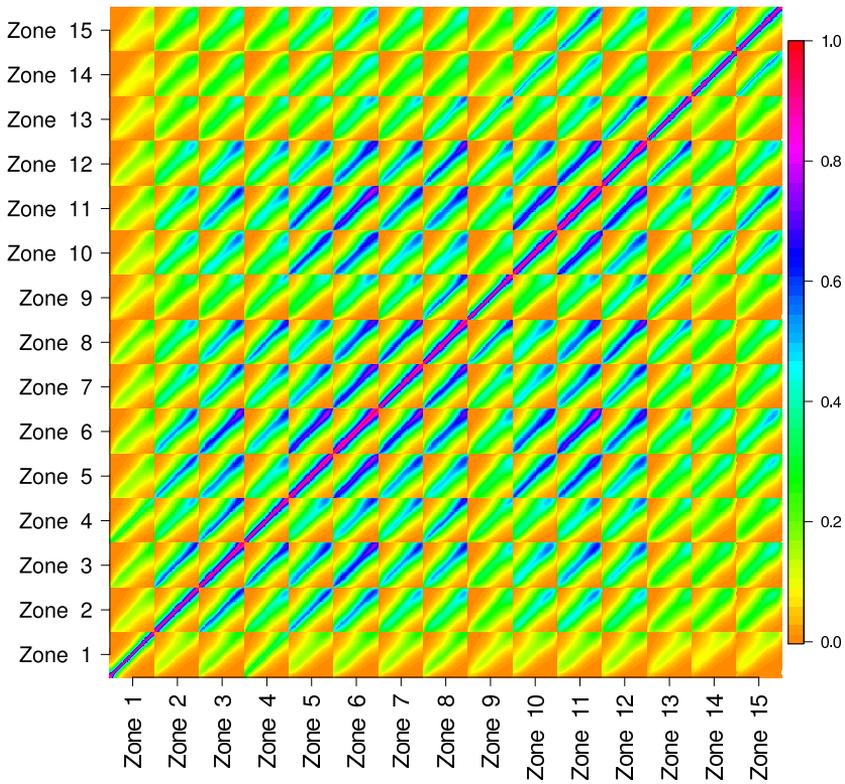


Figure E.8: Estimated correlation matrix

E.6.2 Assessing predictive model performance

In this section focus is on evaluating predictive performance of the derived probabilistic forecasts. While the first year of data covering a period from the 1st of January, 2006 to the 30th of November, 2006 has been used for model estimation, validation is performed using another data subset, which is covering a period from the 30th of November, 2006 to the 24th of October, 2007.

The section starts with a presentation of the benchmark approaches. Further, scores used for the overall quality assessment are discussed. Finally, the empirical results are presented.

E.6.2.1 Considered models

The following models are considered in this study:

1. *Independent*: The corresponding multivariate predictive densities are based on the assumption that the marginal densities are independent. That is:

$$F_t(y_1, y_2, \dots, y_n) = F_{t,1}(y_1)F_{t,2}(y_2) \cdots F_{t,n}(y_n) \quad (\text{E.30})$$

2. *First order time-dependence*: The corresponding multivariate densities are obtained using a Gaussian copula approach. The covariance matrix accounts only for the temporal dependencies while completely ignoring the spatial ones. This is done by constructing the precision matrix \mathbf{Q} as described in Section E.4.3, but setting $\rho = q_1 = q_{43} = 1$ and $b_{-1} = b_0 = b_1 = c_{-1} = c_0 = c_1 = 0$. That is, the precision matrix in this case is described by the parameters a_1 and σ^2 only. This model does not allow for any special arrangement for the boundary points. The conditional precisions are assumed to be constant. In other words, this model corresponds to a conventional stationary GMRF defined by the first order autoregressive process in time.
3. *Separable model with first order decays in time and in space* allowing for non-constant conditional precisions: The corresponding multivariate densities are obtained using a Gaussian copula approach. The precision matrix \mathbf{Q} is parametrized as in Section E.4.3 while setting $c_0 = b_0$, $b_1 = b_{-1} = c_1 = c_{-1} = a_1 * b_0$. That is, the precision matrix in this case is described the first order time-dependence (given by a_1) and the first order spatial dependence (given by b_0). Additionally, the model gives more flexibility compared with the conventional separable covariance structures

by considering non-constant conditional precisions (modelled by ρ , q_1 and q_{43}). The model does not account for the directional influence, and that is why c_j is set to be equal to b_j with all $j = -1, \dots, 1$

4. *Sample correlation:* The corresponding multivariate predictive densities are obtained using a Gaussian copula approach with the correlation structure given by the sample correlation matrix.
5. *Full model:* The first order model which proposed in this study. That is the precision matrix is described by the full parameter vector $\boldsymbol{\theta}$ as given in eq. (E.24).

E.6.2.2 Choosing an appropriate scoring rule for the quality evaluation

In order to evaluate and compare the overall quality of multivariate probabilistic forecasts proper scoring rules are to be employed [31, 32]. An overview of proper scoring rules used for the multivariate forecast verification is given in [33]. In this work the Logarithmic score is used as a lead score for evaluating the performance of the joint predictive densities. The logarithmic scoring rule, s , is defined as

$$s(p(\mathbf{x}), \mathbf{x}_t) = -\ln(p(\mathbf{x}_t)) \quad (\text{E.31})$$

Where $p(\mathbf{x})$ stands for the predictive density, which in our case is given by $\mathcal{N}(\mathbf{0}, \mathbf{Q}(\boldsymbol{\theta})^{-1})$. \mathbf{x}_t denotes the corresponding observation.

Suppose, the verification set consists of H observations, then the overall score, S , is given by the average value of the corresponding $s(p(\mathbf{x}), \mathbf{x}_t)$

$$S(p(\mathbf{x})) = -\frac{\sum_{t=1}^H \ln(p(\mathbf{x}_t))}{H} \quad (\text{E.32})$$

That is, essentially the Logarithmic score is given by the average minus log likelihood derived from the observations. Therefore, this score is negatively orientated.

There are several reasons for choosing the Logarithmic score as the lead evaluation criterion.

Firstly, it is consistent with the optimization criterion used when estimating the model parameters.

Secondly, allowing for some affine transformations, this is the only local proper score (see Theorem 2 in [34]). Locality means that the score depends on the predictive distribution only through the value which the predictive density attains at the observation [31]. An important advantage of using local scores when dealing with multivariate predictive densities comes with the related computational benefits. When dealing with local scores, there is no need to draw random samples from the predictive density in order to make the evaluation.

For instance, an alternative is to use the Energy score (see detailed information on this in [33]). This score is non-local and is based on the expected Euclidean distance between forecasts and the related observations. Most often, closed form expressions for such expectation are unavailable and one needs to employ Monte Carlo methods in order to estimate the score [33]. When dealing with problems of a very high dimension, Monte Carlo techniques result in computational challenges.

On the downside of local scores is their sensitivity to outliers. For instance, the Logarithmic score is infinite if the forecast assigns a vanishing probability to the event which occurs. In practice, when working with the real data, such sensitivity might be a problem.

In this work, we considered both the Energy score and the Logarithmic score for the final density evaluation. In general the results suggested by the two scores were consistent and no contradictions were observed. However, what we noticed is that the Energy score was not very sensitive to the changes in the correlation structure. That is, the changes in the Energy score when moving from the assumption of independence between the marginal predictive densities to models accounting for the dependence structure were rather small (even though they still proved statistically significant based on Diebold-Mariano test statistics [35]). This is caused by low sensitivity of the Energy score to changes in the dependence structure as argued in [36]. This is another reason to focus on the Logarithmic score further in this study.

E.6.2.3 Empirical results

One can appreciate the importance of accounting for the dependence structure from the fact that multivariate predictive densities derived from the independence assumption are shown to be of the lowest quality (see results in Table E.1). The full model proposed in this study outperforms another two

Table E.1: Quality assessment of the predictive densities in terms of the Logarithmic score (S).

Model	Nr. of parameters	S
Independent	0	853.14
First order in time	1	409.98
Separable space-time model	6	357.84
Full model	10	318.07
Sample correlation	207690	267.96

considered dependency structures: first order time-dependence as well as the separable space-time model. Statistical significance of the improvements was verified using a likelihood ratio test [37]. This confirms that letting the related conditional correlations change depending on the direction as well as allowing for non-separable space-time influence results in better quality of the multivariate probabilistic forecasts.

Predictive densities defined by the sample correlation matrix provide the best quality forecasts. This is also expected, since in this study the estimation period consisted of one year of hourly data. Large amount of data made it possible to estimate the covariance structure of the given dimension. However, the main interest in the future is to make the covariance structure dependent on meteorological conditions. In this setup, tracking sample covariance will become impossible. Thus, the proposed parametrization is crucial for further development of the methodology as it significantly reduces the effective problem dimension.

E.6.3 Scenario generation

As an illustration of probabilistic forecasts obtained with the proposed approach Fig. E.9 shows five scenarios describing wind power generation at zones 6 and 7 from 1 to 43 hours ahead issued on the 15th of June, 2007, at 01:00.

One can see that the scenarios generated using the model proposed in this study respect dependencies both in time and in space. Respecting correlations in time ensures that the corresponding scenarios evolve smoothly with time. That is, given that a scenario predicts wind power generation at time t to be far from the marginal expectation, then the power generation at time $t + 1$ is also expected to deviate a lot from its marginal expectation. As an example see scenario 5 for zone 6 for lead times from 22 to 30 hours ahead.

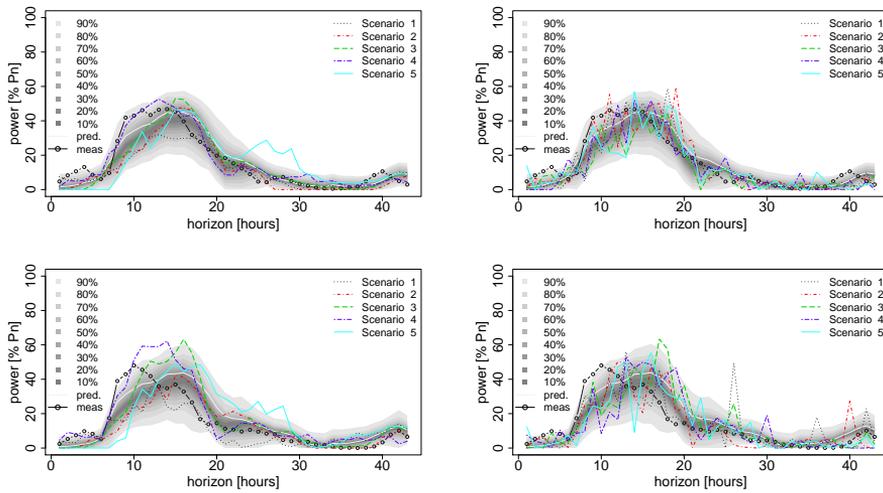


Figure E.9: Scenarios describing wind power generation at zones 6 (top) and 7 (bottom) from 1 to 43 hours ahead issued on the 15th of June, 2007, at 01:00. The scenarios given in the left column correspond to the ones obtained with the model proposed in this study, meanwhile the scenarios on the right are obtained under the assumption of independent marginals, thus, not respecting neither temporal, not spatial dependencies in the data.

Respecting spatial dependency between the zones ensures that when large (small) forecast errors are observed at one zone, the errors at the other zone are also expected to be large (small). This is also visible from Fig. E.9. For example, in the case of scenario 4, wind power generation deviates a lot from the expected value in both zones 6 and 7.

On another hand, one can see that the corresponding scenarios generated using the independent model do not respect neither temporal, not spatial dependencies in the data.

E.7 Conclusions

This study considers the problem of obtaining a joint multivariate predictive density for describing wind power generation at a number of sites over a period

of time from the set of marginal predictive densities, targeting each site and each lead time individually. A Gaussian copula approach has been employed for this purpose. The novelty of the proposed methodology consists in the proposed parametrization of the dependence structure. More specifically, instead of modelling the covariance matrix directly, focus is given to its inverse (precision matrix). This solution results in several benefits.

Firstly, the precision matrix is shown to be very sparse. This puts us in the framework of Gaussian Markov random fields and results in computational benefits due to the faster factorization algorithms available for sparse matrices.

Secondly, the proposed parametrization allows for more flexibility as one can easily obtain non separable in space and in time dependence structures following a more complex pattern than the conventional exponential decay in time (and/or space). Additionally, the study has revealed that the empirical precision matrix is given by the non-constant conditional precisions as well as by the varying conditional correlations. This puts us beyond the framework of the conventional approaches given by the homogeneous stationary Gaussian fields. We propose a way to model the changes in the conditional precisions and we permit for conditional correlations to change with the direction. Accounting for such directional influence is not only clearly necessary when looking at the data, but it is also quite intuitive, provided that wind power forecast errors propagate in time and in space under the influence of meteorological conditions.

All the empirical results were obtained by considering a test case of 15 groups of wind farms covering the territory of western Denmark. The results have shown that the joint predictive densities derived from the proposed methodology outperform the benchmark approaches in terms of the overall quality.

Additionally, the study raised a number of new questions and gave ideas for future work.

Firstly, when considering the same problem setup, the direct extension of the proposed methodology could be given by conditioning the precision matrix on the meteorological conditions. Specifically, we suggest that the precision matrix would change with the prevailing wind direction. The easiest way to account for this would be to employ a regime switching approach by allowing a neighbourhood structure to change with the wind direction. In other words, instead of distinguishing between “West-East” and “North-South” neighbourhood as we did in this study, one could then consider “Up Wind”-“Down Wind” and “Concurrent”-“Concurrent”. Also, it would be interesting to investigate ways to explain the variations in the conditional precisions among the zones. Possibly some clustering techniques could be employed.

Further, an interesting challenge is to move from the lattice setup considered in this study to a fully continuous approach. Based on [38] there is a link between stochastic partial differential equations and some type of precision matrices. Thus, by understanding how the elements of the precision matrix evolve with distance between the zones and prevailing meteorological conditions, one can get a process description via stochastic partial differential equations.

Another interesting challenge comes with the verification of probabilistic forecasts of a (very) large dimension. Already when working with a dimension of 645, we have faced certain challenges when considering the different scoring rules available for multivariate probabilistic forecast verification. In this study the Logarithmic and the Energy score have been considered. Both scores are proper, thus in theory they can both be used for the forecast verification exercise. However, each of them is associated with some challenges.

The Energy score, being a non-local score, comes with associated computational challenges, since its estimation requires Monte Carlo techniques. Furthermore, following [36], this score has low sensitivity to changes in covariance structure.

