Topics in High Dimensional Energy Forecasting

PhD Thesis

Ciaran Gilbert

Supervisors: Jethro Browell & David McMillan

CDT Wind & Marine Energy Systems

Department of Electronic and Electrical Engineering

University of Strathclyde, Glasgow

June 24, 2021

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.

Abstract

The forecasting of future energy consumption and generation is now an essential part of power system operation. In networks with high renewable power penetration, forecasts are used to help maintain security of supply and to operate the system efficiently. Historically, uncertainties have always been present in the demand side of the network, they are now also present in the generation side with the growth of weather dependent renewables. Here, we focus on forecasting for wind energy applications at the day(s)-ahead scale. Most of the work developed is for power forecasting, although we also identify an emerging opportunity in access forecasting for offshore operations. Power forecasts are used by traders, power system operators, and asset owners to optimise decision making based on future generation.

Several novel methodologies are presented based on post–processing Numerical Weather Predictions (NWP) with measured data, using modern statistical learning techniques; they are linked with the increasingly relevant challenge of dealing with high-dimensional data. The term 'high-dimensional' means different things to different people, depending on their background. To statisticians high dimensionaility occurs when the dimensions of the problem are greater than the number of observations, i.e. the classic $p >> n$ problem, an example of which can be found in Chapter 7. In this work we take the more general view that a high dimensional dataset is one with a high number of attributes or features. In wind energy forecasting applications, this can occur in the input and/or output variable space. For example, multivariate forecasting of spatially distributed wind farms can be a potentially very-high dimensional problem, but so is feature engineering using ultra-high resolution NWP in this framework.

Most of the work in this thesis is based on various forms of probabilistic forecasting.

Probabilistic forecasts are essential for risk-management, but also to risk-neutral participants in asymmetrically penalised electricity markets. Uncertainty is always present, it is merely hidden in deterministic, i.e. point, forecasts. This aspect of forecasting has been the subject of a concerted research effort over the last few years in the energy forecasting literature. However, we identify and address gaps in the literature related to dealing with high dimensional data in both the input and output side of the modelling chain.

It is not necessarily given that increasing the resolution of the weather forecast increases the skill, and therefore reduces errors associated with the forecast. In fact and when regarding typical average scoring rules, they often perform worse than smoother forecasts from lower-resolution models due to spatial and/or temporal displacement errors. Here, we evaluate the potential of using ultra high resolution weather models for offshore power forecasting, using feature engineering and modern statistical learning techniques.

Two methods for creating improved *probabilistic* wind power forecasts through the use of turbine-level data are proposed. Although standard resolution NWP data is used, high dimensionality is now present in the output variable space; the two methods scale by the number of turbines present in the wind farm, although to a different extent. A methodology for regime-switching multivariate wind power forecasting is also elaborated, with a case study demonstrated on 92 wind balancing mechanism units connected to the GB network.

Finally, we look at an emerging topic in energy forecasting: offshore access forecasting. Improving access is a priority in the offshore wind sector, driven by the opportunity to increase revenues, reduce costs, and improve safety at operational wind farms. We describe a novel methodology for producing probabilistic forecasts of access conditions during crew transfers.

Contents

Abstract **ii**

List of Figures **vii**

List of Tables **xv**

Acknowledgements **xix**

1 Introduction **2**

1.1 Forecasting in Wind Energy 5

1.2 High Dimensional Forecasting 7

1.3 Objectives of Research & Novelty 9

2 Energy Forecasting: Wind Energy Applications **11**

2.1 Forecast Lead Time 12

2.2 Numerical Weather Prediction 13

2.3 Types of Forecasts 15

2.3.1 Deterministic 17

2.3.2 Probabilistic 18

2.4 State of The Art 24

2.4.1 Feature Engineering & Forecasting Competitions 25

2.4.2 Hierarchical Forecasting 27

2.4.3 Multivariate Forecasting 29

2.4.4 Forecasting in Wind O&M 31

2.5 Summary 33

3 Forecast Verification **35**

3.1 Deterministic . . . 36

3.2 Probabilistic . . . 37

3.2.1 Univariate . . . 38

3.2.2 Multivariate . . . 42

3.3 Comparing Performance . . . 45

3.3.1 Cross-Validation . . . 45

3.3.2 Scoring Uncertainty . . . 46

3.4 Summary . . . 48

4 Statistical Post-processing of Turbulence Resolving Weather Forecasts **49**

4.1 Case study at the Horns Rev offshore wind farm . . . 51

4.1.1 Horns Rev . . . 51

4.1.2 Weather forecasts: from ECMWF to the Whiffle Large Eddy Simulation . . . 52

4.2 Post-processing models . . . 55

4.2.1 Wind speed forecasting . . . 55

4.2.2 Power forecasting . . . 56

4.2.3 Feature Engineering . . . 58

4.3 Results . . . 61

4.3.1 LES results . . . 62

4.3.2 Exploratory Data Analysis . . . 62

4.3.3 Wind Farm Power . . . 68

4.4 Conclusions & Future Work . . . 69

5 Leveraging Turbine-level Data for Probabilistic Wind Power Forecasting **73**

5.1 Forecasting Methodology . . . 74

5.1.1 Benchmark Models . . . 76

5.1.2 Turbine-level Feature Engineering . . . 77

5.1.3 Bottom-Up Probabilistic Method 78
5.2 Case Study 81
5.3 Results 82
5.4 Conclusions & Future Work 90

6 Probabilistic Access Forecasting for Improved Offshore Operations 92
6.1 Access in the Offshore Environment 94
6.2 Sea State Forecasting Methodology 95
6.2.1 Parametric & Non-Parametric Regression 95
6.2.2 Scenario Forecasting 101
6.2.3 Wave Direction Regimes: Clustering & Logistic Regression . . . 102
6.3 Vessel Motion During Transfer 103
6.4 Case Study 103
6.4.1 Sea State Forecasting 105
6.4.2 Vessel Motion 115
6.4.3 Forecasting Vessel Motion during Transfers 118
6.5 Conclusions & Future Work 120

7 Regime Switching Multivariate Wind Power Forecasting 122
7.1 Forecasting Methodology 124
7.1.1 Density Forecasting 124
7.1.2 Scenario Forecasting 126
7.1.3 Regime Switching 131
7.2 Case Study 132
7.2.1 Univariate: Density Forecasting 136
7.2.2 Multivariate: Scenario Forecasting 142
7.3 Conclusions & Future Work 150

8 Conclusions 153

9 Published Work 156

A Statistical Post-processing of Turbulence Resolving Weather Forecasts **158**

B Leveraging Turbine-level Data for Probabilistic Wind Power Forecasting **160**

B.1 Wind Farm B 160

B.2 Wind Farm A 165

C Probabilistic Access Forecasting for Improved Offshore Operations **169**

D Regime Switching Multivariate Wind Power Forecasting **173**

Bibliography **176**

List of Figures

1.1 Annual statistics for UK wind energy [1] until 2018. There has been substantial growth in both installed capacity and energy generation over the last 24 years 3

1.2 The empirical power curves of an example wind farm, where Pn is the nominal capacity of the wind farm. At the basic level, it is the relationship in (b) learned by day-ahead forecasting models, which illustrates why accounting for forecast uncertainty is important, especially in the cubic range of the power curve 7

2.1 Example deterministic forecast, where estCap. is the estimated capacity. Deterministic forecasts provide a 'best guess' at each forecast lead time, although uncertainty information is essentially hidden 17

2.2 Example probabilistic forecasts, where estCap. is the estimated capacity. Here, the full distribution is not communicated, although users can see that the uncertainty grows as wind speed forecasts move into the cubic region of the power curve, for the proceeding two days. Figure 1.2b shows the dominant effect of the cubic region of the power curve at the medium range wind speeds 19

2.3 Example density forecasts, where estCap. is the estimated capacity. Here, the density forecast is constructed with multiple quantile forecasts and monotonic spline interpolation at each lead time; a more complete view of the uncertainty is shown 21

2.4 Example temporal covariance matrix of 'forecast errors' which characterises the dependence structure of the multivariate Gaussian copula and permits the generation of scenario forecasts. Here, the errors persist for longer as the lead time develops, due to degradation of the NWP skill at longer lead times 22
2.5 Example statistically based scenario forecast as a spaghetti plot, where estCap. is the estimated capacity. The uncertainty information can be difficult to interpret in this visualisation, but the scenarios are useful for simulations or further processing 23

3.1 Example PIT histogram. Here, the forecast displays over-dispersion; observations are more concentrated at the mid-region of the distribution and the distribution is on average under-confident 40
3.2 Graphical illustration of the CRPS scoring rule for a single lead time, for two forecasts. The score is the area below the squared difference between the observation (step function) and forecast CDF 41
3.3 Example graphical illustration of the energy score contributions from several bivariate normal scenarios and the corresponding observation. The energy score is a function of both the spread and difference components of the scenarios 43
3.4 Boxplots showing the sampling variation of the ES skill between a benchmark and example multivariate forecast. Here, there is a small but significant improvement in the ES for both bootstrap methods; however the sampling uncertainty is greater using the block-bootstrap 47

4.1 Schematic representation of the LES model setup. ECMWF fields are used to estimate the initial conditions and are applied as dynamic tendencies to a pre-cursor LES with periodic boundary conditions. The precursor LES values are then prescribed on the boundaries of a second LES to prevent the wakes from re-entering the domain 53

4.2 Wind speed fields at hub height for 2016-05-18 02:00 UTC, where the turbine locations and their yaw angles are depicted with the grey bars. The average deficit (percent reduction with respect to free-stream wind speed) is calculated using the preceding two hours 54
4.3 Example time series of raw wind speed forecasts from both NWP sources versus the turbine anemometer measurement. The GRASP and ECMWF variables used are the disk averaged and 70 metre wind speed respectively. Each of the 4 day-ahead forecasts have different issue times, indicated by the jumps in wind speed predictions around midnight each day 59
4.4 Yearly average power from observations (left) and forecast by the GRASP LES model (right) 63
4.5 Wind speed RMSE results with different smoothing windows at 70m height. Forecasts were derived from a truncated linear regression model with different smoothed forecasts as input. At this turbine, ECMWF gives more accurate wind speed forecasts across all the smoothing windows 64
4.6 Wind power RMSE results with different smoothing windows. Forecasts were derived from GBM models with smoothed wind speed input variables at 70m height for ECMWF and the disk averaged wind speed (DA) for GRASP. At this turbine, GRASP gives more accurate wind power forecasts across all the smoothing windows. This is because GRASP gives superior wind speed forecasts, *at regions key to wind power prediction* as shown on Figure 4.7a 64
4.7 RMSE of Whiffle's GRASP model and ECMWF conditioned on 5 equally populated observation bins with a 400 minute smoothing window and at a height of 70m. The shaded areas show the interquartile range from bootstrap score averages 65
4.8 Wind speed RMSE at different model heights with a 500 minute smoothing window. DA and DA-all are the disk average wind speed, and the wind farm spatial average of the disk average wind speed respectively . 66

4.9 Wind power RMSE at different model heights with a 400 minute smoothing window. DA and DA-all are the disk average wind speed, and the wind farm spatial average of the disk average wind speed respectively . 66
4.10 Wind power RMSE where ECMWF features are at 70m height and for GRASP the disk averaged wind speed and wind direction at 70m are used for GRASP. **A** is the raw NWP signal; **B** is the smoothed NWP signal over 400 minutes; **C** is the smoothed NWP and wind direction; **D** is the smoothed NWP, PSD features, and wind direction; **E** is the smoothed NWP, rolling variance, and wind direction 67
4.11 Wind farm power results 70

5.1 Flowchart illustration of the entire forecasting model training methodology. Wind Farm Level Power Data and Turbine Level SCADA data are only required for model training and evaluation. The only inputs required to produce operational forecasts are Numerical Weather Predictions. TB indicates turbine, WF is wind farm, Det. is deterministic, and Reg. is regression 75
5.2 Example density forecast using the parametric Gaussian copula approach at Wind Farm A 82
5.3 Wind Farm A calibration plots 85
5.4 Wind Farm B calibration plots 86
5.5 Boxplots showing the bootstrap sample distributions of mean CRPS for the best benchmark and proposed model at both wind farms 87
5.6 Parametric covariance matrix at Wind Farm B 88
5.7 Parametric covariance plots at Wind Farm A. Note that the latitude and longitude scales of (b) are indicative 89

6.1 Flowchart of the modelling chain in operation and training 93

6.2 Example predictive CDF for significant wave height at a single lead time, based on either multiple quantile regression or distributional regression. In the latter, the conditional distribution family is the Generalised Beta Prime 97
6.3 Example probabilistic forecasts of significant wave height using `gamboostLSS` parametric regression. In 6.3a the intervals plotted cover specified probability level ranges, e.g. the 90%int. is the 5% — 95% quantile range . . 101
6.4 Results for all lead times — significant wave height density forecasting by regression model 107
6.5 Results for all lead times — peak wave period density forecasting by regression model 109
6.6 Evaluation of the sharpness for `gamboostLSS` density forecasts vs lead-time during testing. The width of the prediction intervals increases with lead-time as the forecasts becomes less confident further into the future 110
6.7 Temporal correlation matrices during testing 112
6.8 Boxplots showing the block-bootstrap sample distributions of Energy Score improvement for significant wave height (swh), peak wave period (pwp), and coupled (swh & pwp) scenario forecasting. The benchmark in all cases is the *independence* case, with no temporal correlation. Linked means the temporal inter-dependency between significant wave height and peak period is modelled 114
6.9 Sea-state regime classification plots — regime forecasts are used in the vessel motion model 115
6.10 Plot of vessel motion during push-on and concurrent ocean measurements. The distributional model fit is for the *tr*-T-5 model in the south west regime with a fixed peak period. Note that push-ons only occur in a small range of the possible significant wave height forecast values and the variance is capped at the highest observed significant wave height. See Sections 6.4.3 and 6.4.3 for more detail on dealing with this effect . 116
6.11 Example scenario forecast of mean vessel motion during transfer 119

6.12 Visualisation stage plots 120

7.1 Illustration of the dynamic exponential tails, which are conditional on the upper/lower quantile forecasts for the upper and lower tail respectively. The tail shape becomes more exponential as the reference quantile moves further from the boundary. Note that the x axis here is not power but the i or $1 - i$ variable, the ranges are therefore not-scaled 126

7.2 Density forecasts as CDFs at clyde central wind farm issued at 2017-04-01 UTC, where estCap. is the estimated capacity. The density forecast is constructed with multiple quantiles and monotonic spline interpolation at each lead time. The approximated CDFs show the high resolution of quantile estimates in the tail regions of the distribution 127

7.3 Contour plots of a parametric covariance using equation 7.6 with paramaters a=0.4, $\tau_t = 0.7$, $\tau_k = 60$, $\nu = 0$, and finally $\beta = 0$ for the left plot and $\beta = 1.25$ on the right. The interaction parameter allows for more flexibility when modelling the covariance 129

7.4 Each wind Balancing Mechanism Unit (BMU) location in the GB grid. Note that the larger wind farms are decomposed into several BMUs . . 133

7.5 Univariate evaluation: reliability diagrams over all lead times during both cross validation (CV) and testing datasets. The models are very well calibrated in cross-validation; however not as well calibrated in testing. This is because of the volume of data we have in cross validation, and that we are forecasting a whole year at each site during testing with no re-training 137

7.6 Univariate evaluation: reliability diagrams in lead time groups during both cross validation (CV) and testing datasets 138

7.7 Univariate evaluation: pinball loss over all lead times during both cross validation (CV) and testing datasets. The average pinball loss is very similar in both cases which means that the model generalised well to unseen data 139

7.8 Univariate evaluation: PIT histograms of a subset of the BMUs considered during cross validation. The calibration of the full distribution has a big impact on the quality of the copula model 140
7.9 Univariate evaluation: quantile bias diagrams of the extreme quantiles over all lead times during cross validation. We show $\alpha \in \{0.001, 0.00125, 0.0025, 0.005, \text{and } 0.01\}$ for the lower tail, the maximum of the set is modelled using quantile regression and the rest are found using a dynamic exponential interpolation method. The mirror image of these quantiles are plotted for the upper tail. Given the scale of the bias, the extreme quantiles are relatively well calibrated considering the simplicity of the interpolation method. 141
7.10 Univariate density forecasts as fan plots at the subset of BMU locations issued at 2018-02-25 12:00 UTC. The same legend as Figure 6.3a applies 143
7.11 Multivariate evaluation: boxplots showing the block-bootstrap sample distributions of multivariate score improvement for the different dependencies tested, over the *independence* benchmark. The improvements are relatively small, but significantly different from zero. There is marginal improvement using the regime-switching method over the *spatiotemporal* case; however this is not significant. Note that the variogram score is unweighted, subsequently the improvement is relatively small here as well. Both scores are in agreement in terms of forecast performance . . . 146
7.12 Spatiotemporal covariance matrices on a subset of BMUs, the full matrix used to generate scenario forecasts on the testing dataset is shown on Figure 7.13. The parametric model regularises the covariance structure, compared to the sample estimate. Figure 7.12b is used to learn the coefficients for the parametric model, and white space indicates where the sample covariance is negative. Please refer to Figure 7.14 for higher resolution axis labels 147
7.13 Parametric covariance matrix of all BMUs used to generate scenario forecasts on the testing dataset 148

7.14 Difference between the regime-switching covariance matrices (easterly minus westerly) of a subset of BMUs during the testing dataset. The temporal dependency persists for longer in the easterly regime and the spatial dependency is somewhat similar 150
7.15 Multivariate scenario forecast as a spaghetti plot with 100 samples at the subset of BMU locations. The forecast is issued at 2018-02-25 12:00 UTC, and scenarios are blue and the corresponding measurement black 151

B.1 Wind Farm B: relative influence of input features: WF($\mathbf{x}_t$) benchmark . 162
B.2 Wind Farm B: relative influence of input features: WFT($\mathbf{x}_t^{\text{SUP}}$) feature engineering wind farm probabilistic model 162
B.3 Mapped parametric partial covariance at Wind Farm B of turbine 35 [2]. Latitude and longitude scales are indicative 164
B.4 Wind Farm A: relative influence of input features – WF($\mathbf{x}_t$) benchmark 165
B.5 Wind Farm A: relative influence of input features – WFT($\mathbf{x}_t^{\text{SUP}}$) feature engineering wind farm probabilistic model 168

D.1 Pinball loss in lead time groups during both cross validation (CV) and testing datasets 177
D.2 Comparison between separable and non-separable covariance model fits for the purely temporal and purely spatial cases 178
D.3 PIT histograms of all BMUs considered during cross validation, the number of observations available are overlaid on the plots 179

List of Tables

4.1 Pearson correlation coefficient of raw and smoothed forecasts with wind speed observations. For smoothing the moving average spans 500 minutes 58

5.1 Results at Wind Farm A. %Δ indicates improvement compared to specified benchmark [CRPS in % of max power] 84
5.2 Results at Wind Farm B. %Δ indicates improvement compared to specified benchmark [CRPS in % of max power] 84
5.3 Deterministic forecast performance based on the median (p50) of each predictive distribution — %Δ indicates improvement compared to specified benchmark [MAE as % of max power] 90

6.1 Multivariate forecast evaluation results. Linked means the temporal dependency between significant wave height and peak period is modelled. Best results during testing are in bold. Note that for the joint wave height and period target variable case, the marginals and observations are first standardised before applying scoring rules 111
6.2 Selected formulations for the vessel motion model during transfer dependent on observed sea-state. The symbols *ps* indicates a penalised beta spline, † is a varying coefficient model, *poly* is a fractional polynomial model, H_s is significant wave height, T_p is peak wave period, ω_p is peak wave direction, and *reg.* is the regime membership 117

7.1 Multivariate forecast evaluation results. Best results during testing and cross-validation are in bold. The evaluation results are based on 500 scenarios at each issue time 144
7.2 Multivariate evaluation: Diebold-Mariano test statistics based on the energy score differential on the out-of-sample cross validation data, where p values are shown in bold 145
7.3 Fitted coefficients (est.) for the parametric covariance models based on Equation 7.6, which describes the uncertainty dependency used to issue scenario forecasts over the test dataset, along with approximate 95% confidence intervals 149

A.1 List of input features for wind farm level power prediction using either ECMWF or GRASP data, or both sources as in the combined case. Specific heights and locations are indicated in square brackets for before, after, and in the combined feature selection case. TB is short for turbine, RA means rolling average, and RV means rolling variance. The FFT features were calculated over a 400 minute window on the disk averaged wind speed and the features extracted from the high frequency band. . . 159

B.1 Wind Farm B: list of input features by model - meteorological inputs at 100m (unless specified) 161
B.2 P50 MAE Results at Wind Farm B - %Δ indicates improvement compared to specified benchmark above [MAE in % of max power] 163
B.3 P50 RMSE Results at Wind Farm B - %Δ indicates improvement compared to specified benchmark above [RMSE in % of max power] 163
B.4 Details of Copula Vine selections at Wind Farm B 164
B.5 Wind Farm A: list of input features by model - meteorological inputs at 100m (unless specified) 166
B.6 P50 MAE Results at Wind Farm A- %Δ indicates improvement compared to specified benchmark above [MAE in % of max power] 167

B.7 P50 RMSE Results at Wind Farm A - %Δ indicates improvement compared to specified benchmark above [RMSE in % of max power] 167
B.8 Details of Copula Vine selections at Wind Farm A 168

C.1 List of input features and base learners by model — significant wave height regression 170
C.2 List of input features and base learners by model — peak wave period regression 171
C.3 List of input features and base learners — clustered peak wave direction logistic regression 172

D.1 List of wind balancing mechanisms units (BMUs) used in the case study, with associated details from the UK Government renewable energy planning database [3]. GC means generating capacity, which for the BMUs is subject to change. NC is the site nameplate capacity 173
D.2 Diebold-Mariano test statistics based on the energy score differential on the testing data, where p values are shown in bold 178

Acknowledgements

First I would like to thank my two supervisors Jethro Browell and David McMillan, who have supported me throughout this work by giving me the very best guidance, advice, and motivation. I have really enjoyed the past few years and becoming involved in energy forecasting research, in no small part due to them and the opportunities they provided.

Acknowledgements are due to the EPSRC for funding the (amazing) Wind & Marine Energy Systems CDT, via grant EP/L016680/1. Thanks to Drew Smith, Bill Leithead, and everyone at the CDT who contribute to making the CDT what it is. Special thanks are also due to Pierre Pinson and Jakob Messner at DTU for hosting and mentoring me during a research visit; it was a pleasure to work with you both.

I am grateful to Catriona who is always ready to say that my graphs look great, and was very patient and supportive during the write-up of this thesis. Thanks to my parents, Bernadette and Alastair, who encourage me to do my best and kept asking how my 'wee essay' was going. To my sisters, Clair and AKG, thanks for always making me laugh and being there to wind me up. Last but not least, thanks to my pals Conor, Pete, Martin, Chub, Tumfy, and Hoagy for lots of reasons probably best not to mention here...

[page left intentionally blank]

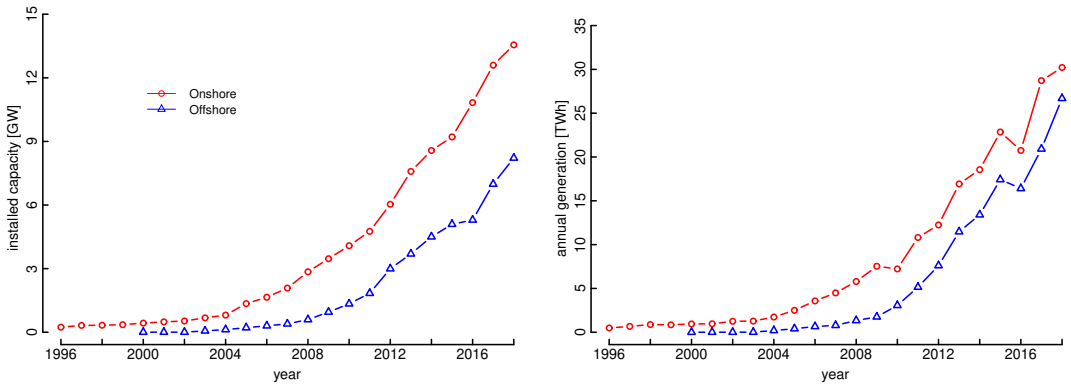
Chapter 1

Introduction

Renewable energy sources contributed a third of UK electricity generation in 2018 [1]. This share has risen from a meagre 2.6% in 2000 [1], and if the UK is to meet it's national and international climate targets should be expected to grow. Wind power, both onshore and offshore, represents 51.8% of this total contribution, with an installed capacity of 21.7 GW [1]. These figures show the seismic shift in the electricity landscape due to renewable technology. An illustration of the accelerated development of wind power in the UK in recent years is shown in Figure 1.1.

Looking globally, total wind power capacity in 2019 reached 650 GW, with 60 GW added during the year. Countries such as as China (236.4 GW), the USA (105.5 GW), Germany (61.4 GW), India (37.5 GW), and Spain (≈25 GW) lead the way in terms of installed capacity [4], but it is Denmark (48%), Ireland (33%), and Portugal (27%), that reported the highest percentages of demand serviced by wind energy in 2019 [5].

Along with the fast rise in installed capacity, there has been a sharp drop in the cost-of-energy associated with both onshore and offshore wind. Encouraging results recently have indicated that wind power is becoming the most economical option for new generation units in several markets. Specifically, in the UK the most recent auction results for Contracts for Difference (strike price) have been allocated, with Doggerbank A and Sofia offshore wind farms both awarded a staggering £39.65/MWh [6], which is cheaper than the average wholesale day-ahead electricity prices in 2018/19 (£58.6MWh) [7]. These new CfD results shift the view of the function of the subsidy;



(a) Installed Capacity

(b) Energy Generation

Figure 1.1: Annual statistics for UK wind energy [1] until 2018. There has been substantial growth in both installed capacity and energy generation over the last 24 years

these are now mainly a risk hedging strategy. Although, time will reveal the actual viability of these cost savings offshore. Also, in Scotland SSE Renewables have recently announced an extension to Gordonbush onshore wind farm, the first subsidy free wind farm in the UK [8].

Electricity is a challenging commodity to trade because of the lack of a viable industrial scale storage technology, meaning supply and demand of electricity must be matched in real time. In the pre-renewables era this challenge was tackled by using deterministic forecasts of demand to manage generation profiles and transmission constraints [9]. Modern-day power systems were originally built to accommodate large centralised thermal power stations stoked with energy-dense fossil fuels, transferring power via high voltage transmission to cities or industrial centres. Until recently, electricity markets were also geared toward accommodating these technologies.

Understandably, the shift in paradigm to decentralised and variable renewable energy sources throws up interesting challenges in transmission and distribution; uncertainties in the power network are now present on the supply-side increasing the difficulty of balancing the network [10]. In order to manage the influence of these weather-dependant renewable sources, it is now necessary to use predictions of future solar and wind generation. Furthermore, probabilistic forecasts are necessary for risk minimisation in both market and operations applications, and for effective economic

participation in electricity markets with asymmetric penalties.

As important as security of supply is the economic operation of the network. Historically the network was operated by a single monopoly, a vertically integrated company governing generation, transmission, distribution, and retail. These monolithic companies were broken-up and privatised during a period of liberalisation, which began with the UK in 1989 [11]. The modern-day system comprises of regional monopolies operating distribution governed by a regulator (e.g. OFGEM), private companies competing for generation and retail opportunities, and finally high voltage transmission managed by a Transmission System Operator (TSO). NationalGridESO, the TSO in the UK, currently cites energy forecasting as an innovation priority for 2020/21 [12].

To facilitate the new electricity landscape, markets such as the day-ahead, balancing, ancillary services, and capacity markets have been created to try and ensure the economic and secure supply of electricity. All participants in these markets, regardless of portfolio make-up, are therefore exposed to the risks involved with a high penetration of variable renewable power, to some degree [10]. Financial penalties are incurred by participants for failing to meet agreed generation levels, and also for over-producing. However there are also financial rewards available to shrewd participants for helping to balance the grid at delivery — if the grid is short of supply then being able to step in and increase generation levels, and vice versa.

Participants in these markets rely on energy forecasts, as well as forecasts of price and demand, to make trading strategies, and given the high penetration of wind and solar on the grid, are now looking toward probabilistic forecasts to hedge risk and balance the generation profile [10]. Within this challenging and substantial field, this thesis is focused on forecasting for wind energy applications. Applications include trading in balancing or day-ahead markets, balancing supply and demand, energy storage management, and other developing research themes such as wind farm operation and maintenance, and the role of predictability in investment decision making [13, 14].

1.1 Forecasting in Wind Energy

Wind itself can be considered as the bulk flow of air particles in the atmosphere. It is the speed and direction of this flow is of most importance for wind energy applications [15]. The flow of air particles is driven by differences in atmospheric pressure across the globe. A simple example of this effect on the local scale, and also important for wind energy forecasting, is the diurnal patterns of sea breezes. During the day, the land heats up to a much greater extent than the sea causing a temperature gradient, which warms air directly above the land. This warm air becomes less dense than in the surrounding area and rises, causing a region of low pressure over land and a more dense high-pressure region at sea. This gradient begins to equalise and wind moves swiftly inland. A similar but opposite effect is observed at night.

Similarly on a global scale, temperature gradients between the hotter equator region and the colder poles causes global wind circulation, which is then complicated by the Coriolis effect. This Coriolis effect deflects the flow of the wind to the right from the direction of travel in the northern hemisphere, which leads to the prevailing west/south-westerly winds across the UK.

Numerical Weather Predictions (NWP) can be used to gain some foresight into the characteristics of the wind (or atmosphere) for the next $\approx$10 days. NWP models are computed on the global scale by many organisations around the world, notably the Met Office in the UK; the European Centre for Medium-Range Forecasts (ECMWF) based in Reading; and the National Oceanic and Atmospheric Administration (NOAA) in the USA. These vendors generate atmospheric forecasts on a three dimensional gridded output covering regions of interest, or the entire globe. Often these organisations run a higher resolution model (grid) around the region which constitute their main customer base.

The wind to power conversion process is also a vital component of the problem. The power available in a 'chunk' of air is proportional to the velocity cubed. This well known relationship can then be modified to include the mechanical and electrical efficiency of a typical modern variable speed pitch regulated wind turbine to give the

power equation

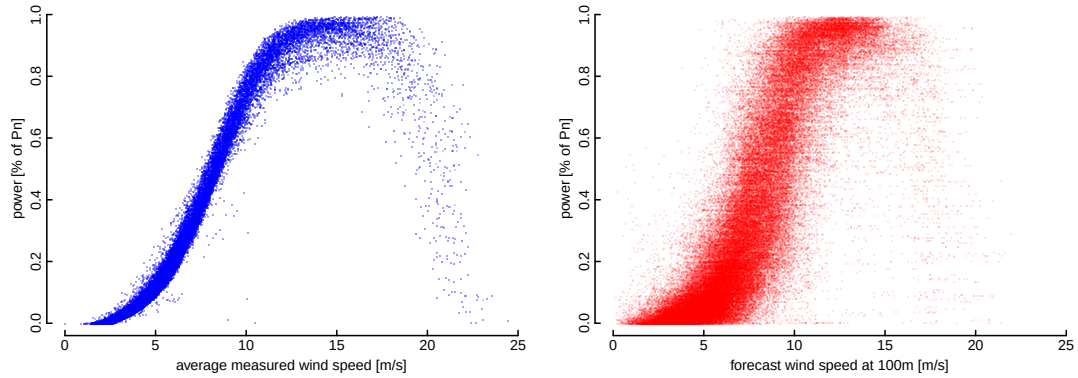
$$P = \frac{1}{2}\rho A C_p(\lambda, \beta) v^3 \tag{1.1}$$

where ρ is the density of air, A is the rotor swept area, C_p is the power coefficient, and v is the wind speed. The power coefficient is related to the aerodynamic efficiency of the specific turbine which depends on the tip speed ratio (λ) and the blade pitch angle (β).

This cubic relationship is only part of the story in the full turbine power curve. Turbines are typically designed to cut-out at low wind speeds (typically $<$ 4m/s), then as wind speed increases track the maximum possible aerodynamic efficiency curve ($C_{p,max}$) until some maximum rated power, where pitching of the blades is used to regulate the power output, to minimise generator and drive-train costs. At extreme wind speeds ($>$ 25m/s) the turbine is designed to cut-out, by pitching the blades out of the wind to protect the electrical and mechanical equipment [16]. Turbine vendors rigorously test turbines to characterise this relationship, which is known as the manufacturers power curve.

In the field, mechanical wear, erosion, yaw misalignment, and general degradation of turbine components can influence the power curve. Therefore, for forecasting applications the empirical power curve is most relevant. An example wind farm empirical power curve is shown on Figure 1.2a versus average wind speed, measured via a cup anemometers on each turbine nacelle. Figure 1.2b also shows the same power data plotted versus concurrent forecast wind speed at the site using ECMWF forecasts, extracted at the closest grid point and interpolated to the required time resolution. At the basic level, it is a version of the latter relationship that is being tackled in the forecasting problem at the day-ahead temporal scale, and the plot is an indicator of why accounting for uncertainty in the power forecast is quite so essential.

Outwith the typical wind speed and power forecasting applications, other target variables can be very important for wind energy applications. It is estimated that 20-30% of the total cost of energy for an offshore wind farm is due to Operations & Maintenance (O&M) in the UK [17]. Since O&M savings can be achieved by operators at any stage of the project life cycle and independently of turbine manufacturers there



(a) Average measured wind speed

(b) Forecast wind speed

Figure 1.2: The empirical power curves of an example wind farm, where Pn is the nominal capacity of the wind farm. At the basic level, it is the relationship in (b) learned by day-ahead forecasting models, which illustrates why accounting for forecast uncertainty is important, especially in the cubic range of the power curve

is a great opportunity to reduce this percentage. Therefore, improving installation, operations, and maintenance practices is a current focus in both industry and academia. Day-ahead forecasting can play a role here for improving decision making around the dispatch and schedule of vessels and technicians. Deterministic forecasts of significant wave height are typically used to forecast turbine accessibility in practice; this is an area which could benefit from probabilistic information as risk management is key for these safety critical decisions.

1.2 High Dimensional Forecasting

There is no strict condition that defines high dimensional data analysis. However, it is generally interpreted as a dataset with a high number of attributes or features. In wind energy forecasting applications, high dimensionality can enter the modelling chain in both the input and output variable space. For the typical example of day-ahead power forecasting, this means that high dimensionality can be present in the NWP feature side and/or in the power variable side. An example of the former is the use of ultra-high resolution NWP models, the research question is: how best to integrate the temporal and spatial information content from the NWP to generate more accurate forecasts?

and: is this high dimensional input feature space worthwhile?

In terms of the target variable side, high dimensionality can occur at multiple temporal and spatial scales. A current research theme in the wider forecasting community is hierarchical forecasting (see Section 2.4). Here, an imposed structure is applied on the target variables such that the physical reality of the system being forecast is respected. For a wind farm, this structure could mean that forecasts of the individual turbine power should sum to the wind farm forecast, which is not guaranteed when modelling components separately. Therefore, the dimensionality of the forecast problem scales with the number of turbines.

Specific types of probabilistic forecasts are required when making time and/or space dependent decisions, such as balancing national generation/demand and managing transmission constraints. These are termed trajectory forecasts and can be either statistical or physically based trajectories. A familiar example of this type of forecast is ensemble members from a NWP model [18]. This research is focused on statistical trajectory (i.e scenario) forecasts. Continuing the wind power example, this problem can be very high dimensional in nature; the method demands that the dependencies in time and/or space are embedded in the forecasting model chain for potentially all transmission connected wind farms.

High dimensional analysis is an area of interest in the wider forecasting community [19–21]. This is driven by the wealth of data now being collected across basically all areas of modern life. Forecasting in such frameworks is computationally demanding and a challenging prospect. Traditional methods for data visualisation, exploration, and manipulation; fitting robust models; and forecast evaluation all become more difficult with the increased dimensions of the problem. In retail, a common high dimensional problem is demand forecasting [22]; this is reflected in the recently completed 'M5' forecasting competition [23]. There, the task is to forecast ≈40k time-series of hierarchical sales data both expected sales and (optionally) uncertainty information.

Research in high dimensional statistics is ongoing outwith forecasting. Areas such as image classification, social networking, online streaming, medicine, banking, entertainment etc. are using high dimensional data to (for example) recommend new

products to customers [24]. Regularisation is essential for regression, dependency, and clustering algorithms using such volumes of data, i.e. sparse solutions help prevent over-fitting [25]. This a recurring theme throughout the chapters of this thesis where boosted/penalised regression models and sparse/parametric dependency structures are applied in the energy forecasting context.

1.3 Objectives of Research & Novelty

The objective of this research is to develop and improve novel forecasting methods for wind energy applications, with a focus on high dimensional problems. This is carried out at a variety of spatial scales, but more generally at the day-ahead temporal scale. Wind power forecasting is the main focus, along with forecasting turbine accessibility, which is an emerging research theme that incorporates metocean aspects. In this section, we discuss the research aims which are explored in Chapters 4, 5, 6, and 7, and outline the contributions to knowledge.

The pursuit of ultra-high resolution forecasts and appraising their value for wind power forecasting is not a new topic and has been discussed in literature with varying success [26, 27]. The situation has evolved substantially over the last few years, thanks to quantity of data being collected, increases in computational power, and advances in data science. Consequently, this enables the novel forecasting framework in Chapter 4, based on computationally demanding approaches to producing ultra high-resolution weather forecasts at the wind farm, and post-processing via advanced statistical learning techniques. In this case, the research aim is simple: to ascertain if it is possible to extract value from a high dimensional input NWP model in the context of wind power forecasting.

The nature of wind power on the network is hierarchical; from the turbine, to the farm, regional, and national level. Tackling the forecasting challenge with this in mind is important when creating forecasts that are coherent (i.e. sum appropriately at all levels of the system). Currently, the topic of combining hierarchical *and* probabilistic forecasting is lacking in literature. It is proposed in Chapter 5 that leveraging informa-

tion from the turbine-level will enable improved wind-farm level forecast performance, particularly since modern utility scale wind farms are often distributed over large areas of complex terrain and as a result, individual turbines can experience different conditions from one another at any given time.

Furthermore, correlations in power forecast errors at the national scale are explored in Chapter 7. Here, the research aim is to quantify the uncertainty in wind power forecasts for the utility scale wind farms in the UK and importantly, exploit the forecast error dependencies in space and time. There are numerous studies published in this topic; however, these are normally based on a few renewable generators or regional generation, and exhibit static or simple time-varying dependencies. Here, we develop novel conditional dependency structures, applied over a substantial spatial scale across large wind generators in the GB network.

Lastly, forecasting turbine accessibility is identified as an emerging opportunity to aid decision making, and an area currently lacking in the literature. In Chapter 6 the aim is to provide a novel an end-to-end framework for generating access forecasts based on vessel monition during transfer. Quantifying the uncertainty due to weather conditions and using the temporal correlation structure of forecast errors is essential for these multi-temporal decisions.

Chapter 2

Energy Forecasting: Wind Energy Applications

This thesis is focused on short-term forecasting, where the prediction horizon is several hours to days ahead. Numerical Weather Predictions are key inputs into wind power forecasting models at such horizons [26]. Best practice in creating these models involves mapping the relationship between meteorological forecasts and the corresponding target variable via a statistical learning technique. Outwith the energy forecasting community this process is more commonly known as statistical post-processing.

The statistical post-processing of raw weather forecasts to a specific location of interest, using measured data from the site, gives significant improvement in forecast performance [28, 29]. In many cases, engineering additional explanatory features from raw NWP data can significantly improve performance, a practice widely adopted in the energy forecasting community [30–32]. Statistical learning tools that perform feature selection and regularisation have been successful in international forecasting competitions [33], improving out-of-sample performance, i.e. the ability of the model predict well on unseen data. Gradient boosted regression trees in particular, which are an ensemble of 'weakly predictive' trees, have been used widely by the winning teams at these competitions; briefly, feature selection and regularisation is achieved by penalising the effect of each individual tree and by selecting only a few variables with high explanatory power during the fitting of each tree.

This chapter summarises key components in energy forecasting including, forecast lead time, NWP, the types of forecasts (i.e. deterministic, probabilistic, univariate, multivariate), as well as a literature review of state-of-the art forecasting methodologies, all with a focus on wind energy applications.

2.1 Forecast Lead Time

The lead time of the forecast determines the suitability of methods and data for creating an informative model. Very short term forecasts — 0 to (approximately) 6 hours ahead — are usually built from purely statistical time-series models using information from individual or spatially distributed power time series [29]. Looking further ahead in the future — greater than approximately 6 hours ahead — the use of Numerical Weather Prediction (NWP) has shown to give higher forecasting skill [26]. This finding extends outwith the typical power forecasting example, such as forecasting significant wave height [34].

Using the output of a NWP it is possible to use a data-driven approach to the short-term forecasting problem. This avoids any assumption of the physical phenomena governing the wind-to-power process. Supervised learning techniques map the input variables (NWP, historical and lagged power measurements, time-of-day, season) to the output wind power variable, and can include diverse techniques such as boosted regression trees, support vector machines, neural networks, time-series models, and regularised linear models among others [35].

Another important consideration is extracting the most amount of value from available resources such as the NWP — called feature engineering — and how to efficiently incorporate and process high dimensional data [31]. This general methodology implicitly accounts for site-specific effects such as turbine degradation, as well as any systematic NWP biases present and is therefore recommended best practice [36]. Turbine availability can also be accounted for using an appropriate normalisation of the power measurements or by using seasonal features in the regression, since most planned maintenance occurs during the summer months.

The superiority of purely statistical models within the very short lead times is due

to a number of factors; the most recent input measurements to a NWP model, used to initialise the physical model, may be several hours old by the time the forecast is issued, and errors introduced by the forecast wind to power process. Therefore, an ideal operational forecasting tool would include a blend of purely statistical and post-processed NWP models for improved accuracy across different lead times.

2.2 Numerical Weather Prediction

NWP models are initialised using the most recent available measurements of the atmosphere around the globe, from satellites and other sources, to estimate the current condition of the atmosphere and oceans. Possible future states of the atmosphere can then be computed at a specific temporal and spatial scale. Typically the spatial scales range from 1-25km, the temporal resolution of 1-3 hours, and forecasts can be issued multiple times each day [18]; the necessary elements of wind speed and direction can then be extracted for a specific region of interest.

Chaos theory plays a significant role in weather forecasting systems [37], meaning small changes in the underlying measurements that initialise the model can lead to significant differences in the resulting forecasts. Ensemble forecasts are used to quantify this uncertainty, where each ensemble member represents a NWP simulation initialised with a perturbed initialised state, or alternatively via multi-model ensembles with different physical representations. The former results in a number of equally likely scenarios of the future weather system, and forms the basis of global probabilistic weather forecasting. These are a type of physically-based scenario forecast, but the underlying structure of the uncertainty is very different to the statistically based counterparts, and they can capture different characteristics of the weather; ensemble forecasts can capture new events that may be present in a weather system, but are typically under-dispersed for a given location.

NWP models have been steadily improving over the last 40 years in terms of both forecast skill and temporal/spatial resolution due to advances in scientific research, computing capability, and the expansion of available observations used to initialise atmospheric models [18]. Currently, ECMWF offers global weather forecasts at a spatial

resolution of ≈9km (0.1°x0.1° grid) at 1-6 hour intervals within the high resolution 10-day forecast model. Wind speed forecasts retrieved directly from the ECMWF fields therefore have no resolved variance below these spatial and temporal scales. A common strategy to improve global weather forecasts is to downscale them using NWP models that cover only a limited area and are referred to as Limited Area Models (LAM) or mesoscale models. Mesoscale models have the advantage of representing the earth's surface in higher detail and resolve physical processes in higher temporal detail, but the formulation of physical processes like turbulence and (cloudy) convection is essentially the same as those of the global NWP models.

A purely physical approach to the wind power forecasting problem aims to improve the weather predictions by describing the flow inside the wind farm, by downscaling via a mesoscale model or CFD simulations. The improved description of the flow around the site can then be used with a manufacturer's power curve to estimate the power output of a farm. However, it is typically necessary to post-process systematic errors in such a model via Model Output Statistics (MOS) [35]. This forecasting method is most useful when there isn't historical data available at the site, for instance after commissioning and more so in the past when global models were more coarse in resolution.

Large-eddy simulations (LES) have a typical resolution of 10m-100m or less and directly resolve turbulence and boundary layer clouds. Furthermore, because a typical LES resolution allows for wind turbine resolving simulations, the use of LES to study flows through wind farms has received a lot of attention in the scientific literature lately [38–40]. Until recently though, the high spatial and temporal (roughly 5-10 seconds) resolution of LES prohibited their use in an operational forecast setting due to their high computational cost and excessively long run-times.

The pursuit of high resolution forecasts and appraising their value for wind power forecasting is not a new topic and has been discussed in literature with varying success [26, 27]. Advances in computing, like the use of GPUs to accelerate LES computations, have drastically shortened the run-times and have paved the way for operational weather forecasting using LES [41, 42]. A turbine resolving operational weather forecast on wind farm scale is currently possible and raises the research question of how to

integrate this high dimensional input feature space by statistical post-processing into a power forecasting tool.

In general, statistical post-processing is a common tool used in the wider weather forecasting community; the goal is to remove systematic bias present in the global NWP model for improved predictions at a specific location [43]. This process is typically carried out using Model Output Statistics, mentioned previously, via multiple linear regression [44], using historical NWP (of varying resolution and skill) and meteorological observations [45]. More recent applications include using machine learning for the regression [46] with varying success, post-processing ensembles for site specific calibration [47, 48], and using multiple forecast source information [28, 49]. Conventional post-processing techniques are difficult to apply to wind power directly because of the non-linear power curve and bounded nature of the time series [50]; this motivates the use of non-linear statistical learning techniques in this thesis.

Importantly, increasing the resolution of the NWP model does not immediately translate to improvements in forecast skill, which is usually measured on the average performance, e.g. root mean square error and mean absolute error — see Section 3. A forecast which may represent better the inherent behaviour of the underlying processes, can then be heavily penalised due to phase and location errors where the exact timing or placement of any particular weather event is missed. As a result, high resolution forecasts can often perform worse in terms of skill scores when compared to low resolution forecasts. This leaves an **open research question**: is possible to extract value from an ultra-high resolution NWP model in the context of wind power forecasting using advanced statistical learning techniques? This is explored in Section 4.

2.3 Types of Forecasts

There are a variety of types of forecast available to decision makers depending on the application and personal preference [35]. Ubiquitous in industry is the deterministic forecast which is simply a 'best guess' forecast involving a single number for a time horizon ahead. Deterministic forecasts for continuous variables are typically either optimised for the expected value or median value; accomplished simply by changing

the loss function in the regression task. The former are generated by minimising the squared loss and the latter the absolute loss. This is an important subtlety, as the loss function should be chosen appropriately for the end-use case [51]. For example, in wind power trading the financial penalty for over/under producing is typically based on absolute differences. Additionally, the type of loss function used dictates the most appropriate error metric for model development.

There are also probabilistic forecasts which quantify the uncertainty around this central 'best guess' in the form of intervals, quantiles, or full density forecasts. Scenario forecasts are a more academic topic for now, but are arguably the most useful for many applications and can be used for time/space dependent decision making [10]. Crucially, they contain information on the spatial-temporal correlations of forecast uncertainty. It is also important to note that forecasts are available commercially at different levels of aggregation: at the wind farm, regional, portfolio, and national scale.

In recent years, academic research has been focused on quantifying the uncertainty associated with energy forecasts due to the underlying chaotic weather systems driving the energy generation [52]. These uncertainty forecasts are useful for applications such operating reserve management [53] and defining optimal bidding strategies [54]. However, currently practitioners prefer deterministic (point) forecasts of future generation which comprise of single valued best estimates. This is due to the ease of interpretation and incorporation into existing decision making systems, as well poor communication of the underlying information content of an uncertainty forecast [36]. Therefore, improving point forecasts is still a relevant research pursuit. For comprehensive reviews of short-term wind power forecasting please refer to [26, 35].

The proceeding two subsections are intended to introduce the different types of forecasts available, touch on general methods available to generate the forecasts, and explore applications. For a detailed review of the state-of-the-art, the reader is referred to Section 2.4. All the example forecasts shown in the proceeding subsections are generated for the west coast offshore wind portfolio in GB using NWP data from ECMWF.

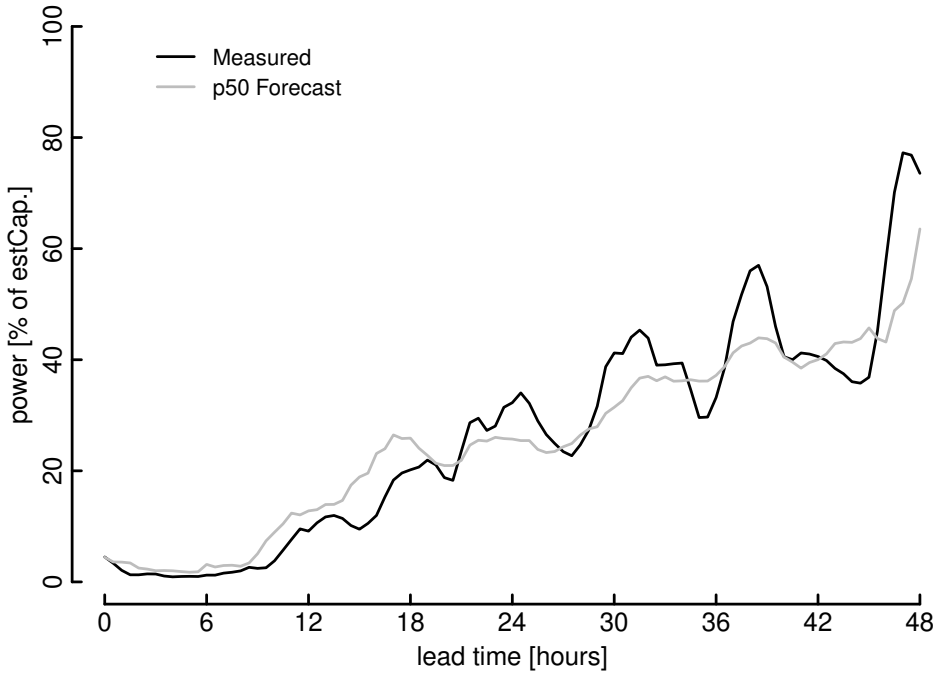


Figure 2.1: Example deterministic forecast, where estCap. is the estimated capacity. Deterministic forecasts provide a 'best guess' at each forecast lead time, although uncertainty information is essentially hidden

2.3.1 Deterministic

Wind power prediction was initially approached as a deterministic problem with research and early commercial products focusing on providing single-valued best estimates of future generation [55]. The ease of interpretation and incorporation into decision making models are clear advantages of this type of forecast. An example deterministic forecast is shown in Figure 2.1. Clearly, these are intuitive and provide moderate insight for end-users. Deterministic forecasting is approaching technological maturity following a concerted research effort reviewed comprehensively in [35, 56]. At present there are many commercial providers offering deterministic wind power forecasts.

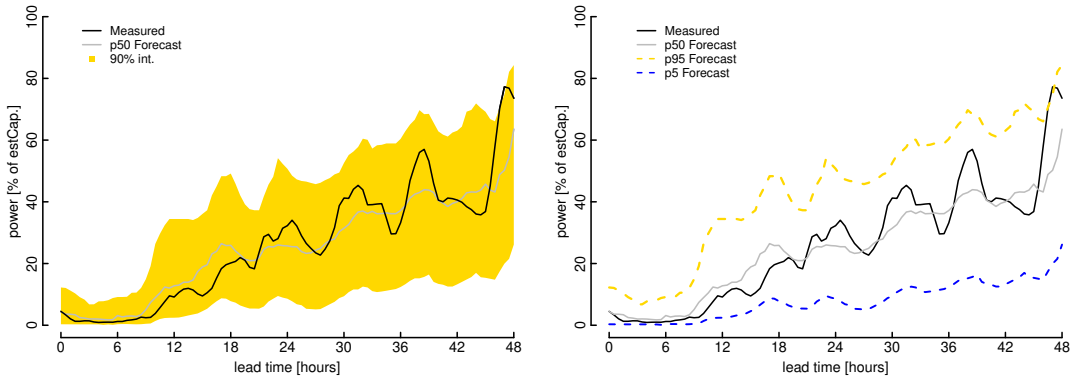
In the 1990s, as wind energy capacity started to reach a noticeable penetration on the grid, one of the first examples of an on-line power prediction tool was developed by Landberg [57]. The method used a high resolution NWP, known as the High Resolution Limited Area Model (HIRLAM), and transformed the wind direction and speed to

the surface height. A downscaling process was then employed to incorporate local effects such as obstacles and surface roughness. Following this a power curve and farm losses model transforms the local forecast wind speed to power predictions. Finally, systematic errors are reduced using MOS. This model averaged mean absolute error metrics of around 15% of installed capacity during a case study in Denmark, and was shown to be most useful in prediction lengths greater than 6 hours ahead.

Also commercially available is WindFor (formerly known as the Wind Power Prediction Tool - WPPT) which was developed at the Danish Technical University (DTU). This tool provides point (and optionally) uncertainty forecasts for a single wind farm, portfolios, or regions. This is done using time-adaptive estimation of conditional parametric models to relate NWP wind speeds with measured power for each lead time. The process has the advantage of being able to cope with a non-stationary process as old information is discounted by down-weighting with a forgetting factor [58]. There are various other commercial vendors of deterministic wind power forecasts, for more information please refer to [35]. The key argument for using probabilistic forecasts is that forecast uncertainty is only hidden using this type of forecast; it is always present.

2.3.2 Probabilistic

Uncertainty information associated with a forecast is essential for risk management and profit maximisation. For example: for a risk-neutral trader operating in a market with asymmetric penalties for over/under providing on wind generation, the optimal bidding strategy will not be the expected generation or deterministic forecast; rather it will be a specific quantile forecast derived from a predictive distribution [54]. Risk-averse traders and those with a risk-appetite require probabilistic forecasts regardless of the nature of the penalty in these markets. Probabilistic forecasts are also useful as inputs to unit-commitment problems, reserve holding, and managing power flows, among others [52]. Probabilistic information is also highly relevant to wind farm operations & maintenance (O&M) practices; risk management is important for vessel dispatch decisions because of the aim of maintaining high levels of crew safety whilst minimising turbine downtime. However, these decisions are typically based on deterministic forecasts of sea-state



(a) Interval forecast

(b) Quantile forecast

Figure 2.2: Example probabilistic forecasts, where estCap. is the estimated capacity. Here, the full distribution is not communicated, although users can see that the uncertainty grows as wind speed forecasts move into the cubic region of the power curve, for the proceeding two days. Figure 1.2b shows the dominant effect of the cubic region of the power curve at the medium range wind speeds

conditions, live measurements, and accrued experience of marine coordinators.

Univariate

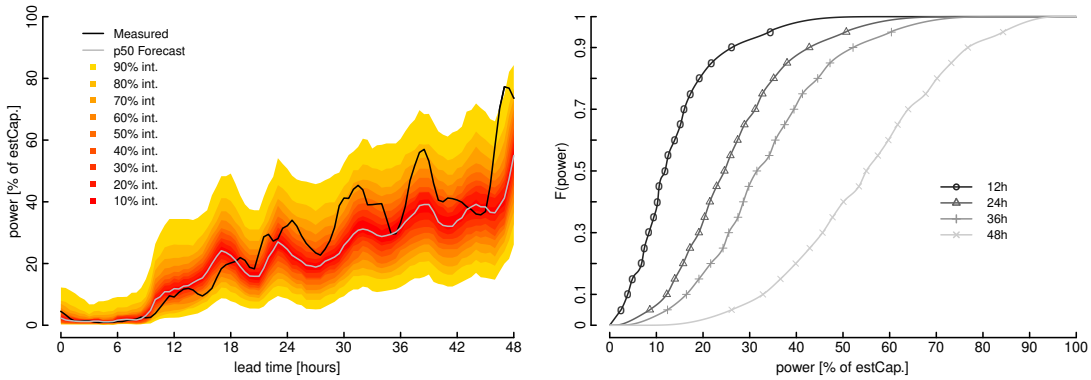
Univariate probabilistic forecasts, those issued independently on a per lead time and per location basis, can be represented in a number of ways, the three most common variations are the quantile, interval, and density forecast. The quantile forecast is any specific quantile of the future predictive distribution and represents the probability that an observation will fall under a certain value. An example of a quantile forecast is given in Figure 2.2b. A prediction interval interval is a range in which future observations will lie with a specific coverage probability, shown in Figure 2.2a. Full information relating to the predictive distribution can be found in a full density forecast, which can be obtained from a predictive Probability Density Function (PDF) or Cumulative Density Function (CDF). It is important to note that quantile and interval forecasts can be obtained from predictive PDFs and CDFs and they therefore provide the most versatile framework for forecasting.

Constructing the predictive distribution can be done using a parametric assumption of the conditional distribution shape, or by using a non-parametric data driven

approach. Parametric approaches have advantage in terms of computational efficiency, however certain assumptions have significant drawbacks in terms of fully characterising the uncertainty around a wind power forecast. A Gaussian assumption of the forecast uncertainty has been shown to be not applicable because the shape of the predictive density changes with time, the uncertainty is highly dependent on the forecast wind speed transformation using a non-linear power curve, and it is a bounded process between cut-in and rated power [59,60]. Therefore, parametric approaches to the problem have been proposed using the bounded Beta distribution [60], the censored normal distribution [50], and using a mixture of generalized logit–normal distributions distributions with two probability masses at the process boundaries [61]. These approaches have the advantage that it is often only one or two parameters that are needed to be predicted to characterise the full predictive distribution.

A more flexible approach is offered using a non-parametric, distribution free method. This requires more data and is typically more computationally expensive, but enable a completely data driven prediction of the forecast uncertainty. Quantile regression (QR) and kernel density estimation, are often used to tackle this problem. QR based methods can be used to calculate the full predictive distribution using a number of discrete quantiles and interpolation. The explanatory variables used to train the models (e.g. NWP and historical power data) are crucially important and feature engineering and selection is often tested rigorously via cross-validation [30]. An example density forecast obtained via multiple QR is shown on Figure 2.3, visualised via a fan plot (2.3a) and as a CDF at four lead times in Figure 2.3b with quantiles prediction points indicated.

Rather than using a deterministic NWP input to a univariate probabilistic forecasting model, it is also possible to use probabilistic weather forecasts (NWP ensemble members). Importantly, ensemble weather forecasts can be post-processed for improvements in reliability [62]; they therefore require calibration to match observed probabilities of occurrence for wind power [63]. Typically this method involves converting ensemble wind speeds to power and then adding a calibration stage [64], although direct quantile regression is equally applicable. In [48], an inverse power curve method,



(a) Density forecast as a fan plot

(b) CDF at four lead times from Figure 2.3a

Figure 2.3: Example density forecasts, where estCap. is the estimated capacity. Here, the density forecast is constructed with multiple quantile forecasts and monotonic spline interpolation at each lead time; a more complete view of the uncertainty is shown

where power observations are transformed into the wind speed domain, is used for both parametric and non-parametric regression with ensemble wind speed data. For more information regarding ensemble post-processing for wind power, please refer to [65].

Multivariate

Specific types of uncertainty forecasts are required to inform multi-temporal and spatial decision making, such as stochastic unit commitment and reserve holding [66–68]. These are termed scenario (or trajectory) forecasts, which maintain the dependency structure between variables and over time [69]. This type of forecast is motivated by the fact univariate energy forecasts are issued on a per-horizon and per-location basis, without accounting for spatio-temporal dependencies that exist between the locations and/or lead times. Intuitively, consider that distributed wind farms can experience the same weather pattern over a number of days as it develops; forecast errors from NWPs are then highly structured [70]. Generating statistical scenario forecasts requires the dependency between the marginal distributions of each lead time to be modelled. The most common method to model the dependency is to use copulas [69].

Pinson et. al [69] describe a method whereby from the marginal predictive distributions of a reliable forecast, forecast errors are transformed to a normally distributed random variable. From this normally transformed variable, the forecast dependency

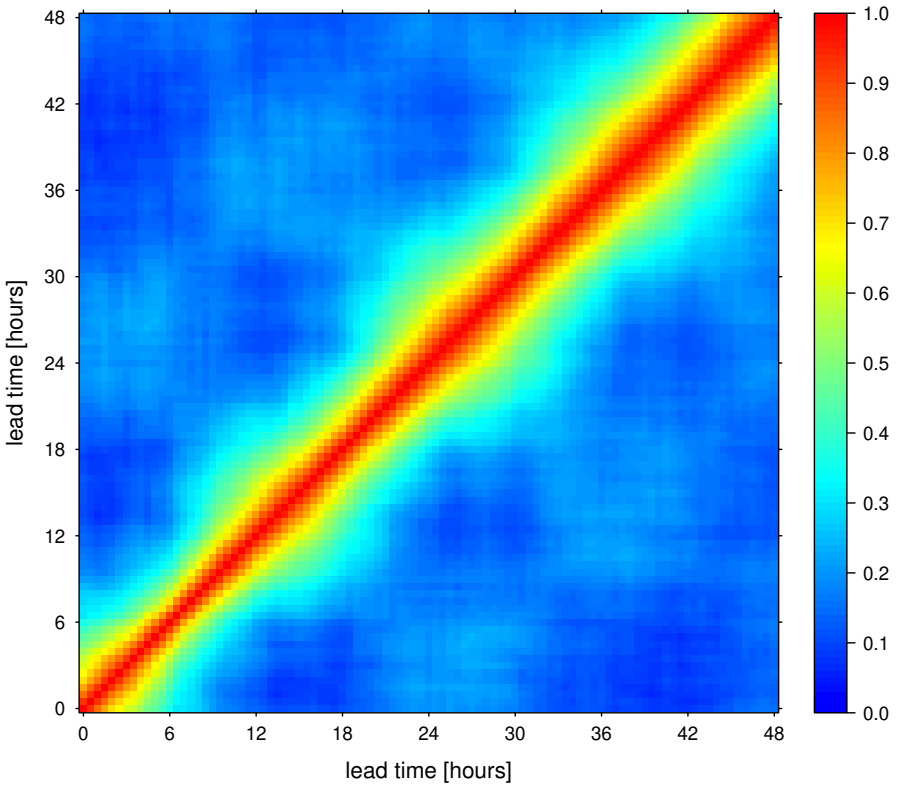


Figure 2.4: Example temporal covariance matrix of 'forecast errors' which characterises the dependence structure of the multivariate Gaussian copula and permits the generation of scenario forecasts. Here, the errors persist for longer as the lead time develops, due to degradation of the NWP skill at longer lead times

structures in space and time can then be fully described by a single covariance matrix. An example temporal covriance matrix is shown on Figure 2.4. Statistical ensembles can then be generated via sampling from the covariance matrix and transforming the samples into the original power domain. An example of a scenario forecast is plotted in Figure 2.5 with 100 scenarios, again generated for the west coast offshore wind portfolio in GB.

The above approach is essentially modelling a multivariate predictive distribution using a Gaussian copula. Copulas are a tool widely used in, e.g. quantitative finance, which provide a framework for describing the joint distribution of random variables and the inter-dependencies from the individual marginal distributions. The advantages

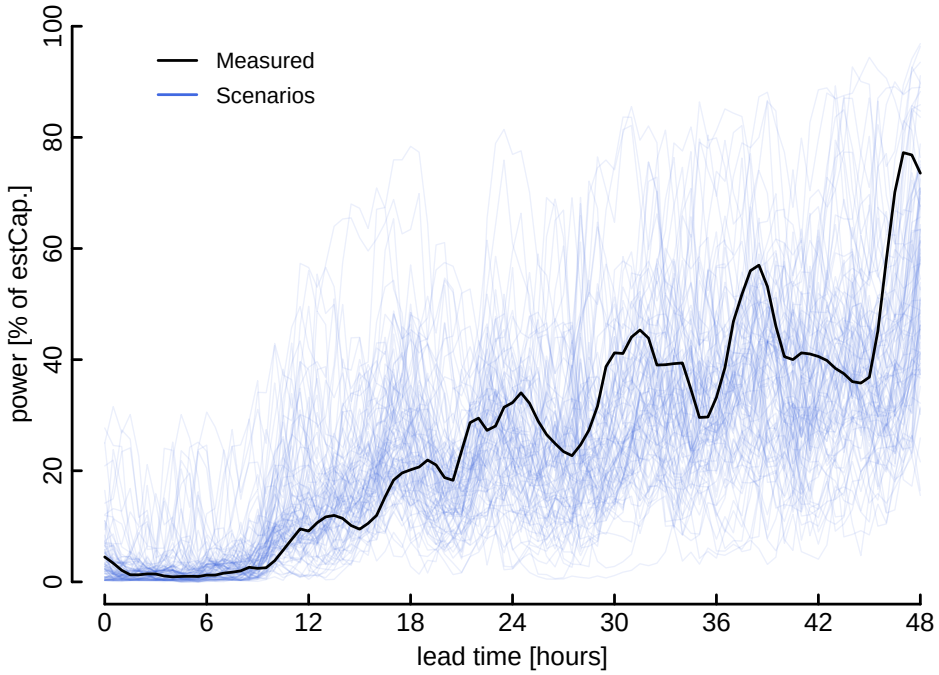


Figure 2.5: Example statistically based scenario forecast as a spaghetti plot, where estCap. is the estimated capacity. The uncertainty information can be difficult to interpret in this visualisation, but the scenarios are useful for simulations or further processing

of this methodology are that it is capable of producing as many scenarios as necessary (as, opposed to physical ensemble methods), it can be based on a single NWP deterministic input, and it is possible to condition the copula on external processes such as weather conditions; however, the process does require large datasets to fully characterise the dependency and it is notoriously difficult to model the tails of the predictive distributions, which are both open research opportunities in the field [36].

The tails are often the most important area of interest for decision makers using these forecasts; consider that it is the tails of the distribution, i.e. large forecast errors, that will influence reserve holding capabilities; TSOs are typically risk-averse, as to maintain security of supply. Additionally if the tails of the distribution are not well defined the calibration of the forecasts (and therefore suitability of the Gaussian copula) is diminished.

Communication of uncertainty forecasts is a current area of research and Bessa et

al. [36] emphasised that the number of ensembles in a scenario forecasts should be tuned to the end-users need, because a large spread of scenarios may not provide clear information to the end user. Also highlighted is the possibility of using simultaneous forecast intervals to represent the uncertainty [71], which leads to essentially quantile trajectories representing a certain probability of observing a wind power measurement lower or equal to that trajectory.

Statistically based scenario forecasts are comparable to ensemble-NWP forecasts, which can also be used in multi-temporal decision making [36]. However, an important subtle difference must be emphasised between the underlying information of the two forecast types. The latter represents the physical uncertainty for the future atmosphere state, whereas former represents the uncertainty dependent on the historical relationship between power and forecast wind speed, direction, and any other explanatory variables used. Ensemble forecasts have the advantage of being able to capture new events present in any given weather forecast; the downside is that it is very computationally demanding to run a NWP simulation, so there are typically only 50 ensembles available at any given issue time, and post-processing for calibration would likely be necessary for any specific site [72].

Other methods have been proposed to generate scenario forecasts of wind power such as stochastic differential equations [73] and multivariate analog ensembles [74]; these topics are discussed in more depth in the proceeding section.

2.4 State of The Art

For a brief history of the origins of wind power forecasting please refer to Costa et al. [55] which accounts for the 30 year history of wind power prediction (in only 20 pages), with a useful distinction between the physical and statistical approaches. Overall, the ANEMOS.plus project [56] and a technical report by C. Monteiro et al. [35] are two very useful and detailed resources which account the concerted research effort in the short-term wind prediction topic over recent years.

Within academia in the past 10 or so years, the focus has been on developing

probabilistic methods for forecasting, which is reviewed comprehensively by Zhang et. al [52]. However, it has been noted that there is a severe disconnect between the research in academia and recent industrial practices. Bessa et al. [36] addresses the blockage in the transfer of knowledge and poses that the main barrier is due to the lack of understanding of the underlying processes enabling uncertainty estimation, and a lack of standardization in uncertainty forecast products. This study is a sub-project within the International Energy Agency (IEA) Wind Task 36 on Forecasting [75], which is a collaborative research project focusing on improving the value of wind energy forecasts to the wind industry.

Recently, C. Sweeny et al. discuss the future of renewable power forecasting, including a literature review of the state-of-the art methodologies for short-term wind power forecasting [29]. Also discussed are promising new forecast products and business models capitalising on newly available high-dimensional power and NWP data, one of which includes the exchange of power and forecast data via decentralised frameworks; the challenge here is in preserving the privacy of commercially sensitive data. This is driven by the idea that having access to more diverse forecasts and data-sources will ultimately improve forecast skill for an end-user.

Here, we focus on three main research themes prevalent in the energy forecasting community at the moment: feature engineering and the results from energy forecasting competitions; hierarchical forecasting; and multivariate forecasting. Note that these research themes are all linked by the challenge of modelling potentially high dimensional data. We also discuss an emerging application of forecasting in wind energy: access forecasting for offshore operations, which incorporates aspects of metocean forecasting.

2.4.1 Feature Engineering & Forecasting Competitions

An area of challenge for wind power prediction, highlighted within the IEA Wind Task 36, is the lack of widely available and publicised benchmark test cases and datasets. The importance is highlighted because many approaches to wind forecasting have been proposed, and it is often difficult to compare forecast skill since results will differ across datasets. For this reason, forecasting competitions such as the two wind power (2012

and 2014) Global Energy Forecasting Competitions (GEFCom) [33, 76] are a valuable pursuit and provide learning for both forecast producers and users. A common theme amongst the top approaches to the two competitions is the use of feature engineering and cross validation to generate robust and powerful prediction models [30,32]. Feature engineering is the process of creating new input variables for a statistical learning technique with greater explanatory power, using domain knowledge of the problem at hand. Examples of these features include: spatially averaged wind speed forecasts from a grid of NWP around a wind farm, temporally smoothed wind speed features, and the wind direction difference between model heights to capture wind veer.

The two winning teams from GEFCom (2012 and 2014) utilised Gradient Boosting Machines (GBM) with tree-based learners, the latter for quantile regression to produce density forecasts, with input features engineered from NWPs [30, 32]. Other entrants also employed GBMs but did not produce as skillful forecasts highlighting the importance of feature engineering in such methods. The k-nearest neighbours method has also been successful in producing non-parametric density forecasts in these competitions [77], which is very similar to the analog ensemble which has been demonstrated in weather [78], wind power [79], and solar power forecasting applications [80].

Interestingly Landry et al. [30] details a probabilistic solution using Gradient Boosted Trees with a quantile loss function using: leading and lagging wind forecast features; independent hyper-parameter tuning for middle and tail quantiles; including hour of the day features; and finally using off-site median power predictions for some of the forecast zones improved performance in the probabilistic models. It is important to note that the five most successful models in the latest competitions employed a fully non-parametric approach [33].

Andrade et al. [31] reported value in extracting features from a spatially distributed NWP grid surrounding the power plant, rather than interpolating to a single value at the coordinates of the plant. The improvement is shown in both a solar and wind power forecasting framework. The methodology combines the gradient boosting tree algorithm, with feature engineering using leading and lagging NWP forecasts; spatial and temporal standard deviations of NWP variables; and many others. Compared

to a benchmark model, based on one NWP point for a specific location, the average probabilistic improvement for the wind power predictions was shown to be 12.06% in terms of Continuous Ranked Probability Score (CRPS).

Although the root cause of the improvement is not explicitly discussed in [31], one possible reason could be due to the systematic errors present in the NWP; it becomes beneficial to generate various smooth NWP input features to hedge against phase and location errors and use the inherent feature selection ability of the algorithm to learn the powerfully predictive inputs and subsequent patterns and interactions. This root cause is therefore currently an **open research question**, and the impact of temporally smoothing NWP input features is explored in Section 4, within the context of leveraging ultra-high resolution NWP forecasts.

Importantly, when engineering a high-dimensional input feature space over-fitting can occur during model training, where the learned relationships match too closely to the training data and the model does not generalise well to new 'unseen' data. This motivates the use of statistical learning techniques with intrinsic feature selection and regularisation capabilities, such as GBMs or alternatively linear regression with sparsity [81].

2.4.2 Hierarchical Forecasting

Hierarchical forecasting has received increased attention in recent years because of the desire from forecast users for coherency (or consistency), i.e. the forecast of each level in a hierarchy should sum together appropriately. For example, in a coherent hierarchical wind power forecast the predictions at the turbine level sum to the forecast at the wind farm level, which sum together with surrounding wind farms to the regional level forecast, and so on. This is motivated by the fact that forecasting individual components in any given system may not be consistent after summation due to the nature of the learning problem. Additionally, including coherency constraints in predictive models can improve performance at all levels of the hierarchy. Currently, most published research on hierarchical forecasting has been in a deterministic framework.

Hierarchies can be both spatial and temporal in nature [82, 83]. There are different

approaches to hierarchical forecasting, the simplest being the bottom-up approach, which forecasts the top level in the hierarchy by summing the constituent lower level forecasts [84, 85]. As discussed in [21], the bottom-up approach can in practice tend to deliver poor performance because of the low signal to noise ratio of the bottom hierarchy in applications such as load forecasting using smart meter data. However, consider that a wind turbine provides a consistent weather dependent signal. Therefore, wind farm power forecasting provides a useful test-case for bottom-up hierarchical forecasting.

The concept of coherent *probabilistic* forecasts is explored in [21, 86] where the importance of this property is emphasised in settings where forecasts from multiple levels of the hierarchy are used in decision-making. In these works, the marginal distributions are determined for nodes in the system and the dependence is modelled using an empirical copula. However, in the wind farm setting the structure of the hierarchy is relatively simple, and the size lends itself to families of parametric copulas rather than the empirical copula, which requires large volumes of data to satisfactorily estimate.

In the wind power forecasting domain, deterministic forecast reconciliation, which is the processing of predictions at all levels of the hierarchy to achieve the coherency constraint, has been explored via generalised least squares [87] and forecast improvement reported at all levels of the hierarchy. Additionally, an online adaptive method is proposed using constrained multivariate regression, motivated by the fact that the optimal reconciliation weights will be non-stationary due to the stochastic nature of the problem; this is also carried out in the deterministic domain [88]. Privacy preserving distributed methods have also been applied to the day-ahead deterministic power forecasting problem [89]. Temporal hierarchies in wind power forecasting have also been evaluated in [90], which to the best of the author's knowledge is the only study published to date on coherent *probabilistic* wind power forecasts. There, they opt for an alternative to the bottom-up approach to the problem. Therefore, incorporating wind turbine level data in a bottom-up probabilistic method is currently an **open research question**, and is explored in Section 5.

Copulas provide a suitable method for generating bottom-up spatially coherent forecasts. A wide variety of copula families exist, several of which have been applied to

model spatial dependency in the wind power forecasting context but not in a hierarchical setting to the best of the authors knowledge [91]. Copulas are typically used in the wind power forecasting community for generating multivariate spatial-temporal scenario forecasts [69, 92].

2.4.3 Multivariate Forecasting

Research in statistical scenario forecasting for day-ahead power forecasting is focused on advancing or developing alternative methods to the Gaussian copula approach described in Section 2.3.2 for describing the temporal interdependence structures [69]. The success of the Gaussian copula in this setting is due to the ability to effectively model high dimensional distributions, and the apparent absence of tail dependencies in the data.

Pinson et al. [72] propose an exponential covariance function, which is essentially a parametric approach for defining the covariance matrix. This is compared to the data-driven empirical copula using multivariate skill scores, and interestingly event based verification tools. Gneiting proposes a non-separable parametric form for modelling spatio-temporal data, with a case study on Irish wind speed data [93]. This means that both the temporal and spatial components of covariance matrices are jointly modelled from the empirical data, instead of independently and then combined. This approach is readily applicable to the case of day-ahead multivariate wind power scenarios, but to the best of the author's knowledge has not yet been addressed in the literature.

Regime-switching is a method to make models dependent on the current behaviour of the target time-series or on exogenous variable(s). The benefit is that the model can react faster to changing conditions, as opposed to having fixed models or by tracking slower changes in behaviour via (for instance) online updates. Mostly, regime switching has been well addressed in the very-short term forecasting power literature [94, 95]. Browell et al. [96] has shown for very short-term spatio-temporal forecasts that a regime-switching model based on large-scale meteorological phenomena or 'atmospheric regimes' can improve forecast skill by an average of 3.1% at a six hour forecast lead time.

Regime-switching probabilistic forecasts of wind energy are also examined in [97,98],

using a correlation and cross correlation analysis at the wind sites studied; again these studies are both based on very short term forecasting, i.e. ≈2 hours ahead. An open research question remains on the use of regime-switching for conditioning the spatial-temporal covariance dependency for multivariate day-ahead power forecasting.

Tatsu et al. provide the only example of conditioning of copula methods at the day-ahead temporal scale, to the best of the author's knowledge [91]. There, the authors expand on the Gaussian copula method to include the spatial dependencies via sparse precision matrices, and show the value of including direction-dependent correlations and non-constant conditional variances. Additionally, to the best of the author's knowledge, scenario forecasting, and the associated dependency characteristics, in the GB network has not been explored at the wind farm level. This **open research question** is explored in more detail in Section 7.

A valid criticism of the Gaussian copula is that only only symmetric dependencies are effectively modelled. Copula vines, which are a series of linked bivariate copula families, offer a more flexible framework for modelling multivariate dependency, and have subsequently been the subject of recent studies in wind power forecasting [92, 99]. Bessa [92] compared temporal scenarios using vine-copulas and compared these with the conventional Gaussian copula method for three wind farms in a case study. Results showed that in-fact the Gaussian approach with an exponential covariance matrix outperformed the vine copula for two out of three wind farms in terms of the energy score and the p-variogram score. One of the main advantages of the exponential covariance matrix method was highlighted; it does not require a large historical dataset, which was posed as a justification of the higher performance.

The high dimensionality of space-time scenario forecasting at the wind farm level in certain regions, like the UK, mean that generating enough scenarios to capture the full behaviour of portfolios becomes computationally demanding, and solving the associated optimisation problems, such as probabilistic power flows, potentially not practical using the raw scenarios. This has motivated recent research on forecasting multivariate regions such as ellipsoids or polyhedra [100, 101]; such types of forecasts make probabilistic optimisation problems more computationally tractable.

Beyond copula methods, alternative approaches have been proposed using Stochastic Differential Equations [73, 102] based on the historical forecast errors to describe the evolution of the forecast error around the deterministic forecast, using epi-spline basis functions which enable users to control for degree to which extreme errors are captured by specifying particular quantiles in the forecast error distributions [103], and by using a rank-reordering method which preserves spatial-temporal characteristics via the Schaake shuffle [74, 104]. Briefly, the Schaake shuffle works by re-ordering ensemble forecasts to reconstruct the space-time variability observed in a sampled historical record of the measurements.

Using post-processed ensemble NWP data it is also possible to recover the spatial-temporal characteristics of the weather data after applying the statistical learning or calibration technique. This is done via a technique known as Ensemble Copula Coupling (ECC), which is demonstrated on meteorological variables [105, 106]. A slightly modified technique known as dual-ECC is demonstrated on wind speed ensemble data, such that temporal error auto-correlations are accounted for [107]. To the best of the author's knowledge, this technique has not been applied to post-processed wind *power* ensembles.

2.4.4 Forecasting in Wind O&M

It is estimated that 20–30% of the total cost of energy for an offshore wind farm is due to Operations & Maintenance (O&M) in the UK [17]. Since O&M savings can be achieved by operators at any stage of the project life cycle and independently of turbine manufacturers there is a great opportunity to reduce this sizeable percentage. Therefore, improving installation, operations, and maintenance practices is a current focus in both industry and academia.

Access forecasting is concerned with predicting conditions for the transfer of technicians to and from offshore structures at the site. This is clearly highly dependent on the local wave climate and sea-state forecasting plays a major role in the current scheduling practices in offshore wind. These forecasts are typically deterministic forecasts of significant wave height. However, this provides limited information into the

state of the weather over the next few days. Probabilistic forecasts, which quantify the uncertainty around future values, provide a route to risk minimisation [10]. For instance, scheduling tools could evaluate the spread of possible metocean conditions for the target day.

O&M planning at different horizons can be classified into strategic, tactical, and operational decision making [108], the latter of which is the horizon of focus of this thesis. Relevant studies in this area of decision support are based on optimising the scheduling problems in a metaheuristic manner [109], or via exact optimisation methods [110, 111], or through a combination of the two [112]. Uncertainties in short term maintenance scheduling stem from incorrect fault diagnosis, human error, repair times, and of course the weather forecast. The latter has been examined in the literature [113, 114], although most published methods are based on deterministic forecasts of the weather [110–112, 115], others have considered climatology from past observations [116] and ensemble NWP [117].

As discussed, at lead times of greater than approximately 6 hours, NWPs are superior to time series forecasting methods for predicting variables such as significant wave height [34] and should be employed in day-ahead scheduling decisions (and longer lead-times). Probabilistic forecasting of wave height using time series models, driven by recent observations only, is explored in [118] and the value of these forecasts is demonstrated via a cost-loss model, but are limited to within day applications. The economic case for improved offshore wind maintenance access forecasts has also been explored [119], where different deterministic models and evaluation metrics are tested; the cost of a sub-optimal configuration of these aspects is estimated to cost up to hundreds of thousands pounds per year.

Offshore operations scheduling is a type of multi-temporal decision making; consider that accessibility at a single point in time is not sufficient as technicians require to be picked-up and returned to port at the end of the shift. Therefore, turbine accessibility is often framed in terms of weather windows, where forecasts are used to specify access conditions throughout a specified time period [117, 120, 121]. The probability of access from weather forecasts is determined in [122], which are coupled to cost-loss decision

model and compared with the deterministic case; this is found to increase the proportion of access windows utilised and reduce operational expenditure.

For forecasting, this multi-temporal decision framework demands that error dependencies in time are accounted for, making this another high dimensional problem amenable to copula modelling. Various copula families have been tested rigorously for hindcast metocean data [123], where the dependency is modelled between variables including asymmetries. Copulas are also used to simulate wave height and wind speed time series, with a case study application for cable installation at an offshore wind farm [124].

To summarise, an **open research question** remains in the development of a novel end-to-end probabilistic forecasting methodology for offshore wind farm operations, at the day-ahead temporal scale. This uncertainty information could be potentially valuable to operations teams that currently typically rely on deterministic forecasts of wave height for scheduling and dispatch decisions. Please refer to Section 6 for more detail.

2.5 Summary

Forecasting is an integral part of of power system operations. From participants in electricity markets to Transmission System Operators, all who make decisions based on the future generation of wind energy require decision-support tools. Probabilistic forecasts are essential for risk-neutral participants in asymmetrically penalised electricity markets and for risk management; uncertainty is always present, it merely hidden in deterministic forecasts.

There has been a vast research effort in probabilistic forecasting in recent years, both univariate and multivariate, at the day-ahead temporal scale. However, there are still gaps in the literature related to dealing with high dimensional data in both the input and output side of the modelling chain. Specifically, open research questions remain in

1. The performance of ultra-high resolution operational NWP models, using ad-

vanced statistical post-processing techniques and feature engineering

2. Leveraging hierarchical probabilistic methods for wind farm power forecasting, using information from individual turbines

3. Probabilistic forecasts for wind farm operations. Currently, wind farm operators rely solely on deterministic forecasts of significant wave height for safety critical decisions, such as maintenance scheduling and vessel dispatch

4. High-dimensional multivariate wind power forecasting. Here, error dependencies in both space and time are used to generate statistical scenario forecasts. These forecasts could be improved by using large scale atmospheric regimes and coupled space-time parametric covariance matrices. Additionally, there has not been a study at the GB wind farm spatial scale.

Before addressing these gaps, a forecast verification framework is formalised.

Chapter 3

Forecast Verification

The evaluation of a proposed forecasting model must be based on a impartial evaluation of performance over an out-of-sample or testing dataset. In an ideal case, the forecast could be evaluated in terms of the ability to inform optimal decisions, such as revenue generated from using a novel forecast, compared to existing methods. For example, an accurate deterministic forecast in terms of conventional statistical evaluations may not lead to more optimal trading decisions, and therefore greater profits [125]. However situation dependent evaluation metrics are difficult to quantify and don't provide a comparable framework to evaluate methods as a research endeavour. Therefore, forecasts are most commonly evaluated on the analysis of out-of-sample errors, and are used for forecast provider selection, model development, and model diagnostics.

Here, we outline the most commonly used forecast verification tools for deterministic and probabilistic forecasts. These are termed evaluation/verification metrics and scoring rules interchangeably throughout. Univariate and multivariate evaluation metrics are discussed, as well as methods for setting up a valid numerical experiment framework. Lastly, we discuss the uncertainty present in evaluation results themselves (not to be confused with forecast uncertainty), and address methods used to deal with this, such as bootstrapping. Please refer to the recent paper by Messner et al. for more detail on forecast evaluation in wind power forecasting [51].

3.1 Deterministic

Madsen et al. propose a standardised protocol for the evaluation of deterministic short-term wind power predictions [126]. This begins with the standard time-series definition of prediction error e_t which is the difference between the measured value y_t and the forecast value $\hat{y}_t$

$$e_t = y_t - \hat{y}_t \tag{3.1}$$

where t is the time-index. The bias of the model is related to the systematic error and is simply as the mean error over the evaluation period

$$\text{bias} = \frac{1}{N}\sum_{t=1}^{N} e_t \tag{3.2}$$

where N is the total number of samples. Therefore, bias does not give any information as to the discrimination ability of a forecast, i.e. if it gives accurate predictions. It reveals the difference between the average forecast value and the average measured value; an ideal bias result is zero. However, bias is a very useful tool for diagnosing systematic problems in the model. For example, in a regional wind power forecasting tool, if the installed capacity of the region increases over time and this effect is not captured by the model, then error analysis will likely reveal a substantial bias.

The two most common scoring rules for evaluating deterministic forecast accuracy, i.e. discrimination ability, are the root mean square error (RMSE)

$$\text{rmse} = \sqrt{\frac{1}{N}\sum_{t=1}^{N} e_t^2} \tag{3.3}$$

and the mean absolute error (MAE)

$$\text{mae} = \frac{1}{N}\sum_{t=1}^{N} |e_t| \quad . \tag{3.4}$$

Importantly, the RMSE is related to the variance of the errors and large errors will have a significant impact on the results. Typically, a forecasting analysis will report both of

these metrics. However, the error metric chosen should relate to the loss function used in modelling. For instance, if a squared error loss function is used to fit the model, then a squared scoring rule (like RMSE) should be used to evaluate the models during development. The above performance metrics are often averaged by forecast lead time to gauge the performance of the model during time-periods of interest.

3.2 Probabilistic

For probabilistic forecasts verification becomes more complex, and a two pronged approach is necessary to assess performance; **calibration** (or reliability) is essentially a measure of probabilistic bias, and **sharpness** is related to the spread of the forecast distribution. A probabilistic forecast is then evaluated in terms of sharpness, subject to calibration [127].

The calibration is a measure of the statistical consistency between the distributional forecasts and the observations. In other words, in a suitably large test dataset the nominal probabilities of the forecast are approximately equal to the empirical (measured) outcomes [128]. For example, for the median quantile — q50 — then approximately half of the observed values (empirical) should be below this median forecast. An poorly calibrated probabilistic forecast is therefore bias and sub-optimal, regardless of any other properties.

Sharpness is a measure of the accuracy of a probabilistic forecast, i.e. the spread of the distribution, that allows for the ranking of competing forecasts, subject to calibration. Scores such as the pinball loss (for discrete quantiles), and the continuous ranked probability score (for continuous distributions) are typically used for scoring rules on univariate variables. These two measures in fact evaluate both calibration and sharpness simultaneously; however, it is often prudent to simultaneously verify calibration, either by decomposition of these accuracy scores [129, 130] or by direct methods. It is helpful to think of an intuitive example of how these scores are related to accuracy; for the q50 quantile specifically, the pinball loss is simply equal to $0.5 * \text{mae}$.