The challenges associated with the likelihood-based inference are given by its sensitivity to outliers which might cause difficulties in practical applications. Moreover, following [30] the log likelihood criterion is not robust to model errors, which may result in inconsistent estimates.

As a conclusion, more research is needed in order to propose better ways (more informative, robust and computationally feasible) to evaluate probabilistic forecasts of multivariate quantities.

Acknowledgement

Acknowledgments are due to Roland Löwe for his valuable editorial comments and corrections.

References E

- [1] T. Ackermann *et al.*, *Wind power in power systems*, vol. 140. Wiley Online Library, 2005.
- [2] L. Jones and C. Clark, “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*, 2011.
- [3] 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*, vol. 12, no. 6, pp. 1725–1744, 2008.
- [4] G. Giebel, R. Brownsword, G. Kariniotakis, M. Denhard, and C. Draxl, “The state-of-the-art in short-term prediction of wind power—A literature overview. 2nd edition,” tech. rep., ANEMOS.plus, 2011.
- [5] T. Gneiting, “Editorial: probabilistic forecasting,” *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, vol. 171, no. 2, pp. 319–321, 2008.
- [6] A. Botterud, Z. Zhou, J. Wang, R. Bessa, H. Keko, J. Sumaili, and V. Miranda, “Wind power trading under uncertainty in LMP markets,” *Power Systems, IEEE Transactions on*, vol. 27, no. 2, pp. 894–903, 2012.
- [7] X. Liu, “Economic load dispatch constrained by wind power availability: A wait-and-see approach,” *Smart Grid, IEEE Transactions on*, vol. 1, no. 3, pp. 347–355, 2010.
- [8] A. Botterud, Z. Zhou, J. Wang, J. Sumaili, H. Keko, J. Mendes, R. J. Bessa, and V. Miranda, “Demand dispatch and probabilistic wind power

- forecasting in unit commitment and economic dispatch: A case study of Illinois,” *Sustainable Energy, IEEE Transactions on*, vol. 4, no. 1, pp. 250–261, 2013.
- [9] C. Liu, J. Wang, A. Botterud, Y. Zhou, and A. Vyas, “Assessment of impacts of PHEV charging patterns on wind-thermal scheduling by stochastic unit commitment,” *Smart Grid, IEEE Transactions on*, vol. 3, no. 2, pp. 675–683, 2012.
- [10] Á. J. Duque, E. D. Castronuovo, I. Sánchez, and J. Usaola, “Optimal operation of a pumped-storage hydro plant that compensates the imbalances of a wind power producer,” *Electric Power Systems Research*, vol. 81, no. 9, pp. 1767–1777, 2011.
- [11] R. J. Bessa, M. A. Matos, I. C. Costa, L. Bremermann, I. G. Franchin, R. Pestana, N. Machado, H. P. Waldl, and C. Wichmann, “Reserve setting and steady-state security assessment using wind power uncertainty forecast: a case study,” *IEEE Transactions on Sustainable Energy*, 2012.
- [12] M. A. Ortega-Vazquez and D. S. Kirschen, “Assessing the impact of wind power generation on operating costs,” *Smart Grid, IEEE Transactions on*, vol. 1, no. 3, pp. 295–301, 2010.
- [13] R. J. Bessa, V. Miranda, A. Botterud, Z. Zhou, and J. Wang, “Time-adaptive quantile-copula for wind power probabilistic forecasting,” *Renewable Energy*, vol. 40, no. 1, pp. 29–39, 2012.
- [14] G. Papaefthymiou and D. Kurowicka, “Using copulas for modeling stochastic dependence in power system uncertainty analysis,” *Power Systems, IEEE Transactions on*, vol. 24, no. 1, pp. 40–49, 2009.
- [15] S. Hagspiel, A. Papaemannouil, M. Schmid, and G. Andersson, “Copula-based modeling of stochastic wind power in Europe and implications for the Swiss power grid,” *Applied Energy*, vol. 96, pp. 33–44, 2012.
- [16] P. Pinson, H. Madsen, H. A. Nielsen, G. Papaefthymiou, and B. Klöckl, “From probabilistic forecasts to statistical scenarios of short-term wind power production,” *Wind Energy*, vol. 12, no. 1, pp. 51–62, 2009.
- [17] H. Louie, “Evaluation of bivariate Archimedean and elliptical copulas to model wind power dependency structures,” *Wind Energy*, 2012.
- [18] P. P., *Estimation of the uncertainty in wind power forecasting*. PhD thesis, Ecole des Mines de Paris, 2006.
- [19] P. Pinson, H. A. Nielsen, J. K. Møller, H. Madsen, and G. N. Kariniotakis, “Non-parametric probabilistic forecasts of wind power: required properties and evaluation,” *Wind Energy*, vol. 10, no. 6, pp. 497–516, 2007.

- [20] M. Sklar, *Fonctions de répartition à n dimensions et leurs marges*. Université Paris 8, 1959.
- [21] M. Hofert, M. Mächler, and A. J. McNeil, “Likelihood inference for Archimedean copulas in high dimensions under known margins,” *Journal of Multivariate Analysis*, 2012.
- [22] G. D., “A note on the multivariate Archimedean dependence structure in small wind generation sites,” *Wind Energy*, 2013.
- [23] C. Genest and A.-C. Favre, “Everything you always wanted to know about copula modeling but were afraid to ask,” *Journal of Hydrologic Engineering*, vol. 12, no. 4, pp. 347–368, 2007.
- [24] D. Simpson, F. Lindgren, and H. Rue, “In order to make spatial statistics computationally feasible, we need to forget about the covariance function,” *Environmetrics*, vol. 23, no. 1, pp. 65–74, 2012.
- [25] J. Besag, “Spatial interaction and the statistical analysis of lattice systems,” *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 192–236, 1974.
- [26] J. Tastu, P. Pinson, and H. Madsen, “Multivariate Conditional Parametric Models for a spatiotemporal analysis of short-term wind power forecast errors,” in *Scientific Proceedings of the European Wind Energy Conference, Warsaw*, pp. 77–81, 2010.
- [27] R. Girard and D. Allard, “Spatio-temporal propagation of wind power prediction errors,” *Wind Energy*, 2012.
- [28] M. Kyung and S. K. Ghosh, “Maximum likelihood estimation for directional conditionally autoregressive models,” *Journal of Statistical Planning and Inference*, vol. 140, no. 11, pp. 3160–3179, 2010.
- [29] H. Rue and L. Held, *Gaussian Markov random fields: theory and applications*, vol. 104. Chapman & Hall, 2005.
- [30] H. Rue and H. Tjelmeland, “Fitting Gaussian Markov random fields to Gaussian fields,” *Scandinavian Journal of Statistics*, vol. 29, no. 1, pp. 31–49, 2002.
- [31] J. Bröcker and L. A. Smith, “Scoring probabilistic forecasts: The importance of being proper,” *Weather and Forecasting*, vol. 22, no. 2, pp. 382–388, 2007.
- [32] T. Gneiting and A. E. Raftery, “Strictly proper scoring rules, prediction, and estimation,” *Journal of the American Statistical Association*, vol. 102, no. 477, pp. 359–378, 2007.

-
- [33] T. Gneiting, L. I. Stanberry, E. P. Gritmit, L. Held, and N. A. Johnson, “Assessing probabilistic forecasts of multivariate quantities, with an application to ensemble predictions of surface winds,” *Test*, vol. 17, no. 2, pp. 211–235, 2008.
 - [34] J. M. Bernardo, “Expected information as expected utility,” *The Annals of Statistics*, vol. 7, no. 3, pp. 686–690, 1979.
 - [35] F. X. Diebold and R. S. Mariano, “Comparing predictive accuracy,” *Journal of Business & Economic Statistics*, vol. 13, no. 3, pp. 253–263, 1995.
 - [36] P. Pinson and J. Tastu, “Discrimination ability of the Energy score,” tech. rep., Technical University of Denmark, 2013.
 - [37] H. Madsen and P. Thyregod, *Introduction to general and generalized linear models*. CRC Press, 2011.
 - [38] F. Lindgren, H. Rue, and J. 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)*, vol. 73, no. 4, pp. 423–498, 2011.

PAPER F

Proper Evaluation of Neural Network and Learning Systems based Prediction Intervals

Authors:

Pierre Pinson, Julija Tastu

Submitted to :

IEEE Transactions on Neural Networks and Learning Systems, June 2013

Proper Evaluation of Neural Network and Learning Systems based Prediction Intervals

Pierre Pinson¹, Julija Tastu²

Abstract

Neural Network and Learning Systems approaches are increasingly used in probabilistic prediction. Forecast evaluation then comprises a complex task for which a number of scores have been proposed, aiming to summarize the assessment of their overall quality with a single number. Such scores ought to be proper though, that is, to effectively reward interval forecasts of higher quality. If not, the ranking of score values does not allow concluding on the actual superiority of a given approach over others, since one may always have the possibility to hedge in order to obtain the best score value. Recently, the Coverage Width-based Criterion (CWC) was proposed and used for an evaluation of the state of the art. The CWC score is shown to be improper based on theoretical considerations, while the consequences are explored.

F.1 Introduction

In different areas of forecasting, substantial developments are on proposing and applying probabilistic approaches, as for instance in economics and finance [1], meteorology [2], as well as in various aspects of power systems management e.g. electric load [3], wholesale market prices [4] and renewable energy production [5]. Among the alternative methodologies for probabilistic prediction, approaches based on Neural Networks and more generally learning systems have gained increased interest over the last two decades, from the early work of [6] to the recent review in [7].

Prediction intervals (also referred to as interval forecasts) are some of the probabilistic forecasts for continuous variables that attracted the most attention since

¹Department of Electrical Engineering at the Technical University of Denmark

²Department of Mathematics and Computer Science at the Technical University of Denmark

providing visual and easily interpretable information about forecast uncertainty. Rigorous methodologies are required for their evaluation and for the comparison of rival approaches. For examples of benchmark exercises, see [8, 9] among others. Evaluation and comparison are to be based on proven scores and diagnostic tools. Such scores ought to be *proper* [10, 11]: propriety is the basic property of a score to ensure that perfect forecasts should be given the best score value, say, the lowest one if the score is negatively oriented. It appears that such a crucial aspect is not always respected.

Emphasis is placed here on the recent proposal of the Coverage Width-based Criterion (CWC) [9] as a score for the comparison of Neural Network and Learning Systems (NNLS) approaches to issuing prediction intervals. The necessary background on proper scores for prediction intervals is first introduced in Section F.2. Subsequently, our original contribution consists in (i) showing in Section F.3 that the CWC score is not proper, and (ii) underlining in Section F.4 the consequences of employing this improper score, based on the example of a simple hedging strategy permitting to always obtain score values better than those of rival approaches. Finally in Section F.5, we conclude on the fact that, owing to the lack of property, the ranking of CWC score values does not allow concluding on the actual superiority of a given approach over another, shedding doubts on the evaluation of the state of the art performed in [9].

F.2 Proper scores for prediction intervals

Probabilistic forecast verification frameworks were proposed over the last 30 years, their main principles being underlined in [11, 12, 13]. They involve the evaluation of calibration (correspondence of nominal and empirical probabilities), as well as sharpness (concentration of probability—the tighter the better) using a set of diagnostic tools and scores.

Consider a stochastic process $\{G_t\}_t$ for which interval forecasts are to be generated, say, for a lead time k . $\hat{I}_{t+k|t}^{(\beta)}$ are central prediction intervals with nominal coverage rate $(1 - \beta)$, issued at time t for lead time $t + k$,

$$\hat{I}_{t+k|t}^{(\beta)} = \left[\hat{q}_{t+k|t}^{(\underline{\alpha})}, \hat{q}_{t+k|t}^{(\overline{\alpha})} \right], \quad (\text{F.1})$$

where $\hat{q}_{t+k|t}^{(\underline{\alpha})}$ and $\hat{q}_{t+k|t}^{(\overline{\alpha})}$ are quantile forecasts whose nominal levels $\underline{\alpha}$ and $\overline{\alpha}$, such that $\underline{\alpha} = 1 - \overline{\alpha} = \beta/2$.

The joint assessment of calibration and sharpness ideally relies on a unique

criterion, a score Sc , that assigns a single value $\text{Sc}(\hat{\mathbf{I}}_{t+k|t}^{(\beta)}, y_{t+k})$ to each forecast-observation pair, then to be averaged over an evaluation set, $t = 1, \dots, T$. If knowing $\{G_t\}_t$, perfect interval forecasts $\hat{\mathbf{I}}_{t+k|t}^{(\beta)*}$ would be

$$\hat{\mathbf{I}}_{t+k|t}^{(\beta)*} = \left[q_{t+k}^{(\underline{\alpha})}, q_{t+k}^{(\overline{\alpha})} \right], \tag{F.2}$$

with $q_{t+k}^{(\underline{\alpha})}$ and $q_{t+k}^{(\overline{\alpha})}$ the actual quantiles of the process at time $t+k$. Throughout the paper, the “*”-symbol will be associated to perfect forecasts and their score value.

Following [10], a score Sc for prediction intervals is said to be proper if for any prediction interval $\hat{\mathbf{I}}_{t+k|t}^{(\beta)}$ and corresponding observation y_{t+k} ,

$$\text{Sc} \left(\hat{\mathbf{I}}_{t+k|t}^{(\beta)*}, y_{t+k}; \boldsymbol{\theta} \right) \leq \text{Sc} \left(\hat{\mathbf{I}}_{t+k|t}^{(\beta)}, y_{t+k}; \boldsymbol{\theta} \right), \quad \forall t, k, \beta, \boldsymbol{\theta}, \tag{F.3}$$

i.e., perfect interval forecasts $\hat{\mathbf{I}}_{t+k|t}^{(\beta)*}$ are to be assigned the lowest possible score value. Better, a score for intervals is strictly proper if having a strict inequality in (F.3).

Employing proper scores ensures consistency in forecast verification and when comparing rival approaches. Perfect interval forecasts are obviously not available when focusing on real-world processes. It is hence impossible to have them as a reference the other approaches should try to get close to. However, the mere idea that already from theoretical considerations perfect forecasts would not be given the optimal score value can only bring discredit on the ranking of rival forecasting methods. Worse, as will be shown and discussed in the following, it discourages competitors to issue their best forecasts, instead turning into an incentive to hedge by playing the score.

Following the pioneering work of [14], it was shown (see e.g. [15]) that a family of proper scores for interval forecasts can be readily obtain from scoring rules for its defining quantiles. For instance, the proper score proposed by [14] is defined as

$$\begin{aligned} \text{Sc} \left(\hat{\mathbf{I}}_{t+k|t}^{(\beta)}, y_{t+k} \right) &= (\hat{q}_{t+k|t}^{(\overline{\alpha})} - \hat{q}_{t+k|t}^{(\underline{\alpha})}) \\ &+ \frac{2}{\beta} (\hat{q}_{t+k|t}^{(\underline{\alpha})} - y_{t+k}) \mathbf{1}\{y_{t+k} \leq \hat{q}_{t+k|t}^{(\underline{\alpha})}\} \\ &+ \frac{2}{\beta} (y_{t+k} - \hat{q}_{t+k|t}^{(\overline{\alpha})}) \mathbf{1}\{y_{t+k} \geq \hat{q}_{t+k|t}^{(\overline{\alpha})}\}. \end{aligned} \tag{F.4}$$

It naturally rewards sharp intervals, while penalizing cases where the observation is not covered. It is to be averaged over an evaluation set, $t = 1, \dots, T$.

F.3 The Coverage Width-based Criterion is not a proper score

F.3.1 Definition of the Coverage Width-based Criterion

The calibration assessment of interval forecasts follows a frequentist approach, by comparing their empirical and nominal coverage rates. The empirical coverage rate is derived based on the indicator variable $\xi_{t,k}$, defined for a prediction interval $\hat{I}_{t+k|t}^{(\beta)}$ and corresponding observation y_{t+k} as

$$\xi_{t,k} = \mathbf{1}\{y_{t+k} \in \hat{I}_{t+k|t}^{(\beta)}\} = \begin{cases} 1, & \text{if } y_{t+k} \in \hat{I}_{t+k|t}^{(\beta)} \\ 0, & \text{otherwise} \end{cases}, \quad (\text{F.5})$$

i.e., as a binary variable indicating if the observation lies or not within the prediction interval. Subsequently, the empirical coverage rate b_k , for nominal proportion $(1 - \beta)$ and lead time k , is obtained by calculating the mean of the time-series $\{\xi_{t,k}\}_t$ over an evaluation set of length T ,

$$b_k = \frac{1}{T} \sum_{t=1}^T \xi_{t,k}. \quad (\text{F.6})$$

The difference $\Delta b_k = (1 - \beta) - b_k$ between nominal and empirical coverage rates is to be seen as a probabilistic bias of the prediction intervals, also referred to as calibration deficit.

Since for probabilistically calibrated intervals, sharpness is a desired property, one may monitor the width of prediction intervals and then derive some summary statistics. The average prediction interval width over an evaluation set of length T is

$$\bar{\delta}_k = \frac{1}{T} \sum_{t=1}^T \hat{q}_{t+k|t}^{(\bar{\alpha})} - \hat{q}_{t+k|t}^{(\underline{\alpha})}. \quad (\text{F.7})$$

Using these measures for probabilistic calibration and sharpness, the CWC score, for a give lead time k and nominal coverage rate $(1 - \beta)$, was introduced by [9] as

$$\text{CWC} = \bar{\delta}_k (1 + \mathbf{1}\{\Delta b_k > 0\} \exp(\eta \Delta b_k)), \quad \eta > 0, \quad (\text{F.8})$$

This score penalizes interval forecasts if their empirical coverage rate is lower than the nominal one, while it rewards sharpness otherwise. At first sight it could be seen as similar in essence to the score in (F.4).

F.3.2 Why is the CWC score not proper?

If the CWC score were proper, there could not be any set of interval forecasts getting a CWC score value lower than that of perfect interval forecasts. In view of (F.8), the CWC score value for perfect interval forecasts would simplify to their average width. Let us denote by $\bar{\delta}_k^*$ that optimal score value, with

$$\bar{\delta}_k^* = \frac{1}{T} \sum_{t=1}^T q_{t+k}^{(\bar{\alpha})} - q_{t+k}^{(\underline{\alpha})} \tag{F.9}$$

In the case where one would want to get a better score than $\bar{\delta}_k^*$, the only way is to issue sharper forecasts, since wider intervals can only lead to higher CWC score values anyway. Sharpening the prediction intervals necessarily comes at the expense of calibration, since empirical coverage would get lower than $(1 - \beta)$, yielding a calibration deficit $\Delta b_k > 0$.

Hedging the CWC score in order to obtain a value lower than $\bar{\delta}_k^*$ translates to

$$\bar{\delta}_k (1 + \exp(\eta \Delta b_k)) < \bar{\delta}_k^*, \quad \eta > 0, \tag{F.10}$$

which, after a little algebra, yields the following inequality on the free parameter η ,

$$0 < \eta < \frac{\ln(\bar{\delta}_k^*/\bar{\delta}_k)}{\Delta b_k}, \quad \text{with} \quad \frac{\ln(\bar{\delta}_k^*/\bar{\delta}_k)}{\Delta b_k} > 0, \tag{F.11}$$

The above inequality demonstrates that in principle, there always exists a value of η such that some forecasts could be given a better score than that for perfect forecasts, since $\bar{\delta}_k^*/\bar{\delta}_k > 1$ (otherwise one would not have sharper forecasts), while the calibration deficit Δb_k is finite, $\Delta b_k \in [0, (1 - \beta)]$.

Based on inequality (F.11), some might say that the CWC is a conditionally proper score, where one simply has to pick a value for η high enough to ensure no imperfect prediction intervals could get a CWC score lower than $\bar{\delta}_k^*$. However, please consider the limit of the right-hand side quantity in inequality (F.11) as $\bar{\delta}_k$ tends towards infinity,

$$\lim_{\bar{\delta}_k \rightarrow 0} \frac{\ln(\bar{\delta}_k^*/\bar{\delta}_k)}{\Delta b_k} = +\infty. \tag{F.12}$$

This is since the numerator necessarily tends towards infinity while the denominator stays finite.

As a consequence, whatever the value chosen for η , any interval forecasts in the form of a probability mass would be given a CWC score values better than $\bar{\delta}_k^*$.

Looking again at the definition of the CWC score itself, one indeed observes that

$$\text{CWC} \geq 0, \quad (\text{F.13})$$

$$\text{CWC} = 0, \quad \text{if } \bar{\delta}_k = 0. \quad (\text{F.14})$$

With this score, no one can outperform prediction intervals defined as probability masses on a single value, whatever this value. As an illustrative example, simply predict your favorite number all the time, and you will win any forecast comparison or competition where the CWC is used as the lead score.

F.4 Illustrating the consequences of not using a proper score

Let us look at this issue of not employing a proper score in a more practical manner here, by concentrating on the example of simple hedging strategies in a benchmark exercise or forecast competition setup where the CWC is used as the lead score. In a such a case the free parameter η would be fixed. For instance in the case of Ref. [9], $\eta = 50$.

A competitor knowing the behavior of the CWC score underlined in the above and aiming to win the benchmark exercise, will not be tempted to issue his best forecasts, but instead to hedge in order to obtain a better score value. The forecast intervals of the best performing competitor consist in a set of prediction intervals $\{\hat{I}_{t+k|t}^{(\beta)\diamond}\}_t$, with average width $\bar{\delta}_k^\diamond$ and calibration deficit Δb_k^\diamond . Their CWC score is

$$\bar{\delta}_k^\diamond (1 + \mathbf{1}\{\Delta b_k^\diamond > 0\} \exp(\eta \Delta b_k^\diamond)). \quad (\text{F.15})$$

Consequently, the hedging strategy is simply to sharpen such intervals. To keep it simple, consider a linear scaling factor to be applied to every individual prediction interval. Similarly to (F.10) and after a little algebra, a simple condition for the CWC score value of any interval to be lower than that of the best performing interval is

$$\delta_k < \nu \bar{\delta}_k^\diamond, \quad (\text{F.16})$$

where

$$\nu = \frac{1 + \mathbf{1}\{\Delta b_k^\diamond > 0\} \exp(\eta \Delta b_k^\diamond)}{1 + \mathbf{1}\{\Delta b_k > 0\} \exp(\eta \Delta b_k)}. \quad (\text{F.17})$$

From the above, since the numerator is necessarily greater than 1, while $\Delta b_k \leq (1 - \beta)$, one has

$$\nu \geq \frac{1}{1 + \exp(\eta(1 - \beta))}, \quad (\text{F.18})$$

where this bounding value is independent of the intervals themselves and of the potential calibration deficit. It is hence straightforward to find a maximum value for the scaling factor. As long as one defines prediction intervals such that

$$\delta_k < \frac{1}{1 + \exp(\eta(1 - \beta))} \delta_k^\diamond, \quad (\text{F.19})$$

the CWC score of that competitor will be best, without having made any attempt to issue high-quality prediction intervals. Note that even if the best performing interval forecasts were not known and shared among participants, the best strategy is simply to issue prediction intervals as sharp as possible, even if leading to an obvious calibration deficit.

F.5 Conclusions

NNLS-based approaches have a great role to play in the development of probabilistic forecasting methodologies, for a wide range of applications of industrial and societal relevance. It is of utmost importance, however, for the state of the art to progress and to be regularly evaluated on a solid theoretical basis. This translates to the mandatory usage of proper scores, which can leave no doubt on the meaning of score values (and corresponding ranking of rival approaches) in forecast competition and benchmark exercises. Our aim here was to insist on this aspect, based on the recent example of the improper CWC score. Without this basic propriety requirement, it is unfortunately impossible to validate the evaluation of the state of the art in interval forecasting performed with this score as a lead criterion in [9], and to trust any other forecast comparison that would rely on the CWC. It is therefore suggested that future work focusing on probabilistic forecasting with any form of artificial intelligence and machine learning approaches rely on proper scores only.

References F

- [1] A. S. Tay and K. F. Wallis, “Density forecasting: a survey,” *Journal of Forecasting*, vol. 19, pp. 235–254, 2000.
- [2] M. Leutbecher and T. N. Palmer, “Ensemble forecasting,” *Journal of Computational Physics*, vol. 227, no. 7, pp. 3515–3539, 2008.
- [3] A. Khosravi, S. Nahavandi, and D. Creighton, “Construction of optimal prediction intervals for load forecasting problems,” *Power Systems, IEEE Transactions on*, vol. 25, no. 3, pp. 1496–1503, 2010.
- [4] J. H. Zhao, Z. Y. Dong, Z. Xu, and K. P. Wong, “A statistical approach for interval forecasting of the electricity price,” *Power Systems, IEEE Transactions on*, vol. 23, no. 2, pp. 267–276, 2008.
- [5] P. Pinson and G. Kariniotakis, “Conditional prediction intervals of wind power generation,” *Power Systems, IEEE Transactions on*, vol. 25, no. 4, pp. 1845–1856, 2010.
- [6] J. T. G. Hwang and A. A. Ding, “Prediction intervals for artificial neural networks,” *Journal of the American Statistical Association*, vol. 92, no. 438, pp. 748–757, 1997.
- [7] M. Paliwal and U. A. Kumar, “Neural networks and statistical techniques: A review of applications,” *Expert Systems with Applications*, vol. 36, no. 1, pp. 2–17, 2009.
- [8] G. Papadopoulos, P. J. Edwards, and A. F. Murray, “Confidence estimation methods for neural networks: A practical comparison,” *Neural Networks, IEEE Transactions on*, vol. 12, no. 6, pp. 1278–1287, 2001.

-
- [9] A. Khosravi, S. Nahavandi, D. Creighton, and A. F. Atiya, "Comprehensive review of neural network-based prediction intervals and new advances," *Neural Networks, IEEE Transactions on*, vol. 22, no. 9, pp. 1341–1356, 2011.
- [10] J. Bröcker and L. A. Smith, "Scoring probabilistic forecasts: The importance of being proper," *Weather and Forecasting*, vol. 22, no. 2, pp. 382–388, 2007.
- [11] T. Gneiting, F. Balabdaoui, and A. E. Raftery, "Probabilistic forecasts, calibration and sharpness," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 69, no. 2, pp. 243–268, 2007.
- [12] A. P. Dawid, "Present position and potential developments: Some personal views: Statistical theory: The prequential approach," *Journal of the Royal Statistical Society. Series A (General)*, pp. 278–292, 1984.
- [13] F. X. Diebold, A. T. Gunther, and A. S. Tay, "Evaluating density forecasts with applications to financial risk management," *International Economic Review*, vol. 39, pp. 863–883, 1984.
- [14] R. L. Winkler, "A decision-theoretic approach to interval estimation," *Journal of the American Statistical Association*, vol. 67, no. 337, pp. 187–191, 1972.
- [15] T. Gneiting and A. E. Raftery, "Strictly proper scoring rules, prediction, and estimation," *Journal of the American Statistical Association*, vol. 102, no. 477, pp. 359–378, 2007.

PAPER G

Discussion of "Prediction intervals for short-term wind farm generation forecasts" and "Combined nonparametric prediction intervals for wind power generation"

Authors:

Pierre Pinson, Julija Tastu

Submitted to :

IEEE Transactions on Sustainable Energy, submitted, June 2013

Discussion of "Prediction intervals for short-term wind farm generation forecasts" and "Combined nonparametric prediction intervals for wind power generation"

Pierre Pinson¹, Julija TASTU²

In a series of recent work published in the *IEEE Transactions on Neural Networks and Learning Systems*, the *IEEE Transactions on Power Systems*, *Electric Power Systems Research* and here in the *IEEE Transactions on Sustainable Energy* (among others), Khosravi and co-authors propose and utilize a new score for the evaluation of interval forecasts, the so-called Coverage Width-based Criterion (CWC). This score has been used for the tuning (in-sample) and genuine evaluation (out-of-sample) of prediction intervals for various applications, e.g. electric load [1], electricity prices [2], general purpose prediction [3] and wind power generation [4, 5]. Indeed, two papers by the same authors appearing in the *IEEE Transactions on Sustainable Energy* employ that score, and use it to conclude on the comparative quality of alternative approaches to interval forecasting of wind power generation.

Probabilistic forecasting is to become a core aspect in modern power systems engineering, with increased penetration of renewable energy sources, and with their inherent variability and lack of predictability, e.g., wind and solar energy. Besides, load patterns are becoming more variable and less predictable due to changes in consumption patterns with the apparition of proactive prosumers. It will overall result in more uncertainty in market-clearing outcomes such as energy volumes and prices. Probabilistic forecasts in the form of quantiles, intervals, predictive densities or more generally trajectories, are optimal inputs to a wide range of decision-making problems defined in a stochastic or robust optimization framework. These forecasts are more difficult to evaluate than the more common single-valued (/deterministic) predictions, owing to their very nature. Scoring rules to be used for the evaluation of probabilistic forecasts are required to be proper [6, 7, 8]: propriety is the basic property of a score to insure that perfect forecasts should be given the best score value, say, the lowest one if the score is negatively oriented. If not the case, one could then hedge the score, by finding tricks that permit to get better score values without

¹Department of Electrical Engineering at the Technical University of Denmark

²Department of Mathematics and Computer Science at the Technical University of Denmark

attempting to issue better forecasts. More generally, employing a score that is not proper makes than one can never be sure of the validity of the results from an empirical comparison or benchmarking of rival approaches. Research on the topic of proper evaluation of probabilistic forecasts, in the form of prediction intervals, can at least be traced back to the work of Winkler [9].