The idea of proper, and strictly proper scoring rules are developed in [131]. For a distribution forecast $\hat{F}$ with realised measurement y from an unobserved governing

distribution F, the average score is defined as

$$\bar{S} = \frac{1}{N} \sum_{t=1}^{N} S(\hat{F}_t, y_t) \tag{3.5}$$

where the score S to each observation/forecast pair. The score is **proper** if

$$S(F_t, y_t) \leq S(\hat{F}_t, y_t) \tag{3.6}$$

holds for the distribution of F and $\hat{F}$. Note that here we imply a negatively orientated score, i.e. better forecasts produce smaller scores. Furthermore, the score is **strictly proper** when equality is true only if $\hat{F} = F$; i.e. the minimum is unique. Strictly proper scoring rules inform on the forecast verification framework used throughout this thesis. Although for multivariate forecasts, proper scoring rules are used as discussed in Section 3.2.2.

3.2.1 Univariate

Here, we establish the definition of quantile forecasts and density forecasts for continuous univariate variables, as well as the associated verification frameworks. For interval forecasts and probability forecasts for discrete events, scores metrics such as the Winkler and Brier score (respectively) are available, but discussion is omitted here for brevity; the reader is referred to [51, 128] for more information.

The predictive cumulative distribution function (CDF) of a random variable Y is

$$\hat{F}(y) = P(Y \leq y) \tag{3.7}$$

where $\hat{F}$ is a strictly increasing function. A quantile forecast can be defined as

$$\hat{y}^{(\alpha)} = \hat{F}^{-1}(\alpha) = \inf\{y \in \mathbb{R} : \alpha \leq \hat{F}(y)\} \tag{3.8}$$

where $\alpha \in [0, 1]$ is the nominal probability of the quantile. As discussed, we can therefore generate a distribution forecast directly using a technique such as conditional

parametric regression, or estimate the distribution forecast using multiple quantile regression and interpolation between the boundaries and the quantile forecasts.

For discrete quantiles, calibration is often measured directly via reliability diagrams. This is is calculated via an indicator variable $\beta_t^{(\alpha)}$

$$\beta_t^{(\alpha)} = \begin{cases} 1 & \text{if} \quad y_t < \hat{y}_t^{(\alpha)} \\ 0 & \text{otherwise} \end{cases} \tag{3.9}$$

which leads to a series of binary outcomes indicating whether the outcome lies below (1) or above (0) the quantile forecasts. The mean of the series of outcomes can be calculated over the whole dataset to give the actual coverage of the quantile

$$a^{(\alpha)} = \frac{1}{N} \sum_{t=1}^{N} \beta_t^{(\alpha)} \tag{3.10}$$

and if the forecast is well calibrated and the sample size sufficient then $a^{(\alpha)} \approx \alpha$. This measure is then calculated for each modelled quantile and results illustrated via a reliability diagram which plots the nominal vrs. empirical proportions [51]. Quantile bias plots are also used in this work to visualise the reliability when inspection of small differences in calibration are necessary, in which case the nominal proportions are plotted against $\alpha - \hat{a}^{(\alpha)}$ [128].

For continuous density forecasts, it makes sense to evaluate the calibration of full distribution, not just discrete quantiles. The probability integral transform (PIT) histogram is used in such cases, which is calculated via

$$u_t = F_t(y_t) \tag{3.11}$$

and if the forecasts are well calibrated and the sample is sufficiently large then $u \sim \mathcal{U}(0, 1)$, which is visually inspected via a histogram with a certain number of (typically 20 [132]) bins. An example of a PIT histogram is shown in Figure 3.1, where the forecast is slightly under-confident; the relatively more frequent observations in the mid range of the distribution show that spread of the distribution is over-dispersed.

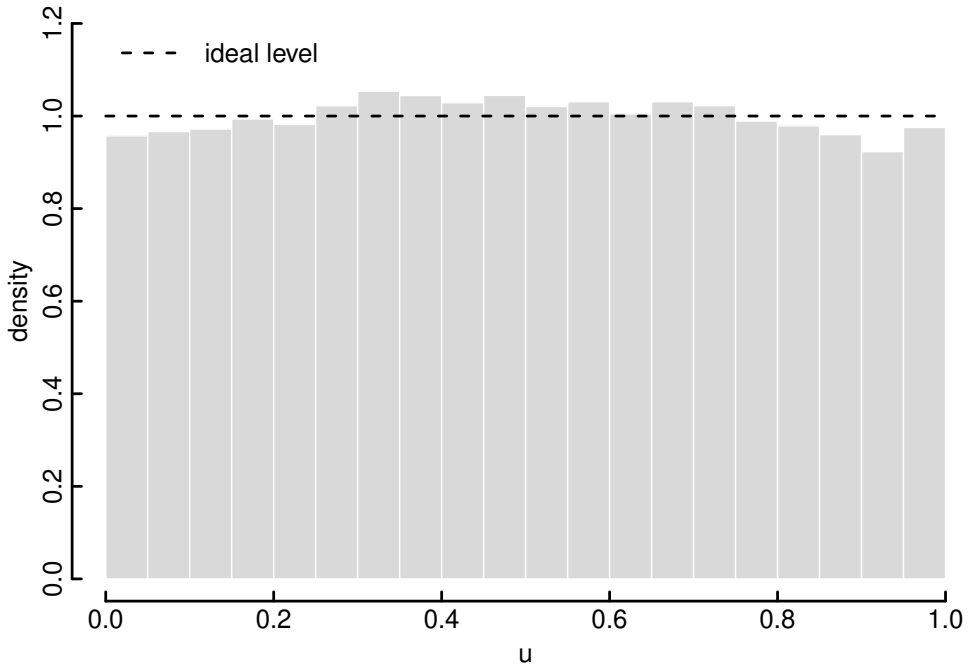


Figure 3.1: Example PIT histogram. Here, the forecast displays over-dispersion; observations are more concentrated at the mid-region of the distribution and the distribution is on average under-confident

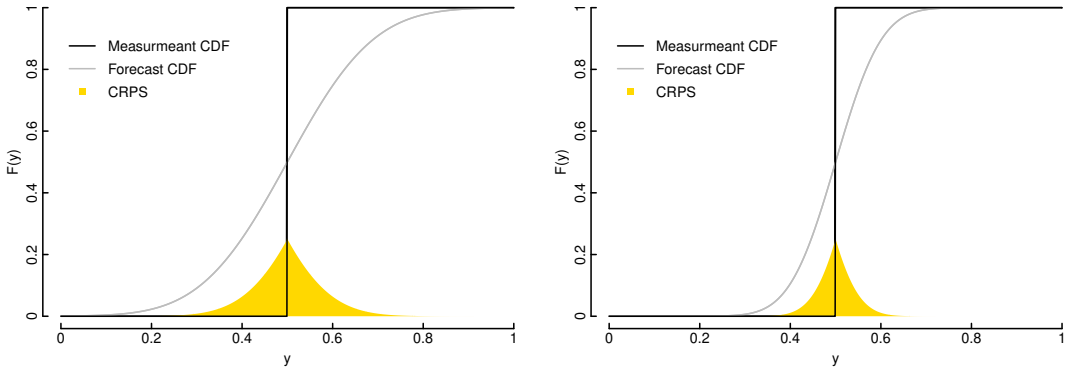
The opposite trend is found for over-confident (i.e. under-dispersed) forecasts, where there are relatively more observations in the tails of the distribution.

For discrete quantiles, the sharpness and calibration is typically measured via the pinball score (PB)

$$\mathrm{pb}_t^{(\alpha)} = \begin{cases} (1-\alpha)(\hat{y}_t^{(\alpha)} - y_t) & y_t < \hat{y}_t^{(\alpha)} \\ \alpha(y_t - \hat{y}_t^{(\alpha)}) & y_t \geq \hat{y}_t^{(\alpha)} \end{cases} \tag{3.12}$$

and again, averaged over the number of samples. This is interchangeably called the pinball or quantile score. The average of multiple quantile pinball scores is often used to gauge the performance of the full predictive distribution [33]; although since the outcome of this score is dependent on both the target variable *and* the nominal probability level, the mid-range of the distribution contributes most to these averaged pinball loss scores.

For the full predictive cumulative distribution, the Continuous Ranked Probability



(a) Distribution forecast

(b) A sharper distribution forecast

Figure 3.2: Graphical illustration of the CRPS scoring rule for a single lead time, for two forecasts. The score is the area below the squared difference between the observation (step function) and forecast CDF

Score (CRPS) [127] is often used to measure both sharpness and calibration

$$\text{crps} = \frac{1}{N}\sum_{t=1}^{N}\int_{-\infty}^{\infty}\{\hat{F}_t(y) - \mathbf{1}(y \geq y_t)\}^2 dy \tag{3.13}$$

where $\mathbf{1}(.)$ is the indicator function. The CRPS for a single forecast observation pair is therefore the area between the squared difference of the forecast and observation CDF, where the latter is a step-function from 0 to 1 at the observed value. Figure 3.2 shows two illustrations of this, where a smaller CRPS is found from a sharper forecast. An advantage of the CRPS is that it reduces to the MAE in the deterministic case [133], which makes the score more intuitive and also consistent when moving from deterministic to probabilistic frameworks. This is clear from equation 3.13; the integrand reduces to either 1 or 0, and the error is found in the interval between the observed and forecast value.

Similarly to the pinball score, most of the contribution of the CRPS score comes from the mid-range of the distribution. When the tail of the forecast distribution is important, e.g. for setting reserve requirements [134], or more generally in extreme event forecasting, weighted scoring rules would be more suited [135]. Specifically, the threshold weighted CRPS proposed in [135] works by emphasising specific regions of interest (e.g. tails) of the distribution with relatively larger weights compared to other

regions, and at the same time preserves the 'properness' of the score.

In fact, the CRPS is the continuous counterpart of the pinball loss. Equation 3.13 can be alternatively described as

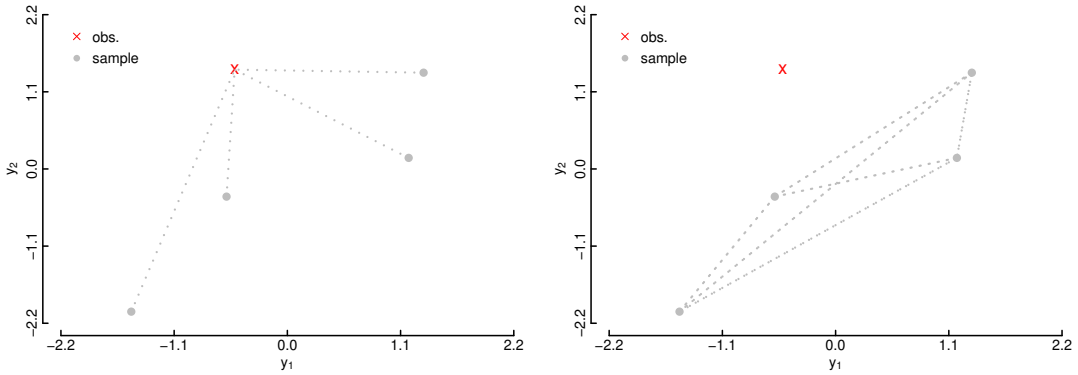
$$\text{crps} = \frac{1}{N}\sum_{t=1}^{N} 2\int_0^1 \text{pb}_t^{(\alpha)} d\alpha \tag{3.14}$$

i.e. the integral of pinball loss over all quantiles [135]. Although both the CRPS and pinball scores evaluate both sharpness and calibration, it is still good practice to also show calibration distinctly, either by reliability diagrams/PIT histograms, or by decomposing the pinball/CRPS scores further into reliability, resolution, and uncertainty components [51].

3.2.2 Multivariate

Here, we establish the definition of multivariate forecasts, as well as the associated verification tools. A multivariate forecast is usually defined by a set of possible scenarios for the forecast issue time, such as in the cases considered in this thesis. These scenarios contain information on the space and/or time correlations and therefore, evaluating the prescribed correlation structure is important in multivariate forecast verification. A multivariate forecast as a set of spatio-temporal scenarios can be defined as

$$\dot{\mathbf{y}}^{J,K} = \begin{bmatrix} \hat{y}_1^{1,1} & \hat{y}_2^{1,1} & \dots & \hat{y}_T^{1,1} \\ \hat{y}_1^{2,1} & \hat{y}_2^{2,1} & \dots & \hat{y}_T^{2,1} \\ \dots & \dots & \dots & \dots \\ \hat{y}_1^{J,1} & \hat{y}_2^{J,1} & \dots & \hat{y}_T^{J,1} \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \hat{y}_1^{1,K} & \hat{y}_2^{1,K} & \dots & \hat{y}_T^{1,K} \\ \hat{y}_1^{2,K} & \hat{y}_2^{2,K} & \dots & \hat{y}_T^{2,K} \\ \dots & \dots & \dots & \dots \\ \hat{y}_1^{J,K} & \hat{y}_2^{J,K} & \dots & \hat{y}_T^{J,K} \end{bmatrix} \tag{3.15}$$



(a) difference component

(b) spread component

Figure 3.3: Example graphical illustration of the energy score contributions from several bivariate normal scenarios and the corresponding observation. The energy score is a function of both the spread and difference components of the scenarios

where J is the number of samples and the maximum lead time is T. Therefore, every row is a scenario at each out of the K locations. However, to simplify the notation we will only consider one location with purely temporal scenario forecasts $\dot{\mathbf{y}}^J$.

Multivariate forecast verification is the subject of current research. The most commonly used scoring rule in wind power forecasting at the moment is the energy score (ES), which is a multivariate extension of the CRPS [131, 136]. The ES for a single issue time is given by

$$\text{es} = \frac{1}{J}\sum_{j=1}^{J}\|\mathbf{y} - \hat{\mathbf{y}}^{(j)}\|_2 - \frac{1}{2J^2}\sum_{i=1}^{J}\sum_{j=1}^{J}\|\hat{\mathbf{y}}^{(i)} - \hat{\mathbf{y}}^{(j)}\|_2 \tag{3.16}$$

where $\|.\|_2$ represents the ℓ_2 norm, $\mathbf{y}$ is the observed trajectory of the measured variable, $\hat{\mathbf{y}}^{(j)}$ is the jth scenario forecast taken from the underlying multivariate distribution. The score is then averaged over all issue times. In Figure 3.3 an illustration of the energy score is shown, generated from an example bivariate normal distribution. The first term of the energy score measures the average euclidean distance between the trajectory forecasts and the observation, and the second term is related to the multi-dimensional spread of the scenarios.

In the literature, it is discussed that the ES has low sensitivity to the multivariate correlations, compared to sensitivity in a mis-specified mean [136, 137]. This means

that different dependency structures governing the multivariate scenarios can have a somewhat similar energy scores, making the assessment of the benefit in a prescribed dependency structure more difficult. The p-variogram score (VS-p) is developed with this in mind [137], and for a single issue time is defined as

$$\text{vs-}p = \sum_{i,j=1}^{T} w_{ij} \left(\left| y_i - y_j \right|^p - \frac{1}{J} \sum_{z=1}^{J} \left| \hat{y}_i^{(z)} - \hat{y}_j^{(z)} \right|^p \right)^2 \tag{3.17}$$

where p is the order of the variogram and $\hat{y}^{(z)}$ is the z^{th} forecast scenario. Typically, the weights, w_{ij}, are set to the inverse distance between the i^{th} and j^{th} components. Results in [137] show that a VS-p with an order of less than one has the best discriminative ability, although typically results are reported for both orders of 0.5 and 1. To summarise, this score finds the weighted sum of differences between pairwise components of the observation trajectory and the average pairwise components from the multivariate scenarios. However, it follows that if we have two competing multivariate forecasts with the same correlation structure, but one set are are offset by some bias in the mean level, they would have the exact same VS-p score because only pairwise differences are considered; this score is proper, but not strictly proper.

Multivariate calibration methods are also used in the wider forecasting literature, particularly when assessing ensemble forecasts of meteorological variables. Minimum spanning tree (MST) rank histograms and variations are typically used [136]. To the best of the authors knowledge, the only examples of these verification tools applied to statistical wind power scenarios are [72, 103]. The former study compares ensemble-based and statistically-based wind power scenarios, and interestingly event based scores. When generating statistically based scenarios in this thesis, the marginals are ensured to be well calibrated so we don't elaborate further here. Please refer to [136] for more detail.

The two discussed multivariate scoring rules have both strengths and weaknesses, so typically both scores are used to verify multivariate scenarios [92], the strictly proper ES is used to verify both the mean level and multivariate correlation, and the proper VS-p is used to discriminate between correlation structures. However, small differences in the

ES may be sufficient to discriminate between models when coupled with significance tests [138], or when using bootstrapping to estimate sampling uncertainty, as discussed in Section 3.3.2.

3.3 Comparing Performance

Any proposed novel methodology should be rigorously compared against fair and competitive benchmark models. Also, comparing evaluation metrics is an important concept, and there is an often overlooked uncertainty associated with scoring results. In this section we discuss the framework of the numerical experiments used in this thesis, and introduce methods for evaluating the uncertainty in scoring results.

3.3.1 Cross-Validation

Cross validation is a tool widely used for the verification of statistical models. The framework is based on the premise that a model should be trained and tested on independent data sets. This has obvious benefits in that evaluation metrics are based on out-of-sample data, somewhat replicating the conditions of an operational model for decision support. This framework prevents models which don't generalise, but learn relationships in the data very well, from giving spuriously better scores.

There are a variety of cross validation routines, but in this work we typically use k-fold cross validation to make decisions on the modelling and evaluate results. This means that the data is first partitioned into training and testing subsets, then on the training subset only, k-fold cross validation is used to tune regression hyper-parameters and develop new forecasting models. This cross validation routine works by sub-setting the training data into k (approximately) equal partitions, then each partition is iteratively left out of the available training data and forecasts are issued for the 'left-out' partition, until each of the k folds have associated out-of-sample forecasts. This means that robust evaluation results can be determined for the entire available dataset.

Setting the k folds is an important consideration in wind power, and generally meterological post-processing tasks. This is because wind power exhibits temporal

correlation; by randomly assigning each sample to a fold, the resulting evaluation can lead to unrealistically low error scores because the data is not independent. Therefore, blocks of data of are used in the following evaluations. There are different cross-validation routines available, as well as using a rolling window to simulate the out-of-sample performance [51]. However, computational expense limits the feasibility of these methods in the following work.

3.3.2 Scoring Uncertainty

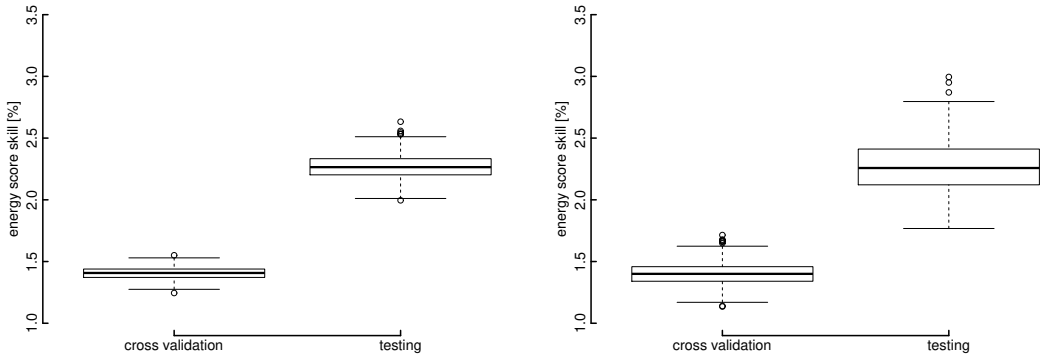
As well as forecast uncertainty, there is also always an inherent uncertainty present in evaluation results due to the fact verification is done on a finite sample of data. This is often overlooked in the wind power forecasting literature which typically presents single-valued evaluation metrics as standard. If, for instance, the the size of the test dataset is constrained, or differences between evaluation metrics is small in magnitude, it is important to evaluate the significance of the results. That is, are the results for the sample also true for the population, or due to pure chance.

Often, forecast performance is reported via relative change to a baseline $\bar{S}^{ref}$, i.e. the skill score. If the perfect score is zero, as in the cases considered here, then the skill score is

$$\text{skill} = \frac{\bar{S}^{ref} - \bar{S}}{\bar{S}^{ref}} \tag{3.18}$$

and in the following the terms skill score, percentage improvement, and relative change are used interchangeably. As discussed, often when evaluating multivariate forecasts the skill score value may be small in magnitude. However, a small relative change alone does not always prove that a forecast is not suitable; the uncertainty associated with the skill should be considered.

The bootstrap is a re-sampling technique for assessing the uncertainty of sample statistics [139]. It is a flexible tool which is used across statistical applications to understand the sampling distribution associated with estimators such as the mean, regression coefficients, correlation coefficients, etc. from the original set of observations. Crucially, the bootstrap can be used as a simple non-parametric method for estimating the significance in forecast improvement [51]. Typically, data-points are randomly



(a) bootstrap

(b) block-bootstrap

Figure 3.4: Boxplots showing the sampling variation of the ES skill between a benchmark and example multivariate forecast. Here, there is a small but significant improvement in the ES for both bootstrap methods; however the sampling uncertainty is greater using the block-bootstrap

sampled with replacement, i.e. individual points can be sampled more than once, with a length equal to the original length of the set, and then skill scores or scoring rules are calculated. This process is repeated a user-defined number of times (typically >100), until the sampling variation of the result is determined, which are then typically presented via boxplots as shown on Figure 3.4a, or confidence intervals. Therefore, evaluation results based on smaller datasets will have an inherently large sampling uncertainty, and a consistent but small positive skill score on a large dataset will be clear.

As with k fold cross validation, the temporal correlation of errors is important to consider here as well; the bootstrap assumes that data is independent, i.e. not serially correlated. This is often not the case in meteorological applications or time-series data and can lead to an underestimation of the sampling variability. However, in these cases the block-bootstrap [140] can be used, as shown on Figure 3.4b, where non-overlapping blocks of length $\approx$7 days are sampled with replacement. There are several types of bootstrap implementations which account for temporal correlation, such as the variable length block bootstrap [141], and several studies in electricity demand forecasting use the block bootstrap to generate future temperature scenarios for model inputs [142, 143].

There are also alternatives to the bootstrap for evaluating the significance of forecast improvement. The use of the Diebold-Mariano test [144,145], coupled with strictly proper scoring rules and large sampling sizes is recommended by Ziel et al. for multivariate scoring [138]. The Diebold-Mariano test also accounts for the temporal correlation of forecast errors. For more information, the reader is referred to Messner et al. in the context of wind power forecasting [51], and Section 7.2. Briefly, the block-bootstrap method has some advantages over statistical tests, in that there is a tendency in the wider academic literature to simply accept or dismiss results that are just below or above an arbitrary probability level in the latter, and bootstrapping is a non-parametric process. However, in the absence of literature demonstrating the suitability of one method over another, both are used in Section 7.2 to gauge the significance of forecast improvement.

3.4 Summary

Forecast verification is a crucial component in model development, and is more complex for probabilistic forecasts. However, the main idea is that they should be as sharp as possible subject to calibration. Calibration is measured directly through techniques such as reliability diagrams and the probability integral transform; sharpness are measured via strictly proper scoring rules, such the energy score and continuous ranked probability score. It is discussed that these strictly proper scores in fact measure both calibration and sharpness. However, it is recommended to evaluate calibration either directly (as in this thesis) or by further decomposition of the scores.

We also discuss the framework in which proposed models can be compared fairly with competitive and suitable benchmark models. This framework consists of using cross validation to generate a large dataset out of sample forecasts, and using bootstrap approaches to show the sampling uncertainty of average scoring rules or of relative change to a benchmark. In the proceeding sections we apply this framework in the pursuit of novel forecasting methodologies.

Chapter 4

Statistical Post-processing of Turbulence Resolving Weather Forecasts

Chapter based on:
Gilbert, C., Messner, J. W., Pinson, P., Trombe, P. J., Verzijlbergh, R., van Dorp, P., & Jonker, H. (2020). Statistical post-processing of turbulence-resolving weather forecasts for offshore wind power forecasting. Wind Energy, 23(4), 884-897.

High-resolution NWP forecasts often exhibit spatial and/or temporal displacement errors, and when regarding typical average scoring rules – e.g. MAE or RMSE, they often perform worse than smoother forecasts from lower-resolution models. Recent computational advances have enabled the use of large eddy simulations in the context of operational weather forecasting, yielding turbulence resolving weather forecasts with a spatial resolution of 100 meters or finer and a temporal resolution of 30 seconds or less; the output from these NWP models therefore represents a high dimensional input feature space for power forecasting models.

As discussed in Section 2.2, the pursuit of high resolution forecasts and appraising their value for wind power forecasting is not a new topic and has been discussed in literature with varying success [26, 27]. The situation has evolved substantially over the

last few years, thanks to quantity of data being collected, increase in computational power (e.g. based on Graphics Processing Units – GPUs), and advances in data science. This enables the high-resolution forecasting framework considered here, based on computationally demanding approaches to producing high-resolution weather forecasts at the wind farm, and post-processing via advanced statistical learning techniques.

At outset of this study, EMCWF offered global weather forecasts at a spatial resolution of ≈16km, i.e. the 0.125° grid, at 1-6 hour intervals within the high resolution (HRES) 10-day deterministic forecast model. This resolution is essentially a limit on the variance and processes resolved in the wind speed fields. The NWP model considered here on the other hand, is based on an operational localised large-eddy simulation (LES), which is driven at the boundaries by the global resolution ECMWF model.

NWP forecasts based on the LES framework have a significantly higher spatial and temporal resolution; in this case, a horizontal resolution of ≈ 64m and temporal resolution of 30s, yielding a very high dimensional input feature space for a utility scale wind farm power forecasting model. The temporal and spatial resolution of these models mean that atmospheric processes important for wind power variability, such as turbulence, are directly resolved. Additionally, the wind turbines are embedded in the simulation, allowing for a more detailed account of the wind wakes and interactions around the wind farm.

However, the new information content of ultra-high resolution models does not necessary translate into improvements in average error metrics. For one thing, there is obviously a lot of uncertainty in the boundary forecasts driving the simulation at the day-ahead scale. In fact, high-resolution forecasts may perform worse compared to standard resolution models due to penalisation of phase and location errors; a forecast that contains a higher, and subsequently more realistic, variability in space and time is heavily penalised when the precise location/timing of a weather event is missed when using typical average error metrics for evaluation.

The research aim and contribution of this chapter is simple: to ascertain if it is possible to extract value from a high resolution NWP model in the context of wind power forecasting. We also explicitly explore the impact of temporal smoothing of

wind speed forecasts using single-input models, with the aim of understanding better the benefits of feature engineering published widely in the literature [30, 31].

We generate post-processed point forecasts of wind speed and power at the iconic Horns Rev I 160 MW wind farm off the coast of Denmark and compare the forecasting ability of the high resolution LES model with forecasts from ECMWF. In the post-processing of the raw NWP data into power and wind speed forecasts, smoothing techniques are used to account for the penalisation of temporal and spatial displacement errors. The disadvantage of smoothing is that high resolution information can be diminished; to remedy this, a simple feature engineering stage is proposed which can account for the variability in the atmospheric model.

Section 4.1 details the specifics of the NWP models used and the case study site, and Section 4.2 includes descriptions of statistical post-processing techniques. Results are presented, evaluated, and compared in Section 4.3 based on measurements from an individual turbine and subsequently the overall wind farm, and finally Section 4.4 details the conclusions and future work.

4.1 Case study at the Horns Rev offshore wind farm

To evaluate the performance of post-processed high resolution forecasts generated from the LES model, results are presented for Horns Rev I and compared with those based on ECMWF predictions. Horns Rev I was one of the first large capacity offshore wind farms and therefore has a large bank of historical data. Both wind speed and wind power post-processing are considered as these are both extremely important for operational tasks such as trading and operation and maintenance (O&M).

4.1.1 Horns Rev

Out of the 3 offshore wind farms located at Horns Rev, we focus on that which was installed first, in operation since 2002, and commonly referred to as Horns Rev I. Horns Rev I consists of 80 2-MW turbines located approximately 18km off the coast of Denmark in the North Sea covering an area of around 21km^2. It was historically

the first large offshore wind farm worldwide. The data used for the following study consists of average power generation and wind speed anemometer measurements from each constituent turbine from January 2015 to December 2016 at a 10-minute temporal resolution. Initially, to reduce the dimensions of this exploratory analysis, much of the post-processing results presented are for an individual turbine located at the north west corner of the park with a final analysis of the wind farm power as whole. It should be noted that a single turbine located at the south west corner of the wind farm is excluded from the analysis, due to a permanent turbine malfunction.

4.1.2 Weather forecasts: from ECMWF to the Whiffle Large Eddy Simulation

The operational ECMWF deterministic forecasts covering the years 2015 and 2016 were retrieved at 3-hour time resolution and approximately 16km spatial resolution. Note, that forecasts are also available at 1-hour resolution, but at the outset of this study ECMWF typically offered only the 3-hour resolution data to commercial customers. To make the forecasts comparable to the turbine data, the forecast fields were interpolated in time using linear interpolation and model levels were interpolated to a hub height of 70 meters by linear interpolation of the closest model levels [146].

Whiffle is a start-up that has focused on the computation of LES on Graphics Processing Units (GPUs), which enables their model GRASP (GPU-Resident Atmospheric Simulation Platform) to compute 48h hour ahead forecast within roughly an hour. The formulation of the model is provided in [147] and more recent features, including the GPU implementation and the method to drive an LES model with large-scale boundary conditions from a NWP model have been described in [42] and [148].

The simulation domain for this study is 8.2km in the horizontal and 5km in the vertical with 128^3 grid points, yielding a horizontal resolution of 64m. In the vertical direction, the grid is stretched, with a resolution of roughly 16m near the surface and 80m near the top of the domain. Periodic boundary conditions are taken from the ECMWF deterministic forecasts and are applied as dynamic tendencies to a so-called concurrent precursor simulation [149]. The values of the precursor LES are

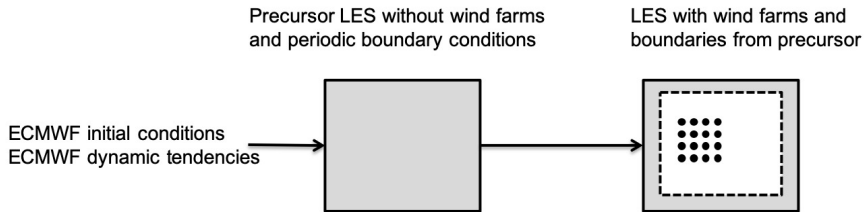
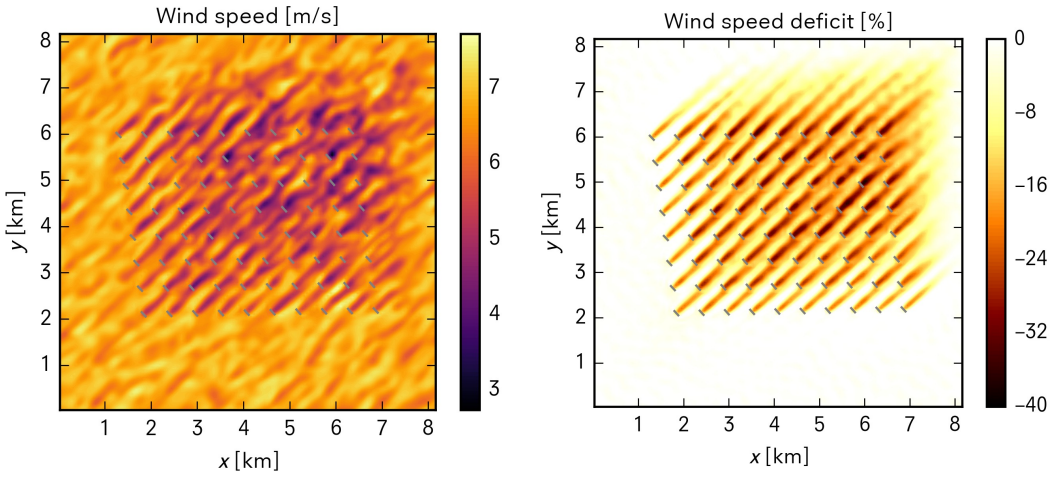


Figure 4.1: Schematic representation of the LES model setup. ECMWF fields are used to estimate the initial conditions and are applied as dynamic tendencies to a pre-cursor LES with periodic boundary conditions. The precursor LES values are then prescribed on the boundaries of a second LES to prevent the wakes from re-entering the domain

then prescribed, during run-time, to the boundaries of a second LES that includes wind turbines. This setup allows for the development of sufficient turbulence, while it prevents the re-circulation of the turbines wakes in the LES domain, due to the periodic nature of the boundary conditions. Figure 4.1 shows a schematic representation of this setup.

Wind turbines have been modelled in the LES with a uniform actuator disk model as described by [150]. The turbine parametrisation applies axial and radial forces to the LES wind fields that are based on the power and thrust curves of the turbine type. Therefore, the torque exerted on the rotor blades can readily be diagnosed from the simulation and the produced power can be exported as output variable with the same time resolution as the LES time-step, which is typically in the order of 5 seconds. All output variables that may be relevant for post-processing, such as power, rotor-disk averaged wind speed and direction and air density have been exported for each individual turbine location with a 30 seconds time resolution. Care was taken to produce a forecast dataset that was as representative as possible for day-ahead forecasting, so a maximum computation time of 1 hour per forecast was observed and the 00UTC ECMWF operational forecasts valid for the next 24-48h (local time) were used as boundary conditions. The 00UTC cycle of the ECMWF high-resolution forecast is available around 06:20 UTC for commercial customers. With the settings applied in this paper, the day-ahead LES forecast can thus be delivered at 07:20 UTC the latest, which corresponds to 08:20 CET and 09:20 CEST. The LES forecast is therefore well in time to be used for day-ahead trading.



(a) Instantaneous wind speed

(b) Two hour average wind speed deficit

Figure 4.2: Wind speed fields at hub height for 2016-05-18 02:00 UTC, where the turbine locations and their yaw angles are depicted with the grey bars. The average deficit (percent reduction with respect to free-stream wind speed) is calculated using the preceding two hours

Figure 4.2a shows a snapshot of the instantaneous wind speed field at hub height for a typical day with south-westerly winds. Using the pre-cursor simulation that has no wind turbines, it is also possible to calculate the difference between the free-stream wind speed and the wind speed in the wind farm. Figure 4.2b shows the wind speed deficit, i.e. the percent reduction in wind speed with respect to the free stream wind, in the simulation domain. The wind speed deficits for this particular day are roughly 20% to 40%, which is in agreement with values that have been reported in the literature [151].

Although the focus of this paper is on the prediction of wind speeds and wind farm power, the framework of a turbine resolving LES model driven by NWP boundary conditions opens more possibilities for future research and practical applications. Currently ongoing research efforts focus of wind resource assessments and annual energy prediction using LES with reanalysis fields as boundary conditions, air-sea interactions in and around wind farms, more advanced turbine parametrizations and in-depth analysis of the interactions of (far) wake effects and their environment.

4.2 Post-processing models

The following describes the statistical post-processing models utilised and the general post-processing strategy for the case study. The approach can be summarised by using statistical learning techniques for both wind speed and power forecasting with a focus on smoothing, feature engineering, and selection. Note that due to the inclusion of both wind speed and power forecasting, the notation y_t signifies the target variable (measured wind speed or power) and x_t is the weather forecast-derived input variable at time t.

Smoothing is an interesting approach in this high-resolution forecasting context, however it has been surprisingly overlooked in the applicable research literature for energy forecasting. Benefits have been reported with temporal and spatial feature engineering in the post-processing of coarse global atmospheric models [30–32] with rolling average and leading/lagging wind speed variables among others. Therefore, the aim here is to explore this relationship more explicitly with single input models and then to capture the value in the high-frequency content of the signal with other engineered features. This approach, compared to using the high resolution wind speed signal directly, aims to retain the value of the smoothed wind speed forecast whilst augmenting the model with selected summary statistics of the high frequency signal.

4.2.1 Wind speed forecasting

For wind speed forecasting the raw NWP signals are statistically post-processed to account for systematic biases between each signal and the measured time series at the turbine nacelle. A truncated linear regression model is used to account for the non-negativity of wind speed [152] which is defined as

$$y_t \sim \mathcal{N}_0(\mu_t, \sigma^2) \tag{4.1}$$

$$\mu_t = \beta_0 + \beta_1 x_t \tag{4.2}$$

where wind speed y_t at time t follows a zero truncated normal distribution with mean μ_t and standard deviation σ while μ_t is a linear function of the predicted wind speed x_t

and β_0, β_1, and σ are regression coefficients. The zero truncated normal distribution has a probability density function of

$$d(y_t) = \begin{cases} \frac{1}{\sigma}\phi(\frac{y_t-\mu_t}{\sigma}) & y_t > 0 \\ 0 & y_t \leq 0 \end{cases} \tag{4.3}$$

where $\phi(.)$ is the probability density function of the standard normal distribution. The regression coefficients are estimated with maximum likelihood estimation as implemented in the `R` software package `crch` [153].

4.2.2 Power forecasting

The Gradient Boosting Machine (GBM) is a supervised learning method whereby a series of individually weakly predictive base-learners (e.g. regression trees) are combined to make a powerfully predictive ensemble [154]. For an excellent and in-depth tutorial on this subject the reader is referred to [155]. The algorithm works by consecutively fitting single base-learners to improve the overall predictive estimate of the target variable; each learner is trained sequentially on the negative gradient of the loss function, with respect to the ensemble constructed so far. Intuitively, the model is grown at each iteration to improve upon the prediction of the ensemble so far; regularisation is extremely important to prevent over-fitting, which necessitates the shrinkage term introduced in the following definition of the algorithm.

Specifically, for base-learner regression trees, at each iteration the available input space is partitioned into disjoint regions, which allows for the direct capture of non-linear relationships, such as the wind power curve, and makes this a powerful tool for energy forecasting applications [30–32]. Additionally, the flexibility of the algorithm in terms of loss functions is a desirable attribute; in this chapter a squared loss function is used. For target variable y and a pool of explanatory variables $\mathbf{x}_t = (x_1, x_2, \ldots)^\top$, the gradient boosting machine [154, 156] $F_N(\mathbf{x}_t)$ is defined as the sum of N base-learners $f_n(\mathbf{x}_t)$

$$y_t = F_N(\mathbf{x}_t) + \epsilon_t = \sum_{n=0}^{N} f_n(\mathbf{x}_t) + \epsilon_t \tag{4.4}$$

where $f_0(\mathbf{x}_t)$ is the initialisation guess, ϵ_t is an error term, and the subsequent ensemble of base-learners is constructed sequentially by estimating the latest via

$$\underset{f_n}{\operatorname{argmin}} \sum_t L\left(y_t, F_{n-1}(\mathbf{x}_t) + f_n(\mathbf{x}_t)\right) \tag{4.5}$$

for some loss function (e.g., squared loss) $L(\cdot)$. To tackle this approach in practice, the base learner chosen here is the regression tree $f_n = h(\mathbf{x}; \theta_n)$, specified by a vector of tree parameters θ_n. Where $L(\cdot)$ is differentiable, the negative gradient $g_n(\mathbf{x})$ is defined as

$$g_n(\mathbf{x}_t) = -\left[\frac{\partial L\left(y_t, F_n(\mathbf{x}_t)\right)}{\partial F_n(\mathbf{x}_t)}\right]_{F_n(\mathbf{x})=F_{n-1}(\mathbf{x})} \tag{4.6}$$

and the regression tree is efficiently fit to this negative gradient by least squares

$$\theta_n = \underset{\theta}{\operatorname{argmin}} \sum_t \left[g_n(\mathbf{x}_t) - h(\mathbf{x}_t; \theta)\right]^2 \quad . \tag{4.7}$$

The ensemble is then updated with

$$F_n(\mathbf{x}_t) = F_{n-1}(\mathbf{x}_t) + \lambda \rho_n h(\mathbf{x}_t; \theta_n) \tag{4.8}$$

where λ which is a user defined regularisation lever termed shrinkage, and ρ_n which is the best gradient step-size found via

$$\rho_n = \underset{\rho}{\operatorname{argmin}} \sum_t L\left(y_t, F_{n-1}(\mathbf{x}_t) + \rho h(\mathbf{x}_t; \theta_n)\right) \tag{4.9}$$

and specifically for regression trees a different ρ_n is computed for each terminal leaf. This model fitting optimisation strategy is then based on two stages: least squares fitting of the base learner followed by the parameter optimisation according to the general loss function [157]. Regularisation is extremely important when deploying a GBM model. Generalisation performance is achieved via the tunable shrinkage parameter (λ) which allows the user to penalise the importance of each individual learner in the overall ensemble, the number of learners, and row/column [158] sub-sampling fractions where a random subset of rows or covariates respectively are used for each training

round. Row and column sub-sampling are really useful for generalisation performance when a large set of input features are available. If the entire dataset is defined as $\{y_t, \mathbf{x_t}\}_1^T$ then a random sub sample of rows of size $\tilde{T} < T$ is given by $\{y_{r(i)}, \mathbf{x}_{r(i)}\}_1^{\tilde{T}}$, where $\{r(i)\}_i^T$ is random permutation of the integers $\{1, \ldots, T\}$ [157]. A random subset of features/columns can be defined similarly.

Other tree-specific tunable hyper-parameters include the depth (or number of splits) of each tree, and the minimum number of observations per terminal leaf of the tree. It is evident that the benefits of the GBM method come at the expense of the necessary fine-tuning stages compared to for instance a linear model. These regression models were created in `R` via the `h2o` [158] software package.

4.2.3 Feature Engineering

Although high-resolution numerical models may be able to better approximate specific weather features, temporal and/or spatial displacement errors can in fact reduce their skill compared to coarser global models that provide a smoother representation of the same events. Figure 4.3 shows an example time series of the raw wind speed forecasts versus the turbine anemometer measurements. Visually, it is clear that neither GRASP or ECMWF can perfectly predict the truth, although GRASP clearly shows a higher and thus more realistic variability.

When regarding the Pearson correlation coefficient between the observations and NWP forecasts in Table 4.1 (top row) it can be seen that the high frequency information actually seems to disturb the forecast signal and that smoother forecasts are of advantage. This observation raises the idea of temporally smoothing the forecasts to obtain forecasts with the same underlying signal but with some of the (potentially disturbing) high frequency information removed. This is most simply done by calculating

	GRASP	**ECMWF**
Raw	0.898	0.910
Smoothed	0.905	0.915

Table 4.1: Pearson correlation coefficient of raw and smoothed forecasts with wind speed observations. For smoothing the moving average spans 500 minutes

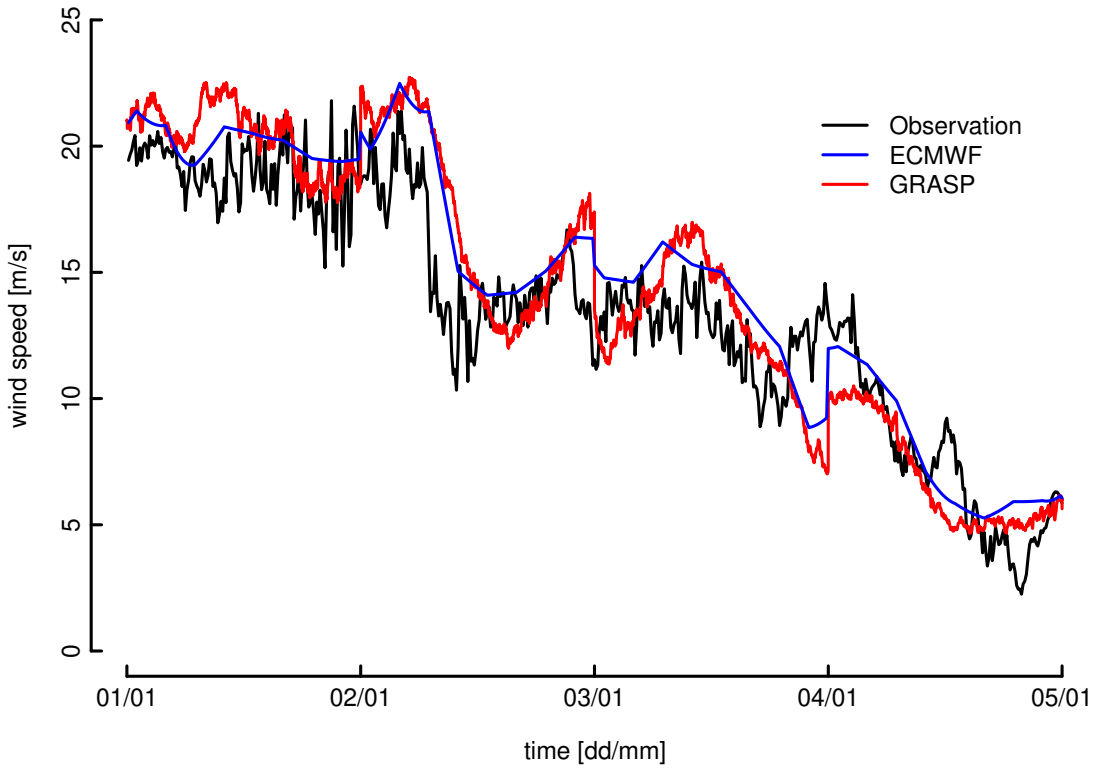


Figure 4.3: Example time series of raw wind speed forecasts from both NWP sources versus the turbine anemometer measurement. The GRASP and ECMWF variables used are the disk averaged and 70 metre wind speed respectively. Each of the 4 day-ahead forecasts have different issue times, indicated by the jumps in wind speed predictions around midnight each day

moving averages of the raw time series

$$\tilde{x}_t = \frac{1}{M+1} \sum_{r=t-M/2}^{t+M/2} x_r \tag{4.10}$$

where M is an even number that controls the degree of smoothing. For this study, we tested moving averages spanning time windows from 30 seconds to 16 hours. Although high frequency information is disregarded, Table 4.1 (bottom row) shows that this smoothing increases the correlation between the observations even in the ECMWF model. The justification of smoothing in post-processing, even for ECMWF forecasts, can be alternatively explained by the fact that the NWP global models resolve processes

at resolutions that are less than that of the grid they are computed on. In addition, high-resolution forecasts expose themselves to the risk of being doubly penalised when predicting the right events, but slightly misplaced in time or in space. Smoothing somewhat dampens this risk. It should be emphasised that the GRASP data retains the high resolution spatial information and one example of spatial smoothing is retained in the analysis and is defined as the spatial average of all the disk averages wind speeds in the wind farm.