Unfortunately in the case of the aforementioned papers on interval forecasting of wind power generation (and other quantities), the CWC score employed is not proper, as will be illustrated below based on a simple example. As a consequence, it is difficult to appraise the quality of the results in these manuscripts. This is while there exists simple known scoring rules that could be readily used instead, for instance inspired by the original proposal of Winkler [9].

Let us first remind the reader about the definition of the CWC score. For a given lead time and nominal coverage rate $(1 - \beta)$, it writes

$$\text{CWC} = \bar{\delta}(1 + \mathbf{1}\{\Delta b > 0\} \exp(\eta\Delta b)), \quad \eta > 0, \quad (\text{G.1})$$

with $\mathbf{1}\{\cdot\}$ an indicator function, returning 1 if the condition between brackets realizes, and to 0 otherwise. In parallel, $\Delta b = (1 - \beta) - b$ is the difference between nominal $(1 - \beta)$ and empirical (b) coverage rates (that is, a form of probabilistic bias), while $\bar{\delta}$ is the average width of the prediction intervals. η is a free parameter that can be set to any positive value. It is argued that based on the above definition, the CWC penalizes intervals that are not probabilistically reliable, while it rewards them for their sharpness (since sharp intervals are intuitively expected to be more informative). The CWC is negatively oriented: lower values indicate prediction intervals of higher quality.

We now introduce a simple example in order to show how the CWC is not proper and may give a better score value to intervals that should actually be deemed of lower quality. Consider a stochastic process $\{X_t, t = 1, \dots, T\}$ defined as a sequence of T independent and identically distributed (i.i.d.) random variables X_t with probability density function (pdf) defined on a compact support, with

$$g(x) = 12 \left(x - \frac{1}{2}\right)^2, \quad x \in [0, 1]. \quad (\text{G.2})$$

We denote by G the cumulative distribution function (cdf) associated to g , given by

$$G(x) = 4 \left(x - \frac{1}{2}\right)^3 + \frac{1}{2}, \quad x \in [0, 1]. \quad (\text{G.3})$$

One can readily verifies that G is an increasing function, with $G(0) = 0$ and $G(1) = 1$.

For this stochastic process consisting of i.i.d. random variables, it straightforward to define the optimal interval forecasts directly based on the density in (G.2). For instance, for a nominal coverage rate of 0.9 (to cover observations 90% of the times), optimal central prediction intervals \mathcal{I}_t^* for any time t are defined by the quantiles with nominal levels 0.05 and 0.95:

$$\mathcal{I}^* = [G^{-1}(0.05), G^{-1}(0.95)]. \quad (\text{G.4})$$

And, based on the expression for G given in (G.3),

$$\mathcal{I}^* = [0.017, 0.983]. \quad (\text{G.5})$$

These intervals are perfectly reliable by definition, and therefore the CWC value assessing their quality is equal to their average width, i.e., $\text{CWC}^* = 0.966$. Since the above prediction intervals are the perfect ones, no other intervals should be given a better score.

Now in order to hedge the score, simply consider generating prediction intervals in a binary manner, although acknowledging that the nominal coverage rate should be respected in practice. Following such a binary approach, intervals are defined as full intervals $[0,1]$ 90% of the times, and as empty intervals (i.e., any single value in $[0,1]$) 10% of the times. This writes

$$\mathcal{I} = \begin{cases} [0, 1], & \text{if } u_t \geq 0.1 \\ 0.5, & \text{otherwise} \end{cases}, \quad (\text{G.6})$$

using 0.5 as an example value for the empty intervals, and where u_t is a realization at time t from a sequence of i.i.d. uniform random variables $U_t \sim \mathcal{U}[0, 1]$. These intervals are clearly not sophisticated ones, and not informative at all. Since covering the actual observations of the process 90% of the times, by construction, their CWC score values is also given by their average width, that is, $\text{CWC} = 0.9$ (significantly lower than the value obtained for the perfect prediction intervals).

In the frame of an empirical investigation comparing the quality of alternative interval forecasting methods for the stochastic process $\{X_t, t = 1, \dots, T\}$, using the CWC score would lead to the conclusion that the binary-type of intervals are better than the optimal ones. Due to the lack of propriety of the CWC score, this type of problem may appear in any type of empirical investigation, making that one can never conclude on the respective quality of the interval forecasts being evaluated.

References G

- [1] A. Khosravi, S. Nahavandi, and D. Creighton, "Construction of optimal prediction intervals for load forecasting problems," *Power Systems, IEEE Transactions on*, vol. 25, no. 3, pp. 1496–1503, 2010.
- [2] A. Khosravi, S. Nahavandi, and D. Creighton, "A neural network-GARCH-based method for construction of prediction intervals," *Electric Power Systems Research*, vol. 96, pp. 185–193, 2013.
- [3] A. Khosravi, S. Nahavandi, D. Creighton, and A. F. Atiya, "Comprehensive review of neural network-based prediction intervals and new advances," *Neural Networks, IEEE Transactions on*, vol. 22, no. 9, pp. 1341–1356, 2011.
- [4] A. Khosravi, S. Nahavandi, and D. Creighton, "Prediction Intervals for Short-Term Wind Farm Power Generation Forecasts," *Sustainable Energy, IEEE Transactions on: available online*, 2013.
- [5] A. Khosravi and S. Nahavandi, "Combined nonparametric prediction intervals for wind power generation," *Sustainable Energy, IEEE Transactions on: available online*, 2013.
- [6] J. Bröcker and L. A. Smith, "Scoring probabilistic forecasts: The importance of being proper," *Weather and Forecasting*, vol. 22, no. 2, pp. 382–388, 2007.
- [7] T. Gneiting and A. E. Raftery, "Strictly proper scoring rules, prediction, and estimation," *Journal of the American Statistical Association*, vol. 102, no. 477, pp. 359–378, 2007.

- [8] T. Gneiting, F. Balabdaoui, and A. E. Raftery, “Probabilistic forecasts, calibration and sharpness,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 69, no. 2, pp. 243–268, 2007.
- [9] R. L. Winkler, “A decision-theoretic approach to interval estimation,” *Journal of the American Statistical Association*, vol. 67, no. 337, pp. 187–191, 1972.

PAPER H

Discrimination ability of the Energy score

Authors:

Pierre Pinson and Julija Tastu

Technical report

Discrimination ability of the Energy score

Pierre Pinson¹, Julija Tastu²

Abstract

Research on generating and verification of multivariate probabilistic forecasts has gained increased interest over the last few years. Emphasis is placed here on the evaluation of forecast quality with the Energy score, which is based on a quadratic scoring rule. While this score may be seen as appealing since being proper, we show that its *discrimination* ability may be limited when focusing on the dependence structure of multivariate probabilistic forecasts. For the case of multivariate Gaussian process, a theoretical upper for such discrimination ability is derived and discussed. This limited discrimination ability may eventually get compromised by computational and sampling issues, as dimension increases.

H.1 Introduction

Probabilistic forecasting has gained increased attention over the last decade, both in terms of theoretical and of more practical developments. This phenomenon touches a wide range of applications, from economics and finance [1, 2], to earthquake prediction [3], while it also has wide appeal in meteorology [4], and for weather-related processes like renewable energy production [5, 6] and floods [7]. Such a focus on probabilistic forecasting is justified by the fact that, even if forecast users may prefer being provided with single-valued forecasts easier to handle in decision-making processes, those should be preferably extracted from probabilistic forecasts in a decision-theoretical framework, by accounting for user-specific loss functions (see, e.g., [8]).

Probabilistic forecasts optimally take the form of predictive densities for the stochastic process considered. If decisions to be made involve a univariate

¹Technical University of Denmark, Electrical Engineering (DTU Elektro), Kgs. Lyngby, Denmark

²Technical University of Denmark, Applied Mathematics and Computer Science (DTU Compute), Kgs. Lyngby, Denmark

stochastic process only, or if they do not require modeling a dependence structure (multivariate or spatio-temporal), then only marginal predictive densities are required. These are referred to as marginal since being issued for each variable, location and lead time, individually. In the more general case of decision-making requiring to account for a dependence structure, probabilistic forecasts then ought to consist of multivariate predictive densities, hence describing both the marginal densities and the dependence structure.

Evaluating probabilistic forecasts is more complex than evaluating single-valued predictions, even though some of the basic principles may be seen as similar. The main lines of probabilistic forecast verification frameworks (and underlying theoretical concepts) can be found in, e.g., [9], [10] and [11] among others. Such verification techniques may rely on scores, diagnostic tools, and possibly hypothesis testing. For the case of predictive densities, both univariate and multivariate, a number of scores have been proposed and discussed, see for instance [9] and [12]. Emphasis is placed here on quadratic scoring rules for multivariate predictive densities, that is, more precisely, on the Energy score introduced by [9]. Our aim is to discuss its discrimination ability, i.e., its ability to give significantly different score values to forecasts of different quality.

The manuscript is organized as follows. Section H.2 recalls the background on probabilistic forecast verification based on scoring rules, while insisting on the fact that propriety of a score does not imply any discrimination ability. This section also illustrates how the Energy score has a substantially higher discrimination ability when misspecifying the mean of multivariate distributions, than in the case of misspecifying their variance or the dependence structure. Subsequently, some theoretical results are given in Section H.3 giving a higher bound on differences in Energy score values for the case of misspecification of the dependence structure of multivariate predictive densities, also accounting for the dimension of these forecasts. Note that the discussion and results are produced for the Gaussian case only, though it is commonly used in practice today, for instance for short-term forecasts of surface wind speeds [13, 14] or for seasonal forecasts of sea-surface temperatures [15]. The necessary mathematical developments for obtaining these results are gathered in an Appendix at the end of the manuscript. Finally, Section H.4 develops into a discussion on how to maximize the discrimination ability of quadratic scoring rules for multivariate probabilistic forecasts, also considering perspectives for future work on multivariate probabilistic forecast verification.

H.2 Discrimination ability of the Energy score

H.2.1 General setup

Let us place ourselves in a framework where a forecaster aims at issuing multivariate probabilistic forecasts in the form of predictive densities. He therefore considers a multivariate random variable $\mathbf{Y} \in \mathbb{R}^n$, $n > 1$. Write G the true distribution of \mathbf{Y} , $\mathbf{Y} \sim G$, while F is the multivariate predictive density issued by the forecaster at some point prior or equal to the current time. Time indices are not used here, since the results are independent of the time when the forecast is issued, and of the lead time considered. As an example, the multivariate random variable may be surface wind speed, expressed in its zonal and meridional components, as in the case of [13], [14], and [16]. More generally in meteorological prediction, it could also consist in a set of meteorological variables, e.g. wind speed, precipitation, etc., as in the case of [17]. In addition, the dependencies may not only be between various variables, but also for various geographical locations, and/or a set of times in the future [18]. Other setups exist in econometrics and finance related prediction problems, as for the example of the simultaneous confidence regions of [19] among others.

H.2.2 From propriety of scoring rules to their discrimination ability

When employing skill scores for probabilistic forecast verification, it is required that they are based on proper scoring rules, to ensure that forecasters really aim at issuing better forecasts, instead of focusing on hedging the score only. A scoring rule Sc is defined as a functional assigning a value to the association of a predictive density F with an observation \mathbf{y} from the real density G of the random variable,

$$\text{Sc} : (F, \mathbf{y}) \rightarrow \text{Sc}(F, \mathbf{y}) \in \mathbb{R} \quad (\text{H.1})$$

Formally, following the presentation by, e.g., [20], a scoring rule Sc (and associated score) for multivariate predictive densities, is said to be proper if and only if

$$\text{Sc}(G, \mathbf{y}) \leq \text{Sc}(F, \mathbf{y}), \quad (\text{H.2})$$

meaning, using simple words, that actual densities for the stochastic process are to be assigned the lowest possible score value. This result is for a negatively-oriented score, for which lowest values are seen as best. For simplicity, we only

consider negatively-oriented scores in the following. Better, the scoring rule is strictly proper if only the actual densities get the lowest score value, i.e.,

$$\text{Sc}(G, \mathbf{y}) < \text{Sc}(F, \mathbf{y}). \quad (\text{H.3})$$

Propriety is a property of scoring rules involving a predictive density and the real density of the stochastic process. In practice, that real density is not available anyway, and one is left with comparing alternative predictive densities, say, $F^{(1)}$ and $F^{(2)}$ generated by two rival forecasters. Propriety does not ensure that a difference in quality between $F^{(1)}$ and $F^{(2)}$ would yield a difference between $\text{Sc}(F^{(1)}, \mathbf{y})$ and $\text{Sc}(F^{(2)}, \mathbf{y})$, for any observation \mathbf{y} drawn from G . Consequently, we refer to as *discrimination* the property of the scoring rule Sc such that

$$F^{(1)} \succ F^{(2)} \iff \text{Sc}(F^{(1)}, \mathbf{y}) < \text{Sc}(F^{(2)}, \mathbf{y}) \quad (\text{H.4})$$

for any observation \mathbf{y} drawn from G . In the above, $F^{(1)} \succ F^{(2)}$ means that $F^{(1)}$ genuinely is of higher quality than $F^{(2)}$. A scoring rule is then said to have a high discrimination ability if differences in quality between predictive densities are equivalent to significant differences in score values. At the opposite, a scoring rule is said to have no discrimination ability in the case where the same score values are assigned to predictive densities of different quality. One notes that proper score values may not need to have any discrimination ability, since possibly assigning the same score values to all predictive densities F , as well as G which is that for the actual random variable \mathbf{Y} . The situation is different for strictly proper scoring rules, since they should at least discriminate locally in the neighborhood of G . For densities F significantly different from G , however, there is no insurance that the score discriminate among predictive densities. It is to be noted that this concept of discrimination is inspired by the work of [21], who introduced some of the key concepts in forecast verification. Here, however, discrimination is a property of the score, not of the forecast themselves.

H.2.3 Characterizing the discrimination ability of the Energy score

Given the predictive density F and corresponding realization \mathbf{y} , the Energy score Es is defined as

$$\text{Es}(F, \mathbf{y}) = \mathbb{E}_F [\|\mathbf{X} - \mathbf{y}\|] - \frac{1}{2} \mathbb{E}_F [\|\mathbf{X} - \mathbf{X}'\|], \quad (\text{H.5})$$

where \mathbf{X} and \mathbf{X}' are independent random draws from F , while $\|\cdot\|$ denotes the Euclidean norm. Computationally efficient estimators for the Energy scores were introduced in [13].