Although average scores may be improved by smoothing the numerical forecasts, the high frequency content still might contain some valuable information. Therefore, moving variances are derived as

$$z_t = \frac{1}{M+1} \sum_{r=t-M/2}^{t+M/2} (x_r - \tilde{x}_t)^2 \tag{4.11}$$

which are used as supplementary engineered features to summarize the variability of the forecast. Additionally, a separate strategy to exploit the higher frequency content of the NWP signals is proposed based on a rolling Fast Fourier Transform (FFT). A smoothed Power Spectral Density (PSD) estimate of the transformed signal is split into a number of bands; the average, sum, range, and variance of the power in these bands is used to inform the models. The rolling discrete Fourier transform of the time-series centred on the window M is defined as

$$X_{t,k} = \sum_{r=t-M/2}^{t+M/2} x_r e^{-i2\pi rk/M}, \quad k = 0, 1, ...M-1 \tag{4.12}$$

at frequency domain point k. The corresponding rolling estimate of the PSD at this point is

$$P_{t,k} = \frac{1}{M}|X_{t,k}|^2 \quad k = 0, 1, ...M-1 \tag{4.13}$$

and since the Fourier transform of this real valued data is symmetric, the bands which are defined to engineer features within a frequency range are split equally within the $M/2$ range. The features retained in the final models are then based on summary statistics of $P_{t,k}$ within the highest frequency band which proved to be most informative.

Other transformations could potentially better track the time-varying properties of the time-series such as the Hilbert-Huang Transform [159], however for this proof of concept study these aspects are reserved for future work.

Therefore, the modelling strategy can be summarised as follows: to use temporally smoothed forecasts as the driving signal for the post-processed models and supplement this with engineered features that inform the model with information on variability of the NWP. As discussed, feature engineering has proven extremely successful in both wind and solar energy forecasting [30–32]. However, the temporal smoothing properties have not been explored explicitly, to the best of our knowledge, by single input models. It should be noted that all of the engineered rolling features are calculated per issue time of the forecast because of potential step changes in the variables across at this point.

4.3 Results

The results of this case study are presented as firstly an analysis of the raw LES output, following an exploratory analysis where a single turbine in the farm is selected for post-processing with single input models to explicitly extract the value in smoothing. Both wind speed and wind power post-processing are considered at this stage. Next, additional features are added to the turbine power models to capitalize on the high frequency content, before finally results for the wind farm level forecasting are presented.

For the post-processing analysis, the data is partitioned into tuning and testing data by 6 and 18 months blocks respectively, where the tuning data is used to improve the GBM models. Out-of-sample forecasts are generated using 5-fold cross-validation on the testing dataset with curtailment around the wind farm flagged and removed from the forecasting exercise. The definition of prediction error $e_t = y_t - \hat{y}_t$ is the difference between the measured y_t and forecast value $\hat{y}_t$. Evaluation of the post-processed forecasts is then based on the Root Mean Square Error (RMSE) over the testing dataset and the results are generally presented via boxplots showing the sampling distribution from bootstrap score averages. Bootstrap sampling can convey the distribution of

the evaluation metric via sampling the errors with replacement, then calculating the average error metric, and repeating the process k (in this instance 100) times.

The power point forecasts generated by the GBM models are constructed to minimize the quadratic loss function in this study. Therefore, the models created will be optimised for the mean squared error and this motivated the use of RMSE as the lead measure of accuracy. Though other scores such as mean absolute error could also be similarly reported, the results will not be optimised for the measure unless the models were modified to carry out a median regression. Additionally, due to the wealth of models tested in this case study the MAE results are omitted for brevity.

4.3.1 LES results

We first present a number of quantities from the unprocessed LES output to verify the implementation of the wind farm in the LES model set-up. Figure 4.4 shows the yearly average power production per turbine from the observations and as forecast by GRASP. For the turbines on the western and southern edge of the wind farm, the yearly average production shows a good agreement between the model and the observations. The deviation is roughly within a 20 kW range, which is 2% of the average yearly production. Deeper inside the wind farm, GRASP underestimates the production, or, equivalently, overestimates the wake effects. This is also confirmed by an initial analysis looking at the RMSE of the power as exported directly from the LES (results not shown here) and forms another justification for statistical post-processing.

4.3.2 Exploratory Data Analysis

In order to reduce the dimensions of the problem, the following presents results of forecasting for a single turbine in the farm, located at the north-east perimeter of the farm in terms of both wind speed and power prediction performance. Here we have a few parameters which require explanation: DA indicates the disk averaged wind speed which is a weighted average of the closest LES grid point wind speeds over the rotor disk plane, and DA-all is the spatial average of these disk averaged variables over all the turbine locations.

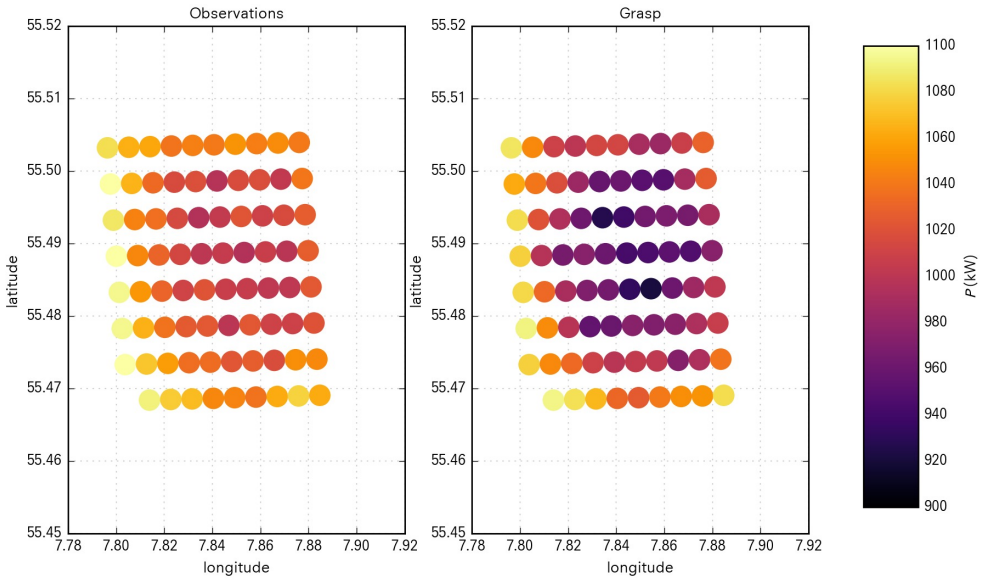
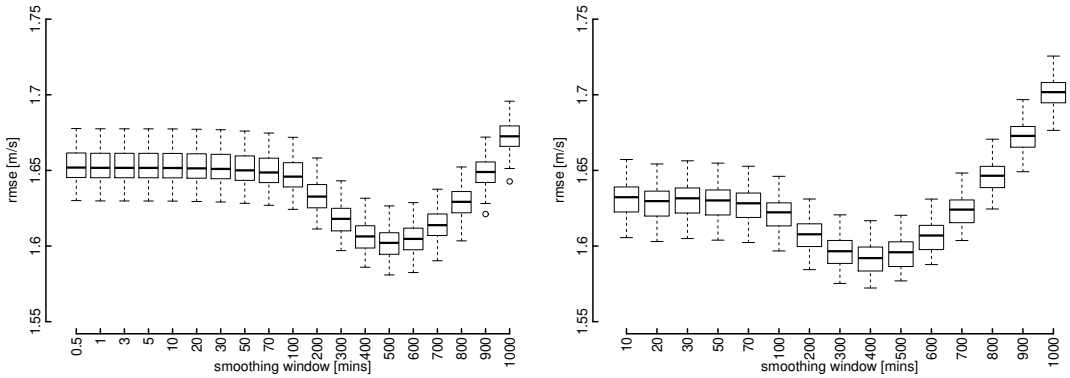


Figure 4.4: Yearly average power from observations (left) and forecast by the GRASP LES model (right)

Single Input Models

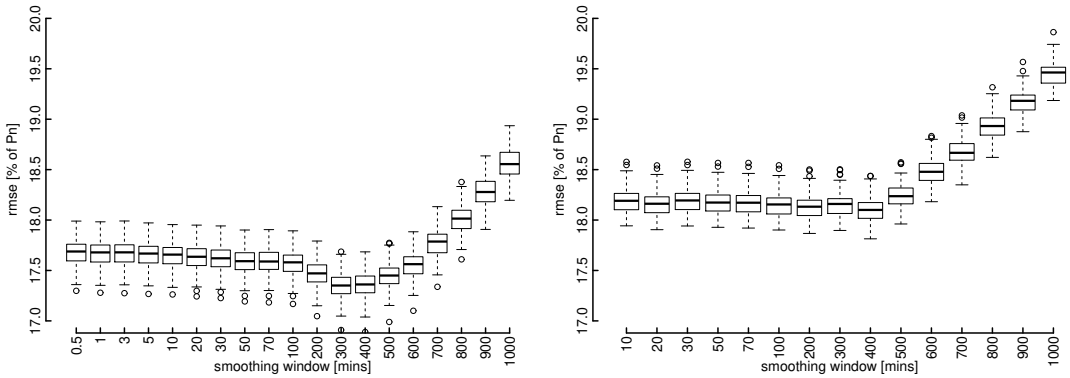
To explicitly investigate the influence of smoothing the wind speed forecast on both wind power and wind speed forecasting single-input models are first considered; a separate model for each NWP wind speed forecast is trained against the measured wind speed or power time-series.

Figure 4.5 shows the results for the wind speed post-processing case at 70m height where 0.5 and 10 minutes indicates no smoothing for GRASP and ECMWF data respectively. It should be noted that the ECMWF is already linearly interpolated to match the temporal resolution of the power measurements. The RMSE for the ECMWF is generally much lower for wind speed across the smoothing window choices, apart from the longest rolling window lengths of >900 minutes. For both models the optimal smoothing window is around 400-500 minutes which is longer than the original temporal resolution of the ECMWF data and implies that smoothing can improve the forecast accuracy of a post-processed model derived from both traditional NWP output as well



(a) GRASP (b) ECMWF

Figure 4.5: Wind speed RMSE results with different smoothing windows at 70m height. Forecasts were derived from a truncated linear regression model with different smoothed forecasts as input. At this turbine, ECMWF gives more accurate wind speed forecasts across all the smoothing windows

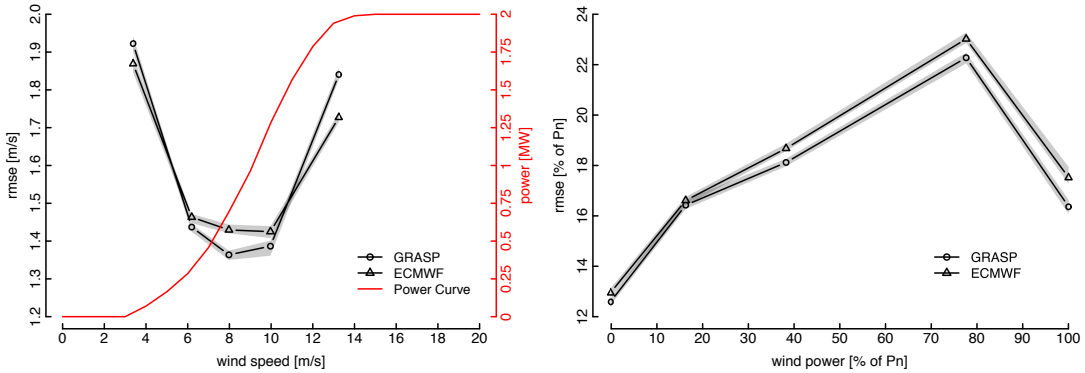


(a) GRASP (b) ECMWF

Figure 4.6: Wind power RMSE results with different smoothing windows. Forecasts were derived from GBM models with smoothed wind speed input variables at 70m height for ECMWF and the disk averaged wind speed (DA) for GRASP. At this turbine, GRASP gives more accurate wind power forecasts across all the smoothing windows. This is because GRASP gives superior wind speed forecasts, *at regions key to wind power prediction* as shown on Figure 4.7a

as high-resolution LES output.

For wind power, shown in Figure 4.6, a similar profile is found with some very important distinctions; the GRASP based models give a lower error than ECMWF across the smoothing windows, and the characteristic dip of the wind speed plots is slightly shifted and not as pronounced. This suggests that from smoothing, the error



(a) Wind speed with power curve

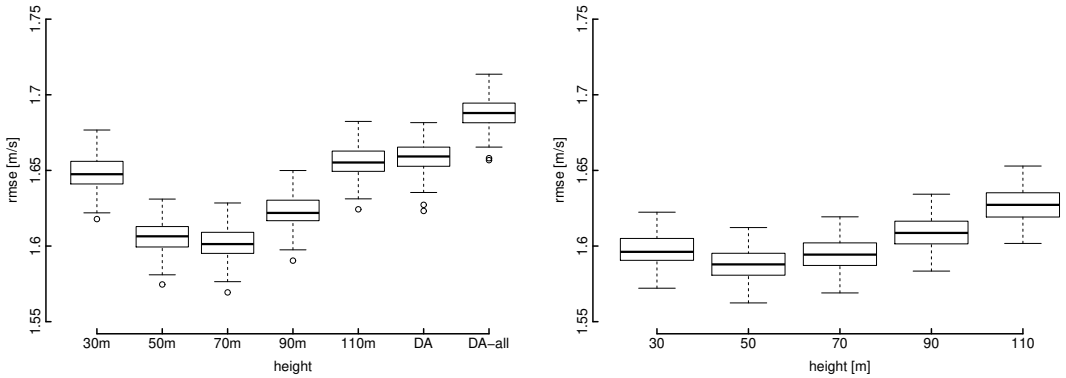
(b) Wind power

Figure 4.7: RMSE of Whiffle's GRASP model and ECMWF conditioned on 5 equally populated observation bins with a 400 minute smoothing window and at a height of 70m. The shaded areas show the interquartile range from bootstrap score averages

improvements at the optimal smoothing window are out-with the below rated region of the power curve because the non-linear effect of the power curve would tend to amplify improvements in this key region. However, for both NWP sources, smoothing NWP forecasts for use in power forecasting does have a beneficial influence at a window of around 400 minutes.

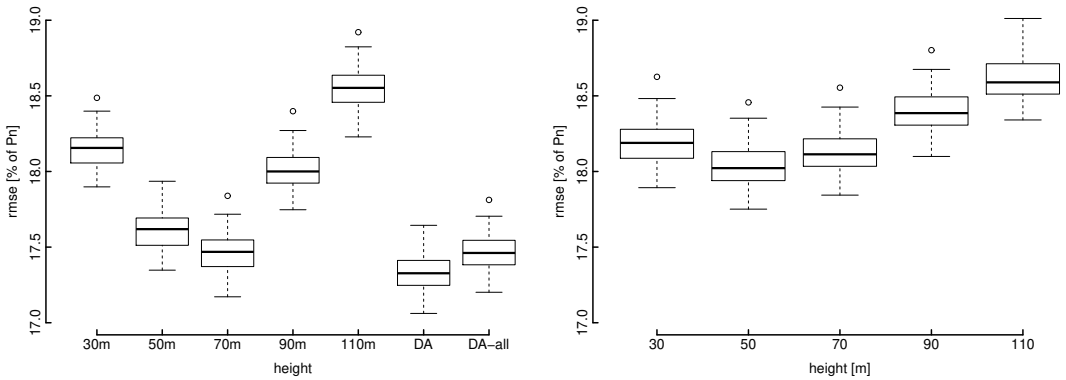
The unusual difference between the wind speed and wind power performance between the NWP sources is explained in Figure 4.7 which shows the performance of four comparable models conditional on the wind speed and power measurements respectively. All models compared here are using a smoothing window of 400 minutes and at 70m height. At the key wind speed ranges for power prediction — the medium range wind speeds — the GRASP post-processed model is more accurate. Due to the influence of the cut-in and rated regions of the turbine power curve, the larger wind speed prediction errors of the GRASP model at the low and very-high wind speed ranges respectively are negated in the power domain. It should be noted that wind speed predictions could be improved at these regions by using (for instance) splines in the regression, however this is out-with the scope of this preliminary analysis which is mainly focused on wind power.

Looking at the performance in terms of the model height is also informative. Figure 4.8 shows that for wind speed forecasting the ECMWF gives the highest accuracy



(a) GRASP (b) ECMWF

Figure 4.8: Wind speed RMSE at different model heights with a 500 minute smoothing window. DA and DA-all are the disk average wind speed, and the wind farm spatial average of the disk average wind speed respectively

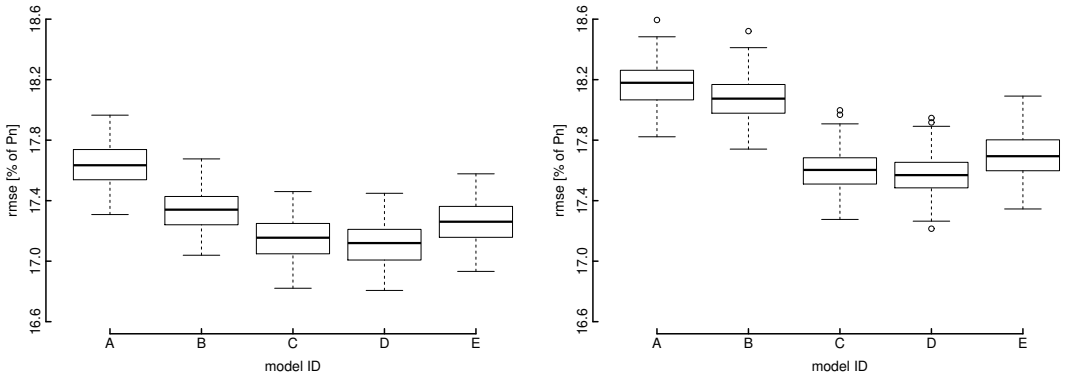


(a) GRASP (b) ECMWF

Figure 4.9: Wind power RMSE at different model heights with a 400 minute smoothing window. DA and DA-all are the disk average wind speed, and the wind farm spatial average of the disk average wind speed respectively

at 50 metres, with broadly similar performance across the heights compared to GRASP; it is important to emphasise that these additional ECMWF height fields are linearly interpolated from the nearest available model heights, roughly 10m, 31m, 54m, 79m and 107m. For the LES model there is a pronounced improvement around the hub height of the turbine, which signals that there is more skill in the vertical profile of the LES compared to that of ECMWF. This is expected because the resolved wake effects are most prominent at hub height.

For power predictions over the different model heights, as shown in Figure 4.9, there



(a) GRASP (b) ECMWF

Figure 4.10: Wind power RMSE where ECMWF features are at 70m height and for GRASP the disk averaged wind speed and wind direction at 70m are used for GRASP. **A** is the raw NWP signal; **B** is the smoothed NWP signal over 400 minutes; **C** is the smoothed NWP and wind direction; **D** is the smoothed NWP, PSD features, and wind direction; **E** is the smoothed NWP, rolling variance, and wind direction

is again a reversal of roles where GRASP gives much improved forecasts at four key model heights. Notably, the disk averaged wind speed variable is the best predictor of wind power. The wind farm spatial average of the disk averaged wind speed forecasts is not the best predictor here, which suggests that for this particular turbine the high spatial resolution is providing some benefit.

Improving the Power Forecast

Thus far, the analysis has essentially excluded the high temporal resolution content. Including the proposed rolling variance and PSD band features aims to capitalize on this available information. Figure 4.10 is the culmination of the exploratory analysis; it shows the progression from the raw NWP forecast, to the smoothed signal, to the smoothed forecast with wind direction, then including PSD band features, or finally rolling variance variables. For GRASP, the forecast used is the disk averaged wind speed and wind direction at 70 meters, and for ECMWF the speed and direction forecasts are based at 70 meters. The smoothing and rolling window used for both is 400 minutes.

Although these single-input models have highlighted the influence and value of temporal smoothing, the errors can be clearly further reduced by incorporating wind

direction and engineered features. As shown in Figure 4.10, incorporating wind direction into the model is valuable for the GRASP. For the ECMWF forecasts, a more significant improvement is observed when including wind direction. It is suggested that could be due to some of the directional effects of the farm, such as wake deficits, are already resolved in the high resolution model, whereas for ECMWF these effects are obviously excluded. The plot also illustrates that incremental benefit is achievable for GRASP and even ECMWF using these PSD engineered features. However, using the simpler rolling variance degrades performance across both cases. Overall at this turbine, Figure 4.10 shows that accuracy improvements are achievable using the high resolution LES data. Additionally, it is clear that to fully utilise the high resolution temporal content of the signal more advanced models are required.

4.3.3 Wind Farm Power

Investigating the overall predictive skill, in terms of the wind farm power, can illustrate the average performance of both source datasets. This section is based on an approach whereby the wind farm power is modelled with the entire set of explanatory variables in each case, and the inherent feature selection ability of the statistical learning technique is relied upon to select the most relevant features. The three cases considered are the separate GRASP and ECMWF datasets with smoothing and features engineered as shown previously, and finally a combined model with access to all available NWP inputs and features.

The dimensionality of the data in these three cases are an issue however because of the large pool of available information; for the combined case we have an input feature space of 820 variables, due to the amount of features engineered at each turbine. Although the boosting algorithm provides an inherent feature selection capability, empirically it is found that a more rigorous selection and feature reduction stage can improve model performance by making them more parsimonious [30].

The feature selection stage employed here involves fitting a regularised GBM model with all the available inputs and retaining only the features which have the highest influence. A second (and final) model is then trained using these selected variables.

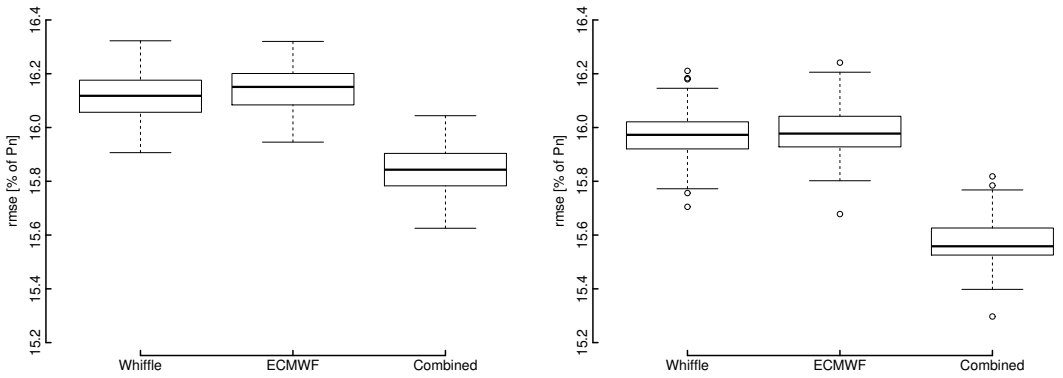
Briefly, the variable influence in this tree-based algorithm is based on the number of times a variable is selected for splitting in the model fitting, as well as the resulting empirical reduction in the loss function as a result of these splits; the reader is referred to [155] for more information. A complete list of the initial and reduced set of features can be found in Appendix A. This process removes the influence of unnecessary predictor variables where no additional valuable information is available but may have been used sparingly in the primary model training. Regularisation levers in the GBM framework provide the necessary framework for this feature selection stage and have been compared to other more familiar algorithms such as penalised regression [156].

Figure 4.11 shows the results of the wind farm forecasting case study. From this it is clear that, even in this context, the high resolution data is competitive with ECMWF. However, the real value comes from using the information from both GRASP and the standard ECMWF forecast, which gives significant improvements in RMSE over both single dataset models. This is understandable given that the model is selecting the best of both worlds in terms of information content. Lastly, a feature selection stage is clearly shown to improve the accuracy of the 3 models tested.

The key result of this study is in demonstrating that it is possible to obtain improvements in typical average error metrics using a combination of ultra-high resolution NWP data, modern state-of-the-art regression techniques, and typical weather forecasts available to commercial operators. The GRASP model clearly adds new information via the spatial and temporal resolution which, combined with the regression technique and feature engineering, leads to improved forecast skill. However, this is a proof-of-concept study and further work is required to both maximise the value of the high resolution data, and benchmark against different techniques; For instance, assessing the value of using GRASP in a typical commercial forecasting system, with an ensemble of established NWP sources. More future work avenues are elaborated in Section 4.4.

4.4 Conclusions & Future Work

The value in post-processing ultra high resolution weather forecasts for power prediction has been demonstrated, by improving the accuracy of point forecasts at Horns Rev I.



(a) All available NWP variables (b) After feature selection

Figure 4.11: Wind farm power results

The case study evaluates the performance of wind speed and power forecasting at a single turbine in the array to first explore the dataset and then investigates the performance of overall wind farm power prediction. These predictions are compared against forecasts generated purely from the standard resolution ECMWF model.

The exploratory investigation uses single-input models to characterise the value in temporal smoothing of both NWP sources which is shown to improve performance. Generally, ECMWF proved to be better at prediction of wind speeds, and Whiffle's LES model GRASP superior at power prediction at this single turbine. It is revealed that this is due to the high resolution model providing more accurate wind speed forecasts in the cubic region of the turbine power curve, which is expected because the wake effects resolved by the LES model are most prominent in this regime. Improving the power forecast by including wind direction and engineered features to capitalise on the high frequency content of the GRASP signals is investigated. Whilst wind direction measurably improves both forecasts, further work is necessary to fully exploit the temporal information in the LES model.

For wind farm power forecasting using information from both ECMWF and GRASP proved to give significant increases in accuracy, especially with a feature selection stage, compared to using information content solely from either single source, which are comparable to each other. This can be explained by the statistical learning technique model selecting the most relevant information content from a diverse pool.

Moving from the standard ECMWF forecast to the ultra-high resolution turbulence resolving model represents a significant jump in resolution. For future work looking to benchmark against all possible methods, it would be prudent to include a comparison of a middle-ground between the two, and evaluate performance against a mesoscale model [160, 161], typically used in commercial power forecasting systems. Additional benchmark comparisons should certainly include direct comparisons with ensemble members or an ensemble of weather prediction sources, however, it should be noted that a robust numerical comparison in this context would necessarily require inclusion of ultra-high resolution ensembles generated by the LES simulation. Such a study would currently require significant computational power and time.

Another possible benchmark for future work is the Model Output Statistics (MOS) method, which form the basis of the early wind power prediction tools [162]. As discussed in Chapter 2.2, conventional MOS techniques are difficult to apply to wind power directly because of the non-linear power curve and bounded nature of the time series [50]. However, it is particularly relevant to the wind speed forecasting aspect of this work. Furthermore, wind speed predictions from such models could be used as features for forecasting the power variable in a modern statistical learning framework.

In general, statistical post-processing is a common tool used in the wider weather forecasting community; the goal is to remove systematic bias present in the global NWP model for improved predictions at a specific location [43]. This process is typically carried out using Model Output Statistics, mentioned previously, via multiple linear regression [44], using historical NWP (of varying resolution and skill) and meteorological observations [45]. More recent applications include using machine learning for the regression [46] with varying success, post-processing ensembles for site specific calibration [47, 48], and using multiple forecast source information [28, 49]. Conventional post-processing techniques are difficult to apply to wind power directly because of the non-linear power curve and bounded nature of the time series [50]; this motivates the use of non-linear statistical learning techniques in this thesis.

In the context of this proof-of-concept study it is clear that it is possible to use the high resolution data with machine learning post-processing models to improve on

conventional wind farm forecasting by combining information content from GRASP and ECMWF. To realise further improvements in the LES wind forecasts, a number of promising venues for future research include evaluating the influence of the LES domain size and the time-resolution of the NWP boundary conditions. The latter should be updated to match the hourly resolution data recently made available from ECMWF for commercial use. Furthermore, a probabilistic framework in which the LES model is used to derive probability distributions of the forecast variables can have added value for wind energy applications.

The high resolution LES weather model is shown to approximate much more closely the underlying behaviour of the wind speed signal than the benchmark NWP. However, the double penalisation of spatial and temporal errors mean that average error metrics are perhaps not the best framework for evaluating these forecasts; event-based metrics and applications in ramp forecasting of wind power could be very useful for optimally leveraging the GRASP information content. Future work should consider this as well as utilising the high spatial and temporal information in a modelling framework more suitable to the data.

Engineered features such as the rolling variance which quantify the variability of the signal could be more valuable in probabilistic forecasting for modelling the upper and lower ends of the distribution via quantile regression [31, 163]. Variability features extracted from the power measurement time-series have also proven successful for very-short term probabilistic forecasting, i.e. 15 minutes to 6 hours ahead [164]. Leveraging temporal variability features extracted from the high resolution NWP then represents a promising avenue for future work in terms of improving the skill of probabilistic forecasts at extended lead times. A more in depth study focused on extracting value from the temporal content of the wind forecast signal such as deep-learning [165], instantaneous frequency transforms [159], or wavelet decomposition [166] techniques should be explored. Finally, a hierarchical model where each turbine is used to generate a consistent wind farm forecast could be an optimal way of using the high spatial content of the data.

Chapter 5

Leveraging Turbine-level Data for Probabilistic Wind Power Forecasting

Chapter based on:
Gilbert, C., Browell, J., & McMillan, D. (2019). Leveraging turbine-level data for improved probabilistic wind power forecasting. IEEE Transactions on Sustainable Energy, 11(3), 1152-1160.

In the previous chapter, point forecasts of wind power were generated from a high dimensional NWP feature space, and future work highlighted the possibility of using a hierarchical framework for the optimal use of the input data. This chapter describes two methods for creating improved *probabilistic* wind power forecasts through the use of turbine-level data. Although standard resolution NWP data is used, high-dimensionality is now present in the output variable space; the two methods scale by the number of turbines present in the wind farm, although to a different extent.

The first method proposed is a feature engineering approach, where deterministic power forecasts for individual turbines are used as predictor variables when producing non-parametric wind farm forecasts. This is a hierarchical method in the sense that information from the turbine-level is used to supplement the available information set.

However, forecast coherency, as discussed in Section 2.4.2 is not guaranteed. Therefore, a second approach is proposed based on hierarchical coherency. In this second bottom-up approach, density forecasts are produced for all turbines and the spatial dependence between them is modelled in a copula framework to allow aggregation to the wind farm level.

The **turbine-level feature engineering method** aims to improve the wind farm forecast by generating new covariates from individual turbine data. Whereas the **bottom-up probabilistic method** reflects the physical reality of the problem — that the total wind farm power output is the sum of individual turbine generation — and therefore have the added benefit of coherency. The main contributions of this chapter are the two proposed forecasting methods and their evaluation. We hypothesise that leveraging information from the turbine-level will enable us to improve forecast performance at the day-ahead lead times, particularly since modern utility scale wind farms are often distributed over large areas of complex terrain and as a result, individual turbines can experience different conditions from one another at any given time. The advantages of the proposed hierarchical method are improved accuracy and coherency between turbine-level and wind farm total, however the results suggest that nature of the wind farm (terrain, layout, size, ...) has a bearing on the extent of this improvement.

This chapter is organised as follows: Section 5.1 details the forecasting methods and benchmark models, Section 5.2 describes the case study based on two utility scale wind farms in the UK, Section 5.3 presents and evaluates the results, and conclusions are drawn in Section 5.4. Supplementary information provides additional detail and results in Appendix B.

5.1 Forecasting Methodology

This section covers the two tested methods for leveraging turbine level data, the benchmark models, and the statistical learning techniques employed. The entire forecasting methodology is summarised in Figure 5.1, which details the training process, input data, and output forecast of each model. The turbine level feature engineering model

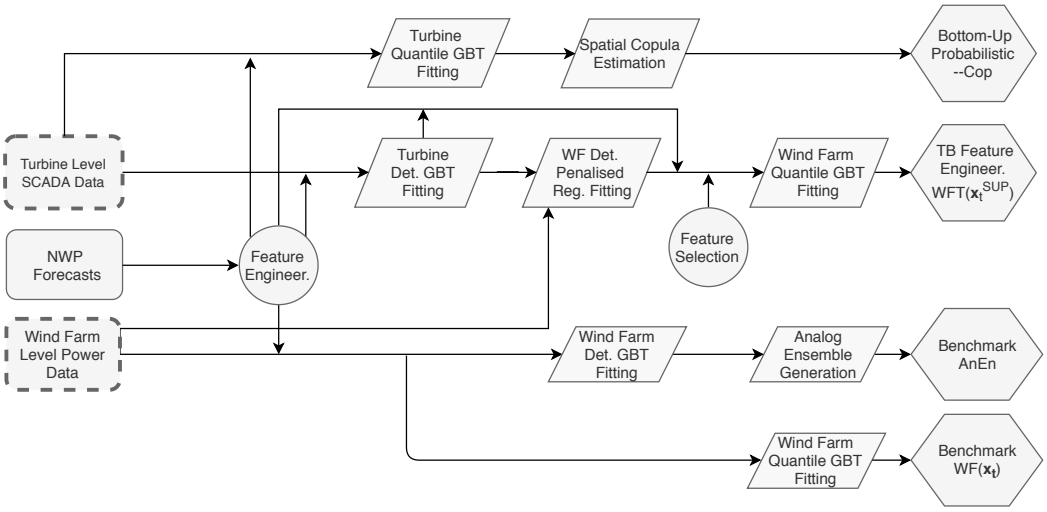


Figure 5.1: Flowchart illustration of the entire forecasting model training methodology. Wind Farm Level Power Data and Turbine Level SCADA data are only required for model training and evaluation. The only inputs required to produce operational forecasts are Numerical Weather Predictions. TB indicates turbine, WF is wind farm, Det. is deterministic, and Reg. is regression

is generated using quantile regression, where NWP predictions are supplemented with additional features; these include deterministic forecasts of individual turbine generation and wind farm-level generation.

The bottom-up probabilistic method involves estimating the full multivariate predictive distribution of generation from all turbines. To this end, the marginal distribution of each turbine is determined via quantile regression and the spatial dependency structure is modelled via a copula. The wind farm-level density forecast is then generated by sampling from the multivariate distribution and taking the empirical distribution of the aggregated turbine-level samples. The Gaussian copula with both empirical and parametric covariance matrices is examined, due to its simplicity and successful use in similar studies [69, 72, 91]; vine copulas with a range of copula families are also considered [92, 99].

Explanatory variables $\mathbf{x}_t$ common to both proposed methods and benchmarks are derived from NWP wind speed and direction outputs at 10m and 100m. Features that capture wind shear, veer, and phase errors in NWP are engineered inspired by [30, 31]. Cubic spline basis functions are also included to capture diurnal bias in the NWP at the

specific sites along the lines of [81]. Full details of all features are listed in Appendix B.

We use the GBM statistical learning technique to map the relationship between the input features derived from the NWP and the target measured time series, i.e. individual turbine or wind farm power measurements, as discussed in Section 4.2.2. Turbine-level deterministic forecasts used as features in this study are produced by GBMs fit with a squared loss function, and density forecasts are produced using GBTs via multiple quantile regression, with spline interpolation to estimate the predictive CDF; knots are placed at each predicted quantile and the boundaries, 0 and nominal power.

5.1.1 Benchmark Models

Two highly competitive benchmark models are implemented based on wind farm level power measurements and input features $\mathbf{x}_t$ derived solely from NWPs. These features include temporal averaging, shear and others; a full list is provided in Appendix B. The first benchmark is a wind farm-level GBT quantile regression model, $\mathbf{WF(x}_t\mathbf{)}$, and the second is an Analog Ensemble method, **AnEn**, described below. These benchmarks represent the state-of-the-art in wind power forecasting and were informed by [30,32,77] in particular.

The Analog Ensemble is a non-parametric algorithm that ranks similarity between the current forecast and a training dataset of historical forecasts with concurrent measurements. The k most similar concurrent measurements are used to construct an ensemble, assumed to be equally likely, from which empirical quantiles can be extracted. In this case, a mean GBM benchmark forecast is used as the explanatory variable and the model searches for the most similar out-of-sample mean power forecasts in the training dataset. The AnEn is also conditioned by lead time and the ranking metric used is euclidean distance. This algorithm is similar to the k-Nearest-Neighbours regression solution used in the second placed entry to the GEFCom2014 wind track [77]. For more information, the reader is referred to [78].

5.1.2 Turbine-level Feature Engineering

Here, we present the method to engineer features based on individual wind turbines to feed into the wind farm-level forecast. This approach comprises of two layers: in the first layer, deterministic forecasts for individual wind turbines and the wind farm as a whole are produced; then in the second layer, density forecasts for the wind farm are produced by quantile regression using features from both NWP and the first layer. The deterministic forecasts for individual wind turbines $y_{i,t}$ are produced using the same explanatory variables $\mathbf{x}_t$ as for direct wind farm-level forecasting benchmark. These forecasts are combined via a weighted sum over all D turbines to produce the deterministic wind farm forecast

$$z_t = \sum_{i=1}^{D} \omega_i y_{i,t} + \epsilon_t \quad , \tag{5.1}$$

which completes the constitution of the supplementary feature set

$$\mathbf{x_t^{SUP}} = [\mathbf{x}_t, y_{1,t}, ..., y_{D,t}, z_t] \quad . \tag{5.2}$$

The weights ω are estimated via elastic net regression motivated by the necessity to regularise turbine forecasts because they are highly correlated. The weights are calculated via

$$\boldsymbol{\omega} = \underset{\boldsymbol{\omega}}{\operatorname{argmin}} \left\{ \frac{1}{2N} ||\mathbf{Z} - \mathbf{Y}\boldsymbol{\omega}||_2^2 + \lambda \left[(1-\alpha)\frac{1}{2}||\boldsymbol{\omega}||_2^2 + \alpha||\boldsymbol{\omega}||_1 \right] \right\} \tag{5.3}$$

where α and λ are hyper parameters requiring tuning, $\mathbf{Z}$ and $\mathbf{Y}$ are matrices of vertically stacked instances of z_t and $\mathbf{y}_t$ [167]. The hyper parameter $0 \leq \alpha \leq 1$ controls the weighting of the two penalty terms, in effect trading off between ridge ($\alpha = 0$) and lasso ($\alpha = 1$) regression. Total regularisation is controlled by $\lambda \geq 0$. The optimal values of α and λ are determined through grid search and k-fold cross validation.

The final wind farm level density forecast, $\mathbf{WFT(x_t^{SUP})}$, is produced using quantile regression in the same way as the benchmark model but with the expanded feature set $\mathbf{x}_t^{\text{SUP}}$. In order to refine the forecast skill, a reduced feature set selected from $\mathbf{x}_t^{\text{SUP}}$ is used. This selection process involves fitting a regularised GBM model with

all the available inputs from $\mathbf{x}_t^{\text{SUP}}$, then selecting and retaining only the features that have the greatest influence. Briefly, the variable influence is based on the number of times a variable is selected for splitting in the tree-ensemble model fitting, as well as the resulting empirical reduction in the loss function as a result of these splits; the reader is referred to [155] for more information. This additional selection stage removes superfluous predictors which provide no additional information and only deteriorate forecast performance. The final variables retained in each model, and their relative importance, can be found in Appendix B. Low shrinkage and interaction depth hyper-parameter choices for the GBT algorithm provide a degree of regularisation and feature selection from which the dimensions of the problem can be reduced substantially [156].

5.1.3 Bottom-Up Probabilistic Method

Here, we propose a novel approach to forecast the power from the wind farm by estimating the joint predictive distribution of production from all wind turbines in the farm in a copula framework. The marginals of the copula comprise of density forecasts which are produced for each turbine using quantile regression and spline interpolation from the collection of quantiles. A range of copula functions are explored.

Let the random variable Y_i denote the wind power generation at the i^{th} turbine, and y_i the corresponding realisation (time indices are dropped to avoid notational clutter). The predictive CDF of the i^{th} turbine is

$$F_i(y_i) = P(Y_i \leq y_i) \tag{5.4}$$

for $i = 1, 2, ..D$ turbines. Sklar's theorem [168] states that for any D-dimensional cumulative distribution $F(\cdot)$ with continuous marginals $F_i(\cdot)$ there exists a unique copula function $C(\cdot)$ such that

$$F(y_1, y_2, ..., y_D) = C\left(F_1(y_1), F_2(y_2), ..., F_D(y_D)\right) \quad , \tag{5.5}$$

which separates the marginal distributions and dependency structures between the marginals. This is useful because it decouples the problem into two constituent parts:

1) estimating the marginal distributions for each turbine, and 2) estimating the dependence structure via a copula function. Note that the copula function links uniformly distributed marginals $u_i = F_i(y_i)$ and therefore the calibration of the density forecasts that form the marginal distributions is critical. Equation 5.5 can be alternatively written as

$$C(u_1, u_2, ..., u_D) = F\left(F_1^{-1}(u_1), F_2^{-1}(u_2), ..., F_D^{-1}(u_D)\right) \tag{5.6}$$

where $F_i^{-1}(\cdot)$ is the inverse of the marginal distribution $F_i(\cdot)$. Therefore, via sampling from the multivariate copula, pseudo-observations can be back transformed into the original domain to produce spatial scenario forecasts of power generation [91]. Next we introduce a range of options for the copula function.

Gaussian Copula

The Gaussian copula is given by

$$C(F_1(y_1), F_2(y_2), ..., F_D(y_D)) = \Phi_\Sigma\big(\Phi^{-1}(F_1(y_1)), \Phi^{-1}(F_2(y_2)), ..., \Phi^{-1}(F_D(y_D))\big) \tag{5.7}$$

where $\Phi^{-1}(\cdot)$ indicates the inverse standard normal distribution function and $\Phi_\Sigma(\cdot)$ the D-dimensional normal distribution function with covariance matrix Σ and zero mean. In this context, the covariance matrix encodes the spatial dependence structure for the D-turbines which illustrates one of the reasons why the Gaussian copula is so popular: the dependency structure is characterised by a single covariance matrix. It should be noted that $v_i = \Phi^{-1}\left(F_i(y_i)\right)$ constitutes the transformation of the uniformly distributed marginals into the Gaussian domain where $v_i \sim \mathcal{N}(0, 1)$. Therefore, we can estimate the copula by calculating the sample covariance matrix for the transformed normally distributed variables.

Using this framework, it is simple to sample from the multivariate distribution and generate D-spatial scenarios of the future generation. Each of the samples are back-transformed $\hat{u}_i = \Phi(\hat{v}_i)$, and then transformed into the original power domain using

the inverse CDF for the i^{th} turbine

$$\hat{y}_i = F_i^{-1}(\hat{u}_i) \tag{5.8}$$

which are summed over the D-turbines to give a snapshot of the wind farm forecast generation $\hat{z}_j$ for a j^{th} out of K ordered samples $j = 1, 2, ..., K$. Using the empirical distribution function the wind farm forecast with the correct underlying spatial dependence structures is finally given by

$$\hat{F}(z) = \frac{1}{K} \sum_{j=1}^{K} \mathbf{1}(\hat{z}_j \leq z) \quad . \tag{5.9}$$

We refer to this approach (based on the empirical covariance matrix) as **EGCop** in the proceeding text. From observing the often noisy empirical covariance estimates in this and other studies based on temporal scenarios forecasting [72, 91], we also consider a parametric exponential covariance structure – **PGCop**. This approach has shown to be effective in increasing forecast skill by smoothing the empirical covariance matrix. The parametric spatial covariance between two turbines is

$$\Sigma_{i,j} = \text{cov}(v_i, v_j) = \exp\left(- \frac{\Delta s}{\eta} \right) \tag{5.10}$$

where Δs is the spatial distance between turbines i and j in kilometres which are extracted from a distance matrix, and the parameter η is found using using empirical covariance and distance information; the lower triangle, including the on-diagonal terms, are extracted from the empirical matrix and distance matrix and the coefficient estimation is formulated as a non-linear least squares regression problem, where the coefficients are found by numerical optimisation. The parametric covariance matrix is then constructed using the full spatial distance matrix between the turbines and the coefficient estimate.

Copula Vine

The vine copula, **VCop**, is a series of bivariate copulas in which a different distribution family may be used for each pair. This allows for more complex dependency structures with asymmetry and tail dependencies to be captured, at the expense of added computational cost compared to the Gaussian method. This flexibility has encouraged recent studies considering vine copulas in the wind power forecasting context [92,99]. The vine method works by factorising the D-dimensional density into the $d(d-1)/2$ product of bivariate copulas where each pair copula is estimated via maximum likelihood from a set of distribution families (Gumbel, Gaussian, Student-t etc.). The optimal family for each pair-copula is chosen by minimisation of the Akaike Information Criteria (AIC). The implementation here follows [92] and for more detail please refer to [169].

5.2 Case Study

The proposed methodologies and benchmarks are tested on two large UK wind farms, Wind Farm A (128MW capacity, 56 turbines) and Wind Farm B (70MW capacity, 35 turbines), which cover an area of approximately 20km^2 and 15km^2 respectively. Training and testing data are partitioned at Wind Farm A into 12 and 4 month blocks respectively and at Wind Farm B 15 and 6 month blocks, due to differences in data availability. The test dataset covers the months of December to March for Wind Farm A and April to September for Wind Farm B. Both test periods contain periods of high, low, and variable wind speed, and results based on the shortest test dataset (Wind Farm A) covers the most challenging period for forecasters. An example density forecast at Wind Farm A using the parametric copula method is shown in Figure 5.2.

Generation data from individual turbine SCADA systems and the wind farm power export meter are used at 30-minute resolution with instances of curtailment flagged and excluded from the forecasting exercise. Data is also adjusted for availability so the impact of outages on evaluation results is minimised. NWP data from the European Centre for Medium-Range Weather Forecasts is extracted at the closest grid point to each wind farm from 0 to 48 hours ahead in hourly intervals, with 2 issue times per

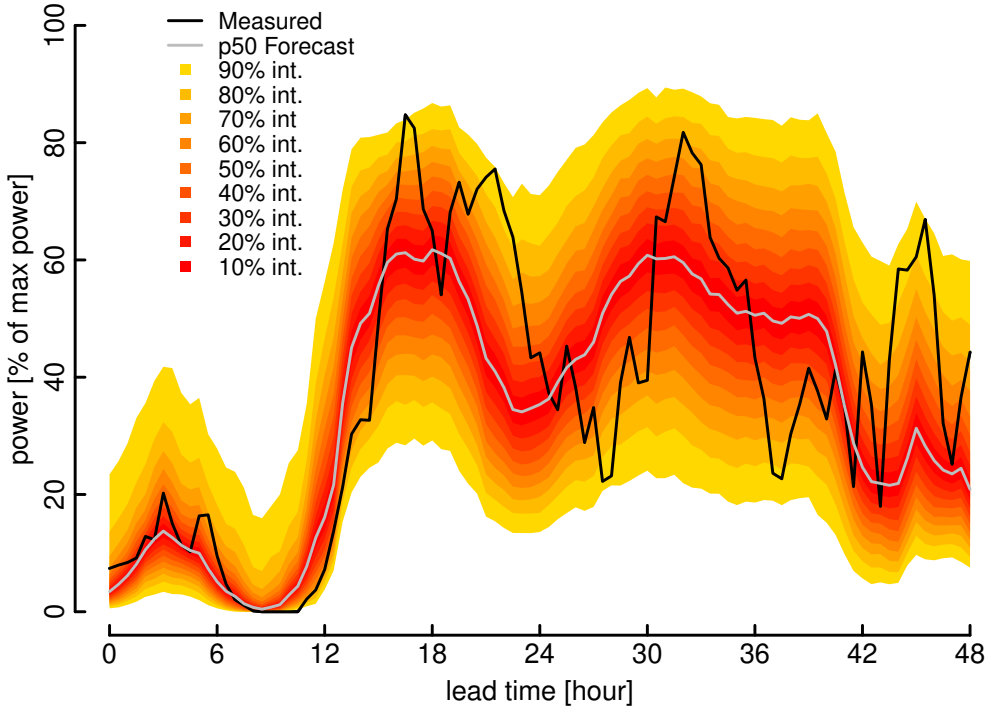


Figure 5.2: Example density forecast using the parametric Gaussian copula approach at Wind Farm A

day. Linear interpolation is used to match the resolution of the hourly forecasts and half hourly power data. The methodologies described are implemented in `R` using the packages `glmnet`, `VineCopula`, `kknn`, and `gbm` [170–173].

5.3 Results

The hyper-parameters of the GBT and AnEn models considered here are tuned in order to minimise CRPS, subject to reliability. However, it is beneficial to tune hyper-parameters for different quantiles separately. Here, we produce 19 GBT models for quantiles from 0.05 to 0.95 in steps of 0.05. To minimise the burden of hyper-parameter selection, only hyper-parameters for the 0.05, 0.3, 0.5, 0.7, and 0.95 quantiles are optimised and then used for neighbouring quantiles. The shrinkage and tree depth hyper-parameters are selected using k-fold cross validation and a grid search of the parameter space on the training data. The number of trees is kept constant at 500, as is the

minimum number of observations in each terminal node at 30, and the bag fraction at 75%. For the AnEn benchmark, the number of members in the ensemble is selected by minimising the CRPS on the training data via k-fold cross validation.

For VCop, C-vine and R-vine structures were both tested. The C-vine, which uses a star shaped configuration for each tree in the vine to connect the bivariate copulas, consistently provided lower error metrics than the R-vine structure, so only results from that structure are detailed here for brevity. The reason for this difference could be that the R-vines, which are a hybrid of the star-shaped (C-vine) and simple path-shaped (D-vine) dependency [169], are simply too complex and over-parameterised. The performance of the parametric covariance model, elaborated below, supports the suitability of simpler and regularised dependency structures. However, vine-copula models are not the sole-focus of this chapter and more work is needed to verify the reason for the difference in skill between the vine structures. Each bivariate copula is selected using the AIC on the training data and then used to produce forecasts on the test data. Full details of copula family selections are given in Appendix B.

At Wind Farm A, all of the proposed methods show improvements over the two benchmarks across all lead times. The CRPS and improvement over benchmark metrics at Wind Farm A are detailed in Table 5.1. The feature engineering method reduces CRPS by 3.95% and 5.46% compared to direct wind farm-level forecasting using WF($\mathbf{x}_t$) and AnEn respectively. The only difference between the WF($\mathbf{x}_t$) benchmark and this method is the incorporation of features derived from turbine-level information. The copula-based methods also consistently outperform the benchmarks, and the Gaussian copula with parametric covariance matrix give the best performance of all models across all lead-times with reductions of 5.01% and 6.50% over WF($\mathbf{x}_t$) and AnEn respectively.

The calibration plots in Figure 5.3 reveal that the turbine-level feature engineering and copula methods also marginally improve the reliability of the forecast compared to the WF($\mathbf{x}_t$) benchmark, and that these methods are all well calibrated, indicating that reductions in CRPS are mainly due to increased sharpness.

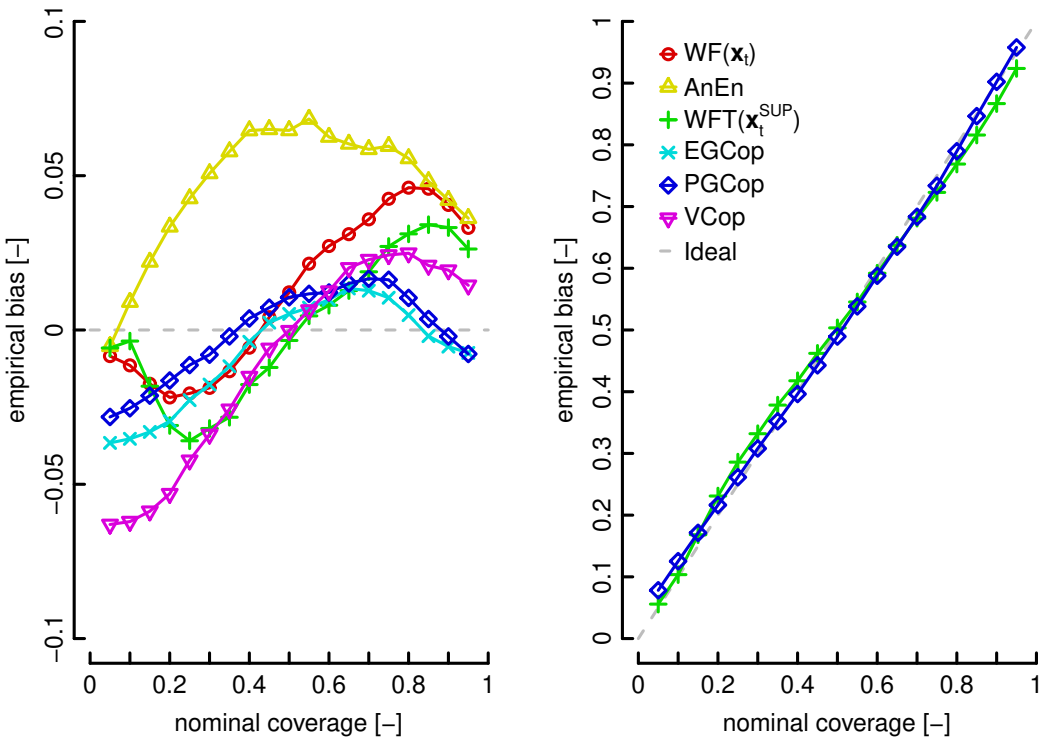
At Wind Farm B, as detailed in Table 5.2, all proposed methods outperform the benchmarks, though to a lesser extent than Wind Farm A. Unlike Wind Farm A, the

Table 5.1: Results at Wind Farm A. %Δ indicates improvement compared to specified benchmark [CRPS in % of max power]

Lead time	Benchmarks		Feature Engineering			Gaussian Copula, Empirical Cov.			Gaussian Copula, Parametric Cov.			Copula Vine		
	WF($\mathbf{x}_t$) [30]	AnEn [78]	WFT($\mathbf{x}_t^{\text{SUP}}$)	WF($\mathbf{x}_t$)	AnEn	EGCop	WF($\mathbf{x}_t$)	AnEn	PGCop	WF($\mathbf{x}_t$)	AnEn	VCop	WF($\mathbf{x}_t$)	AnEn
Hour	CRPS	CRPS	CRPS	%Δ	%Δ	CRPS	%Δ	%Δ	CRPS	%Δ	%Δ	CRPS	%Δ	%Δ
0-6	5.23	5.38	5.02	3.93	6.55	5.00	4.49	7.09	**4.93**	5.75	8.31	4.96	5.13	7.71
7-12	6.10	6.22	5.89	3.36	5.25	5.88	3.58	5.47	**5.81**	4.70	6.56	5.88	3.52	5.41
13-18	6.08	6.17	6.04	0.57	2.01	5.92	2.48	3.90	**5.86**	3.50	4.91	5.93	2.37	3.79
19-24	6.89	7.03	6.53	5.18	7.07	6.53	5.20	7.09	**6.47**	6.09	7.97	6.55	4.88	6.78
25-30	7.16	7.24	6.87	3.95	5.00	6.89	3.79	4.84	**6.84**	4.48	5.53	6.92	3.26	4.32
31-36	7.76	8.06	7.43	4.28	7.81	7.41	4.55	8.08	**7.35**	5.30	8.79	7.46	3.88	7.43
37-42	8.07	8.07	7.73	4.25	4.14	7.69	4.76	4.65	**7.65**	5.27	5.16	7.75	4.08	3.96
43-48	8.90	8.92	8.45	5.12	5.35	8.51	4.40	4.63	**8.47**	4.84	5.07	8.59	3.47	3.70
All	7.01	7.12	6.74	3.95	5.46	6.72	4.21	5.71	**6.66**	5.01	6.50	6.75	3.81	5.32

Table 5.2: Results at Wind Farm B. %Δ indicates improvement compared to specified benchmark [CRPS in % of max power]

Lead time	Benchmarks		Feature Engineering			Gaussian Copula, Empirical Cov.			Gaussian Copula, Parametric Cov.			Copula Vine		
	WF($\mathbf{x}_t$) [30]	AnEn [78]	WFT($\mathbf{x}_t^{\text{SUP}}$)	WF($\mathbf{x}_t$)	AnEn	EGCop	WF($\mathbf{x}_t$)	AnEn	PGCop	WF($\mathbf{x}_t$)	AnEn	VCop	WF($\mathbf{x}_t$)	AnEn
Hour	CRPS	CRPS	CRPS	%Δ	%Δ	CRPS	%Δ	%Δ	CRPS	%Δ	%Δ	CRPS	%Δ	%Δ
0-6	7.14	7.23	**7.01**	1.77	3.06	7.03	1.46	2.75	7.02	1.69	2.98	7.02	1.64	2.93
7-12	7.13	7.17	**7.05**	1.13	1.74	7.10	0.40	1.01	7.08	0.73	1.34	7.10	0.48	1.09
13-18	7.91	8.06	**7.78**	1.71	3.47	7.82	1.13	2.91	7.81	1.27	3.04	7.83	0.98	2.76
19-24	7.48	7.55	**7.38**	1.26	2.23	7.46	0.23	1.22	7.44	0.55	1.53	7.47	0.15	1.14
25-30	8.67	8.80	**8.56**	1.32	2.82	8.58	1.01	2.53	8.57	1.10	2.62	8.61	0.68	2.20
31-36	8.30	8.37	**8.19**	1.35	2.14	8.23	0.88	1.68	8.21	1.09	1.88	8.25	0.66	1.45
37-42	9.38	9.50	**9.28**	1.06	2.32	9.29	0.95	2.22	9.29	0.91	2.17	9.33	0.55	1.82
43-48	8.90	8.99	**8.86**	0.42	1.41	**8.86**	0.44	1.43	**8.86**	0.47	1.46	8.89	0.08	1.07
All	8.10	8.20	**8.00**	1.24	2.39	8.04	0.82	1.98	8.02	0.97	2.13	8.05	0.62	1.81



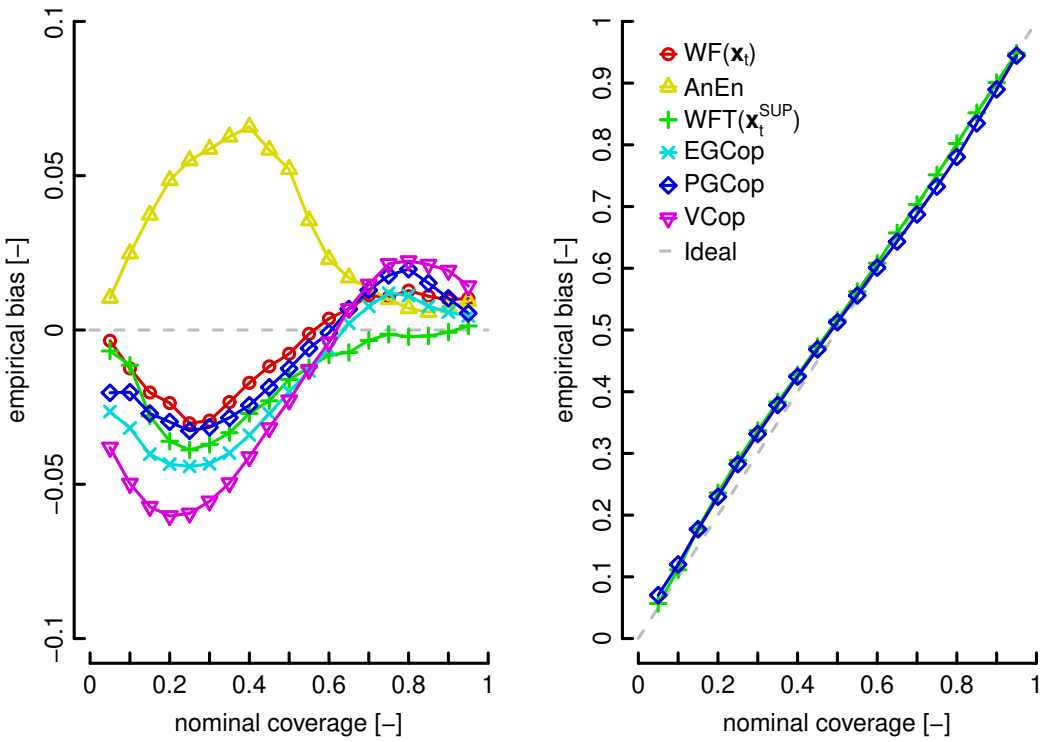
(a) Quantile bias

(b) Reliability of selected methods

Figure 5.3: Wind Farm A calibration plots

feature engineering approach provides the greatest improvement reducing CRPS by 1.24% and 2.39% compared to the WF($\mathbf{x}_t$) and AnEn benchmarks respectively. This improvement is also consistent across lead-times. The quantile bias plots, shown in Figure 5.4a, illustrate that the model calibration is slightly diminished when compared to the WF($\mathbf{x}_t$) benchmark from the 15th-60th percentile, but otherwise provides improvement outside this range. The reliability diagram in Figure 5.4b reveals that the proposed models are well calibrated and that variations between the models are small.

Bootstrapping [139] is used here to estimate the uncertainty of evaluation results. The CRPS values from the test datasets are re-sampled with replacement (number of samples equal to the size of the test dataset) and averaged 1000 times in order to estimate the sampling variation of the average scores in Tables 5.1 and 5.2. The results of this process are presented via boxplots in Figure 5.5 and show that improvement in CRPS compared to benchmarks is pronounced at both sites.



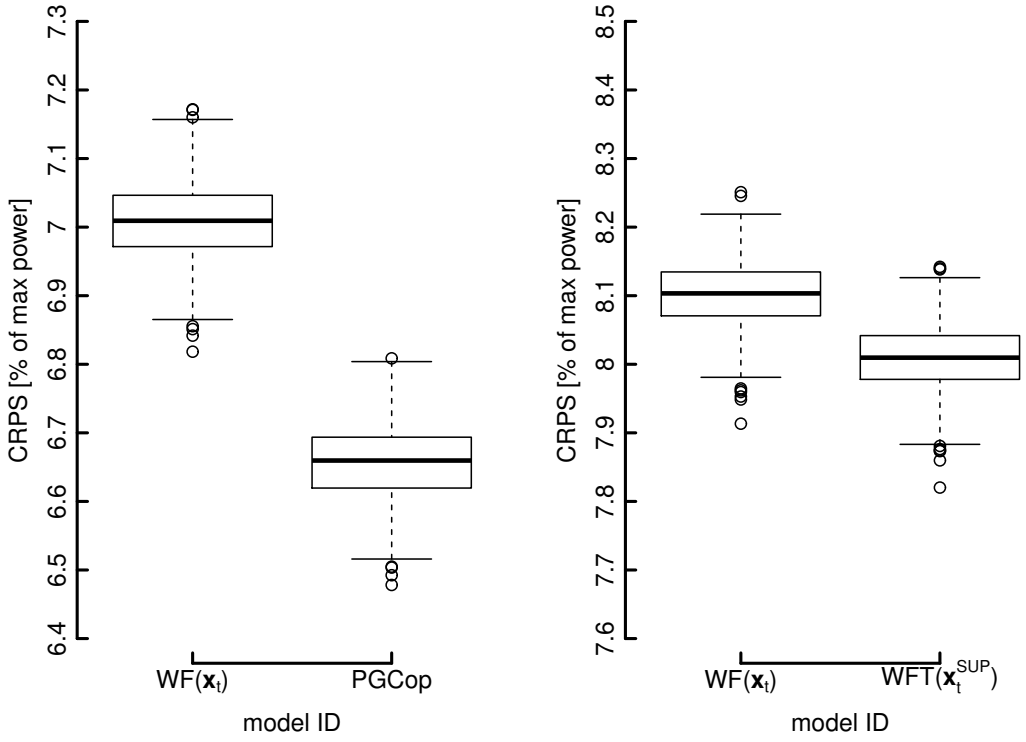
(a) Quantile bias

(b) Reliability of selected methods

Figure 5.4: Wind Farm B calibration plots

Comparing the copula methods at both wind farms, the Gaussian copula with parametric covariance matrix produces forecasts with lower CRPS and superior calibration, supporting parametrisation of the covariance matrix to produce a smooth spatial dependency structure. The more detailed and flexible dependency structure of the copula vine does not lead to further improvements in the forecast skill, and neither does the Gaussian copula with empirical covariance suggesting that both of these models are over parametrised given the volume of training data. The calibration of the vine copula in particular is poor compared to the WF($\mathbf{x}_t$) benchmark, which is again reflective of this over-fitting issue.

The regular layout of turbines at Wind Farm B is evident in the covariance matrix for that wind farm, shown in Figure 5.6 and the layout of the farm can be found in Appendix B. The block pattern is consistent with the evenly spaced rows of turbines. The covariance is relatively high across the wind farm with only 6% of values below



(a) Wind Farm A (b) Wind Farm B

Figure 5.5: Boxplots showing the bootstrap sample distributions of mean CRPS for the best benchmark and proposed model at both wind farms

0.7, which implies that there is little information to be gained by considering individual turbines as forecast errors are very similar across the site. At Wind Farm A, as shown in Figure 5.7, the covariance structure is more complex because of the wind farm's irregular layout and terrain; covariance is high within small areas of the wind farm but weak between regions.

Deterministic forecast performance is summarised in Table 5.3. The median (p50 in Figure 5.2) of each predictive distribution is taken as the deterministic forecast and evaluated in terms of Mean Absolute Error. As expected, the behaviour of the results is very similar to the probabilistic case; at Wind Farm B, which is characterised by a relatively simple layout and terrain, a modest improvement in forecast accuracy is achieved by using turbine level information, whereas at Wind Farm A the parametric Gaussian Copula results in a significant improvement over the benchmark models.

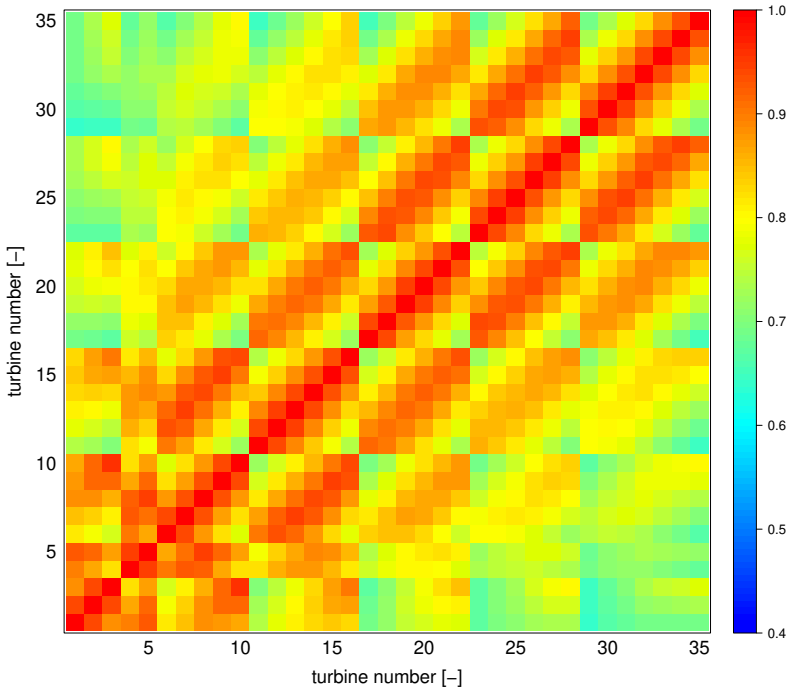
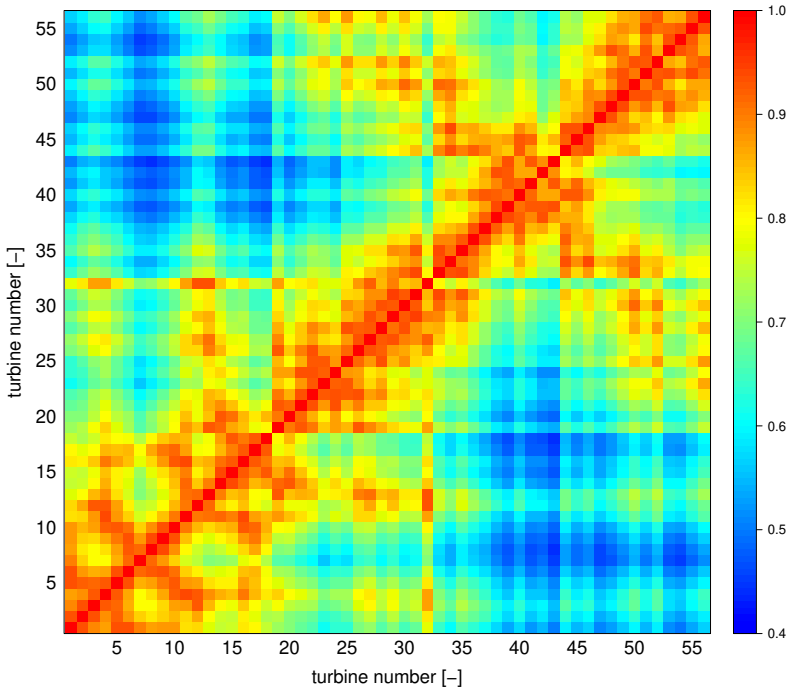


Figure 5.6: Parametric covariance matrix at Wind Farm B

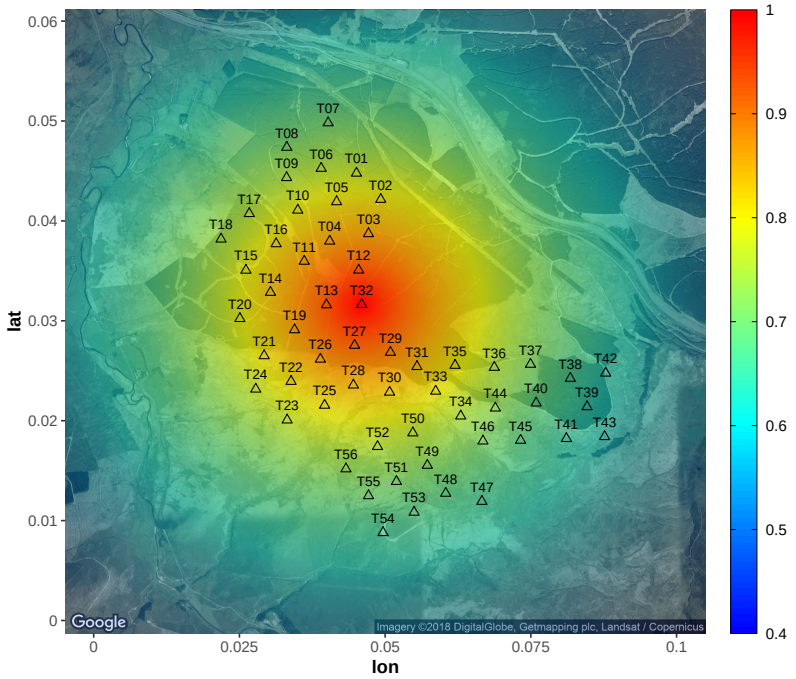
Performance evaluations separated by forecast lead time and in terms of Root Mean Square Error are available in Appendix B.

One feature of the bottom-up probabilistic method is the extended computational time required to train all the models. In this study, with a desktop computer (8 virtual cores, 3.6GHz CPU, 16GB RAM) it takes approximately 10.5 minutes to fit the required 19 quantile regression models using parallelization. This is the length of the model training phase for the WF($\mathbf{x}_t$) benchmark. The feature engineering method will take 10.5 minutes plus an additional 3.5 minutes multiplied by the number of turbines. The bottom-up hierarchical method training duration is 10.5 minutes multiplied by the number of turbines. However, significant additional time is required to determine the vine copula structure. Operationally the time required to issue a forecast is negligible for all but the VCop method and re-training models would be required infrequently.

The case study results indicate that turbine-level data can be leveraged to improve forecast skill, although the characteristics of the wind farm also have a bearing on the performance of the different methods. At a site with simple layout where the response



(a) Covariance matrix



(b) Partial covariance of turbine 32 [2]

Figure 5.7: Parametric covariance plots at Wind Farm A. Note that the latitude and longitude scales of (b) are indicative

Table 5.3: Deterministic forecast performance based on the median (p50) of each predictive distribution — %Δ indicates improvement compared to specified benchmark [MAE as % of max power]

Model	Wind Farm A			Wind Farm B		
	MAE	%ΔWF($\mathbf{x}_t$)	%ΔAnEn	MAE	%ΔWF($\mathbf{x}_t$)	%ΔAnEn
WF($\mathbf{x}_t$) [30]	9.69	–	–	11.39	–	–
AnEn [78]	9.88	–	–	11.49	–	–
WFT($\mathbf{x}_t^{\text{SUP}}$)	9.27	4.25	6.09	**11.21**	1.61	2.41
EGcop	9.16	5.47	7.28	11.26	1.12	1.92
PGcop	**9.11**	5.92	7.72	11.26	1.12	1.92
Vcop	9.19	5.09	6.92	11.27	1.09	1.90

of all turbines to the weather is similar, and therefore forecast errors are similar, only a modest improvement in forecast skill is realised by considering turbine-level information. In this situation there is no advantage in modelling the full spatial dependency structure between forecast errors at individual wind turbines; it is sufficient to supplement a conventional forecasting method with turbine-level features. However, at a complex site modelling the spatial covariance structure provides greater improvement — 5% greater in this case study — than feature engineering alone.

Importantly, these improvements come at very low cost. Turbine-level SCADA data is routinely collected and stored by operators, and only modest computational power is required to realise the benefits of the methods proposed here. Furthermore, turbine-level data is only required for training, not in real-time operation, so there is no need for new communications or data feeds, and third party forecast providers could enhance their forecasts for individual wind farms with a static dataset of historic turbine-level data. Importantly, the proposed framework is not constrained to GBTs as these can be readily substituted with any other method of producing density forecasts.

5.4 Conclusions & Future Work

Turbine-level data provides valuable information about how a wind farm responds to different weather conditions, and the nature of forecast errors, which is not accessible when only considering a wind farm's total power production. Two methods for improving wind power forecasting by leveraging data from individual wind turbines are

evaluated. The first is a feature engineering approach whereby deterministic forecasts for individual turbines are aggregated and used as supplementary input variables to a conventional wind farm-level model. The second is a novel bottom-up probabilistic approach which forecasts the joint predictive distribution of generation from all turbines in a copula framework, which is then used to produce a wind farm-level forecast.

Both methods are shown to increase forecast skill compared to two highly competitive benchmarks, particularly at the site with complex terrain. At Wind Farm A, the Gaussian copula method with parametric covariance matrix reduces CRPS by 5% compared to the best performing benchmark while the feature engineering approach provides a 4% improvement. At Wind Farm B, both methods improve forecast skill by approximately 1%. These improvements come at almost no cost as turbine-level data is routinely recorded by SCADA systems and this data is only required for training forecast models; no additional communications or data flows are required operationally. Therefore, both utilities producing in-house power forecasts and third party forecast providers could enhance their forecast performance using a static dataset of turbine-level data.

Future work should explore the benefits of turbine-level data in spatio-temporal forecasting and the dynamic evolution of covariance structures. For example, Pinson et al. [69] propose an adaptive update scheme to track slow changes in temporal covariance, but fast changes require dependency structures to be conditional on suitable explanatory variables or regimes. The turbine level probabilistic forecasts would also benefit from a time-adaptive model which would enable better tracking of adjustments in the power curves; this could be achieved by regular re-training of the multiple quantile regression models in a rolling window style framework or via an online update methodology [174]. The latter would require a stream of operational data which could be very valuable, however would present different challenges such as dealing with missing or poor quality data in real-time; the very-short term forecasting literature provide potential avenues to address this challenge [175].

Chapter 6

Probabilistic Access Forecasting for Improved Offshore Operations

Chapter based on:

Gilbert, C., Browell, J., & McMillan, D. (2021). Probabilistic access forecasting for improved offshore operations. International Journal of Forecasting, 37(1), 134-150.

In the previous chapter we introduced copulas in a hierarchical wind power forecasting framework to generate a multivariate spatial distribution. Here, we instead focus on temporal correlations of forecast errors in the pursuit of statistical scenario forecasts for a different application: offshore access forecasting. This chapter involved modelling a high dimensional input feature space via feature engineering for the marginal forecasts, as well as high dimensional multivariate distribution for generating temporal scenario forecasts. Improving access is a priority in the offshore wind sector, driven by the opportunity to increase revenues, reduce costs, and improve safety at operational wind farms. This chapter describes a novel methodology for producing probabilistic forecasts of safety-critical access conditions during crew transfers.

As discussed in Section 2.4.4, access prediction is an emerging application in energy forecasting. This is because operators are looking to improve safety around the transfer decisions, and the industry-wide drive to operate more efficiently. The ability to access turbines is dominated by the local wave climate, and sea-state forecasting plays a major

role in the current scheduling practices in offshore wind, which is typically based on deterministic forecasts of significant wave height.

The contribution of this chapter is in providing an end-to-end framework for generating access forecasts based on vessel monition during transfer up to 5 days ahead, including quantifying the uncertainty due to weather conditions. To this end, a method is developed to produce probabilistic forecasts of significant wave height and peak wave period using statistical post-processing of NWP. Temporal and cross-variable dependency is modelled in a copula setting to generate scenario forecasts. These are converted to vessel specific forecasts using a data-driven vessel motion model which captures the displacement of the vessel during transfer. An option for transformation of the vessel motion forecast into an 'access score' is also presented. This score enables simple visualisation and communication of uncertainty information for decision-makers. A flowchart summarising the entire modelling chain in both training and operation phases is shown in Figure 6.1.

This chapter is organised as follows: Section 6.1 covers some background on main-

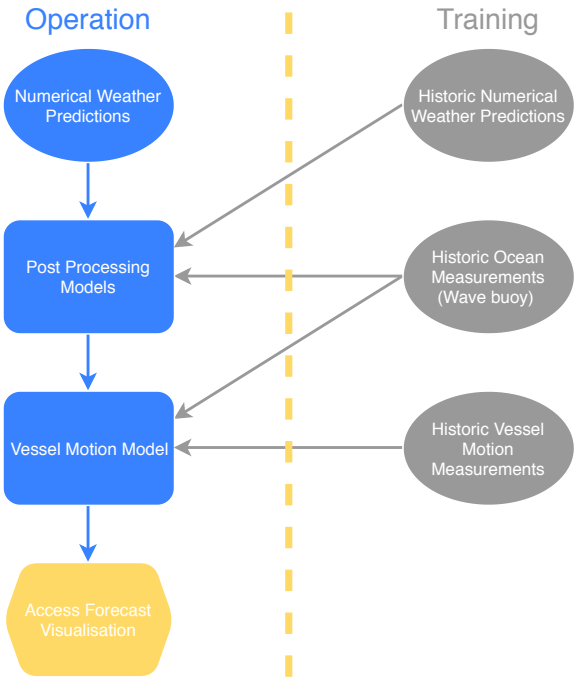


Figure 6.1: Flowchart of the modelling chain in operation and training

tenance access and the metocean environment, Section 6.2 details the forecast post-processing method, Section 6.3 describes the data driven vessel motion model, followed by Section 6.4 where results are presented and discussed for a UK offshore wind farm. Future work is outlined and conclusions are drawn in Section 6.5.

6.1 Access in the Offshore Environment

Typically, vessel dispatch is managed by a marine coordinator in a control room at the operations base. This work aims to innovate in the decision space where coordinators makes the scheduling/dispatch decision depending on the weather forecast, any available live measurements, accrued experience of the site, and the list of work orders. This schedule is typically made first thing in the morning and then updated at night for the next day accounting for the completed work, any new tasks, and updated weather forecasts. As the day progresses the marine coordinator has to deal will deviations as a result of turbine inaccessibility, technician sea-sickness, or delays; these first two issues are clearly partly due to metocean forecast errors. Innovations in probabilistic access forecasting are then useful for both marine coordinators and schedulers.

Crew transfer referrers to the process of transferring an individual from a vessel to an offshore structure. In the offshore wind industry, it is routine practice for technicians to transfer from dedicated Crew Transfer Vessel (CTV) to a wind turbine via a specially designed ladder which the CTV pushes up against. CTVs are equipped with a rubber fender shaped to fit the ladder, and the vessels propulsion system is used to create friction between the fender and the ladder in order to stabilize the vessel. Once stable, the crew member may proceed with the transfer and climb the ladder to the wind turbine. Safety is paramount in this dangerous environment and the individual transferring, the vessel master, and marine coordinator all have the power to stop operations if they are deemed unsafe. In contrast, it is common for contractual levers to be in place between asset owners and operators specifying a significant wave height threshold below which CTVs are expected to attempt transfers [176].

In order to plan and execute maintenance operations, including crew transfer, forecasts of the sea state are then required. In practice, these typically comprise of signifi-

cant wave height at 1- or 3-hour intervals for a single location in space, representative of the wind farm, for the next 48 hours. In this work three crucial environmental factors for access quality are forecast: significant wave height, peak wave period, and peak wave direction at 1 hour intervals up to 5 days ahead. The spatial resolution is for a single location in space representative of the wind farm. Other factors, such as lightning risk and visibility, are reserved for future work, as well as incorporating information from wave spectra.

6.2 Sea State Forecasting Methodology

Here, the method for generating scenarios of significant wave height and peak wave period are detailed. The method for post-processing wave direction for the vessel motion model is also described. The NWP outputs used here include significant wave height, peak wave period, and mean wave direction. It is common practice in contemporary regression models to engineer additional features and use cross-validation for algorithm-specific parameter tuning. For example, the NWP error characteristics suggest that including leading and lagged lead-times will be beneficial.

We also consider rolling averages, rolling variances, diurnal, and seasonal effects. However, the significantly increased dimensions of the input space necessitates feature selection techniques and/or regularisation, which is discussed in the proceeding sections. All engineered temporal features are calculated per issue time. A full list of features used in each model can be found in Appendix C.

6.2.1 Parametric & Non-Parametric Regression

Three methods are considered for producing density forecasts of significant wave height or peak wave period. All are based on post-processing NWP, i.e. learning the relationship between historical observations and concurrent weather predictions. One non-parametric and two parametric density forecasting methods are compared. Parametric techniques assume that the predictive distribution follows a parametric distribution and the forecasting task is to predict the parameters of that distribution, whereas

non-parametric techniques make no assumptions of this sort. Here, gradient boosting machines are used for quantile regression, and two variations of generalised additive models for location, scale, and shape are used to produce parametric density forecasts, which are made distinct by the method of model fitting.

Gradient Boosting Machines

The use of gradient boosting machines, discussed in Section 4.2.2, as a statistical learning technique is motivated by the success of this algorithm in the energy forecasting arena [30,31]. The predictive distribution is constructed for each variable and lead-time from multiple quantile forecasts (probability levels: 0.01, 0.05, 0.1,..., 0.95, 0.99) using cubic spline interpolation with knots at each predicted quantile and fixed boundaries at zero and the maximum value observed in the training data. Monotone cubic spline interpolation is used to guarantee a valid Cumulative Distribution Function (CDF) at each lead time [177]. An example predictive CDF of significant wave height for a single lead time is shown in Figure 6.2 using the multiple quantiles and spline interpolation method, as well as a parametric predictive CDF which is discussed in the following section.

Modelling extreme quantiles, those in the tails of the predictive distribution, is challenging due to high estimation error of quantile regression at the extremes and motivates the use of extreme value theory (e.g. Pareto-type tails) and parametric methods. Here we found negligible difference between the described approach and use of Generalised Pareto tails in terms of calibration, which require an additional parameter to be estimated and does not remove the need to impose boundaries. Further work is necessary to generate sharp and calibrated non-parametric models for the tail of the distribution of metocean variables, which may well be more suited to sites with different tail characteristics than the case study considered here. Recent advances in wind power forecasting provide a suitable starting point for such analysis [178].

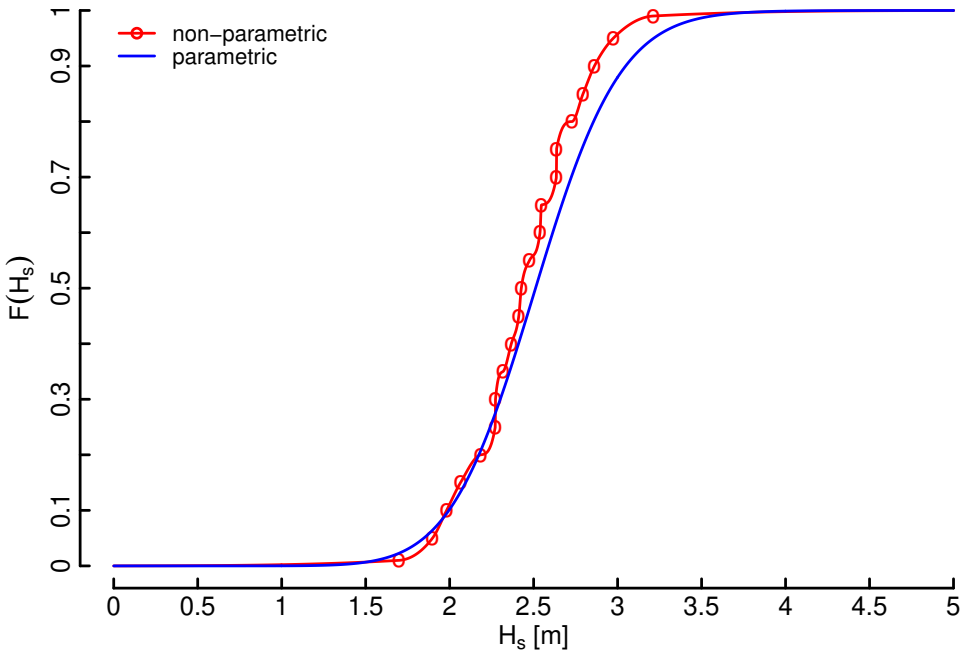


Figure 6.2: Example predictive CDF for significant wave height at a single lead time, based on either multiple quantile regression or distributional regression. In the latter, the conditional distribution family is the Generalised Beta Prime

Generalised Additive Models for Location, Scale, and Shape

Since the main objective of the sea-state forecasting stage is to produce scenario forecasts, parametric regression models are considered because the tails of the distribution are well defined compared to quantile regression, where tails require special treatment. The tails of the distribution have a large impact on dependency structure estimation and scenario forecast production. Additionally, the full distribution is described by fewer parameters. To this end, two variations of generalised additive models for location, scale, and shape are used. One uses maximum likelihood to optimise the model fit and the other uses boosting; these are termed the `gamlss` [179] and `gamboostLSS` [180, 181] models respectively. The use of boosting in this case is employed to allow for the use of feature engineering as discussed above.

`Gamlss` models are termed 'semi-parametric' models, because a parametric distribution is assumed for the target variable and the parameters of that distribution may include non-parametric smoothing functions of explanatory variables — this should not

be confused with non-parametric probabilistic forecasts in the form of quantiles. This framework is an extension of the more familiar generalised additive models [156] in that any parameter of the distribution can be a function of the explanatory variables, not just the conditional mean.