The corresponding expected Energy score, $\overline{\text{Es}}$, can be calculated as the expectation of the Energy score in Eq. (H.5) over all potential observations of \mathbf{Y} , i.e.,

$$\overline{\text{Es}}(F, G) = \mathbb{E}_G \left[\mathbb{E}_F [\|\mathbf{X} - \mathbf{Y}\|] - \frac{1}{2} \mathbb{E}_F [\|\mathbf{X} - \mathbf{X}'\|] \right]. \quad (\text{H.6})$$

In order to analyze the discrimination ability of the Energy score, we define a metric to be used in the following, corresponding to relative changes in Energy score values, induced by differences between predicted and actual multivariate density of the stochastic process. Considering multivariate Gaussian processes, such differences may relate to prediction errors in the mean, variance, or interdependence structure. The relative change in expected Energy score is defined based of expected Energy score values for F and G ,

$$\Delta \text{Es} = \frac{\overline{\text{Es}}(F, G) - \overline{\text{Es}}^*}{\overline{\text{Es}}^*}, \quad (\text{H.7})$$

where the Energy score value $\overline{\text{Es}}^* = \overline{\text{Es}}(G, G)$ directly comes from the inherent uncertainty of the random variable \mathbf{Y} .

H.2.4 Illustrating the discrimination ability of the Energy score for multivariate Gaussian processes

To illustrate this concept of discrimination ability we make the following simulation study. We assume that a real process generating density is G , corresponding to a bivariate Gaussian. The process is simulated by considering 1000 instances: at each of these instances the process realization \mathbf{y} is given by a single draw from G .

Suppose, there are two competing forecasters. One of them issues forecasts based on the real process generating density G — the perfect forecast. In parallel, the other forecaster issues alternative predictive density F . In order to compare the forecasting approaches, we estimate the corresponding Energy score values and compare them. For the calculation of these Energy score values, 1000 random draws from both densities are used, then employing the computationally efficient estimator proposed by [13]. Our main interest is to see how differences between G and F reflect in the corresponding Energy score values.

The process generating density, G , is bivariate Gaussian with a mean given by $\boldsymbol{\mu} = [\mu \ \mu]^\top$ and a covariance structure

$$\boldsymbol{\Sigma} = \sigma^2 \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}. \quad (\text{H.8})$$

Then at every time step t a single process realization $\mathbf{y} = [y_1 \ y_2]^\top$ of \mathbf{Y} is such that

$$\mathbf{Y} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}). \quad (\text{H.9})$$

The following differences between G and F have been considered:

- *Error in mean* corresponds to the case where the second forecaster makes an error in centering the predictive density only. In this case F is given by a bivariate Gaussian with a well specified covariance structure and a misspecified mean. That is, for every time step this forecaster issues a forecast describing the multivariate density for a random variable \mathbf{X} such that

$$\mathbf{X} \sim \mathcal{N}(\hat{\boldsymbol{\mu}}, \boldsymbol{\Sigma}), \quad (\text{H.10})$$

where $\hat{\boldsymbol{\mu}} = [\hat{\mu} \ \hat{\mu}]$. The resulting difference between F and G is depicted in Fig. H.1(a).

- *Error in variance* corresponds to case where the forecaster makes an error while specifying the variance only. More specifically, for every time step the forecaster issues a forecast describing the multivariate density for a random variable \mathbf{X} such that

$$\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \hat{\boldsymbol{\Sigma}}), \quad (\text{H.11})$$

where

$$\hat{\boldsymbol{\Sigma}} = \hat{\sigma}^2 \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}. \quad (\text{H.12})$$

The resulting difference between F and G is depicted in Fig. H.1(b).

- *Error in correlation* corresponds to cases where the forecaster makes an error in describing the dependency structure, while well specifying the mean and the variance of the process. More specifically, for every time step the forecaster issues a forecast describing the multivariate density for a random variable \mathbf{X} such that

$$X \sim \mathcal{N}(\mu, \hat{\boldsymbol{\Sigma}}) \quad (\text{H.13})$$

where

$$\hat{\boldsymbol{\Sigma}} = \sigma^2 \begin{bmatrix} 1 & \hat{\rho} \\ \hat{\rho} & 1 \end{bmatrix} \quad (\text{H.14})$$

The resulting difference between F and G is depicted in Fig. H.1(c).

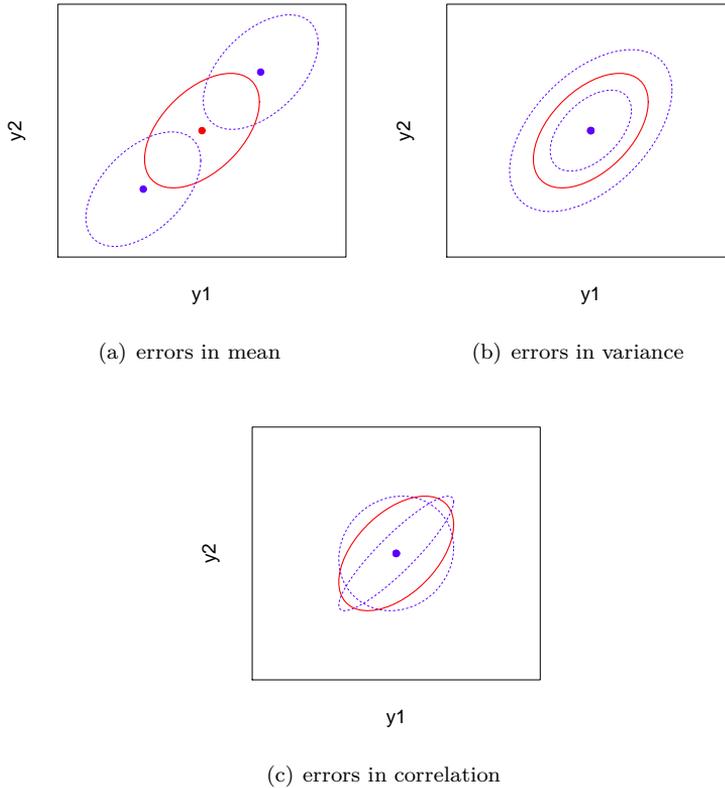
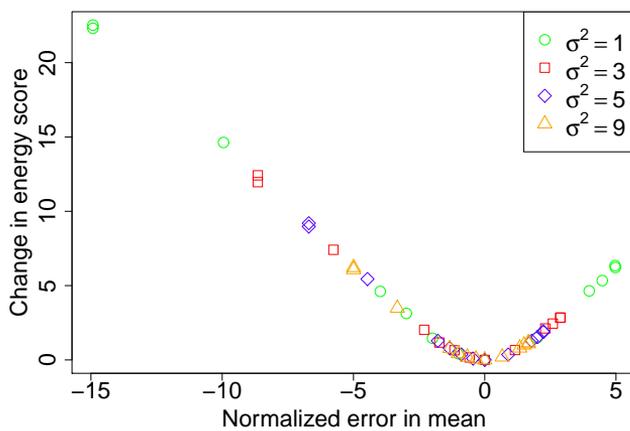


Figure H.1: Illustration of different misspecification in the process generating density. Each density is represented by a single iso-density contour. For every density the volume over the area bounded by the ellipse equals α . Here α denotes a pre-defined probability that a random draw from the corresponding density falls inside the ellipse. Red solid lines represents the real process generating density, while dotted blue lines correspond to predictive densities. Errors in mean are shown in (a). They correspond to predictive densities being shifted variants of the real process generating density. The shift along the major axis of the ellipse (45°) has been considered in the simulation work. Errors in variance are shown in (b). They correspond to an inflation ($\hat{\sigma}^2 > \sigma^2$) or a deflation ($\hat{\sigma}^2 < \sigma^2$) of the ellipse. Finally, errors in correlation are shown in (c). They correspond to stretching the ellipse in the direction of its major axis ($\hat{\rho} > \rho$) or its minor axis ($\hat{\rho} < \rho$)

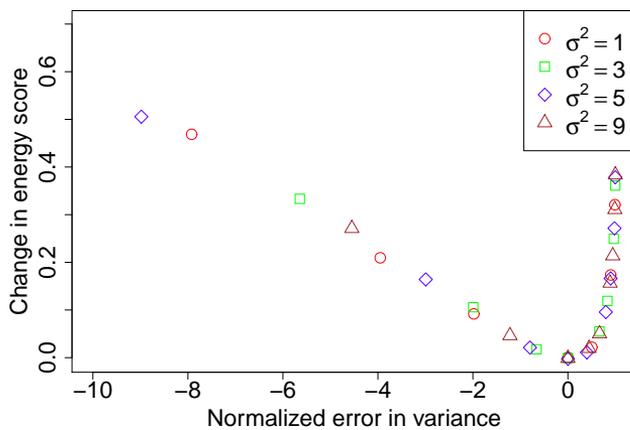
Let us first look at the discrimination ability of the Energy score for errors in mean. μ is set to $\mu = 5$, while the correlation value is fixed to $\rho = 0.5$ (any other values would yield qualitatively similar results). The relative change in Energy score ΔE_s is evaluated as a function of the normalized error in predicting the mean parameter for \mathbf{Y} (See Fig. H.2(a)). That normalized prediction error is defined as $(\mu - \hat{\mu})/\sigma$. This assessment is performed for a number of values of σ^2 ($\sigma^2 \in \{1, 3, 5, 9\}$), in order to characterize the sensitivity to the process variance. A key result here is that, independently of the process variance, the effect of the relative prediction error for μ remains the same. Besides, since the Energy score is based on an Euclidean distance, the relative change in Energy score only depends on the magnitude of the normalized error in mean and not on its direction along the given translation axis (see Fig. H.1(a)).

For the case of errors in variance, the setup is fairly similar, with mean and correlation parameters fixed to $\mu = 0$ and $\rho = 0.5$. A set of values for σ^2 are considered, i.e., $\sigma^2 \in \{1, 3, 5, 9\}$. Predictive densities F there only differ in terms of process variance, where the relative prediction error in variance is defined as $(\sigma^2 - \hat{\sigma}^2)/\sigma^2$. A plot of the relative change in Energy score ΔE_s as a function of that relative prediction error in variance is depicted in Fig. H.2(b). Here again, the relative change in Energy score follows similar patterns, independently of the actual process variance. The discrimination ability of the score is not symmetric, since the score increase for sharper densities is steeper than that for predictive densities that are too wide. Finally, comparing the discrimination ability of the Energy score for the mean and variance parameters, it is clear that the scale of variations in Figs. H.2(a) and H.2(b) are totally different (by a factor of 50), the score being clearly more sensitive to differences in the mean parameter than for the variance.

We finally look at errors in correlation. A plot for the relative change in Energy score as a function of predicted correlation is depicted in Fig. H.3. The results were with $\sigma^2 = 1$ and $\mu = 0$. The results obtained for other values of σ^2 and μ were qualitatively similar. ΔE_s describes how much the Energy score changes when instead of the real process generating density, the forecaster issues the predictive density F . The largest ΔE_s is obtained when the real process generating density is perfectly correlated ($\rho = 1$), while the forecaster totally neglect the correlation by setting $\hat{\rho} = 0$. This is the extreme case for which, visually, $\Delta E_s \approx 0.07$. A theoretical value for this upper bound may be derived analytically, as will be done in the following section. In practice this upper bound is seldom reached, since it corresponds to a very special case of a perfectly correlated bivariate Gaussian process. This means that for any realization \mathbf{y} for G , $y_1 = y_2$. In such an extremes case, the only way for the forecaster to obtain an Energy score value (in expectation) of only 7% worse than if issuing perfect forecasts, is to totally ignore this strong dependency and assume independence of the components of \mathbf{Y} .



(a) errors in mean



(b) errors in variance

Figure H.2: Discrimination ability of the Energy score assessed with ΔE s, in terms of its sensitivity to prediction errors in mean (a) and variance (b) for bivariate Gaussian predictive densities (hence for $n = 2$).

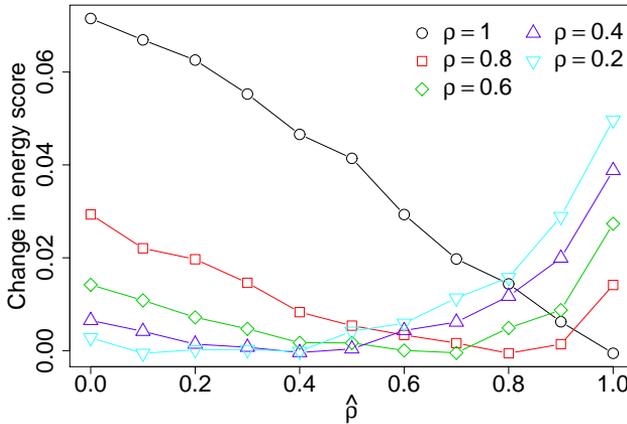


Figure H.3: Discrimination ability of the Energy score assessed with ΔE s, in terms of its sensitivity to errors in correlation for bivariate Gaussian predictive densities (hence for $n = 2$).

In practice, bivariate Gaussian processes are seldom perfectly correlated. One can notice that when the actual correlation, ρ , is less than 1, the maximum penalty (in expectation) stemming from errors in correlation becomes substantially less than 0.07. Already with $\rho = 0.8$, the maximum ΔE s is not even reaching 0.04. A conclusion from this plot is that the upper bound that may be derived by considering perfectly correlated generating densities, and predictive densities with independence of individual components, would be rarely met in most practical applications. For instance, if $\rho = 0.8$ and the forecaster tries predicts of a correlation $\hat{\rho} = 0.4$, then ΔE s < 0.02 only. Another important factor is that the increase in ΔE s is steeper in the case for which $\hat{\rho} > \rho$. This also reduces the motivation of forecasters to move from the assumption of independence to try and capture the actual ρ .

H.3 Some theoretical results on the discrimination ability of the Energy score

In this section, emphasis is placed on our main result, which consists in an upper bound on the discrimination ability of the Energy score for multivariate Gaussian processes for the case of errors in correlation. Such a theoretical upper bound is of importance, since justifying the limited differences in Energy score

values reported in various recent works focusing on multivariate probabilistic forecasts and the predictive modeling of interdependence structures, e.g. [13], [14], also giving some insight on results for Gaussian-copula based modeling of multivariate predictive densities as in [17]. Some other works, e.g., [16], appear to report results that go significantly beyond this theoretical upper bound, for reasons we cannot explain. Simulations studies performed with different variances did not show a significant change in this theoretical upper bound.

Our result is given for the multivariate Gaussian case, for any dimension n , then discussing the specific case of $n = 2$, which may still be the most common in practice.

Let us consider that the generating process G is distributed n -variate Gaussian, $\mathbf{Y} \sim \mathcal{N}(0, \Sigma)$ with same variance σ^2 on all dimensions, and a correlation of 1 between any of these dimensions, i.e.,

$$\Sigma = \sigma^2 \mathbf{1}_{(n \times n)}, \tag{H.15}$$

where $\mathbf{1}_{(n \times n)}$ is a $n \times n$ matrix of ones. This definition of the generating process implies that, at time t , a process observation \mathbf{y} is such that $\mathbf{y} = y \mathbf{1}_n$, where $\mathbf{1}_n$ is a n -dimensional vector of ones, and with $y \sim \mathcal{N}(0, \sigma^2)$.

In the following, we compare the two cases where (i) the forecaster issues a *perfect* forecast $F = G$, and where (ii) the forecaster issues a so-called *naive* forecast that ignores the interdependence structure of \mathbf{Y} , though with appropriate mean and variance on all dimensions. In that latter case, the predictive density F is a n -variate Gaussian density with zero mean and diagonal covariance matrix, $\hat{\Sigma} = \hat{\sigma} \text{diag}(\mathbf{1}_n)$. It is referred to as naive for simplicity only, since already rightly predicting mean and variance of n -variate random variables would be a nice achievement. In both cases, a closed-form formula for the Energy score is provided. They will be denoted by Es^* and Es_i . Looking at this case yield on upper bound on the discrimination ability of the Energy score for varying dependence structures, since comprising a worst case on the distance between multivariate Gaussian densities (as discussed in the above).

H.3.0.1 Expected Energy score for the naive forecast

For the naive forecast, one can directly work with the expression in (H.6). The computation of $\overline{\text{Es}}_i$ is split into that of $\mathbb{E}_G [\mathbb{E}_G [||\mathbf{X} - \mathbf{Y}||]]$ and that of $\mathbb{E}_G [\mathbb{E}_G [||\mathbf{X} - \mathbf{X}'||]]$.

After some algebra described in Appendix H.4.1, one obtains that

$$\mathbb{E}_G [\mathbb{E}_F [||\mathbf{X} - \mathbf{Y}||]] = \frac{\Gamma\left(\frac{n}{2} + \frac{1}{2}\right) \sqrt{2}\sigma}{\Gamma\left(\frac{n}{2}\right) \sqrt{n+1}} {}_2F_1\left(\frac{n+1}{2}, \frac{1}{2}; \frac{n}{2}; \frac{n}{n+1}\right), \quad (\text{H.16})$$

where ${}_2F_1$ is the hypergeometric function. In parallel,

$$\mathbb{E}_G [\mathbb{E}_F [||\mathbf{X} - \mathbf{X}'||]] = 2\sigma \frac{\Gamma\left(\frac{n}{2} + \frac{1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}. \quad (\text{H.17})$$

The final formula for $\overline{\text{Es}}_i$ therefore reads

$$\overline{\text{Es}}_i = \sigma \frac{\Gamma\left(\frac{n}{2} + \frac{1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \left\{ \frac{\sqrt{2}}{\sqrt{n+1}} {}_2F_1\left(\frac{n+1}{2}, \frac{1}{2}; \frac{n}{2}; \frac{n}{n+1}\right) - 1 \right\}. \quad (\text{H.18})$$

H.3.0.2 Expected Energy score for the perfect forecast

In the case where $F = G$, the expression for the expected Energy score in (H.6) is such that

$$\overline{\text{Es}}^* = \mathbb{E}_G \left[\mathbb{E}_G [||\mathbf{X} - \mathbf{Y}||] - \frac{1}{2} \mathbb{E}_G [||\mathbf{X} - \mathbf{X}'||] \right]. \quad (\text{H.19})$$

Consequently, the calculation of the above can be split into that of $\mathbb{E}_G [\mathbb{E}_G [||\mathbf{X} - \mathbf{Y}||]]$ and that of $\mathbb{E}_G [\mathbb{E}_G [||\mathbf{X} - \mathbf{X}'||]]$.

After some algebra described in Appendix H.4.2, one obtains that

$$\mathbb{E}_G [\mathbb{E}_G [||\mathbf{X} - \mathbf{Y}||]] = 2\sigma \sqrt{\frac{n}{\pi}}, \quad (\text{H.20})$$

while, similarly,

$$\mathbb{E}_G [\mathbb{E}_G [||\mathbf{X} - \mathbf{X}'||]] = 2\sigma \sqrt{\frac{n}{\pi}}. \quad (\text{H.21})$$

The final formula for $\overline{\text{Es}}^*$ therefore reads

$$\overline{\text{Es}}^* = \sigma \sqrt{\frac{n}{\pi}}. \quad (\text{H.22})$$

H.3.1 An upper bound on the discrimination ability of the Energy score in the multivariate Gaussian case

As a consequence of the developments in the above, the upper bound ΔEs^\dagger on the discrimination ability of the Energy score in the multivariate Gaussian case is obtained as

$$\Delta Es^\dagger = \frac{\overline{Es}_i - \overline{Es}^*}{\overline{Es}^*} = \sqrt{\frac{\pi}{n}} \left\{ \frac{\Gamma\left(\frac{n}{2} + \frac{1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \left(\frac{\sqrt{2}}{\sqrt{n+1}} {}_2F_1\left(\frac{n+1}{2}, \frac{1}{2}; \frac{n}{2}; \frac{n}{n+1}\right) - 1 \right) \right\}, \tag{H.23}$$

solely depending on the dimension n of the multivariate Gaussian process considered.

The evolution of this upper bound ΔEs^\dagger is depicted in Fig. H.4 as a function of the dimension of the multivariate Gaussian process of interest, from $n = 2$ to $n = 300$. This upper bound increases with n , even though it reaches an asymptote (of less than 15%) for higher dimensions. Taking the example of bivariate processes, the maximum improvement in Energy score that could even be observed is of 7.4% of the Energy score value if predicting the actual distribution G of \mathbf{Y} . This upper bound is considerably less than if making errors in predicting the variance parameter for that multivariate Gaussian process, hence also very small compared to the case of prediction errors in the mean parameter.

Looking at Fig. H.4, one could think that since the upper bound is increasing with the problem dimension, then the Energy score has a better ability to discriminate between covariance structure in higher dimensions. However, as can be seen from Fig. H.3 such a theoretical upper bound is substantially higher than the differences observed under more realistic conditions for correlation values of the generating process. As the dimension of the problem grows, this upper bound may then become substantially higher than the practical discrimination ability of the Energy score.

Another important aspect to be mentioned relates to computational issues. Estimation of Energy score calls for Monte Carlo techniques, since no closed-form expression exists, even for multivariate Gaussian processes. That estimation hence becomes computationally expensive. Being more precise, the cost of sampling from a Gaussian distribution (with a covariance matrix not being restricted to any particular pattern) is cubic in the dimension. This means that estimating the Energy score is hampered by the “big n ” problem.

For example in a practical application to probabilistic forecasting of wind power generation, we have considered a problem with dimension $n=645$ [22]. A year

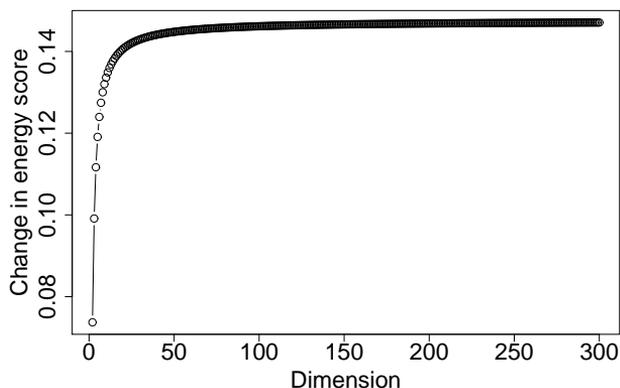


Figure H.4: Upper bound on discrimination ability (i.e., ΔE_s^\dagger) of the Energy score for a multivariate Gaussian process with well-predicted means and variances, as a function of the dimension n .

of hourly data was used (8760 time steps) and for each time step 10 scenarios (which is very small given the dimension of 645) were in order to evaluate the score. Given this setup, it took more than 12 hours to estimate the average Energy score when using 8 parallel cores. Such computational issues also translate to limiting the number of samples used for estimating the Energy score, therefore leading to a certain level of uncertainty in the score values obtained. This uncertainty becomes especially importance, given that a rather low sensitivity of the score to the changes in the covariance structure.

H.4 Discussion

The field of probabilistic forecasting is developing rapidly, and with increasing focus on multivariate processes, often of relatively low dimensions (say, $n = 2, \dots, 5$). Considering higher dimensions will be natural for instance for applications related to energy, meteorology and climate sciences, with focus on different variables, locations and lead times. As a consequence, it is important to further develop and analyze frameworks for probabilistic forecast verification, permitting to draw useful and practical conclusions on forecast quality.

For the case of multivariate probabilistic forecasts, the Energy score is a relevant candidate for becoming a lead score for evaluation of such forecasts. As of today,

our understanding of its inherent properties for various types of processes and their varying dimensions is somewhat fairly limited. Also, the properties of related estimators, in terms of their potential bias, and sensitivity to sampling and correlation effects, are to be studied.

Our aim here was to focus on the discrimination ability of the Energy score, i.e., its ability to assign different score values to predictive densities of different quality. The most simple case of multivariate Gaussian processes and predictive densities was considered, still providing interesting insight on some of the properties of this score. Indeed, the Energy score is known to be proper, but this does not insure that it has a high discriminatory power. While it may nicely discriminate predictive densities with different mean parameters, it was discussed that differences in score values would be much less when looking at differences in their variance parameters. Also, for the case of the interdependence structure of predictive densities, an upper bound on score differences that may be observed (in expectation) was derived. Our conjecture is that, comparatively, the Energy score may hardly allow to discriminate among predictive densities with different interdependence structures. Maybe its discrimination ability could be maximized by slightly altering its definition, and using other forms of distance, better considering the structure of predictive densities. Besides, additional consideration should be given to other scoring rules defining scores for verifying multivariate probabilistic forecasts, since they may give a different discrimination ability, while having additional computational advantages.

Acknowledgements

Acknowledgments are due to Tilmann Gneiting for various discussions on probabilistic forecast verification and properties of scoring rules.

Appendix

H.4.1 Necessary calculations to evaluate $\overline{\mathbf{E}s}_i$

H.4.1.1 Evaluating $\mathbb{E}_G [\mathbb{E}_F [||\mathbf{X} - \mathbf{Y}||]]$

Given a single process realization $\mathbf{y} = y\mathbf{1}_n$,

$$||\mathbf{X} - \mathbf{y}|| = \sqrt{(x_1 - y)^2 + (x_2 - y)^2 + \cdots + (x_n - y)^2}, \quad (\text{H.24})$$

where a realization of \mathbf{X} is $\mathbf{x} = [x_1 \dots x_n]^\top$.

In parallel, a known result is such that

$$\left. \begin{aligned} (x_1 - y) &\sim \mathcal{N}(-y, \sigma^2) \\ (x_2 - y) &\sim \mathcal{N}(-y, \sigma^2) \\ &\dots \\ (x_n - y) &\sim \mathcal{N}(-y, \sigma^2) \\ x_1, x_2, \dots, x_n &\text{ are i.i.d.} \end{aligned} \right\} \Rightarrow z = \sum_{i=1}^n \frac{(x_i - y)^2}{\sigma^2} \sim \text{Non-central Chi-squared.}$$

Consequently, following [23], the parameters of the non-central Chi-squared distribution are given by n and λ , where

$$\lambda = \frac{1}{2} \sum_{i=1}^n \frac{y^2}{\sigma^2}. \tag{H.25}$$

Then,

$$\mathbb{E}_F [||\mathbf{X} - \mathbf{y}||] = \sigma \mathbb{E}_F \left[z^{\frac{1}{2}} \right]. \tag{H.26}$$

That is, in order to evaluate $\mathbb{E}_F [||\mathbf{X} - \mathbf{y}||]$ we need to know a fractional moment of order 1/2 of the non-central Chi-squared distributed variable z .

Following [23],

$$\begin{aligned} \mathbb{E}_F \left[z^{\frac{1}{2}} \right] &= \sqrt{2} \frac{\Gamma\left(\frac{n}{2} + \frac{1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} {}_1F_1\left(-\frac{1}{2}; \frac{n}{2}; -\lambda\right) \\ &= \sqrt{2} \frac{\Gamma\left(\frac{n}{2} + \frac{1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} {}_1F_1\left(-\frac{1}{2}; \frac{n}{2}; -\frac{n y^2}{2 \sigma^2}\right), \end{aligned} \tag{H.27}$$

where ${}_1F_1$ denotes the confluent hypergeometric function of the first kind. This yields

$$\mathbb{E}_F [||\mathbf{X} - \mathbf{y}||] = \sigma \sqrt{2} \frac{\Gamma\left(\frac{n}{2} + \frac{1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} {}_1F_1\left(-\frac{1}{2}; \frac{n}{2}; -\frac{n y^2}{2 \sigma^2}\right). \tag{H.28}$$

As a consequence,

$$\begin{aligned}
 \mathbb{E}_G [\mathbb{E}_F [|\mathbf{X} - \mathbf{Y}|]] &= \sqrt{2} \frac{\Gamma\left(\frac{n}{2} + \frac{1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \int_{-\infty}^{\infty} {}_1F_1\left(-\frac{1}{2}; \frac{n}{2}; -\frac{ny^2}{2\sigma^2}\right) \frac{1}{\sigma\sqrt{2\pi}} \exp\frac{-y^2}{2\sigma^2} dy \\
 &= \frac{\Gamma\left(\frac{n}{2} + \frac{1}{2}\right)}{\sigma\sqrt{\pi}\Gamma\left(\frac{n}{2}\right)} \int_0^{\infty} \exp\left(-\frac{1}{2\sigma^2}y^2\right) {}_1F_1\left(-\frac{1}{2}; \frac{n}{2}; \frac{-ny^2}{2\sigma^2}\right) (y^2)^{-\frac{1}{2}} dy^2 \\
 &= \frac{\Gamma\left(\frac{n}{2} + \frac{1}{2}\right)}{\sigma\sqrt{\pi}\Gamma\left(\frac{n}{2}\right)} \int_0^{\infty} \exp\left(-\frac{1}{2\sigma^2}t\right) {}_1F_1\left(-\frac{1}{2}; \frac{n}{2}; \frac{-nt}{2\sigma^2}\right) (t)^{-\frac{1}{2}} dt
 \end{aligned}$$

To integrate the expression further we use (4) on page 822 of [24] stating that

$$\int_0^{\infty} \exp(-st)t^{b-1} {}_1F_1(a; c; kt)dt = \Gamma(b)(s - k)^{-b} F(c - a, b; c; \frac{k}{k - s}), \quad (\text{H.29})$$

if $|s - k| > |k|$ and $Re(b) > 0, Re(s) > \max(0, Re(k))$. In the above F is the Gauss hypergeometric function.