If we have observations y, in this case significant wave height or peak wave period, the conditional density typically $f_d(y|\boldsymbol{\theta})$ depends on up to four parameters; these are the location (θ_1), scale (θ_2), and shape parameters (θ_3, θ_4). Distributions with less than four parameters are supported. Note that the time index of the observation from the above is dropped to avoid notational clutter. So, an additive regression predictor η_{θ_i} is generated for each distribution parameter θ_i for $i = 1, \ldots, 4$. Let $\boldsymbol{x}_i$ be the pool of N_i explanatory variables in the sub-model for θ_i, and $g_i(.)$ the link function, then the model formulation of `gamlss` is

$$g_i(\theta_i) = \eta_{\theta_i} = \beta_{0\theta_i} + \sum_{n=1}^{N_i} f_{n\theta_i}(x_{i,n}), \quad i = 1, \ldots, 4 \tag{6.1}$$

where the function $f_{n\theta_i}$ is the effect of explanatory variable n on the distribution parameter θ_i. This can be linear or non-linear effects such as penalised splines; $\beta_{0\theta_i}$ are the intercepts of each sub-model. Typically, these models are fitted iteratively using a combination of maximum likelihood, transformation of distribution parameters $\boldsymbol{\theta}$ using the inverse link function, and successive back-fitting of the predictor functions in each sub-model η_{θ_i} [179].

However, when $\boldsymbol{x}_i$ becomes large, feature selection procedures should be carried out to avoid over-fitting and the computational expense of repeated model estimation for feature selection can increase significantly. Model fitting based on component-wise gradient boosting is an attractive solution to this problem [181]. Formally, given y_t observations and $\boldsymbol{\eta_t}$ additive predictors of the four sub-models, the `gamboostLSS` [180, 181] algorithm minimises the loss function $L(\cdot)$ i.e the negative log likelihood

$$\frac{1}{T}\sum_{t=1}^{T} L(y_t, \eta_{\theta_1 t}, \eta_{\theta_2 t}, \eta_{\theta_3 t}, \eta_{\theta_4 t}) = \frac{1}{T}\sum_{t=1}^{T} L(y_t, \boldsymbol{\eta}_t) \quad . \tag{6.2}$$

Similarly to the `gamlss` model, for each distribution parameter a set of base learners $h_{i,n}(\cdot)$ (penalised splines, cyclical splines, linear effects etc.) are specified for each explanatory variable, and the model formula can be different for each predictor. Where $L(\cdot)$ is differentiable, the vector of the negative gradient $\boldsymbol{r}_i$ is defined as

$$\boldsymbol{r}_i^{[m-1]} = -\left[\frac{\partial L(y_t, \boldsymbol{\eta}_t)}{\partial \eta_{\theta_i}}\right]_{t=1,\ldots,T} \tag{6.3}$$

where the boosting iteration is m and $\boldsymbol{\eta}_t = \hat{\boldsymbol{\eta}}_t^{[m-1]}$ are the current estimates of the additive predictors. To begin the algorithm, additive predictors $\hat{\eta}_{\theta_i}^{[0]}$ are initialised with offset values. The base learners $h_{i,n}(\cdot)$ are fit to this negative gradient and *only the best base learner* (n^*), according to the least squares error, is used to update the additive predictor

$$\hat{\eta}_{\theta_i}^{[m-1]} := \hat{\eta}_{\theta_i}^{[m-1]} + \lambda h_{i,n^*}(\cdot) \tag{6.4}$$

where λ is a shrinkage parameter, or step-length, which is included for regularisation. The additive predictor $\hat{\eta}_{\theta_i}^{[m]}$ is then set equal to $\hat{\eta}_{\theta_i}^{[m-1]}$ and the process is repeated for the remaining θ parameters in this boosting iteration. Following this, the boosting process is repeated until the user specified $m = m_{stop}$ is reached. Therefore, k-fold cross-validation is used in this case to tune the total number of boosting iterations and the value of the shrinkage term. This process is known as component-wise gradient boosting, enabling an intrinsic feature selection capability (as some features are never the *best learner* n^* and therefore do not form part of the model), which performs well with high dimensional input data.

For both parametric additive models explored, the selection of the base learner is important, and types of learners available are very similar. Taking the `gamlboostLSS` model notation, the typical base learner $h_{i,n}(\cdot)$ specification in this case is a penalised B-spline (i.e. the p-spline), with cyclical splines used for direction variables, and a bivariate p-spline for the seasonal terms — time-of-day and day-of-year — to include the smooth interaction of these variables. Taking the most commonly used base learner

as an example, the p-spline is defined as

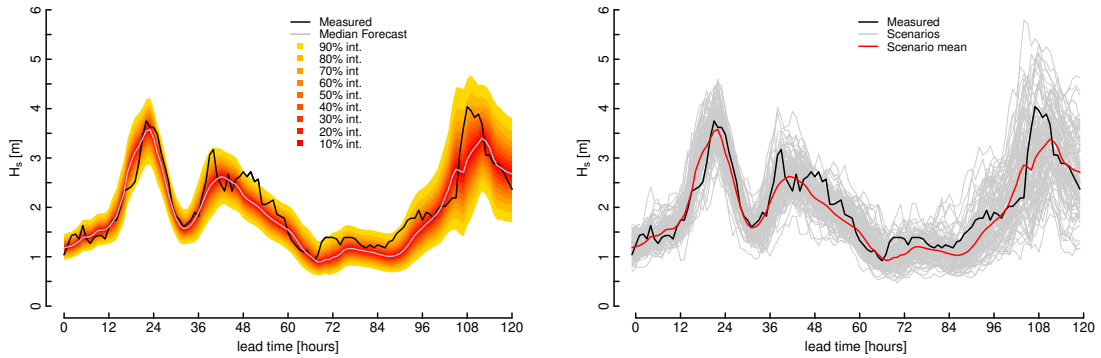
$$h_{i,n}^{\text{p-s}}(x_{i,n}) = \sum_{k=1}^{K} a_k B_k(x_{i,n}) \tag{6.5}$$

where the kth B-spline basis function is $B_k(x_{i,n})$, and a_k are the associated spline coefficients. The coefficients are estimated with penalisation to enforce a degree of smoothness to the fit [182]. The exact method of penalised coefficient estimation varies across the two implementations tested here; the `gamboosLSS` implementation is expanded by [183], with information on the constrained cases, such as circular variables. Various penalised spline implementations are explored and compared by [184], along with details on the `gamlss` procedure. For a full description of each model formulation used in the entire analysis, the readers are refereed to the supplementary material.

A key component of distributional regression is choosing an appropriate conditional distribution family for the target variable. In the case of wave height and period regression the distribution should support values on the positive real line. A number of candidate families meeting this criteria are tested and the best performing candidate identified by evaluating resulting forecasts in a cross-validation exercise. An example predictive CDF of significant wave height using a candidate distribution family is shown in Figure 6.2 for a single lead time, and an example density forecast is shown in Figure 6.3a.

Benchmark Models

Two benchmark methods are included as both a 'naive' and 'smart' comparison. In both cases the target variable is related to a corresponding single input from the NWP source, e.g. significant wave height to significant wave height, and the target variable is assumed to follow a Gamma distribution, as this was found to be a competitive model during exploratory analysis and in related work [185]. These models are implemented in `gamlss`, in which the variance of the distribution is also influenced by the mean due to the parameterization. Therefore, the shape of the distribution is not constant and both models are very competitive. The naive benchmark is a generalised linear model



(a) Density Forecast

(b) Scenario Forecast

Figure 6.3: Example probabilistic forecasts of significant wave height using `gamboostLSS` parametric regression. In 6.3a the intervals plotted cover specified probability level ranges, e.g. the 90%int. is the 5% — 95% quantile range

with a single linear base learner, this is termed the *benchmark — glm*. The second benchmark is a simple generalised additive model with a single penalised spline as the base learner, termed *benchmark — gam*.

6.2.2 Scenario Forecasting

Where probabilistic forecasts are used for multi-temporal decision making, scenario forecasts are required [36]. Here, the Gaussian copula, introduced in Section 5.1.3, is used although the marginals of the copula are now density forecasts separated by lead time, i.e. this is now temporal dependency rather than spatial. An example scenario forecast of significant wave height based on this technique is shown in Figure 6.3b, where the temporal scenarios are those defined in Equation 5.8.

Here three configurations of the dependency are tested: 1) *Independence* — the benchmark where no temporal correlation is embedded in the high dimensional dependence, 2) *Linked* — the full temporal inter-dependency between significant wave height and peak wave period is modelled across the lead times, and 3) *Temporal* — the dependency is modelled for each variable separately across the forecast lead times. The *linked* case is motivated by the idea that the uncertainty is linked between the variables, because they are summary statistics from the same source wave spectrum forecast; again the cross-variable dependency matrix is simply estimated using the sample covariance

matrix of the transformed normally distributed variables.

6.2.3 Wave Direction Regimes: Clustering & Logistic Regression

Wave direction can have a significant impact on vessel motion and on the characteristics of NWP errors, particularly if wave direction is associated with different physical processes. This section describes the wave direction post-processing strategy. For the purpose of the vessel motion model, a small number of distinct directional regimes are considered rather than incorporating direction as a continuous variable.

Peak wave direction is incorporated by first clustering the wave buoy measurements into two distinct regimes, motivated by the fact that the wave climate at the case study location is dominated by locally driven wind waves or waves from the swell, though this technique could easily be extended to wave climates with more than two distinct regimes. Logistic regression is then used with NWP to predict the cluster membership at any forecast lead time. Therefore, wave direction prediction is simplified into a straightforward classification problem. Note, here we are using a forecasting model to predict regime membership based on a clustering model. Alternatively, it is possible to directly use distance metrics from the NWP forecasts as they become available and the cluster centres based on the buoy measurements. However, this would not account for biases present between the site measurements and NWP forecasts. Comparing this alternative approach is reserved for future work.

To cluster the measured variables, we define the input space $z_t = (\omega_{p_t}, T_{p_t}, H_{s_t})$, where the three environmental factors are peak wave direction, peak wave period, and significant wave height respectively. The k-means clustering algorithm is used to define the two regimes [156]. This algorithm generates disjoint regions R_k that collectively cover the input space spanned by z_t. Note that all input variables are scaled and the wave direction variable is linearised. Since, in this case there are only two regimes defined, logistic regression is used to determine the probability of regime membership. The gradient boosting machine described above is used as the logistic regression tool with inputs features engineered similar to the continuous target variable regression case. Again, a full list of input features for the wave direction regime forecasting

can be found in the supplementary material. For more information on the regression technique, please refer to [154].

6.3 Vessel Motion During Transfer

Accessibility is constrained by vessel motion during push-on and transfer. Therefore, in order to provide forecasts of accessibility to both wind farm and vessel operators it is necessary to forecast the sea conditions and understand how individual vessels will respond in those conditions. Here a data-driven approach to vessel motion modelling is undertaken.

Distributional models using generalised additive models for location, scale and shape, described in Section 6.2, are used to learn the relationship between met-ocean observations and the vessel motion data (heave peak-to-peak) during push-on instances. The main mode of movement during push-ons which impede transfers is vertical displacement of the vessel fender and the turbine transition piece due to the oncoming wave field, this can result in a 'slip' event which can have serious safety implications. This motivated the use of the heave motion of the vessel as a key transfer quality and safety indicator. The other degrees-of-freedom can clearly have an impact on transfer quality, but this is reserved for future work.

In operation, forecasts from Section 6.2, i.e. the scenarios of wave height and period, as well as the forecast regime membership, are used as inputs to drive the vessel motion model; this process generates vessel-specific scenarios of motion during transfer. A visualisation stage, discussed briefly in Section 6.4 completes the forecasting process, as shown in Figure 6.1.

6.4 Case Study

The methodology is tested at an east coast offshore wind farm in the UK. Ocean measurements are collected from a Centre of Environment Fisheries and Aquaculture Science wave-buoy within the site boundary. NWPs of the wave climate from the European Centre for Medium-Range Weather Forecasts are extracted at the closest

grid point to the site from 0-120 hours ahead at hourly intervals, with 2 issue times per day.

Vessel telemetry data is from two purpose built offshore wind service vessels with the same specification: length 19.2m, width 8.2m, maximum draft 2m, passengers 12, aluminium catamaran; this data is collected during the construction phase of the wind farm, which contains around 700 push-on instances alongside concurrent wave buoy measurements. Transfer events are identified using the measured push-on force as well as time-stamped *swipes* from technicians' ID cards when transferring. The vessel's average peak-to-peak heave is determined during each push-on attempt and used in the following analysis. The time resolution is reconciled by matching transfer events to the closest buoy measurement in time.

For the regression problems the data is partitioned into 4.5 (January 2013 - June 2017) and 1 (July 2017 - June 2018) year(s) for training and testing respectively. This allows for sufficient data in the modelling of the copula and an entire year to evaluate the subsequent forecasts in out-of-sample tests. To make decisions on the best configuration for each forecasting task and to tune the algorithm specific hyperparameters, 4-fold cross-validation is used on the training data only. To refresh, the four main forecasting tasks are: 1) significant wave height regression, 2) peak wave period regression, 3) wave regime clustering & logistic regression, and 4) copula dependency modelling.

The method is implemented in `R` [170] using the package `ProbCast`, which is in development, although a 'beta version' is available with accompanying scripts for this methodology [186]. `ProbCast` is developed for the modelling, evaluation, and plotting of probabilistic forecasts, using `gbm`, `gamlss`, and `gamboostLSS` for the regression models [170, 179, 180].

Probabilistic density forecasts of significant wave height and peak wave period are evaluated according to the principle that the forecast should be optimally sharp subject to calibration [127]. This concept and the appropriate univariate and multivariate scoring rules are discussed extensively in Section 3.

For the univariate density forecasts, implementing quantile and parametric regression for density forecasting introduces a compromise when using evaluation metrics for

comparison. Here the Probability Integral Transform (PIT) is used to measure the calibration of the full distribution, motivated by the direct impact of this variable on the dependency structure, and the pinball loss score is used to measure the sharpness and calibration at discrete quantiles of the distribution averaged over all lead times.

It is also important to understand the growth in the uncertainty of the density forecasts as a function of lead time. Here, sharpness is tested in terms of average interval width plots [128]. For an interval with a nominal coverage rate of 1-β the interval size is

$$\delta_t^{(\beta)} = \hat{y}_t^{(1-\beta/2)} - \hat{y}_t^{(\beta/2)} \tag{6.6}$$

and this measure is averaged over all cases, grouped by each lead time. It is important to note that this measure of sharpness is a final illustrative layer to the forecast which is proven to be sharp and reliable via the pinball loss and PIT histogram respectively.

Scenario forecasts are evaluated via multivariate probabilistic forecast verification methods. Two metrics capable of evaluating the trajectories are the Energy Score (ES) and the p-Variogram Score (VS-p); both are evaluated per issue time of the forecast. For more information regarding these scores, the reader is referred to Section 3.2.2.

The regime classification forecasts are evaluated in terms of the Area Under the Receiver Operator Curve (AUROC) [187]. In evaluating logistic regression, the ROC curve plots the true positive rate versus the false positive rate for different threshold values at which the predicted probability is cut to define the two prediction classes. The maximum area AUROC is equal to one, and optimal forecasts are as close to this value as possible.

6.4.1 Sea State Forecasting

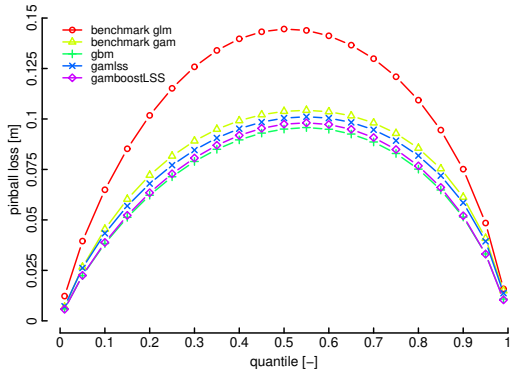
Here we detail the results of the wave height, period, and direction forecasting tasks including density, scenario, and regime membership forecasting.

Parametric & Non-Parametric Regression

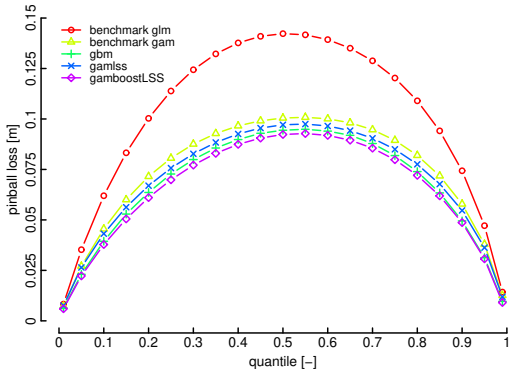
To make decisions on the best regression technique for significant wave height and peak wave period, the calibration and sharpness are compared by 4-fold cross-validation on the training data. Cross-validation is also used to tune the model hyperparameters of the boosting methods and to refine the model formula for the `gamlss` model. For the parametric regression techniques, the conditional distribution of the measured variable is also defined using cross-validation. A large range of distributions with $(0, \infty)$ support were tested. For both significant wave height and peak wave period the Generalised Beta Prime distribution [179, 188] produced the best forecasts in terms of sharpness, subject to calibration. It is a flexible four parameter distribution, which nests other common distributions. For regression model selection in the case study, the calibration of the density forecasts is crucial due to the direct impact on the dependency structure and therefore scenario generation quality.

For the significant wave height density forecasting task, Figures 6.4a and 6.4b show the pinball loss in both cross-validation and testing respectively, which reveals that the two boosting models reduce the pinball loss across the quantiles evaluated, compared to the `gamlss` model, although to a lesser extent at the tails of the distribution. Comparing the `gbm` and `gamboostLSS` techniques the former gives lower error scores in cross-validation and the latter in testing, although the differences are generally minor and only evident in the p30 — p70 range. The *benchmark — gam* model is very competitive, and provides significant improvement over the naive *benchmark — glm*; the only difference in these models is a simple change of base learner from a linear effect to a penalised spline.

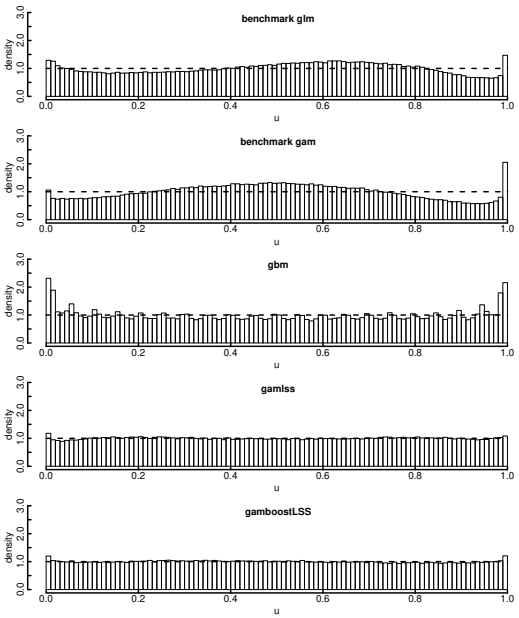
In Figures 6.4c and 6.4d the PIT histograms are presented; comparatively, the advanced parametric regression techniques result in better calibrated forecasts, especially in the tails of the distribution because of the difficulty in estimating quantile regression models in this region. Both benchmark models here show poor calibration, specifically under-confidence. Comparing the *benchmark—gam* and the `gamlss` model, the reliability is much improved for latter even though the models are somewhat similar; this validates the choice of the Generalised Beta Prime distribution. However, the two also



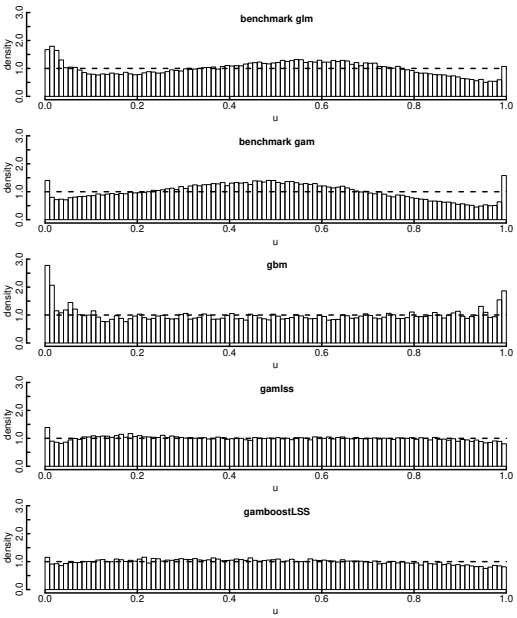
(a) Pinball loss — cross-validation data



(b) Pinball loss — test data



(c) PIT histogram — cross-validation data



(d) PIT histogram — test data

Figure 6.4: Results for all lead times — significant wave height density forecasting by regression model

have different input features, which can be found in the supplementary material.

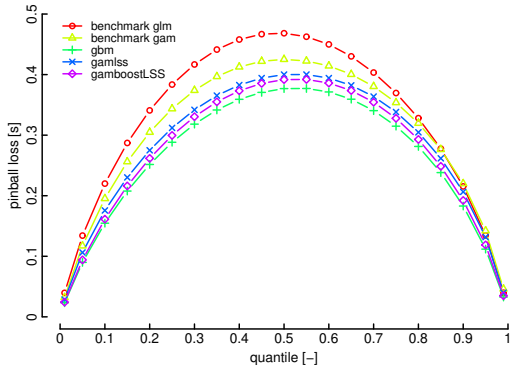
Based on the cross-validation results, the density forecasts of significant wave height obtained via `gamboostLSS` regression are selected for use in the later stages of the access forecasting process, based on the principle that density forecasts should be sharp subject to calibration. The `gbm` approach has a lower pinball loss, but the calibration is poor, notably in both tails of the predictive distribution, and is excluded as a result. The `gamlss` approach is well calibrated but has a higher pinball loss than `gamboostLSS`.

For peak wave period regression, the pinball loss scores are shown in Figures 6.5a and 6.5b for cross-validation and testing. In cross-validation, the lowest pinball loss is more clearly defined using the gradient boosting machine regression technique. However, in testing the two boosting models return very similar pinball loss scores across all the tested quantiles. It should be emphasised that the `gbm` quantile models are directly optimised to minimise this score. A similar behaviour to the wave height case is found for two benchmark models.

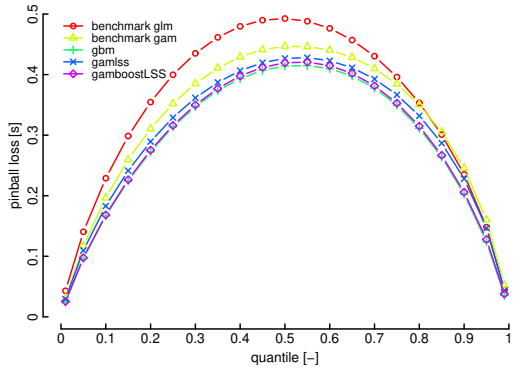
Figures 6.5c and 6.5d detail the PIT histograms for the peak wave period regression models under testing and cross-validation. Clearly again in this case, the calibration of the `gbm` model in the tail region is comparatively poor against the two advanced parametric regression techniques. However, in testing all of the models here present deviations from uniformity. The calibration of two benchmark models is again found to be poor. Based on the cross-validation results, the forecasts based on the `gamboostLSS` regression are selected for implementation due to producing sharp forecasts subject to calibration.

In Figure 6.6 the sharpness, or average interval width, is plotted against forecast lead-time for both significant wave height and peak wave period during testing. The forecast models evaluated are the final `gamboostLSS` densities chosen for further implementation. They show that the uncertainty grows with lead time, which is to be expected. This is particularly pronounced for significant wave height, shown in Figure 6.6a, where the average interval size grows considerably; the 90% interval more than doubles in average width from 0.5m at issue time to over 1.2m at 120 hours-ahead. For this reason, and because we are motivated by day-ahead decision-making, we only consider lead-times of up to 120h despite NWP with lead-times of 120h–240h being available. For wave period in Figure 6.6b, the growth of the interval size is not as pronounced, however the interval widths have a greater spread at the earliest lead times.

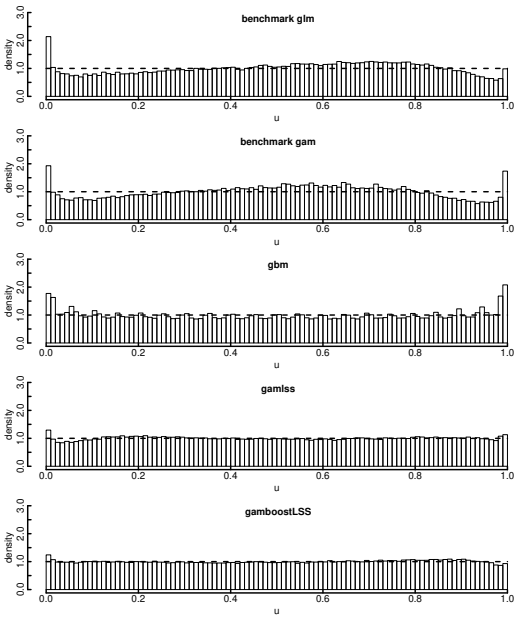
An important aspect of using the `gamboostLSS` method, is that it is very memory intensive; it was not possible to model the density forecasts using a conventional desktop computer (8 virtual cores, 3.6GHz CPU, 16GB RAM), a cloud instance was used



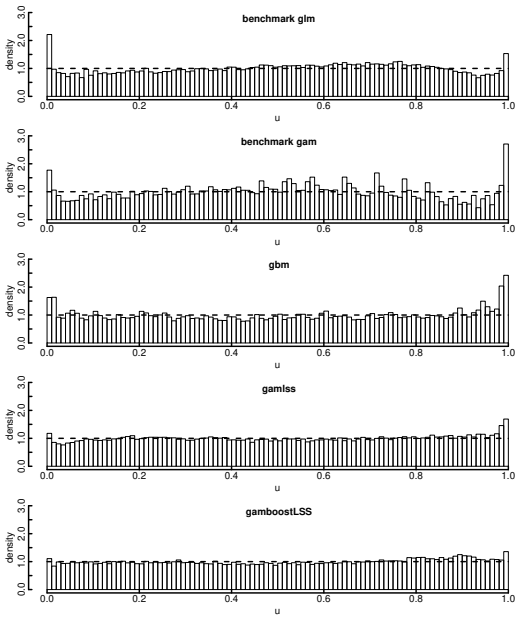
(a) Pinball loss — cross-validation data



(b) Pinball loss — test data



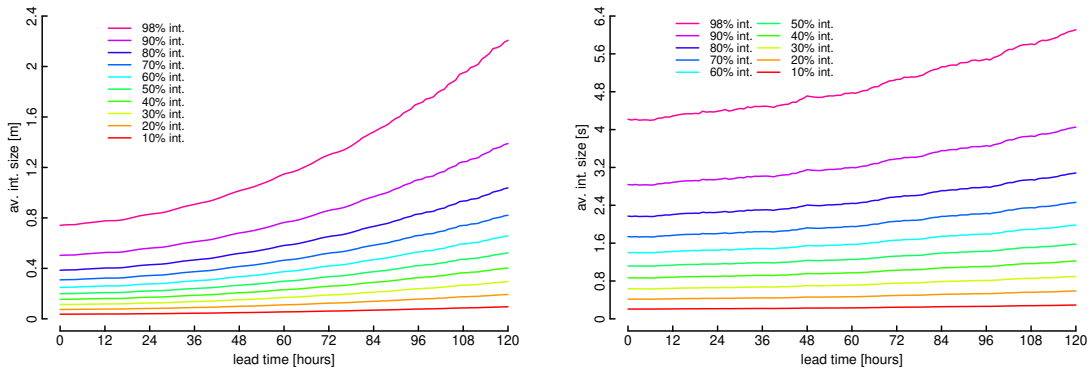
(c) PIT histogram — cross-validation data



(d) PIT histogram — test data

Figure 6.5: Results for all lead times — peak wave period density forecasting by regression model

instead. However, this problem can be significantly reduced by reducing the number of input features, reducing the number of cross validation folds, reducing the number of boosting iterations, and by using a distribution defined by fewer parameters. The `gbm` models were fit using a desktop, although 21 quantile models are required for each fold. Finally the `gamlss` model is computationally cheap, due to the reduced number input variables. Operationally, the time required to issue a forecast is negligible and re-training models would be required infrequently.



(a) Significant Wave Height

(b) Peak Wave Period

Figure 6.6: Evaluation of the sharpness for `gamboostLSS` density forecasts vs lead-time during testing. The width of the prediction intervals increases with lead-time as the forecasts becomes less confident further into the future

Scenario Forecasting

As discussed in Section 6.2.2, for scenario forecasting three configurations are tested: *independence* is a benchmark, *linked* indicates that the full temporal inter-dependency between the significant wave height and peak wave period is modelled, and *temporal* is where the dependency is modelled for each variable separately across the lead times. Again, 4-fold cross-validation is used to determine the most appropriate dependency structure.

In Table 6.1 the scenario forecast scores are presented, where the scores are first separated by the multivariate target variable. For significant wave height it is evident in training and cross-validation there are improvements in modelling the temporal dependency structure across all scores. Comparing the scores, as expected the improvement of modelling the temporal dependency is more significant in the case of both the variogram scores than the energy score because of the better discriminative ability of covariance structures of the former. For evaluation of these forecasts 1000 scenarios are used because, empirically, the mean and standard deviation of the energy score were found to stabilise around this point. In the case of temporal scenarios for significant wave height it takes approximately 0.6 seconds per issue time to generate 1000 samples and transform them into the original domain using the desktop described previously.

Table 6.1 also details the results of the peak wave period scenario forecasting; these

follow a similar profile to the significant wave height case. The temporal dependency provides improvements against the independent case, especially when measuring improvement via the variogram scores.

So far, the results discussed are only for temporal dependency on a per-variable case. The full multivariate distribution of wave height and period is also evaluated in Table 6.1, where the value in modelling the linked dependency is compared against the temporal and independent cases. Note, that the measurements and forecasts are first standardised before calculation of these scores due the different scales of the two quantities. Interestingly, accounting for the cross-variable correlations result in an improved energy score in cross-validation compared to both cases, but the temporal correlation performs best in testing. This is due a change in the cross-variable error correlations in the test data; an ECMWF model update (cycle 43r3) occurs at the beginning of our test period, with statistically significant improvements in both the wave height and period variables, which likely causes this change.

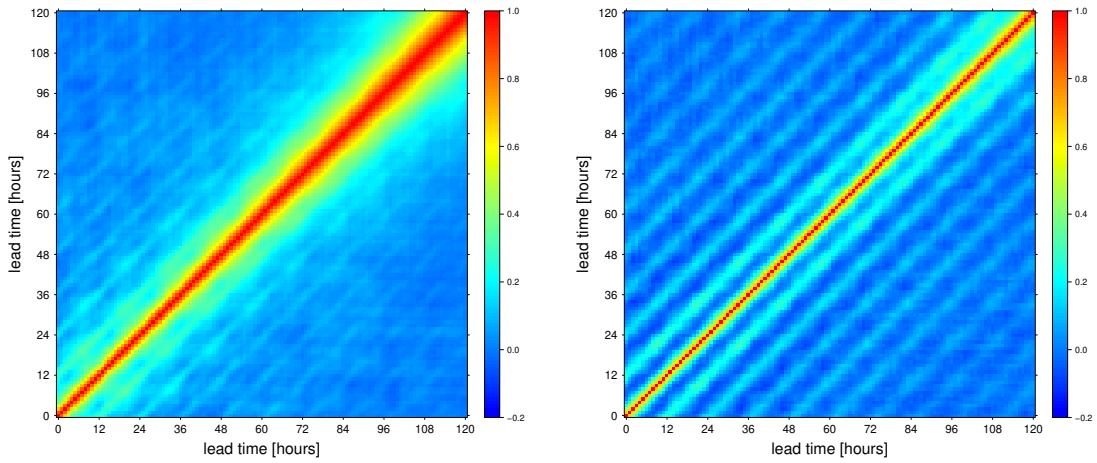
The temporal correlation matrices for both wave height and period scenario fore-

Table 6.1: Multivariate forecast evaluation results. Linked means the temporal dependency between significant wave height and peak period is modelled. Best results during testing are in bold. Note that for the joint wave height and period target variable case, the marginals and observations are first standardised before applying scoring rules

Target Variable	Dependency	Data	ES	VS-1	VS-0.5
Wave Height	Independent	CV	2.086	12.706	15.926
		Test	2.009	13.618	15.494
	Temporal	CV	2.059	3.106	5.763
		Test	**1.975**	**3.317**	**5.476**
Wave Period	Independent	CV	7.872	169.469	51.650
		Test	8.396	190.586	56.933
	Temporal	CV	7.844	105.130	34.515
		Test	**8.376**	**130.679**	**40.209**
Wave Height & Wave Period	Independent	CV	6.437	326.113	125.504
		Test	6.317	321.320	125.481
	Temporal	CV	6.419	272.391	97.766
		Test	**6.303**	273.202	**99.671**
	Linked	CV	6.415	268.780	97.119
		Test	6.305	**272.617**	100.001

casting cases are plotted in Figure 6.7; these are the matrices used to generate scenarios over the test dataset. Here, the dependency characteristics of the uncertainty across the lead time differs between the variables, in that the correlation between consecutive lead times is clearly more prominent in the wave height case. However, they also share some subtle characteristics; the correlation strength persists to a greater extent further ahead in the forecast lead times, and there is clearly a strong diurnal pattern, especially in the peak wave period case. Note that time-of-day, seasonal, and interaction effects are included in the marginal distributions regression formulation.

The energy score improvements for the wave height scenario forecasting task are plotted in Figure 6.8a, compared against the independence benchmark. Here, a simple block-bootstrapping approach is used to estimate the significance in the forecast improvement [139]. The scores are split into non-overlapping blocks of 7-days length, to account for correlation in the governing weather patterns. These blocks are resampled with replacement and then forecast improvement is determined. This process is repeated 1000 times to estimate the sampling variation of the score improvement in Table 6.1. The results are presented via boxplots in Figure 6.8a and clearly illustrate that modelling the temporal dependency is valuable; the sampling variation is greater during testing because of the smaller size of the dataset. The improvement in peak wave period forecasting against independence is also plotted on Figure 6.8b; it is reduced



(a) Significant Wave Height (b) Peak Wave Period

Figure 6.7: Temporal correlation matrices during testing

compared to the significant wave height case.

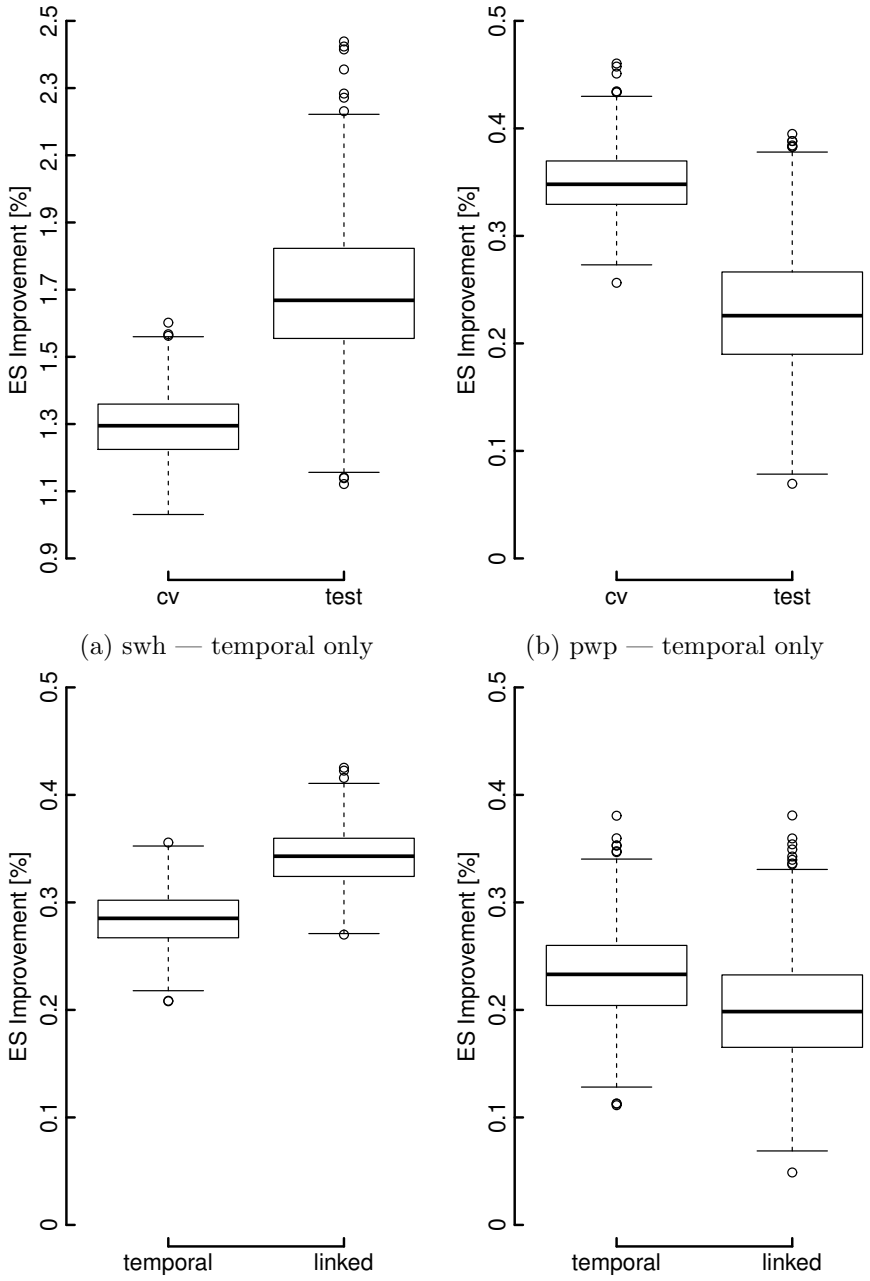
A comparison of the energy score improvements for the wave height *and* period case is shown on Figures 6.8c and 6.8d, against the independence benchmark. The results confirm that in cross validation the linked dependency structure provides improvements over the temporal case, however during testing the cross variable error dependency changes and the simpler temporal dependency structure performs better. Again, this is likely due to the ECMWF model update (cycle 43r3) which changes the structure of the uncertainty inter-dependency between the variables. Time-adaptive estimation of the covariance matrix could track changes such as this in the long-term, however this behaviour is closer to a regime change.

Clustering & Logistic Regression

Wave direction forecasting for input to the vessel motion model is reduced into two stages, clustering the measured wave buoy data and then applying logistic regression to predict regime membership based on NWP data. Please refer back to Section 6.2.3 for more details. When using k-means clustering, the random assignment of the cluster centres at the start of the algorithm can lead to different results if the data-set is small or not amenable to a clustering algorithm. Therefore, in this case the algorithm was set with different random seeds over 4 tests and 98.8% of data points in a large data-set (>140,000 rows) are assigned to the same cluster.

The defined regimes are plotted in Figure 6.9a via a parallel coordinate plot which allows for visualisation of circular variables [189]. Clearly the two regimes are separated mostly by peak wave direction; waves from the north east which are driven by the swell and waves driven by the prevailing south westerly winds. The swell driven waves also on average have longer periods and both regimes have similar average wave heights. The repeatability of the clustering using the measured data and the physical explanation give confidence in the defined regimes at the site.

Post-processing results for the logistic regression are shown in Figure 6.9b via the ROC curve for the testing and cross-validation phases; the AUROC is very close to one for both cases, being 0.97 and 0.95 in cross-validation and testing respectively. This



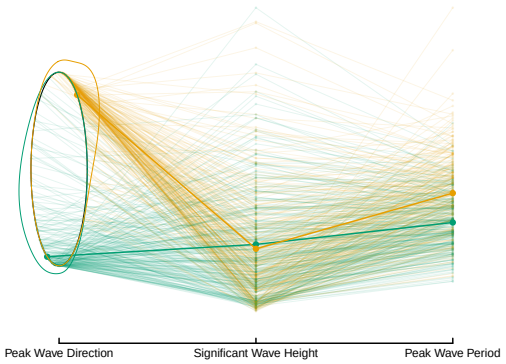
(a) swh — temporal only

(b) pwp — temporal only

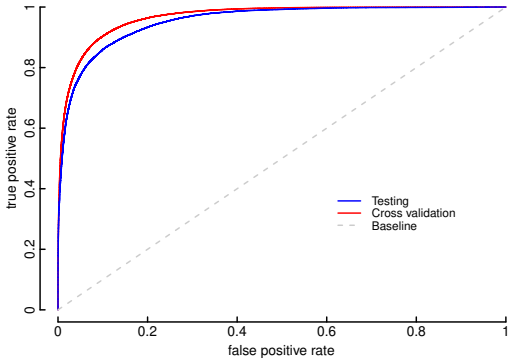
(c) swh & pwp — cross validation

(d) swh & pwp — testing

Figure 6.8: Boxplots showing the block-bootstrap sample distributions of Energy Score improvement for significant wave height (swh), peak wave period (pwp), and coupled (swh & pwp) scenario forecasting. The benchmark in all cases is the *independence* case, with no temporal correlation. Linked means the temporal inter-dependency between significant wave height and peak period is modelled



(a) Parallel coordinate plot showing the regime membership of the measured data after k-means clustering [189]. Clearly the regime separation is mostly defined by peak wave direction



(b) ROC curve for the logistic regression forecasts of the cluster membership for all lead times. Results for the AUROC are 0.97 and 0.95 for the cross-validation and testing respectively

Figure 6.9: Sea-state regime classification plots — regime forecasts are used in the vessel motion model

means that for an optimally defined threshold probability, the classification provides a high rate of positive predictions when the measured value is positive, and a low rate of positive predictions when the measured value is negative. The optimal threshold probability which is used to split the predicted probability space into the predicted regimes is defined by the one which maximises accuracy, so assuming there is an equal weighting to false positive predictions and the false negative predictions.

6.4.2 Vessel Motion

Here, we explore the mapping of the relationship between the measured significant wave height at the buoy and vessel motion measurements during transfer. Several vessel motion models are tested, starting from basic linear regression models and leading to truncated regression models with penalised smooth base learners in the `gamlss` environment. Truncated regression is used to respect the reality that vessel heave peak to peak measurements are always positive. The Akaike Information Criterion (AIC) is used to measure goodness of fit

$$\text{AIC} = 2k - ln(\hat{L}) \quad , \tag{6.7}$$

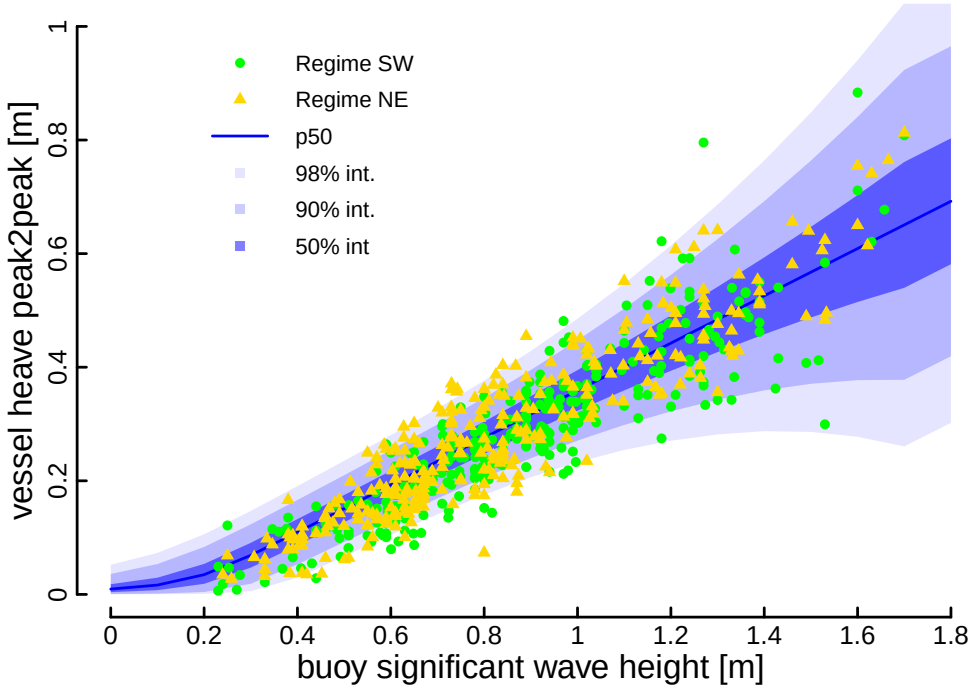


Figure 6.10: Plot of vessel motion during push-on and concurrent ocean measurements. The distributional model fit is for the *tr*-T-5 model in the south west regime with a fixed peak period. Note that push-ons only occur in a small range of the possible significant wave height forecast values and the variance is capped at the highest observed significant wave height. See Sections 6.4.3 and 6.4.3 for more detail on dealing with this effect

which rewards the model with the highest likelihood function $\hat{L}$, penalised by the number of parameters k used to estimate the model; overfitting is then less likely for the model with the minimum AIC. As shown in Table 6.2, the minimum AIC of the models tested is found using a student-t distribution, truncated at 0.

The marginal effect plot of significant wave height is shown in Figure 6.10 for the best model (*tr*-T-5) with peak period and regime membership held constant; the motivation behind using a model with conditional heteroscedasticity is clear, as the uncertainty grows with significant wave height. Importantly, this model uses penalised varying coefficient splines for the location and shape parameters, where the coefficients vary by the regime membership. This allows for the regime membership to influence the model fit more flexibly than varying the intercept terms and the resulting shape of the uncertainty to change depending on the regime membership.

Table 6.2: Selected formulations for the vessel motion model during transfer dependent on observed sea-state. The symbols *ps* indicates a penalised beta spline, † is a varying coefficient model, *poly* is a fractional polynomial model, H_s is significant wave height, T_p is peak wave period, ω_p is peak wave direction, and *reg.* is the regime membership

		Formula & Explanatory Variables			
ID	Distribution	Location	Scale	Shape	AIC
N-1	Normal	$\sim H_s$	~ 1	n/a	-1653
N-2	Normal	$\sim H_s + T_p$	~ 1	n/a	-1687
N-3	Normal	$\sim H_s + T_p * H_s$	~ 1	n/a	-1695
N-4	Normal	$\sim H_s + T_p * H_s$	$\sim H_s$	n/a	-1787
tr-N-1	Tr. Normal $[0, \infty)$	$\sim H_s + \text{poly}(T_p + T_p^2)$	$\sim H_s + \mathbf{1}(\omega_p < 120)$	n/a	-1823
tr-T-1	Tr. *t*-family $[0, \infty)$	$\sim H_s + \text{poly}(T_p + T_p^2)$	$\sim H_s + \mathbf{1}(\omega_p < 120)$	$\sim H_s$	-1827
tr-T-2	Tr. *t*-family $[0, \infty)$	$\sim H_s + \text{ps}(T_p)$	$\sim H_s + \mathbf{1}(\omega_p < 120)$	$\sim H_s$	-1829
tr-T-3	Tr. *t*-family $[0, \infty)$	$\sim \text{ps}^\dagger(H_s, \text{reg.}) + \text{ps}(T_p)$	$\sim H_s + \text{reg.}$	~ 1	-1830
tr-T-4	Tr. *t*-family $[0, \infty)$	$\sim \text{ps}^\dagger(H_s, \text{reg.}) + \text{ps}(T_p) + T_p * H_s$	$\sim H_s + \text{reg.}$	~ 1	-1838
tr-T-5	Tr. *t*-family $[0, \infty)$	$\sim \text{ps}^\dagger(H_s, \text{reg.}) + \text{ps}(T_p) + T_p * H_s$	$\sim \text{ps}^\dagger(H_s, \text{reg.})$	~ 1	**-1846**

6.4.3 Forecasting Vessel Motion during Transfers

Here, the outputs of the access forecasting methodology are described. From Figure 6.1 the last stage of the modelling is a visualisation stage, which is described. This involves transforming the raw vessel forecast scenarios to make the uncertainty information more interpretable.

Raw Output

To generate the vessel motion scenarios, the forecast sea-state scenarios and forecast regime membership are used to drive the vessel motion model. This allows for a forecast estimation of the heave motion of the vessel during transfer. In the presented case, the mean output of the vessel motion model is taken for each scenario input by sampling each generated distribution due to the asymmetric nature of the truncated distribution. A scenario forecast, generated via the mean output of the vessel motion model, is presented in Figure 6.11 and illustrates the motivation behind the visualisation stage; this uncertainty information is difficult to interpret by any decision maker. However, the raw output could be useful for driving scheduling optimisation tools for instance.

An important subtlety of the vessel motion model presented in Figure 6.10 is that data is only collected in a sub-range of the significant wave height marginal distribution at the site; this is a result of push-ons only being attempted in conditions conducive to safe transfer. For forecast values outside this range a threshold significant wave height feature (equal to the maximum observed significant wave height during push-on) is used for the shape parameter in the vessel motion model. This means that stable predictions are obtained across the full distribution of forecast significant wave height values at the site. However, the vessel forecast scenarios above this threshold must be then processed to represent a zero chance of transfer.

Uncertainty information during transfers is explored here, although the conditional mean of displacement is used for the operational forecast. Future work should consider the incorporation of uncertainty in the vessel motion model, for instance by sampling the vessel motion distribution.

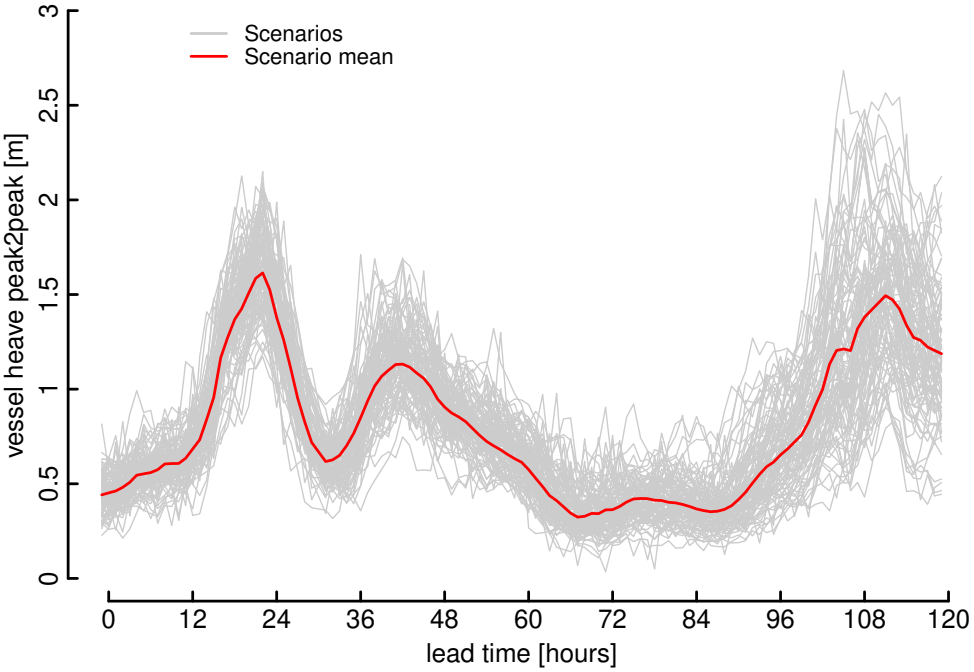


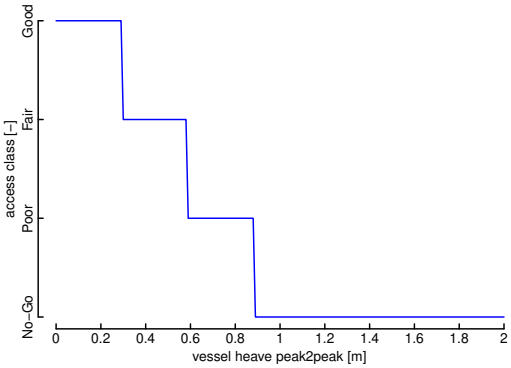
Figure 6.11: Example scenario forecast of mean vessel motion during transfer

Visualisation

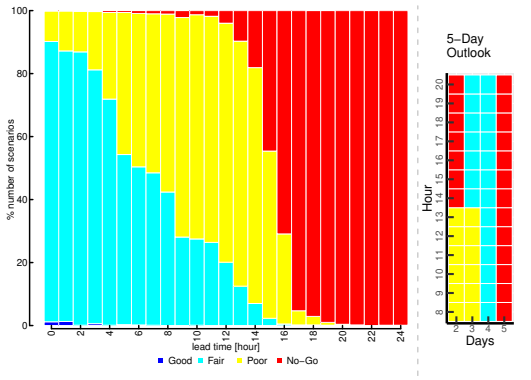
Numerous visualisation options are possible based on the forecast from Figure 6.11. The motivation and end-user requirements must be considered before visualisation options are explored; end-users require the forecast and associated uncertainty to be communicated as simply as possible, and the motivation is to make the forecast interpretable, as well as processing the scenarios so that those which have a zero chance of success, according to the historical data, are conveyed as such.

Here, a user-defined function for transforming the vessel scenarios into classes is shown in Figure 6.12a. This transformation is flexible and can be based on the vessel capabilities, specific mission, experience of the site, and appetite for risk. Additionally, to distil the information content the detailed forecast visualisation focuses on the upcoming day with extended lead times summarised to the right of the main plot.

An example of the resulting transfer quality forecast is shown in Figure 6.12b. Here, the bars quantify the percentage of scenarios belonging to each class at every time step. Details on conditions further into the horizon are shown via a panel plot



(a) User-defined function for transforming mean vessel motion scenarios into a transfer quality forecast



(b) Example visualisation option of Figure 6.11 as a transfer quality forecast with a focus on the upcoming day on the left panel and further horizons summarised in the right panel

Figure 6.12: Visualisation stage plots

where the colour of each panel indicates the dominant class over the next 4 work-days. The advantage of the classification method is that every scenario is accounted for and therefore the end user views a complete picture of the spread of possibilities at every time step. Note that the economic impact of the forecasts is not currently considered, which highlights an important avenue of future work arising from this research.

6.5 Conclusions & Future Work

This chapter describes a novel forecasting solution for predicting safety-critical conditions during transfer for offshore operations with a case study at an east coast wind farm in the UK. The proposed access forecasts predict vessel motion during transfer, accounting for weather uncertainty, up to 5-days ahead. Sharp and calibrated density forecasts of peak wave period and significant wave height are generated by post-processing Numerical Weather Predictions, with boosted generalised additive models for location, scale and shape outperforming non-parametric methods. Scenario forecasts of these variables have then been produced using the Gaussian copula to model temporal dependence and used as inputs to a data-driven vessel motion. Modelling cross-variable dependency added value in terms of the multivariate skill scores during

cross validation, but not in testing. A method of visualisation of these forecasts is also suggested to best communicate the information content for end users.

Future work on the methodology should consider the feasibility of a turbine (or region) specific forecast for large wind farms where access is constrained by local bathymetry. Alternatives to the vessel motion model where data is not available should be considered, as well as investigating the value in transforming the motion to the point of contact with the fender and ladder. Embedding the forecasts into a schedule optimisation tool or a cost/loss model with a corresponding power forecast could further support offshore wind farm operations and be used off-line to demonstrate the value in accounting for the uncertainty in access conditions. Future visualisation options should also incorporate the results of any cost/loss model, which would enable decision makers to anticipate the economic impact of dispatch decisions based on the forecasts.

Extending the forecast horizon beyond 5 days is of interest, though it may be necessary to consider lower temporal resolution forecasts, e.g. daily accessibility, due to the reduced skill in NWP at these lead-times.

Some important meteorological factors that restrict offshore access have not been considered here, such as lightning, visibility, and surface wind speeds; forecasts of these should be provided to decision-makers. The forecasting in this paper is based on summary statistics of the wave field; an interesting extension would be to use the forecast wave spectrum from the NWP, which would enable a more complete picture of expected conditions.

Regarding sea-state forecasting, the high-dimensional nature of the dependency structures mean that alternative copulas are somewhat limited. Options include: copula vines, though this would increase computation cost significantly for both fitting and sampling; the empirical copula, in which training data are re-sampled to produce scenarios; or using the Gaussian copula with parametric covariance matrices, though the diurnal patterns observed in the empirical covariance (Figure 6.7) suggest that this would not be trivial. Lastly, the dependency structure of the scenario forecasts could be made conditional on the dominant forecast direction regime at the site, which could improve the quality of the scenarios.

Chapter 7

Regime Switching Multivariate Wind Power Forecasting

In the previous chapter we discussed the potential in future work for regime-switching covariance dependencies to be leveraged in an offshore access forecasting context. Here, we investigate their potential for day-ahead multivariate forecasting of wind power; we forecast simultaneously 92 utility scale wind farms over 97 lead times in the GB network accounting for temporal and spatial correlations. Therefore, we model a high dimensional distribution in which we have the classic $p >> n$ problem, i.e. the dimension is much greater than the sample of observations. To this end, calibrated density forecasts of wind power are produced and evaluated using non-parametric methods at each wind farm, the temporal and spatial dependencies are then mapped using a Gaussian copula, and a regime switching method is elaborated. The multivariate forecasts are finally evaluated via proper scoring rules.

As discussed in Section 2.4.3, current literature on multivariate wind power forecasting is currently focused on advancing the Gaussian copula method first published in [69]. The reason this technique is so successful is because of the ability to model very high-dimensional problems effectively, and the apparent absence of tail dependencies in the data. Another advantage is that the dependency is described by a single covariance matrix; this matrix can also be parameterised and/or regularised which is essential for very-high dimensional problems.

Parameterising the covariance matrix has been shown to improve the multivariate scoring rules, even for relatively low dimensional problems [72, 92]. This is due to the fact that empirical covariance estimates can often be noisy, and the parametric description effectively acts to regularise the estimate; the covariance only depends on a handful of estimated parameters, whereas the empirical covariance dependency is described by $d(d-1)/2$ estimates, where d is the dimension. Additionally, non-separable parametric descriptions of covariance have been applied to wind speed data [93], where the interaction in between the spatial and temporal covariance is modelled. However, to the best of the author's knowledge have not been applied for day-ahead multivariate wind power forecasting.

In the literature there are also few studies which explore multivariate wind power forecasting at a significant spatial and temporal scale. In [91] the authors explore multivariate wind power forecasting in Denmark at 15 regions over 43 lead times and model the spatial-temporal dependencies via sparse precision matrices and direction-dependent correlations. Here, we model 92 of the large-scale wind balancing mechanism units (BMUs) in the GB grid, over 97 lead times. We also evaluate the potential for regime-switching dependency, where the surrounding literature is discussed in Section 2.4.3.

The contribution of this chapter is in the demonstration of the methodology at such high dimensions of space and time, because this is similar to the scale faced by future end-users of the forecasts, such as power system operators for reserve sizing or managing transmission power flows. We also explore a non-separable parametric estimations of the space-time covariance, and demonstrate a simple regime switching method based on clustering of the average day-ahead wind vectors into westerly and easterly dominated regimes. All multivariate forecasts are evaluated using proper and strictly proper scores and the uncertainty associated with the results explored. Lastly, we explicitly measure the improvement due to accounting for the temporal only and spatio-temporal covariances.

In Section 7.1, the forecasting methodology is outlined, including density forecasting, dependency modelling, and regime switching components. Results are outlined and

discussed for the case study in Section 7.2, followed by conclusions and future work in Section 7.3.

7.1 Forecasting Methodology

Here, we detail the method for: generating the marginal density forecasts via multiple quantile regression and the tail shape estimation; defining and sampling from a multivariate distribution; the non-separable covariance functions used to parameterise the spatio-temporal dependency; and finally the regime switching method used.

7.1.1 Density Forecasting

To generate the density forecasts at each BMU a small set of inputs are used for the multiple quantile regression models to keep the computational time reasonable. Specifically, we extract NWPs at the closest grid point to each location and use inputs of wind speed and direction at 10m, as well as time-of-day and day-of-year. Also, three spatial engineered features are extracted from a grid of NWP data around each location: the mean, maximum, and minimum wind speeds at 100m, which is shown to be successful in the literature [31]. The NWP data sourced from ECMWF is on a 0.1° grid covering the UK, and the spatial features are calculated using a 0.4° by 0.4° box, i.e. using 16 grid points.

The goal of the density forecasting task is to generate calibrated and sharp probabilistic forecasts for each BMU and lead time. We approximate the cumulative distribution function (CDF) for $k \in \{1, 2, ..., K\}$ locations and $t \in \{1, 2, ..., T\}$ lead times

$$\hat{F}_{k,t}(y_{k,t}) = P(Y_{k,t} \leq y_{k,t}) \quad , \tag{7.1}$$

where $\hat{F}$ is a strictly increasing function, by multiple quantile regression. For a given probability level α, a quantile forecast can be defined as

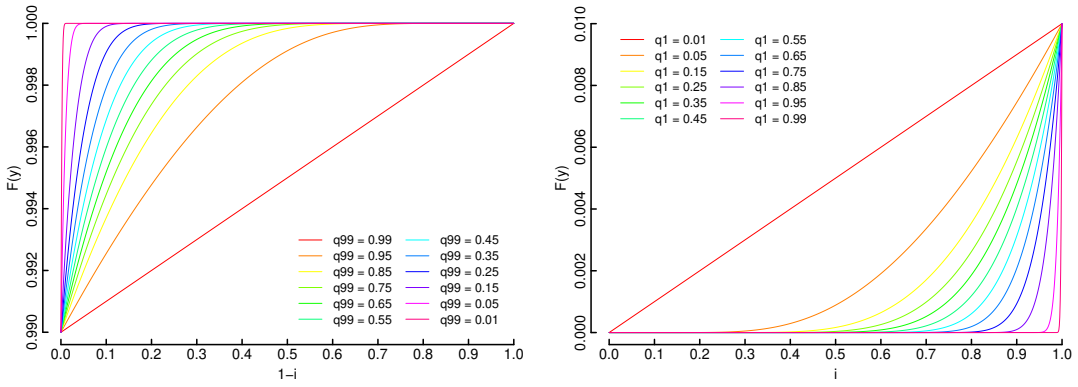
$$\hat{y}_{k,t}^{(\alpha)} = \hat{F}_{k,t}^{-1}(\alpha) \tag{7.2}$$

where $\alpha \in [0, 1]$. By using a collection of quantiles, in this case probability levels of: $\alpha_j \in$ (0.01, 0.02, 0.03, 0.04, 0.05, 0.1,..., 0.95, 0.96, 0.97, 0.98, 0.99), monotone cubic spline interpolation is used to guarantee a valid CDF at each lead time [177]. A higher resolution of quantiles in the tail regions of the distribution is used because empirically, it was found that modelling more quantiles improved the calibration. This is because we rely less on interpolation in a region which can have a high degree of change in power values, and we have a large bank of historical data (5 years of forecasts with two issue times per day) in the case study to estimate the tail regions relatively well via quantile regression, as shown in Section 7.2.

Outside the tail quantile estimates we use a dynamic exponential interpolation to join the upper and lower quantile estimates to the boundary values of maximum and zero power respectively, relatively similar to [190]. The interpolation is dynamic in that the left and right tail shape depends on the current lower (q01) and upper (q99) estimates respectively from the multiple quantile regression. Since the power is $y \in [0, 1]$, where 1 is the normalised maximum power, the quantiles are constructed as

$$\hat{y}^{(\alpha)} = \begin{cases} 0 & \alpha = 0 \\ \hat{y}^{(0.01)} i & \alpha = i\hat{y}^{(0.01)} \left(\frac{0.01}{\hat{y}^{(0.01)}}\right)^{i^{-\frac{1}{1-\hat{y}^{(0.01)}}}}, \forall i \\ \hat{y}^{(\alpha)} & \alpha \in \alpha_j \\ \hat{y}^{(0.99)} + (1-i)(1-\hat{y}^{(0.99)}) & \alpha = 1 - i(1-\hat{y}^{(0.99)}) \left(\frac{0.01}{1-\hat{y}^{(0.99)}}\right)^{i^{-\frac{1}{\hat{y}^{(0.99)}}}}, \forall i \\ 1 & \alpha = 1 \end{cases} \tag{7.3}$$

where $i \in (0, 1)$, and if $\hat{y}^{(0.01)} < 0.01$ or $1 - \hat{y}^{(0.99)} < 0.01$ we simply set these values to 0.01 in the interpolation. Note the lead time and location index is dropped to avoid notational clutter. This construction means that the tail shape is dependent on the current estimates of the upper and lower quantile regression forecasts. If we take the lower tail as an example, the tail shape is linear when the $\hat{y}^{(0.01)}$ estimate is very close to the zero-boundary, and as $\hat{y}^{(0.01)}$ moves further away from the zero-boundary the shape become more exponential. An example of the tail shapes for different values of upper



(a) upper tail (b) lower tail

Figure 7.1: Illustration of the dynamic exponential tails, which are conditional on the upper/lower quantile forecasts for the upper and lower tail respectively. The tail shape becomes more exponential as the reference quantile moves further from the boundary. Note that the x axis here is not power but the i or $1-i$ variable, the ranges are therefore not-scaled

and lower quantiles is shown on Figure 7.1. This aspect of the modelling is definitely a potential area of improvement for future work, by using applicable techniques such as generalised pareto tails; although parametric techniques are complicated by the boundaries, and balancing the correct upper/lower cut-off for the quantile predictions to generate enough training data to fit coefficients.

For multiple quantile regression we use gradient boosting machines, introduced in Section 4.2.2, as the statistical learning technique to map the input NWP feature to prediction relationship. This motivated by the success of this algorithm in the literature [30, 31]. An example predictive CDF for several lead times is shown in Figure 7.2 using the multiple quantiles, spline interpolation, and tail method described.

7.1.2 Scenario Forecasting

Ultimately, the end goal of this study is to generate spatially and temporally coherent scenario forecasts at the BMU level in the GB grid. Again, the Gaussian copula, introduced in Section 5.1.3, is used although the marginals of the copula are now density forecasts separated by lead time and location, i.e. we model the spatio-temporal dependency. Since we have a significantly different framework than in Sections 5 and 6,

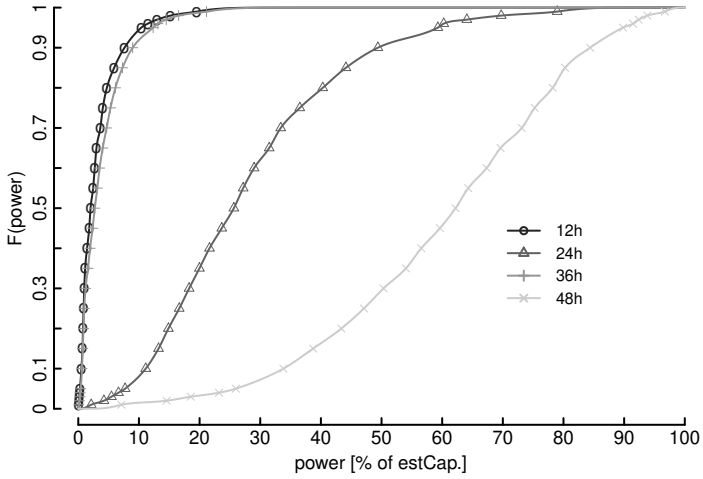


Figure 7.2: Density forecasts as CDFs at clyde central wind farm issued at 2017-04-01 UTC, where estCap. is the estimated capacity. The density forecast is constructed with multiple quantiles and monotonic spline interpolation at each lead time. The approximated CDFs show the high resolution of quantile estimates in the tail regions of the distribution

we will update the definition of the copula framework. Sklar's theorem [168] states that for a KT-dimensional cumulative distribution $F(\cdot)$ with continuous marginals $F_{k,t}(\cdot)$ there exists a unique copula function $C(\cdot)$ such that

$$F(y_{1,1}, y_{1,2}, ..., y_{K,T}) = C\left(F_{1,1}(y_{1,1}), F_{1,2}(y_{1,2}), ..., F_{K,T}(y_{K,T})\right) \quad , \tag{7.4}$$

where K is the number of locations and T is the number of lead times, which decouples the problem into modelling the marginal distributions for each lead time and location, and estimating the dependence between all the marginals via a copula function; based on the uniformly distributed marginals from the Probability Integral Transform, $u_{k,t} = F_{k,t}(y_{k,t})$. Therefore, the calibration of the density forecasts is strongly related to the quality of the dependence modelling. As introduced in Section 5.1.3, after defining the copula we can sample from the multivariate distribution and issue scenario forecasts, by transforming the samples into the original power domain using the inverse CDF at each location and lead time $F^{-1}_{k,t}(\cdot)$.

Analogous to the previous case, the Gaussian copula in this case is given by

$$C(F_{1,1}(y_{1,1}), F_{1,2}(y_{1,2}), ..., F_{K,T}(y_{K,T})) = \Phi_{\Sigma}\big(\Phi^{-1}(F_{1,1}(y_{1,1})), \Phi^{-1}(F_{1,2}(y_{1,2})), ..., \Phi^{-1}(F_{K,T}(y_{K,T}))\big) \quad (7.5)$$

where $\Phi_{\Sigma}(\cdot)$ is the KT-dimensional normal distribution function with covariance matrix Σ and zero mean. Therefore, even for this potentially very high dimensional distribution, the spatio-temporal dependency is characterised by a single covariance matrix. Again, the covariance matrix can be estimated by calculating the sample covariance matrix of the normally distributed marginal variables $v_{k,t} = \Phi^{-1}(F_{k,t}(y_{k,t}))$, where where $\Phi^{-1}(\cdot)$ is the inverse standard normal distribution function, and $v_{k,t} \sim \mathcal{N}(0,1)$. However, in the case study considered here, the dimension of the covariance matrix KT is much greater than the number of samples available to estimate the matrix, without even considering practical issues such as missing data, which means that the sample covariance matrix is singular and estimates are noisy.

The singular sample covariance effectively means that it is not possible to sample from the multivariate Gaussian and transform samples into the original power domain. Therefore, we consider a parametric covariance structure which reduces the number of parameters in estimation of the dependence drastically, and has the added benefit of regularising the covariance estimate which is essential in very-high dimensional problems. This approach has been shown successful in related studies based on temporal scenario forecasting [72, 92] and spatio-temporal forecasting [191]. The parametric covariance between lead time t at location k and lead time t' at location k' is defined as

$$\Sigma_{(k,t),(k',t')} = \mathrm{cov}(v_{k,t}, v_{k',t'}) = \frac{1-\nu}{1+a|\delta t|^{2\tau_t}}\left[\exp\left(-\frac{|\delta k|}{\tau_k(1+a|\delta t|^{2\tau_t})^{\beta/2}}\right) + \frac{\nu}{1-\nu}\mathbf{1}(|\delta k| = 0)\right] \quad (7.6)$$

where δt is the time difference between lead time t and t' in hours, δk is the distance between location k and k' in kilometres, and $\mathbf{1}(.)$ is the indicator function. This is a non-

separable parametric description of the space-time covariance [192], as a combination of a Cauchy temporal function, parameterised by the a and τ_t components, and an exponential spatial function with a nugget effect, governed by the τ_k and ν parameters. The parameter β describes the interaction between the spatial and temporal covariance, and setting $\beta = 0$ results in a separable description of the covariance. An example surface plot of the difference between a separable and non-separable covariance function is shown on Figure 7.3. This method is flexible in that alternative functions can be used to fit empirical covariance matrices with different characteristics. For instance, the nugget effect can be excluded and the temporal and/or spatial functions could be replaced by any valid parametric covariance function [192].

The nugget effect is included because empirically it was found that the spatial covariance drops significantly around the origin, then tails off slowly. In [93, 192], the

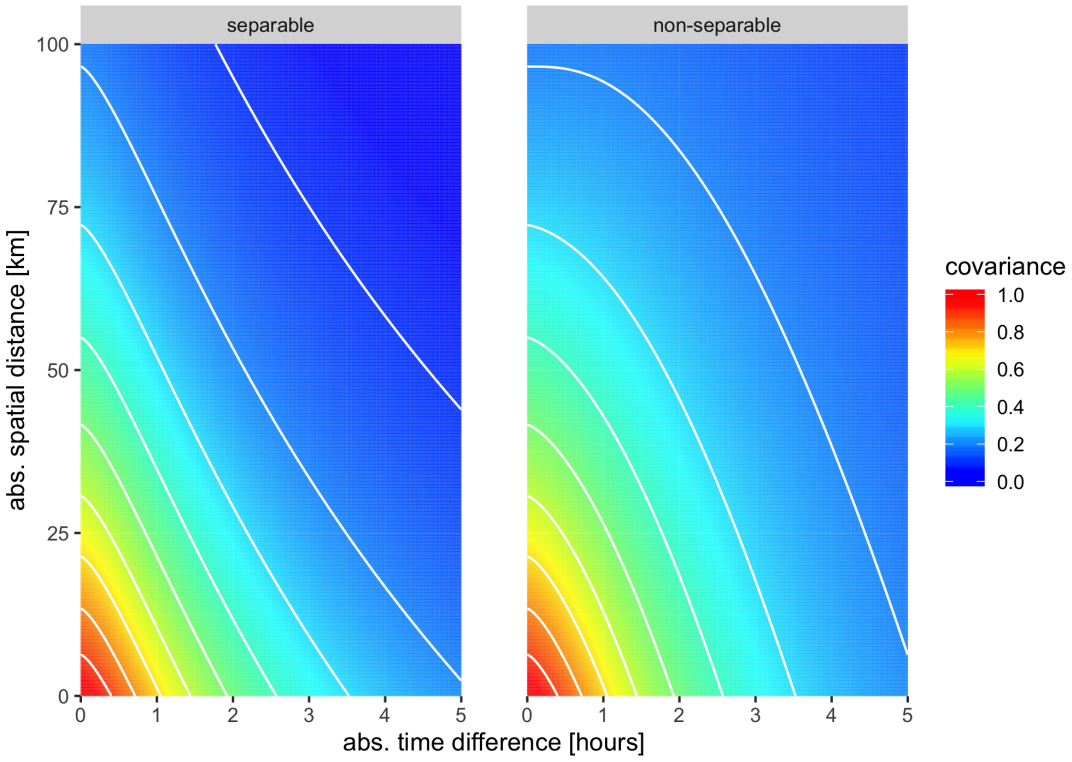


Figure 7.3: Contour plots of a parametric covariance using equation 7.6 with paramaters a=0.4, $\tau_t = 0.7$, $\tau_k = 60$, $\nu = 0$, and finally $\beta = 0$ for the left plot and $\beta = 1.25$ on the right. The interaction parameter allows for more flexibility when modelling the covariance

authors argue that this discontinuity corresponds to measurement error and/or small scale spatial variability in wind speeds, based on a case study on Irish wind speed data. Since the power data collected here is subject to grid code governance, it is most likely the latter in this case.

All the coefficients which govern the dependency are estimated directly using non-linear least squares regression since the covariance function is clearly not linear in terms of the parameters. The coefficients are estimated using iterative optimization to minimise the residual sum of squares, where the target values are extracted from the lower triangle of the empirical covariance matrix and paired with the associated time-distance information for each BMU and lead time. The parametric covariance matrix can then be constructed using the time-distance information and coefficient estimates. Note that the number of parameters describing the spatio-temporal dependency drops from $KT(KT-1)/2$ estimates to 5 coefficients.

There are other methods to regularise covariance estimates, or to find positive-definite matrices in very-high dimensional problems, such as the Ledoit-Wolf shrinkage estimator [193], the nearest valid matrix method [194], and the graphical lasso [195]. The latter requires a non-negative definite matrix, which was not the case in this study. Additionally, wind power forecasting is at an advantage compared to for instance, financial modelling, in that the covariance is governed by physical spatial and temporal interactions, it is particularly suitable to the approach described here.

Four configurations of the dependency are tested: 1) *Independence* — the benchmark where no correlation is embedded in the high dimensional dependence, 2) *Temporal* — where only the temporal correlation is considered between lead times, 3) *Spatiotemporal* — where the full dependence between each lead time and location is included, and finally 4) *RS-Spatiotemporal*, where simple regimes, elaborated in the following subsection, are defined for the forecast issue times and different spatio-temporal correlations are estimated in each.

The *Temporal* covariance matrix in this case was found by setting the spatial components of the *Spatiotemporal* matrix to zero; there was very little difference between this and re-estimating the covariance matrix for the purely temporal case during the case

study, i.e. assuming a separable covariance function, as shown in Appendix Figure D.2. This framework also allows for a explicit quantification of the value in including spatial correlations, because the temporal components of the *Spatiotemporal* and *Temporal* dependency structures are identical.

7.1.3 Regime Switching

As discussed in Section 2.4.3, regime-switching is a method to make models dependent on the current behaviour of the target time-series or on exogenous variable(s). The benefit is that the model can react faster to changing conditions, as opposed to having fixed models or by tracking slower changes in behaviour via (for instance) online updates. It is hypothesised that the spatial and temporal dependence between wind BMUs and lead time changes due to large-scale atmospheric regimes.

Regime-switching models in short-term wind forecasting have been employed to capture structural differences in wind power time series due to localised weather phenomena and characteristics of wind turbine power curve [95, 96, 98]. These models utilise exogenous variables, such as wind direction [98], or atmospheric regimes [96], or model some unobserved hidden-Markov process [95] to condition the forecasts.

In this case, we define the regimes by clustering the 100m meridional and zonal wind vector forecasts, averaged over each issue time (l) and all locations $\theta_l = (\tilde{u}_l^{100}, \tilde{v}_l^{100})$. The k-means algorithm [156] is then used to define 2 regimes. The algorithm generates two disjoint regions R that collectively cover the input space spanned by θ_l. Distinct covariance matrices $\Sigma^{(s_l)}$ are calculated, using the non-separable parametric approach described previously, and scenario forecasts are generated according to the defined regime for each lead time, where

$$s_l = \begin{cases} 1 \text{ for } \theta_l \in R_1 \\ 2 \text{ for } \theta_l \in R_2 \end{cases}, \tag{7.7}$$

which are dominated by either easterly (R_1), i.e. winds travelling from the east, or westerly (R_2) average wind speed forecasts in the case study.

This approach is obviously very simplistic, and we lose a lot of valuable information when averaging the wind speed vectors to this sort of resolution; however, using the scenario forecasting method described above requires the matrices to be defined for a complete issue time. Future work should consider defining the regimes using approaches similar to [96,196,197], leveraging techniques such as principle component analysis, self-organising maps, and different meteorological variables.

7.2 Case Study

The proposed methodology is tested on almost all the large wind generating units in the GB network, during the time-period of available data. Specifically we forecast at the BMU level, therefore some of the larger wind farms are separated into several units. Measured power data availability covered the period of Jan 2015 — Sep 2019, and wind farms which have less than 28 months of available data are removed from the exercise to allow for sufficient data present in both training and testing phases. We also remove the first 6 months of data from new BMU units as they are brought online to account for commissioning, and 5 BMUs were removed completely from the analysis because of data issues during the testing period. This left 92 BMUs which are included in the study, as shown on Figure 7.4 which maps each unit location, and the full list of BMU identification codes and associated site names can be found Appendix Table D.

Generation data from the BMUs are at 30-minute resolution with instances of curtailment flagged and excluded from the forecasting exercise. Data is also normalised and adjusted for availability so the impact of outages on evaluation results is minimised. This is an important step for the larger offshore wind farms, where the availability has a significant impact on the above-rated wind speed region of the forecast (and measured) power curve. The availability is estimated using the measured data and wind speed forecasts by finding the maximum power during periods when the BMU is in the rated region of the power curve, i.e. during high wind speeds with no change in the power measurements; the estimated availability is carried forward until a new estimate is available. At some BMUs, e.g. Whitelee (T_WHILW-1), this method doesn't work well because the unit doesn't reach rated capacity often, in which case a rolling

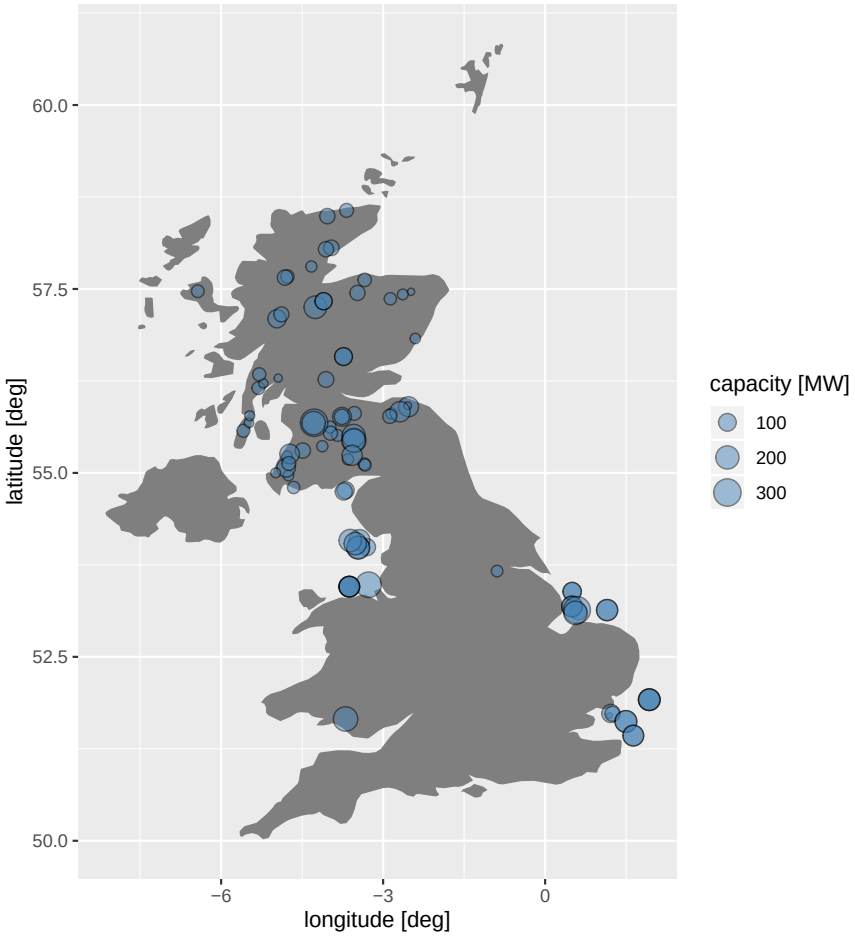


Figure 7.4: Each wind Balancing Mechanism Unit (BMU) location in the GB grid. Note that the larger wind farms are decomposed into several BMUs

maximum is used with a two month window.

We also re-define boundary values, so that $y \in (0,1)$, i.e. perfect boundary values are redefined by sampling approximately 0 or 1 values; for the lower boundary values y_0 we sample from a uniform distribution $y_0 \sim \mathcal{U}(0.001, 0.005)$, and the mirror image for boundary values at 1. This vastly improved the estimation of the lower tail region of the density forecasts, but also has an impact on the reliability results. Reliability, as defined in equation 3.10, is specifically for continuous variables, and at sites with a significant proportion of observations at the boundaries, it becomes difficult to determine calibration without processing or excluding the boundary values.

NWP data from the European Centre for Medium-Range Weather Forecasts is

extracted in a grid around each unit from 0 to 48 hours ahead in hourly intervals, with 2 issue times per day. As mentioned previously, the NWP data sourced from ECMWF is on a 0.1° grid, and the spatial features are calculated using a 0.4° by 0.4° box. Linear interpolation is used to match the resolution of the hourly forecasts and half hourly power data. Although spline interpolation could equally be used, an analysis on the impact of the interpolation method on forecast skill is not the within the scope of this study. The methodologies described are implemented in `R` [170] using the package `ProbCast` [198]; `ProbCast` functions to implement this methodology were written and expanded on by the author.

For the regression and dependency modelling the data is partitioned into approximately 3.75 years for training (Jan 2015 – Sep 2018) and 1 year for testing (Oct 2018 – Sep 2019), which allows for sufficient data for training density forecasts and modelling the spatio-temporal dependence, and an entire year to evaluate the subsequent forecasts in out-of-sample tests. Note that the dimensions of this case study (92 sites with 97 lead times) mean that we would need over 12 years of historical data at each site to estimate an invertible sample covariance matrix in the scenario forecasting stage, not accounting for missing data, with two NWP issue times per day.

We base our development of the forecasting models using 4-fold cross-validation over the training data only; the folds are assigned in blocks of weekly issue times to preserve the temporal characteristics of the time-series, as discussed in Section 3. This gives 4.75 years of out-of-sample forecasts on which to verify the forecasts and base our conclusions, including 1 year of blind testing data not used in any of the modelling. The marginal density forecasts are first evaluated to demonstrate calibration and sharpness at each site, and then the multivariate forecasts are evaluated.

The marginal density forecasts of power are evaluated according to the principle that the forecast should be optimally sharp subject to calibration [127]. This concept and the appropriate univariate and multivariate scoring rules are discussed extensively in Section 3. We use the pinball loss metric and reliability diagrams to measure quantile sharpness and calibration, and the Probability Integral Transform (PIT) validate the calibration of the full distribution, motivated by the direct impact of this variable on

the dependency structure quality.

Scenario forecasts are evaluated via multivariate probabilistic forecast verification methods. Two metrics capable of evaluating the trajectories are the Energy Score (ES) and the p-Variogram Score (VS-p); both are evaluated per issue time of the forecast. For more information regarding these scores, the reader is referred to Section 3.2.2. In this case, calculating the multivariate scores is very computationally expensive due to the large dimensions of the problem. In fact, it was a bottle neck in the whole process. Therefore, we only consider the energy score and an unweighted version of the variogram score; due to wind farm curtailment and data cleaning the weight matrix of the variogram score in practice has to be re-estimated at every issue time, and adds a substantial time to the already lengthy calculation.

To test the significance of forecast improvement the Diebold-Mariano (DM) test is used [144], in parallel to block-bootstrap resampling of skill scores. In the former case, if we define the ES at issue time l, the mean loss differential is

$$\bar{d} = \frac{1}{L}\sum_{l=1}^{L} \mathrm{es}_l^1 - \mathrm{es}_l^2 \tag{7.8}$$

between forecast model 1 and 2. Then the DM test statistic is

$$\mathrm{DM} = \frac{\bar{d}}{\sqrt{\frac{1}{L}(\gamma_o + 2\sum_{h=1}^{H-1}\gamma_h)}} \quad , \tag{7.9}$$

where γ_h is the autocovariance of the loss differential time series at lag h, and the total number of lags H should be large enough to properly account for autocorrelations. The basis of the DM test is that under the null hypothesis, i.e. if there is no difference between forecast performance, then asymptotically $\mathrm{DM} \sim \mathcal{N}(0,1)$ as the total number of samples L increases. The significance of the resulting DM statistic is then found by calculating the associated p-value and comparing it to a defied confidence level. Typically, the null hypothesis can be rejected when the p-value is below a 0.05 confidence level.

7.2.1 Univariate: Density Forecasting

Here, we detail the results of the univariate density forecasting task at the BMU level which constitute the marginals of the Gaussian copula. These are the forecasts generated by post-processing NWP data and site level measurements using multiple quantile regression and tail interpolation. Although these forecasts don't contain spatio-temporal information it is important to demonstrate calibration of the marginals, especially in the tails, since it has a large impact on the quality of the multivariate dependency structure.

The hyper-parameters of the quantile regression models were set to generate sparse tree-ensembles, by reducing the depth of the trees and penalising the importance of each tree to a high extent at every BMU. This reduces the likelihood of over-fitting and empirically was found to generate well calibrated quantile forecasts in cross-validation across all BMUs. Although there's no doubt that using cross-validation at each site and each quantile would improve the sharpness of the density forecasts, this is not computationally feasible with the available resources in this case.

The reliability diagrams are shown in Figure 7.5 for both the cross validation and testing dataset over all lead times considered. The out-of-sample forecasts are very well calibrated in cross validation, due to the model formulation and the volume of data we have at each site to determine the reliability. However, in the testing individual BMUs are seen to deviate from calibration and on average there is a tendency to over-forecast the quantiles in the mid-range of the distribution. It is suggested that this is because, at least in part, that a whole year of forecasts are issued without re-training, and the dynamics of some wind farms are changing. For instance, at the beginning of our test set the new Walney extension is brought online, and the older Walney BMUs are among the worse calibrated in testing. This aspect of the modelling could be improved by considering a rolling window for instance.