In our case: $a = -0.5, c = 0.5n, k = \frac{-n}{2\sigma^2}, b = 0.5, s = \frac{1}{2\sigma^2}$. Then:

$$|s - k| = \frac{n + 1}{2\sigma^2} > \frac{n}{\sigma^2} = |k|$$

$$Re(b) = 0.5 > 0$$

$$Re(s) = \frac{1}{2\sigma^2} > 0 = \max(0, Re(k))$$

All the conditions are fulfilled, therefore we can apply the formula given in [24] and as a result:

$$\begin{aligned}
 \mathbb{E}_G [\mathbb{E}_F [|\mathbf{X} - \mathbf{Y}|]] &= \frac{\Gamma\left(\frac{n}{2} + \frac{1}{2}\right)}{\sigma\sqrt{\pi}\Gamma\left(\frac{n}{2}\right)} \Gamma\left(\frac{1}{2}\right) \left(\frac{n + 1}{2\sigma^2}\right)^{-\frac{1}{2}} {}_2F_1\left(\frac{n + 1}{2}, \frac{1}{2}; \frac{n}{2}; \frac{n}{n + 1}\right) \\
 &= \sigma\sqrt{\frac{2}{n + 1}} \frac{\Gamma\left(\frac{n}{2} + \frac{1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} {}_2F_1\left(\frac{n + 1}{2}, \frac{1}{2}; \frac{n}{2}; \frac{n}{n + 1}\right) \quad (\text{H.30})
 \end{aligned}$$

H.4.1.2 Evaluating $\mathbb{E}_G [\mathbb{E}_G [||\mathbf{X} - \mathbf{X}'||]]$

As a starting point one has

$$\mathbb{E}_G [||X - X'||] = \sqrt{(x_1 - x'_1)^2 + (x_2 - x'_2)^2 + \cdots + (x_n - x'_n)^2}, \quad (\text{H.31})$$

with $x_i, x'_i \sim \mathcal{N}(0, \sigma^2)$, while being mutually independent for all $i = 1, 2, \dots, n$.

Let us introduce $z = \sum_{i=1}^n (x_i - x'_i)^2$. Since $\frac{(x_i - x'_i)}{2\sigma^2} \sim \mathcal{N}(0, 2\sigma^2)$, z follows a non-central Chi-squared distribution with parameters n and $\lambda = 0$.

Therefore following [23],

$$\begin{aligned} \mathbb{E}_G \left[z^{\frac{1}{2}} \right] &= \sqrt{2} \frac{\Gamma\left(\frac{n}{2} + \frac{1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} {}_1F_1\left(-\frac{1}{2}; \frac{n}{2}; 0\right) \\ &= \sqrt{2} \frac{\Gamma\left(\frac{n}{2} + \frac{1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}. \end{aligned} \quad (\text{H.32})$$

Consequently,

$$\mathbb{E}_G [||\mathbf{X} - \mathbf{X}'||] = \mathbb{E} \left[z^{\frac{1}{2}} \right] = 2\sigma \frac{\Gamma\left(\frac{n}{2} + \frac{1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}, \quad (\text{H.33})$$

then also defining $\mathbb{E}_G [\mathbb{E}_G [||\mathbf{X} - \mathbf{X}'||]]$.

H.4.2 Necessary calculations to evaluate $\overline{\mathbf{E}s}^*$

H.4.2.1 Evaluating $\mathbb{E}_G [\mathbb{E}_G [||\mathbf{X} - \mathbf{Y}||]]$

First of all, one has

$$\mathbf{X} - \mathbf{y} = [x_1 - y \quad x_2 - y \quad \cdots \quad x_n - y]^\top. \quad (\text{H.34})$$

Since both \mathbf{X} and \mathbf{Y} are distributed $\mathcal{N}(0, \Sigma)$, with Σ as defined in (H.15), then $x_n = x_{n-1} = \cdots = x_1$ (which we write x) and $y_n = y_{n-1} = \cdots = y_2 = y_1$

(which we write y). Thus,

$$\|\mathbf{X} - \mathbf{y}\| = \sqrt{n(x - y)^2} = \sqrt{n}|x - y|. \quad (\text{H.35})$$

Subsequently, since $x \sim \mathcal{N}(0, \sigma^2)$, then given y , $(x - y) \sim \mathcal{N}(-y, \sigma^2)$. Following this, the variable $\|\mathbf{X} - \mathbf{y}\|$ follows a folded Normal with parameters $-y$ and σ^2 . Using the analytical expression for the mean of a folded Normal distribution (see Section H.4.3.1), we obtain

$$\mathbb{E}_G [\|\mathbf{X} - \mathbf{y}\|] = \sigma \sqrt{2n/\pi} \exp\left(\frac{-y^2}{2\sigma^2}\right) + y \left(1 - 2\Phi\left(\frac{-y}{\sigma}\right)\right). \quad (\text{H.36})$$

Then given that $y \sim \mathcal{N}(0, \sigma^2)$,

$$\begin{aligned} \mathbb{E}_G [\mathbb{E}_G [\|\mathbf{X} - \mathbf{Y}\|]] &= \int_{-\infty}^{\infty} \sqrt{n} \left(\sigma \sqrt{2/\pi} \exp\left(\frac{-y^2}{2\sigma^2}\right) + y \left(1 - 2\Phi\left(\frac{-y}{\sigma}\right)\right) \right) \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-y^2}{2\sigma^2}\right) dy \\ &= \int_{-\infty}^{\infty} \frac{\sqrt{n}}{\pi} \exp\left(\frac{-y^2}{\sigma^2}\right) dy \\ &\quad + \sqrt{n}\sigma \int_{-\infty}^{\infty} \frac{y}{\sigma} \phi\left(\frac{y}{\sigma}\right) d\frac{y}{\sigma} \\ &\quad + -2\sqrt{n}\sigma \int_{-\infty}^{\infty} \frac{y}{\sigma} \phi\left(\frac{y}{\sigma}\right) \Phi\left(\frac{-y}{\sigma}\right) d\frac{y}{\sigma} \\ &= \frac{\sqrt{n}\sigma}{\sqrt{\pi}} + \frac{2\sqrt{n}\sigma}{2\sqrt{\pi}} \\ &= \frac{2\sqrt{n}\sigma}{\sqrt{\pi}}, \end{aligned} \quad (\text{H.37})$$

based on integrals given in Section H.4.3.2.

H.4.2.2 Evaluating $\mathbb{E}_G [\mathbb{E}_G [\|\mathbf{X} - \mathbf{X}'\|]]$

Similarly to the above, one starts with

$$\|\mathbf{X} - \mathbf{X}'\| = \sqrt{(x_1 - x'_1)^2 + (x_2 - x'_2)^2 + \cdots + (x_n - x'_n)^2}. \quad (\text{H.38})$$

Then, since $\mathbf{X}, \mathbf{X}' \sim \mathcal{N}(0, \Sigma)$, with Σ as defined in (H.15), the above can be reformulated as

$$\|\mathbf{X} - \mathbf{X}'\| = \sqrt{n(x - x')^2} = \sqrt{n}|x - x'|, \quad (\text{H.39})$$

with x and x' independent draws from \mathbf{X} and X' such that $\mathbf{X}, \mathbf{X}' \sim \mathcal{N}(0, \sigma^2)$. Consequently, following an argument similar to that in the previous section, it is that $|\mathbf{X} - \mathbf{X}'|$ follows a folded Normal distribution with parameters 0 and $2\sigma^2$.

By applying the formula for finding the expected value of a folded Normal density (Section H.4.3.1), we finally obtain

$$\mathbb{E}_G [|\mathbf{X} - \mathbf{X}'|] = 2\sigma\sqrt{\frac{1}{\pi}}, \quad (\text{H.40})$$

and then

$$\mathbb{E}_G [\mathbb{E}_G [|\mathbf{X} - \mathbf{X}'|]] = \frac{2\sqrt{n}\sigma}{\sqrt{\pi}}. \quad (\text{H.41})$$

H.4.3 Some results on relevant distributions and integrals

Below are given some basic definitions and results for some relevant probability distributions and integrals used in the above mathematical derivations.

H.4.3.1 Folded Normal distribution

The folded Normal distribution is directly linked to the Normal distribution. Indeed, in the case for which X is distributed Gaussian, $X \sim \mathcal{N}(\mu, \sigma^2)$, then $Y = |X|$ follows a folded Normal distribution, $Y \sim \mathcal{N}^f(\mu, \sigma^2)$. For such a distribution, the expectation of Y is given by

$$\mathbb{E}[Y] = \sigma\sqrt{\frac{2}{\pi}} \exp\left(-\frac{1}{2}\left(\frac{\mu}{\sigma}\right)^2\right) + \mu\left(1 - 2\Phi\left(-\frac{\mu}{\sigma}\right)\right). \quad (\text{H.42})$$

H.4.3.2 Some known relevant integrals

$$\int_{-\infty}^{\infty} x\phi(x)\Phi(bx) dx = \int_{-\infty}^{\infty} x\phi(x)\Phi(bx)^2 dx = \frac{b}{\sqrt{2\pi(1+b^2)}} \quad (\text{H.43})$$

$$\int_{-\infty}^{\infty} e^{-ax^2} dx = \sqrt{\frac{\pi}{a}} \quad (a > 0) \quad (\text{H.44})$$

References H

- [1] A. S. Tay and K. F. Wallis, “Density forecasting: a survey,” *Journal of Forecasting*, pages=235–254, year=2000.
- [2] A. Timmermann, “Density forecasting in economics and finance,” *Journal of Forecasting*, vol. 19, no. 4, pp. 231–234, 2000.
- [3] Y. Y. Kagan and D. D. Jackson, “Probabilistic forecasting of earthquakes,” *Geophysical Journal International*, vol. 143, no. 2, pp. 438–453, 2000.
- [4] M. Leutbecher and T. N. Palmer, “Ensemble forecasting,” *Journal of Computational Physics*, vol. 227, no. 7, pp. 3515–3539, 2008.
- [5] P. Bacher, H. Madsen, and H. A. Nielsen, “Online short-term solar power forecasting,” *Solar Energy*, vol. 83, no. 10, pp. 1772–1783, 2009.
- [6] P. Pinson, “Wind energy: Forecasting challenges for its operational management,” *Statistical Science*, in press, 2013.
- [7] H. L. Cloke and F. Pappenberger, “Ensemble flood forecasting: a review,” *Journal of Hydrology*, vol. 375, no. 3, pp. 613–626, 2009.
- [8] T. Gneiting, “Making and evaluating point forecasts,” *Journal of the American Statistical Association*, vol. 106, no. 494, pp. 746–762, 2011.
- [9] T. Gneiting, F. Balabdaoui, and A. E. Raftery, “Probabilistic forecasts, calibration and sharpness,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 69, no. 2, pp. 243–268, 2007.
- [10] A. P. Dawid, “Statistical theory: The prequential approach,” *Journal of the Royal Statistical Society. Series A (General)*, pp. 278–292, 1984.

-
- [11] F. X. Diebold, A. T. Gunther, and A. S. Tay, "Evaluating density forecasts with applications to financial risk management," *International Economic Review*, vol. 39, pp. 863–883, 1984.
- [12] R. Benedetti, "Scoring rules for forecast verification," *Monthly Weather Review*, vol. 138, no. 1, pp. 203–211, 2010.
- [13] T. Gneiting, L. I. Stanberry, E. P. Gritmit, L. Held, and N. A. Johnson, "Assessing probabilistic forecasts of multivariate quantities, with an application to ensemble predictions of surface winds," *Test*, vol. 17, no. 2, pp. 211–235, 2008.
- [14] P. Pinson, "Adaptive calibration of (u, v)-wind ensemble forecasts," *Quarterly Journal of the Royal Meteorological Society*, vol. 138, pp. 1273–1284, 2012.
- [15] D. S. Wilks, "Probabilistic canonical correlation analysis forecasts, with application to tropical Pacific sea-surface temperatures," *International Journal of Climatology*, 2013.
- [16] N. Schuhen, T. L. Thorarinsdottir, and T. Gneiting, "Ensemble model output statistics for wind vectors," *Monthly Weather Review*, available online, 2012.
- [17] A. Möller, A. Lenkoski, and T. L. Thorarinsdottir, "Multivariate probabilistic forecasting using ensemble Bayesian model averaging and copulas," *Quarterly Journal of the Royal Meteorological Society*, 2013.
- [18] P. Pinson and R. Girard, "Evaluating the quality of scenarios of short-term wind power generation," *Applied Energy*, vol. 96, pp. 12–20, 2012.
- [19] Ò. Jordà and M. Marcellino, "Path forecast evaluation," *Journal of Applied Econometrics*, vol. 25, no. 4, pp. 635–662, 2010.
- [20] J. Bröcker and L. A. Smith, "Scoring probabilistic forecasts: The importance of being proper," *Weather and Forecasting*, vol. 22, no. 2, pp. 382–388, 2007.
- [21] A. H. Murphy, "What is a good forecast? An essay on the nature of goodness in weather forecasting," *Weather and forecasting*, vol. 8, no. 2, pp. 281–293, 1993.
- [22] J. Tastu and P. Pinson, "Space-time scenarios of wind power generation produced using a Gaussian copula with parametrized precision matrix," tech. rep., Technical University of Denmark, 2013.
- [23] J. R. Harvey, *Fractional moments of a quadratic form in noncentral normal random variables*. PhD thesis, North Carolina State University at Raleigh, 1965.

-
- [24] A. Jeffrey and D. Zwillinger, *Table of integrals, series, and products*. Academic Press, 2007.

Bibliography

- [1] “World Energy Outlook,” 2010.
- [2] P. Capros, “PRIMES Energy system model, Manual E3MLab, National Technical University of Athens,” tech. rep.
- [3] “Pure Power. Wind energy targets for 2020 and 2030. A report by the European Wind Energy Association,” tech. rep., 2011.
- [4] “Danish Ministry of Climate, Energy and Building. Energy policy report,” tech. rep., 2012.
- [5] T. J. Hammons, “Integrating renewable energy sources into European grids,” *International Journal of Electrical Power & Energy Systems*, vol. 30, no. 8, pp. 462–475, 2008.
- [6] G. Czisch, “Low cost but totally renewable electricity supply for a huge supply area - a European/Trans-European example-,” tech. rep., 2006.
- [7] S. Faias, J. Sousa, and R. Castro, “Embedded energy storage systems in the power grid for renewable energy sources integration,” *Renewable Energy*, pp. 63–88, 2009.
- [8] “Europe’s onshore and offshore wind energy potential. An assessment of environmental and economic constraints,” tech. rep., 2009.
- [9] D. W. Bunn, “Forecasting loads and prices in competitive power markets,” *Proceedings of the IEEE*, vol. 88, no. 2, pp. 163–169, 2000.
- [10] G. Gross and F. D. Galiana, “Short-term load forecasting,” *Proceedings of the IEEE*, vol. 75, no. 12, pp. 1558–1573, 1987.

- [11] I. Moghram and S. Rahman, "Analysis and evaluation of five short-term load forecasting techniques," *Power Systems, IEEE Transactions on*, vol. 4, no. 4, pp. 1484–1491, 1989.
- [12] B. F. Hobbs, S. Jitprapaikulsarn, S. Konda, V. Chankong, K. A. Loparo, and D. J. Maratukulam, "Analysis of the value for unit commitment of improved load forecasts," *Power Systems, IEEE Transactions on*, vol. 14, no. 4, pp. 1342–1348, 1999.
- [13] G. Giebel, R. Brownsword, G. Kariniotakis, M. Denhard, and C. Draxl, "The state-of-the-art in short-term prediction of wind power—a literature overview. 2nd edition," tech. rep., Technical University of Denmark, 2011. available at: <http://orbit.dtu.dk>.
- [14] M. Ortega-Vazquez and D. Kirschen, "Assessing the impact of wind power generation on operating costs," *Smart Grid, IEEE Transactions on*, vol. 1, no. 3, pp. 295–301, 2010.
- [15] T. Ackermann *et al.*, *Wind power in power systems*, vol. 140. Wiley Online Library, 2005.
- [16] L. Jones and C. Clark, "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*, 2011.
- [17] S.-E. Thor and P. Weis-Taylor, "Long-term research and development needs for wind energy for the time frame 2000–2020," *Wind Energy*, vol. 5, no. 1, pp. 73–75, 2002.
- [18] C. L. Vincent, P. Pinson, and G. Giebel, "Wind fluctuations over the North Sea," *International Journal of Climatology*, vol. 31, no. 11, pp. 1584–1595, 2011.
- [19] I. Sanchez, "Short-term prediction of wind energy production," *International Journal of Forecasting*, vol. 22, no. 1, pp. 43–56, 2006.
- [20] 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*, vol. 12, no. 6, pp. 1725–1744, 2008.
- [21] C. Monteiro, R. Bessa, V. Miranda, A. Botterud, J. Wang, G. Conzelmann, *et al.*, "Wind power forecasting: state-of-the-art 2009.," tech. rep., Argonne National Laboratory (ANL), 2009.
- [22] G. Kariniotakis, I. Martí, D. Casas, P. Pinson, T. S. Nielsen, H. Madsen, G. Giebel, J. Usaola, I. Sanchez, A. M. Palomares, *et al.*, "What performance can be expected by short-term wind power prediction models

- depending on site characteristics,” in *Proceedings of the European Wind Energy Conference*, pp. 22–25, 2004.
- [23] P. Pinson, “Estimation of the uncertainty in wind power forecasting,” *Ph.D. thesis, Paris, France: Ecole des Mines de Paris*, 2006.
- [24] H. Madsen, *Models and methods for predicting wind power*. 1996.
- [25] T. S. Nielsen, H. Madsen, H. A. Nielsen, L. Landberg, and G. Giebel, “Zephyr—the prediction models.,” in *European Wind Energy Conference*, pp. 868–871, 2001.
- [26] T. S. Nielsen, H. Madsen, H. Parbo, H. Palmelund, A. Grud, H. S. Christensen, G. Agersbaek, J. K. Vesterdal, and B. Tøfting, J. amd Korshøj, *Using meteorological forecasts in on-line predictions of wind power*. 1999.
- [27] T. S. Nielsen, “Online prediction and control in nonlinear stochastic systems,” *Department of Mathematical Modeling Technical University of Denmark Ph. D. Thesis*, no. 84, 2002.
- [28] T. R. Stewart, “Uncertainty, judgment, and error in prediction,” *Prediction: Science, Decision Making, and the Future of Nature*. Island Press, Washington, DC, pp. 41–57, 2000.
- [29] H. A. Nielsen, T. S. Nielsen, and H. Madsen, “An overview of wind power forecast types and their use in large-scale integration of wind power,” in *Proceedings of the 10th International Workshop on Large-Scale Integration of Wind Power into Power Systems*, 2011.
- [30] T. Gneiting, “Quantiles as optimal point forecasts,” *International Journal of Forecasting*, vol. 27, no. 2, pp. 197–207, 2011.
- [31] T. Gneiting, “Editorial: probabilistic forecasting,” *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, vol. 171, no. 2, pp. 319–321, 2008.
- [32] A. Botterud, Z. Zhou, J. Wang, R. Bessa, H. Keko, J. Sumaili, and V. Miranda, “Wind power trading under uncertainty in LMP markets,” *Power Systems, IEEE Transactions on*, vol. 27, no. 2, pp. 894–903, 2012.
- [33] X. Liu, “Economic load dispatch constrained by wind power availability: A wait-and-see approach,” *Smart Grid, IEEE Transactions on*, vol. 1, no. 3, pp. 347–355, 2010.
- [34] A. Botterud, Z. Zhou, J. Wang, J. Sumaili, H. Keko, J. Mendes, R. J. Bessa, and V. Miranda, “Demand dispatch and probabilistic wind power forecasting in unit commitment and economic dispatch: A case study of Illinois,” *Sustainable Energy, IEEE Transactions on*, vol. 4, no. 1, pp. 250–261, 2013.

- [35] C. Liu, J. Wang, A. Botterud, Y. Zhou, and A. Vyas, "Assessment of impacts of PHEV charging patterns on wind-thermal scheduling by stochastic unit commitment," *Smart Grid, IEEE Transactions on*, vol. 3, no. 2, pp. 675–683, 2012.
- [36] Á. Duque, E. Castronuovo, I. Sánchez, and J. Usaola, "Optimal operation of a pumped-storage hydro plant that compensates the imbalances of a wind power producer," *Electric Power Systems Research*, vol. 81, no. 9, pp. 1767–1777, 2011.
- [37] R. Bessa, M. Matos, I. Costa, L. Bremermann, I. Franchin, R. Pestana, N. Machado, H. Waldl, and C. Wichmann, "Reserve setting and steady-state security assessment using wind power uncertainty forecast: a case study," *IEEE Transactions on Sustainable Energy*, 2012.
- [38] B. G. Brown, R. W. Katz, and A. H. Murphy, "Time series models to simulate and forecast wind speed and wind power," *Journal of climate and applied meteorology*, vol. 23, pp. 1184–1195, 1984.
- [39] M. Lange, "On the uncertainty of wind power predictions—analysis of the forecast accuracy and statistical distribution of errors," *Journal of Solar Energy Engineering*, vol. 127, p. 177, 2005.
- [40] H. A. Nielsen, H. Madsen, T. S. Nielsen, J. Badger, G. Giebel, L. Landberg, K. Sattler, and H. Feddersen, "Wind power ensemble forecasting," in *Proceedings of the 2004 Global Windpower Conference and Exhibition*, 2004.
- [41] J. W. Taylor, P. E. McSharry, and R. Buizza, "Wind power density forecasting using ensemble predictions and time series models," *Energy Conversion, IEEE Transactions on*, vol. 24, no. 3, pp. 775–782, 2009.
- [42] J. Jeon and J. W. Taylor, "Using conditional kernel density estimation for wind power density forecasting," *Journal of the American Statistical Association*, vol. 107, no. 497, pp. 66–79, 2012.
- [43] A. Luig, S. Bofinger, and H. G. Beyer, "Analysis of confidence intervals for the prediction of regional wind power output," in *Proceedings of the European Wind Energy Conference, Copenhagen, Denmark*, pp. 725–728, 2001.
- [44] P. Pinson and G. Kariniotakis, "Conditional prediction intervals of wind power generation," *Power Systems, IEEE Transactions on*, vol. 25, no. 4, pp. 1845–1856, 2010.
- [45] J. B. Bremnes, "Probabilistic wind power forecasts using local quantile regression," *Wind Energy*, vol. 7, no. 1, pp. 47–54, 2004.

- [46] J. Møller, H. Nielsen, and H. Madsen, “Time-adaptive quantile regression,” *Computational Statistics & Data Analysis*, vol. 52, no. 3, pp. 1292–1303, 2008.
- [47] P. Pinson, H. Nielsen, J. Møller, H. Madsen, and G. Kariniotakis, “Non-parametric probabilistic forecasts of wind power: required properties and evaluation,” *Wind Energy*, vol. 10, no. 6, pp. 497–516, 2007.
- [48] 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)*, 2012.
- [49] A. H. Murphy, “What is a good forecast? an essay on the nature of goodness in weather forecasting,” *Weather and forecasting*, vol. 8, no. 2, pp. 281–293, 1993.
- [50] H. Madsen, P. Pinson, G. Kariniotakis, H. A. Nielsen, and T. S. Nielsen, “Standardizing the performance evaluation of shortterm wind power prediction models,” *Wind Engineering*, vol. 29, no. 6, pp. 475–489, 2005.
- [51] T. Gneiting, “Making and evaluating point forecasts,” *Journal of the American Statistical Association*, vol. 106, no. 494, pp. 746–762, 2011.
- [52] L. J. Savage, “Elicitation of personal probabilities and expectations,” *Journal of the American Statistical Association*, vol. 66, no. 336, pp. 783–801, 1971.
- [53] T. Gneiting, F. Balabdaoui, and A. E. Raftery, “Probabilistic forecasts, calibration and sharpness,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 69, no. 2, pp. 243–268, 2007.
- [54] F. X. Diebold, T. A. Gunther, and A. S. Tay, “Evaluating density forecasts with applications to financial risk management,” *International Economics Review*, vol. 39, no. 4, pp. 863–883, 1998.
- [55] J. Bröcker and L. A. Smith, “Scoring probabilistic forecasts: The importance of being proper,” *Weather and Forecasting*, vol. 22, no. 2, pp. 382–388, 2007.
- [56] T. Gneiting and A. E. Raftery, “Strictly proper scoring rules, prediction, and estimation,” *Journal of the American Statistical Association*, vol. 102, no. 477, pp. 359–378, 2007.
- [57] H. Hersbach, “Decomposition of the continuous ranked probability score for ensemble prediction systems,” *Weather and Forecasting*, vol. 15, no. 5, pp. 559–570, 2000.

- [58] T. Gneiting and R. Ranjan, "Comparing density forecasts using threshold- and quantile-weighted scoring rules," *Journal of Business & Economic Statistics*, vol. 29, no. 3, 2011.
- [59] T. Gneiting, L. I. Stanberry, E. P. Gritmit, L. Held, and N. A. Johnson, "Assessing probabilistic forecasts of multivariate quantities, with an application to ensemble predictions of surface winds," *Test*, vol. 17, no. 2, pp. 211–235, 2008.
- [60] P. Pinson and R. Girard, "Evaluating the quality of scenarios of short-term wind power generation," *Applied Energy*, vol. 96, pp. 12–20, 2012.
- [61] J. M. Bernardo, "Expected information as expected utility," *The Annals of Statistics*, vol. 7, no. 3, pp. 686–690, 1979.
- [62] I. G. Damousis, M. C. Alexiadis, J. B. Theocharis, and P. S. Dokopoulos, "A fuzzy model for wind speed prediction and power generation in wind parks using spatial correlation," *Energy Conversion, IEEE Transactions on*, vol. 19, no. 2, pp. 352–361, 2004.
- [63] K. A. Larson and K. Westrick, "Short-term wind forecasting using off-site observations," *Wind energy*, vol. 9, no. 1-2, pp. 55–62, 2006.
- [64] T. Gneiting, K. Larson, K. Westrick, M. G. Genton, and E. Aldrich, "Calibrated probabilistic forecasting at the stateline wind energy center: The regime-switching space–time method," *Journal of the American Statistical Association*, vol. 101, no. 475, pp. 968–979, 2006.
- [65] A. S. Hering and M. G. Genton, "Powering up with space-time wind forecasting," *Journal of the American Statistical Association*, vol. 105, no. 489, pp. 92–104, 2010.
- [66] R. Girard and D. Allard, "Spatio-temporal propagation of wind power prediction errors," *Wind Energy*, 2012.
- [67] H. A. Nielsen, T. S. Nielsen, A. K. Joensen, H. Madsen, and J. Holst, "Tracking time-varying-coefficient functions," *International Journal of Adaptive Control and Signal Processing*, vol. 14, no. 8, pp. 813–828, 2000.
- [68] P. Pinson, L. E. A. Christensen, H. Madsen, P. E. Sørensen, M. H. Donovan, and L. E. Jensen, "Regime-switching modelling of the fluctuations of offshore wind generation," *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 96, no. 12, pp. 2327–2347, 2008.
- [69] P.-J. Trombe, P. Pinson, and H. Madsen, "A general probabilistic forecasting framework for offshore wind power fluctuations," *Energies*, vol. 5, no. 3, pp. 621–657, 2012.

- [70] U. Focken, M. Lange, K. Mönnich, H.-P. Waldl, H. G. Beyer, and A. Luig, “Short-term prediction of the aggregated power output of wind farms—a statistical analysis of the reduction of the prediction error by spatial smoothing effects,” *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 90, no. 3, pp. 231–246, 2002.
- [71] A. Lau, *Probabilistic Wind Power Forecasts: From Aggregated Approach to Spatio-temporal Models*. PhD thesis, Mathematical Institute, University of Oxford, 2011.
- [72] F. X. Diebold and R. S. Mariano, “Comparing predictive accuracy,” *Journal of Business & economic statistics*, vol. 20, no. 1, 2002.
- [73] P. Pinson, “Adaptive calibration of (u, v)-wind ensemble forecasts,” *Quarterly Journal of the Royal Meteorological Society*, vol. 138, no. 666, pp. 1273–1284, 2012.
- [74] A. Möller, A. Lenkoski, and T. L. Thorarinsdottir, “Multivariate probabilistic forecasting using ensemble Bayesian model averaging and copulas,” *Quarterly Journal of the Royal Meteorological Society*, 2012.
- [75] F. Thordarson, H. Madsen, H. A. Nielsen, and P. Pinson, “Conditional weighted combination of wind power forecasts,” *Wind Energy*, vol. 13, no. 8, pp. 751–763, 2010.
- [76] F. Lindgren, H. Rue, and J. 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)*, vol. 73, no. 4, pp. 423–498, 2011.