More reliability results are presented in Figure 7.6, this time separated into lead time groups. The results here show that at the earlier lead times in both cross-validation and testing that the upper quantiles are over-forecast, although to a lesser extent in the cross-validation. Note that the typical shape of the reliability diagrams are found

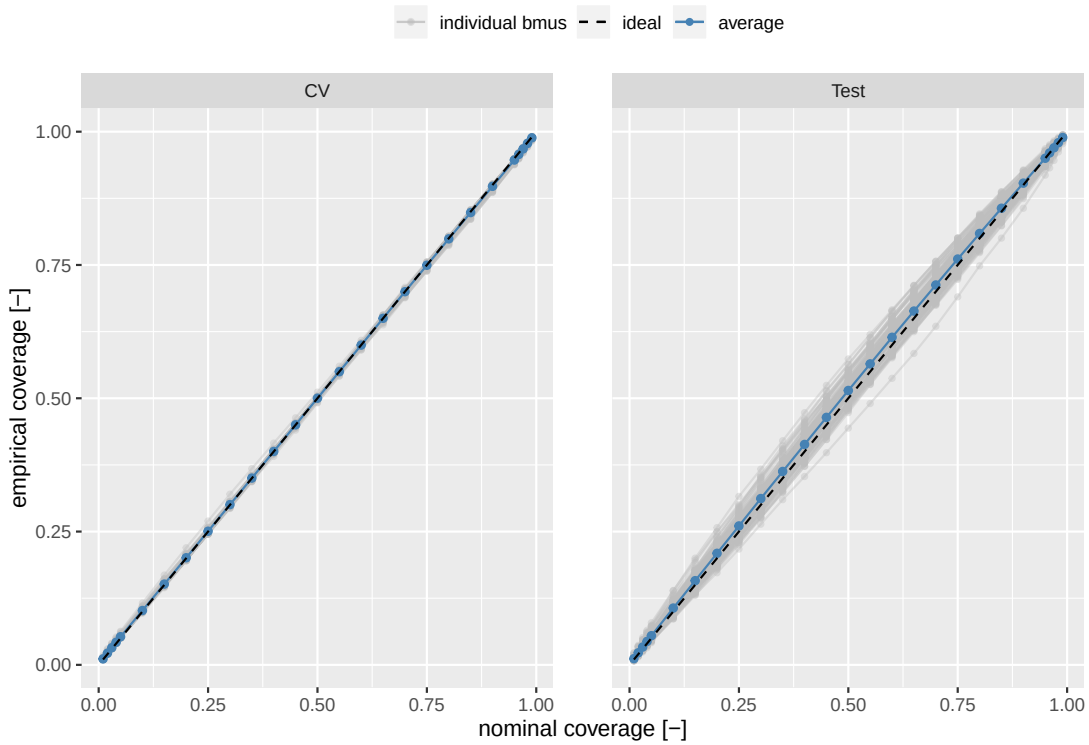


Figure 7.5: Univariate evaluation: reliability diagrams over all lead times during both cross validation (CV) and testing datasets. The models are very well calibrated in cross-validation; however not as well calibrated in testing. This is because of the volume of data we have in cross validation, and that we are forecasting a whole year at each site during testing with no re-training

at the later lead time groups, where the lower quantiles are marginally above optimal levels and the upper quantiles slightly below. The earlier lead time forecasts could be improved by including lagged-observations, or by including a lead time dependent variable in the regression, although this is outwith the remit of this study.

The accuracy at the discrete quantiles is measured via pinball loss, and is shown on Figure 7.7 for both the cross validation and testing dataset over all lead times considered. There is clearly a wide range of sharpness characteristics across the BMUs, which is down to the local wind speed characteristics of each BMU. For example, sites characterised by complex and highly variable wind conditions will obviously be more difficult to forecast. However, looking at the average results the characteristics are very similar for the cross-validation and testing case which shows that the models on average

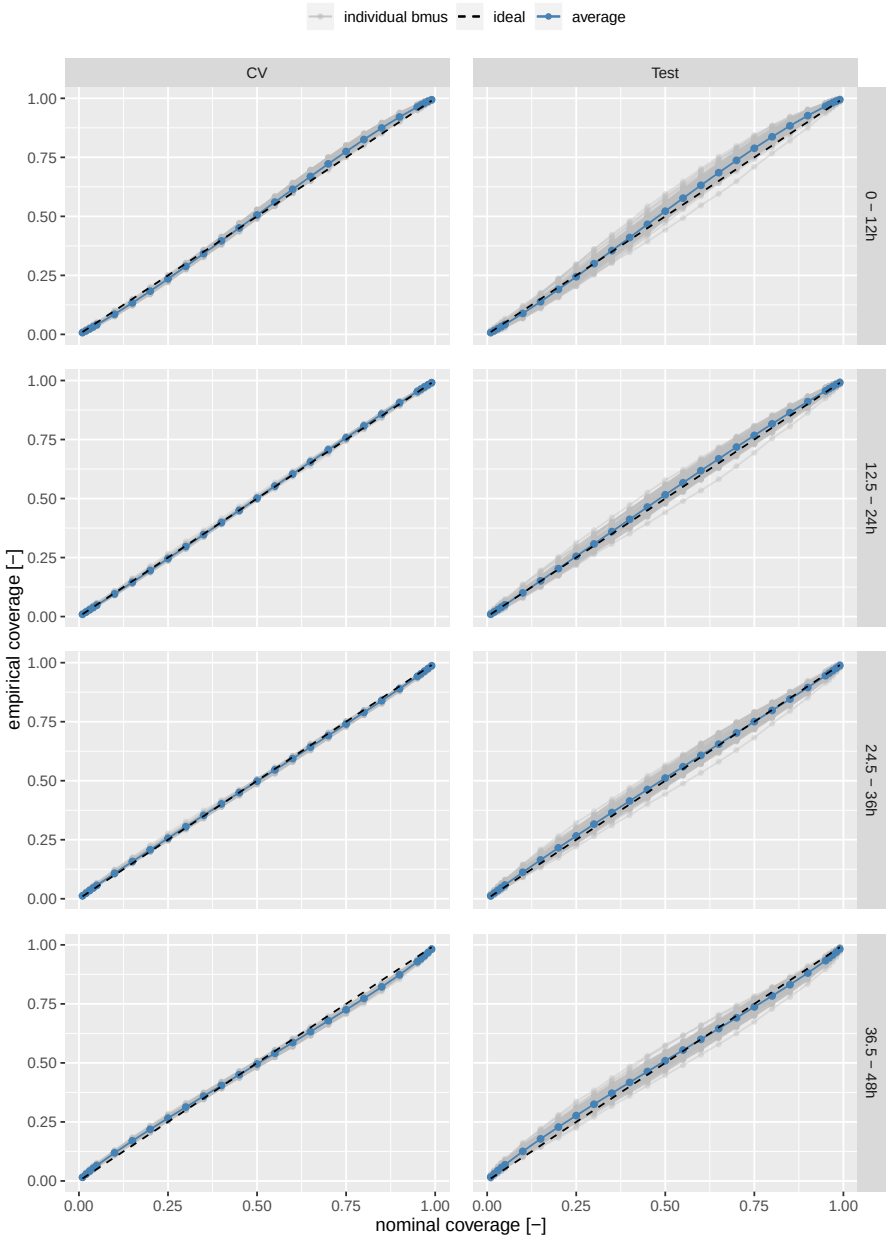


Figure 7.6: Univariate evaluation: reliability diagrams in lead time groups during both cross validation (CV) and testing datasets

generalise well to unseen data. Taking the median quantile in both cases we can see that the average pinball loss is around 5% of estimated capacity, which translates to a mean absolute error of 10% for the median quantile over all lead times. As expected, the average pinball loss increases by lead time group, the results of which can be found

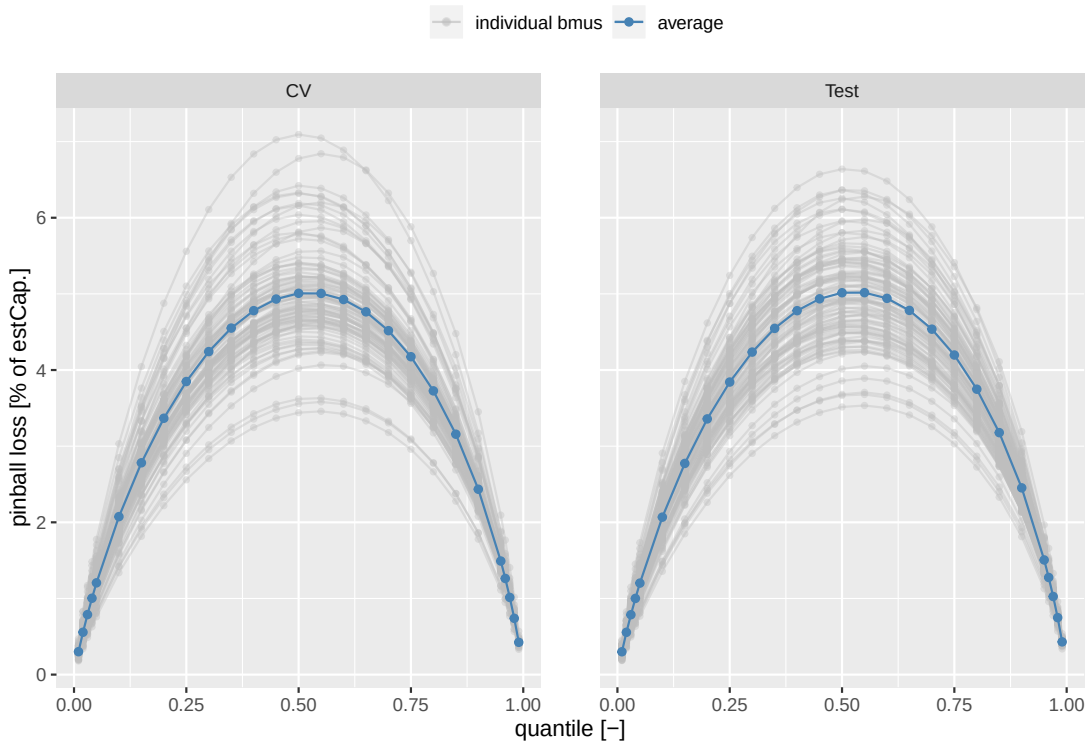


Figure 7.7: Univariate evaluation: pinball loss over all lead times during both cross validation (CV) and testing datasets. The average pinball loss is very similar in both cases which means that the model generalised well to unseen data

in Appendix Figure D.1.

As discussed previously, the PIT transformed variables are key to the quality of the copula model used to generate scenario forecasts. Therefore, we also evaluate the calibration of the approximated CDFs, i.e. the continuous distribution, using PIT histograms. These are presented in Figure 7.8 for a subset of BMUs during cross validation. It is these PIT transformed variables that are used to define the copula for issuing scenario forecasts over the test dataset. The figure shows that the distribution forecasts are well calibrated at these sites; there are only minor deviations from the ideal level; these results are mirrored by the reliability diagrams discussed previously, but importantly show the calibration of the tail regions more clearly. The subset of BMUs are chosen so we have a view of the results throughout the GB network: onshore in Scotland and offshore off the east and west coast of GB. The PIT histograms on a

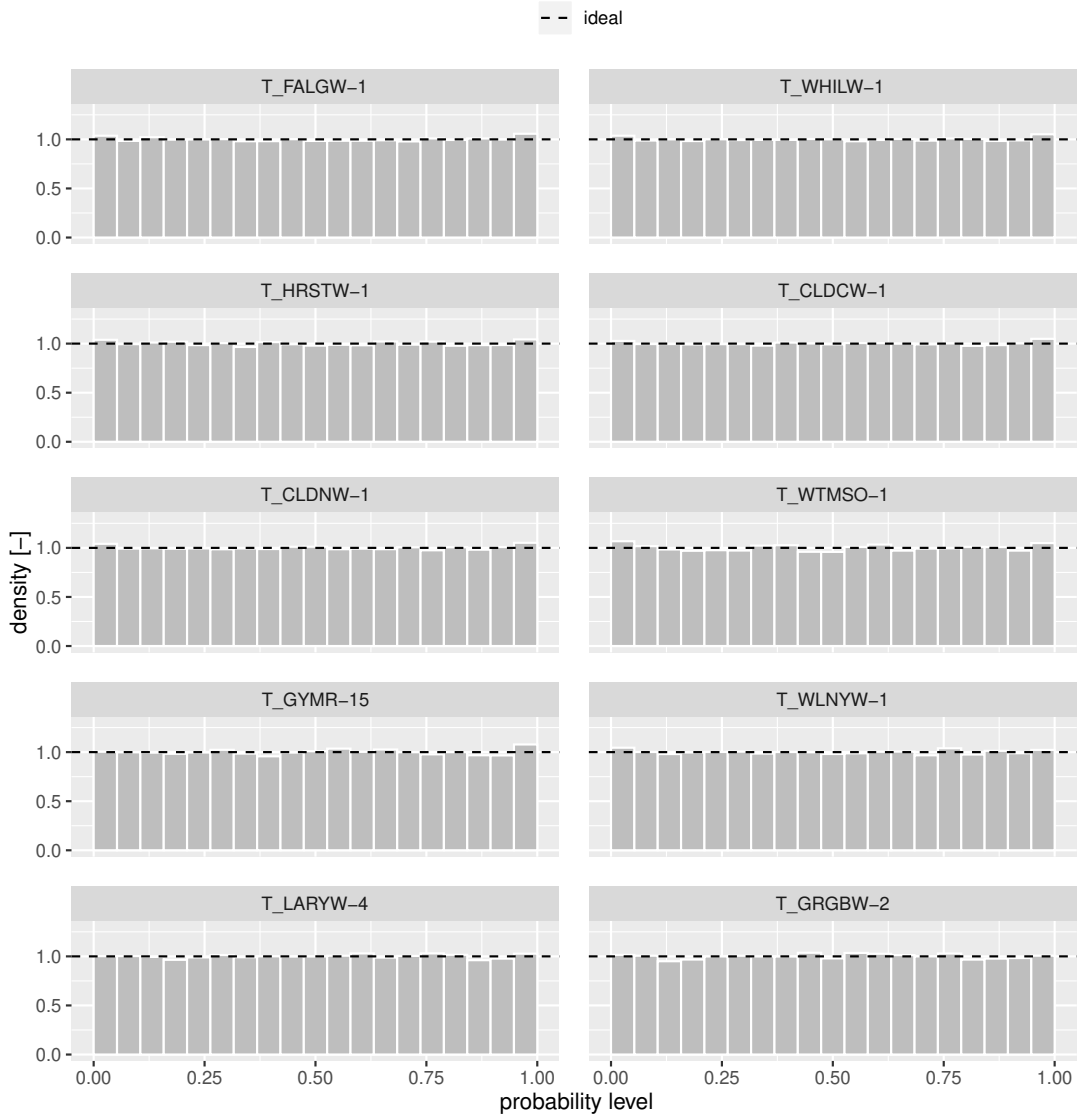


Figure 7.8: Univariate evaluation: PIT histograms of a subset of the BMUs considered during cross validation. The calibration of the full distribution has a big impact on the quality of the copula model

complete per-location basis can be found in Appendix Figure D.3.

Another important evaluation point is the calibration the extreme quantiles located in the tails of the distribution. Although these are included implicitly in the PIT histogram, we focus explicitly on this because, as discussed in Section 2.3.2, the tails are often the most important area of interest for decision makers. In Figure 7.9 quantile bias plots are given for both the upper and lower tail region over cross validation; the

dynamic tail interpolation method results in relatively well calibrated extreme quantiles considering the simplicity of the method, when compared with calibration results from other state-of-the-art approaches [178]. However, it is important to emphasise that the tails in the current work are merely interpolated and will likely be improved by using techniques from extreme value theory. Additionally, in this case we focused only on the calibration of the extreme quantiles (beyond the p1 and p99 quantiles); future work should also consider sharpness as well, where the conditional truncated generalized Pareto distribution has been very recently proven successful [178].

Example density forecasts visualised as fan plots are shown in Figure 7.10 for the same subset of BMUs at a single issue time. This plot intuitively shows why mod-

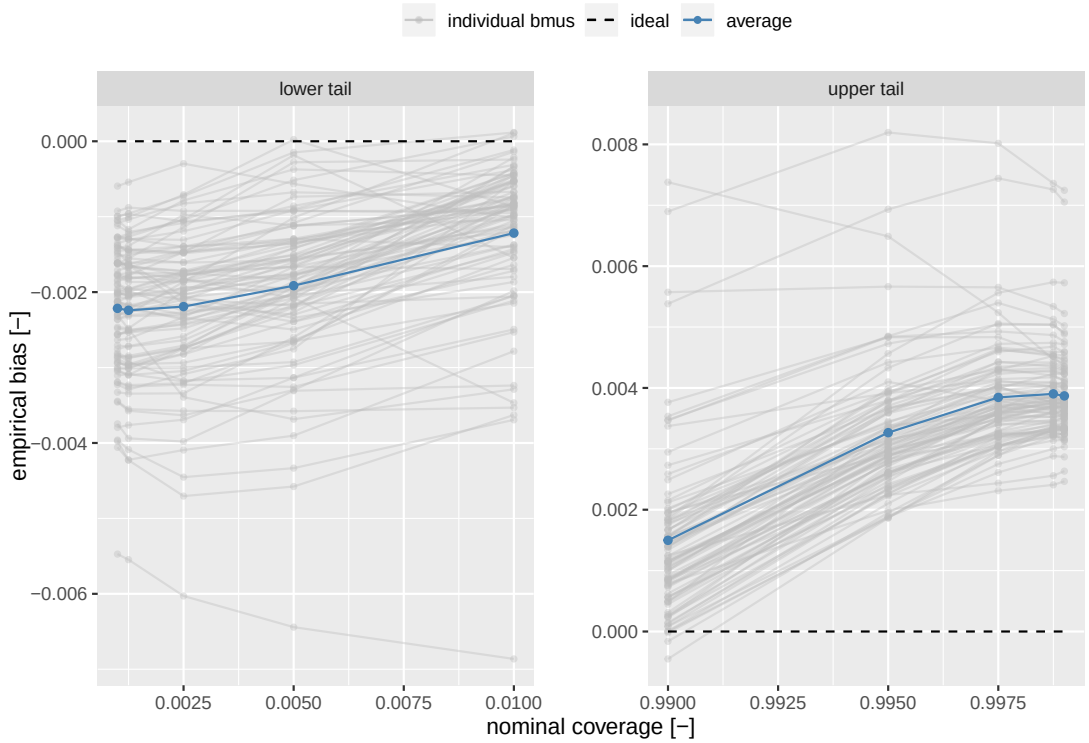


Figure 7.9: Univariate evaluation: quantile bias diagrams of the extreme quantiles over all lead times during cross validation. We show $\alpha \in \{0.001, 0.00125, 0.0025, 0.005, \text{and } 0.01\}$ for the lower tail, the maximum of the set is modelled using quantile regression and the rest are found using a dynamic exponential interpolation method. The mirror image of these quantiles are plotted for the upper tail. Given the scale of the bias, the extreme quantiles are relatively well calibrated considering the simplicity of the interpolation method.

elling the dependencies in the multivariate distribution is essential; there is clearly high spatio-temporal relationship between the BMUs in close proximity at this issue time, as expected. Additionally, there is also high uncertainty at the Westermost Rough BMU (T_WTMSO-1), where the forecast wind speeds are in the mid-range of the wind farm power curve throughout the lead times. In fact, this BMU is gives the highest pinball loss out of all considered in both cross validation and testing, indicating that the local atmospheric conditions are not captured well by the global weather models at this site.

7.2.2 Multivariate: Scenario Forecasting

Here, we detail the results of the multivariate power forecasting task, i.e. the scenario forecasts generated via sampling from the Gaussian copula. To refresh, four dependency models are tested: *Independence* is the benchmark model, *Temporal* is where only the temporal correlation is considered between lead times, *Spatiotemporal* is where the full dependence between each lead time and location is modelled, and finally *RS-Spatiotemporal* is the regime switching covariance model where the dependency is conditional on two wind direction dominated regimes — easterly or westerly.

For each issue time 500 samples are taken, which is not a great amount considering the dimension of the distribution to fully explore the space. However, applying the scoring rules with more samples was not possible given time and computing resource constraints. This practical restraint is discussed in [191], where it is discussed that there is therefore an associated uncertainty contribution to the score when the number of samples is small relative to the dimension. This is another motivating factor in the use of bootstrapping to estimate the sampling distribution of the score in this chapter. Also, results in [138] show that the number of samples and the nature of the difference between the true and modelled correlation structure impacts the stabilisation of the average energy score results; among other points, the authors recommend taking as many samples as computationally feasible to achieve the best discriminatory power.

In Table 7.1 the multivariate scores are presented. It is important to emphasise here, that the differences between the models are solely in the dependency structure, and therefore the difference in evaluation scores is down to this and random sampling

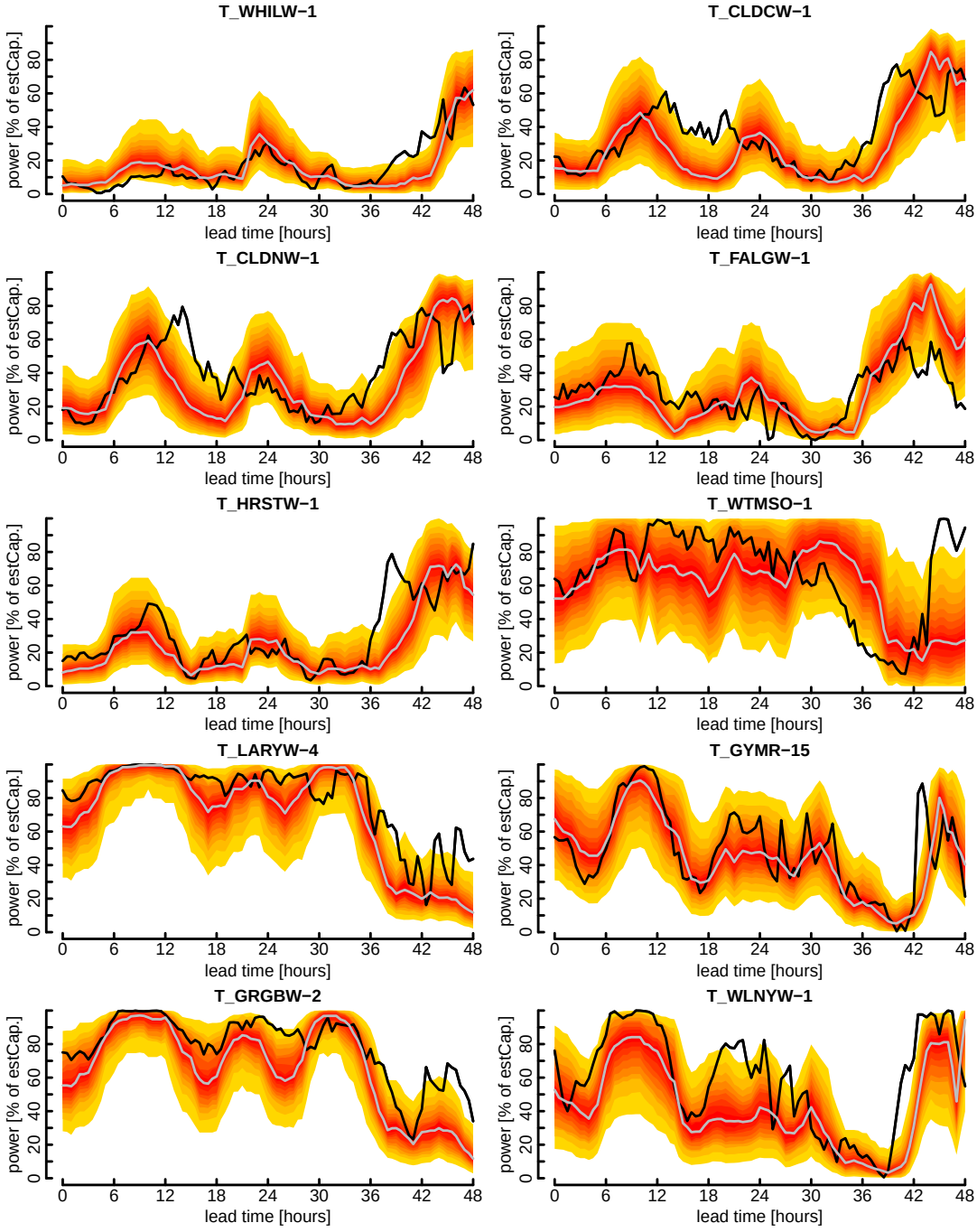


Figure 7.10: Univariate density forecasts as fan plots at the subset of BMU locations issued at 2018-02-25 12:00 UTC. The same legend as Figure 6.3a applies

Table 7.1: Multivariate forecast evaluation results. Best results during testing and cross-validation are in bold. The evaluation results are based on 500 scenarios at each issue time

Dependency	Data	ES	VS-0.5
independence	CV	8.2270	1252757
	Test	9.3489	1929561
temporal	CV	8.2252	1252010
	Test	9.3471	1928611
spatiotemporal	CV	8.2201	1250580
	Test	9.3422	1926398
RS-spatiotemporal	CV	**8.2198**	**1250480**
	Test	**9.3413**	**1926087**

variation. The first thing to note is that there is very little relative change between the scores for the different dependency models. It has been discussed in the literature that the energy score has low discriminatory power, especially in high dimensions, which is clearly evident here [136, 191]; the *Spatiotemporal* model generates completely different trajectories, with more realistic temporal and spatial characteristics, compared to the *Independence case*, but the relative change is very small.

For the variogram score the small relative change is likely due to the fact we employ no weighting. Therefore, a lot of the contributions to the score comes from pairwise differences that we know are not correlated, e.g. pairwise difference between components of the vector that are far away in time and/or space.

However, as discussed in [138] a small relative change does not exclude the possibility that the forecasts are significantly different, in a statistical sense. The score improvements are plotted in Figure 7.11 against the independence benchmark. Here, a simple block-bootstrapping approach is used to estimate the significance in the forecast improvement [139], where issue times are split into non-overlapping blocks of 7-days length. The blocks are then re-sampled with replacement and then forecast improvement is determined. This process is repeated 1000 times to estimate the sampling variation of the score improvement in Table 7.1.

Figure 7.11 shows that the skill scores are indeed significantly different from zero, for the three methods of dependence compared to the independence case. There is also marginal improvement using the regime-switching method over the *spatiotemporal* case;

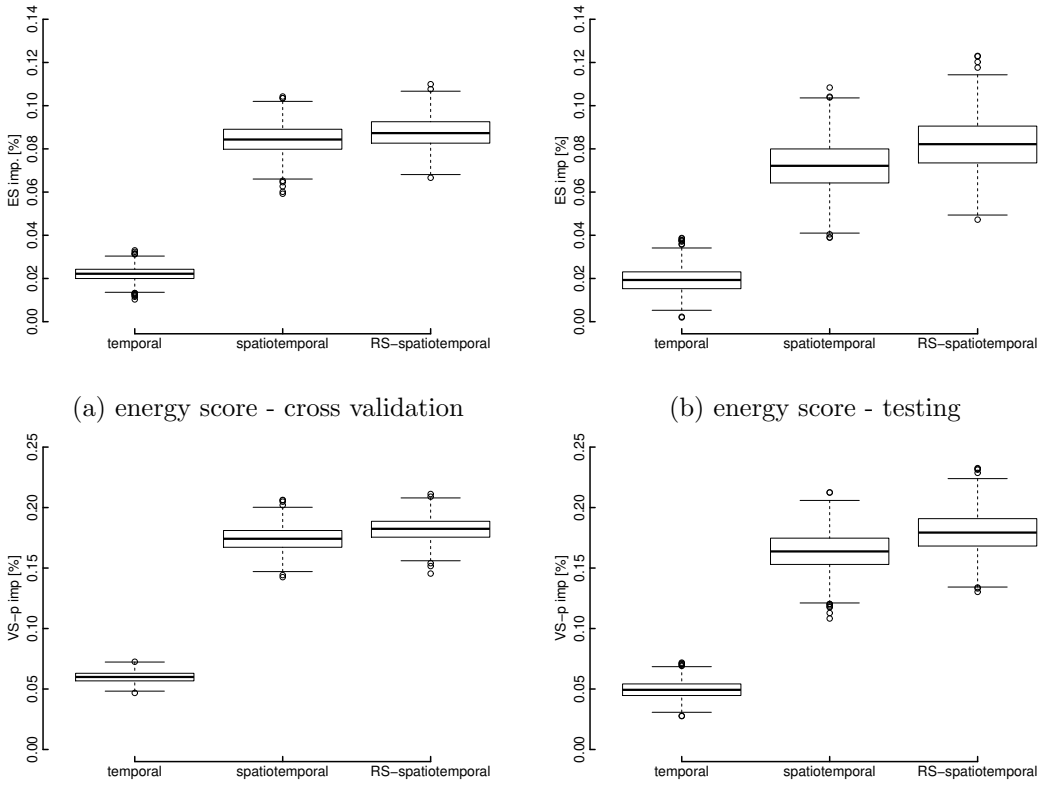
Table 7.2: Multivariate evaluation: Diebold-Mariano test statistics based on the energy score differential on the out-of-sample cross validation data, where p values are shown in bold

	independence	temporal	spatiotemporal	RS-spatiotemporal
independence	-	6.465 **<0.001**	12.447 **<0.001**	12.282 **<0.001**
temporal	-6.465 **<0.001**	-	8.144 **<0.001**	8.455 **<0.001**
spatiotemporal	-12.447 **<0.001**	-8.144 **<0.001**	-	0.442 **0.659**
RS-spatiotemporal	-12.282 **<0.001**	-8.455 **<0.001**	-0.442 **0.659**	-

however this is clearly not significant. Although results are promising, and perhaps higher fidelity definition of the governing regimes would be beneficial. Accounting for the spatial components in the multivariate distribution is seen to be significantly better than the purely temporal case. This is due to the nature of the considered sites on the GB network; wind farms are typically clustered together in groups, as shown in Figure 7.4, with complex terrain and atmospheric conditions. The two scoring rules are also in agreement with the evaluation of the different dependencies, with the variogram score consistently giving a marginally higher improvement over independence. Lastly, we see a greater sampling uncertainty in the test dataset due the smaller sample size.

These conclusions are also supported by the Diebold-Mariano test, where the number of issue time lags considered for calculating the autocovariance of the error differential is 7 days. The test statistic and corresponding p-value are given in Table 7.2 over the out-of-sample forecasts in the cross-validation dataset. The results show that the null hypothesis of equal predictive performance is rejected in all combinations of the dependency, except for the *spatiotemporal* and *RS-spatiotemporal* test, which concurs with the results in Figure 7.11a. The corresponding results for the test evaluation are similar in characteristic and can be found in Appendix Table D.2.

The covariance matrix that describes the dependency structure in the Gaussian copula is shown on Figure 7.12a for the *spatiotemporal* case, again over a subset of BMUs to show the resulting characteristics. The parametric model is clearly successful in regularising the dependency, most of the covariance is set to approximately zero.



(a) energy score - cross validation

(b) energy score - testing

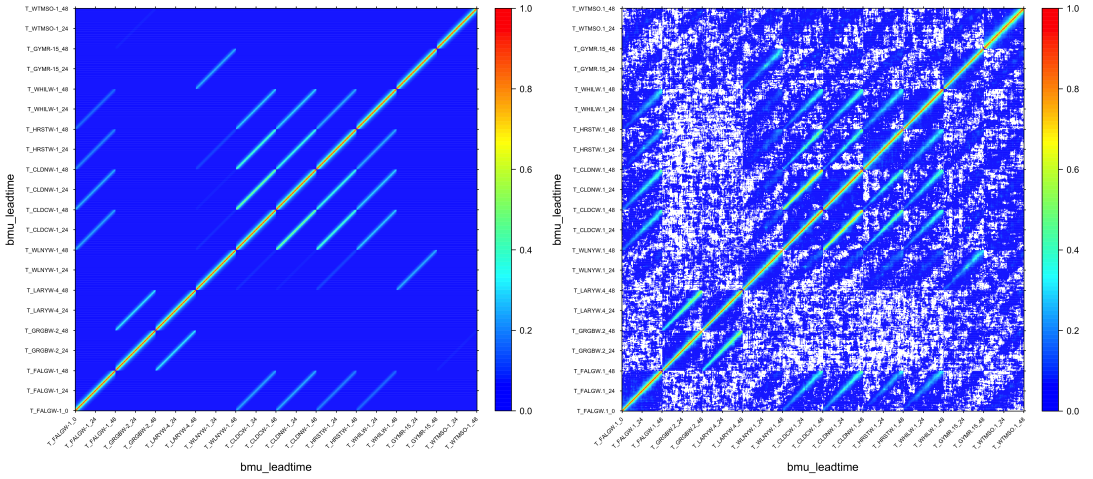
(c) variogram score - cross validation

(d) variogram score - testing

Figure 7.11: Multivariate evaluation: boxplots showing the block-bootstrap sample distributions of multivariate score improvement for the different dependencies tested, over the *independence* benchmark. The improvements are relatively small, but significantly different from zero. There is marginal improvement using the regime-switching method over the *spatiotemporal* case; however this is not significant. Note that the variogram score is unweighted, subsequently the improvement is relatively small here as well. Both scores are in agreement in terms of forecast performance

However, we see that there is a strong temporal correlation of errors in consecutive lead times, as expected. Also, the spatial correlations relatively are strong at concurrent lead times, shown by the block of onshore wind farms based in central Scotland.

The sample covariance estimate is shown in Figure 7.12b for comparison, and we can see that clearly the sample estimate is noisy, and often results in (slightly) negative estimates. This also shows that the parametric model captures the characteristics of the empirical covariance relatively well, although, as configured cannot capture the clear growth in correlation as the lead times progress at each site. Also, we can see



(a) Parametric estimate

(b) Sample estimate

Figure 7.12: Spatiotemporal covariance matrices on a subset of BMUs, the full matrix used to generate scenario forecasts on the testing dataset is shown on Figure 7.13. The parametric model regularises the covariance structure, compared to the sample estimate. Figure 7.12b is used to learn the coefficients for the parametric model, and white space indicates where the sample covariance is negative. Please refer to Figure 7.14 for higher resolution axis labels

that the spatiotemporal correlation between the two Clyde BMUs is slightly under that of the empirical findings.

The full covariance matrix for the *spatiotemporal* case across all lead times and BMUs is shown on Figure 7.13. This plot also shows perhaps why we see a small relative change in the energy score; at this resolution we can see that most of the pairwise lead times and locations considered are uncorrelated. However, the correlations that are present are very important for generating realistic and useful scenario forecasts.

To further understand the differences between the spatiotemporal and the regime-switching dependency models, the fitted coefficients from Equation 7.6 over the cross-validation dataset are given in Table 7.3, along with (approximate) 95% confidence intervals based on linearised standard errors. Although it is difficult to gauge the impact of the differences in the absolute values of these coefficients, we can draw a number of conclusions from the table. Firstly, there is clear separation in the coefficient confidence interval ranges; only the confidence intervals for the β parameter in the westerly regime

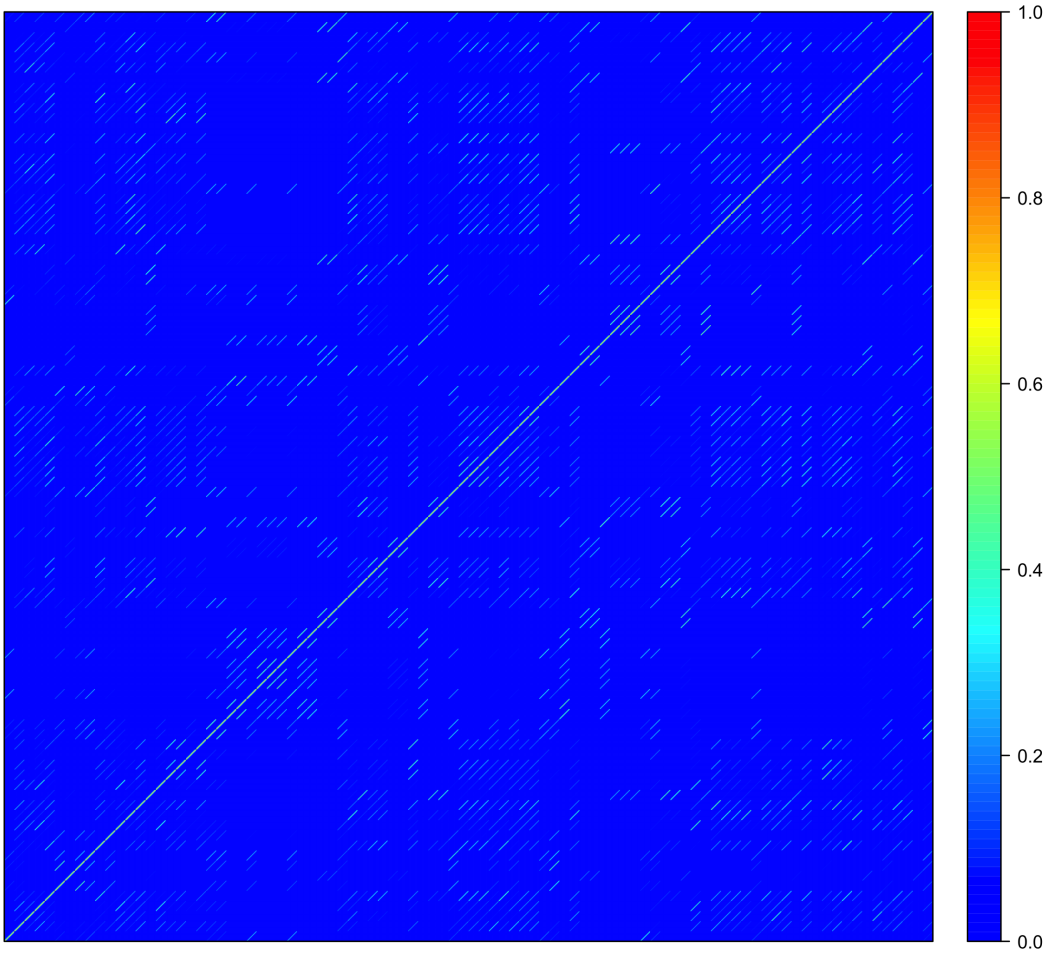


Figure 7.13: Parametric covariance matrix of all BMUs used to generate scenario forecasts on the testing dataset

and the typical spatiotemporal case overlap. However, it is important to note that the intervals are only intended as a guide; it was not possible to explore the sampling distribution of the coefficients, via bootstrapping for instance, due to the associated computational expense.

More specifically, Table 7.3 also shows that over the testing dataset the temporal coefficients, τ_t and a, are greater in the westerly regime, meaning that temporal correlation strength of the errors is weaker than in easterly regime because of the model

Table 7.3: Fitted coefficients (est.) for the parametric covariance models based on Equation 7.6, which describes the uncertainty dependency used to issue scenario forecasts over the test dataset, along with approximate 95% confidence intervals

coefficient	spatiotemporal			RS-spatiotemporal easterly			RS-spatiotemporal westerly		
	2.5%	est.	97.5%	2.5%	est.	97.5%	2.5%	est.	97.5%
a	0.4612	0.4617	0.4623	0.4033	0.4041	0.4050	0.4788	0.4795	0.4802
β	1.1816	1.1828	1.1840	1.2261	1.2280	1.2299	1.1800	1.1813	1.1827
τ_k	77.7534	77.8096	77.8659	79.7072	79.8041	79.9010	76.9494	77.0130	77.0766
τ_t	0.7220	0.7223	0.7226	0.6717	0.6722	0.6727	0.7507	0.7511	0.7515
ν	0.4136	0.4139	0.4143	0.4474	0.4479	0.4484	0.3981	0.3985	0.3989

formulation. The spatio-temporal interaction term β is also stronger in the easterly regime, and the purely spatial correlation strength is marginally weaker. These results are also found in Figure 7.14, again on the subset of BMUs, where the difference between the covariance in the easterly and westerly regimes is visualised.

An important aspect to emphasise is the computational cost of calculating the multivariate skill scores. Although components of the entire methodology are potentially computationally expensive, such as fitting the marginals or issuing the scenario forecasts in this sort of high dimensional problem, really it was the multivariate forecast evaluation that was the bottleneck in the process; it took approximately 1 hour to evaluate *a single issue time* with 500 samples on one CPU, using the ES and the unweighted VS-p. Parallelization and cloud computing were used throughout the study to manage such aspects, e.g. by evaluating unique issue times in parallel, however research is required in the area of multivariate scoring rules for metrics with better discriminatory power and faster computational characteristics.

Finally, example scenario forecasts visualised in spaghetti plots are shown in Figure 7.15, for the same subset of BMUs and lead time as Figure 7.10. Although these plots are difficult to interpret directly, they are useful for further processing and for driving stochastic power system optimisation tools, where preserving the spatiotemporal characteristics of the generation forecasts is essential.

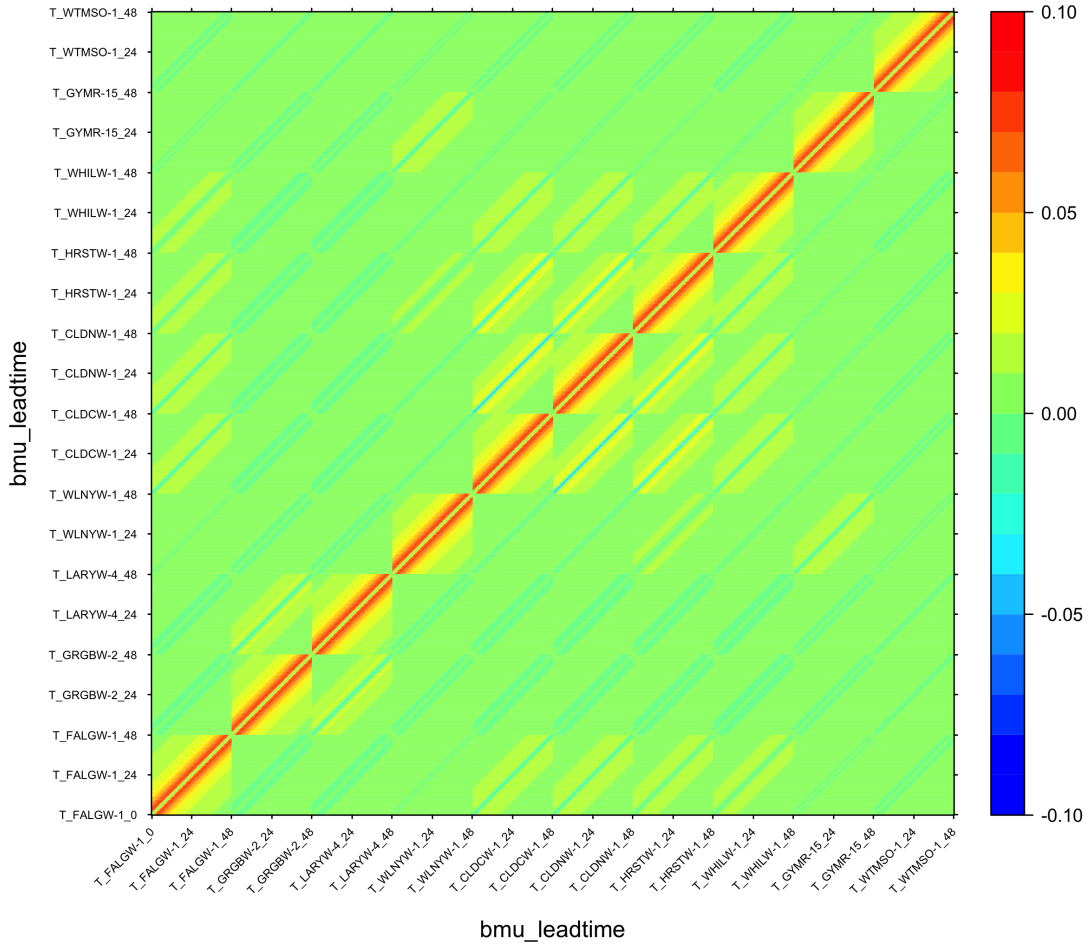


Figure 7.14: Difference between the regime-switching covariance matrices (easterly minus westerly) of a subset of BMUs during the testing dataset. The temporal dependency persists for longer in the easterly regime and the spatial dependency is somewhat similar

7.3 Conclusions & Future Work

This chapter demonstrates a methodology for multivariate wind power forecasting, with a novel very-high dimensional case-study. We leverage parametric models of covariance to model a very-high dimensional distribution, generating scenario forecasts for 92 of the wind balancing units on the GB network up to two days ahead, in a Gaussian copula framework. There is clear significant improvement in the multivariate skill

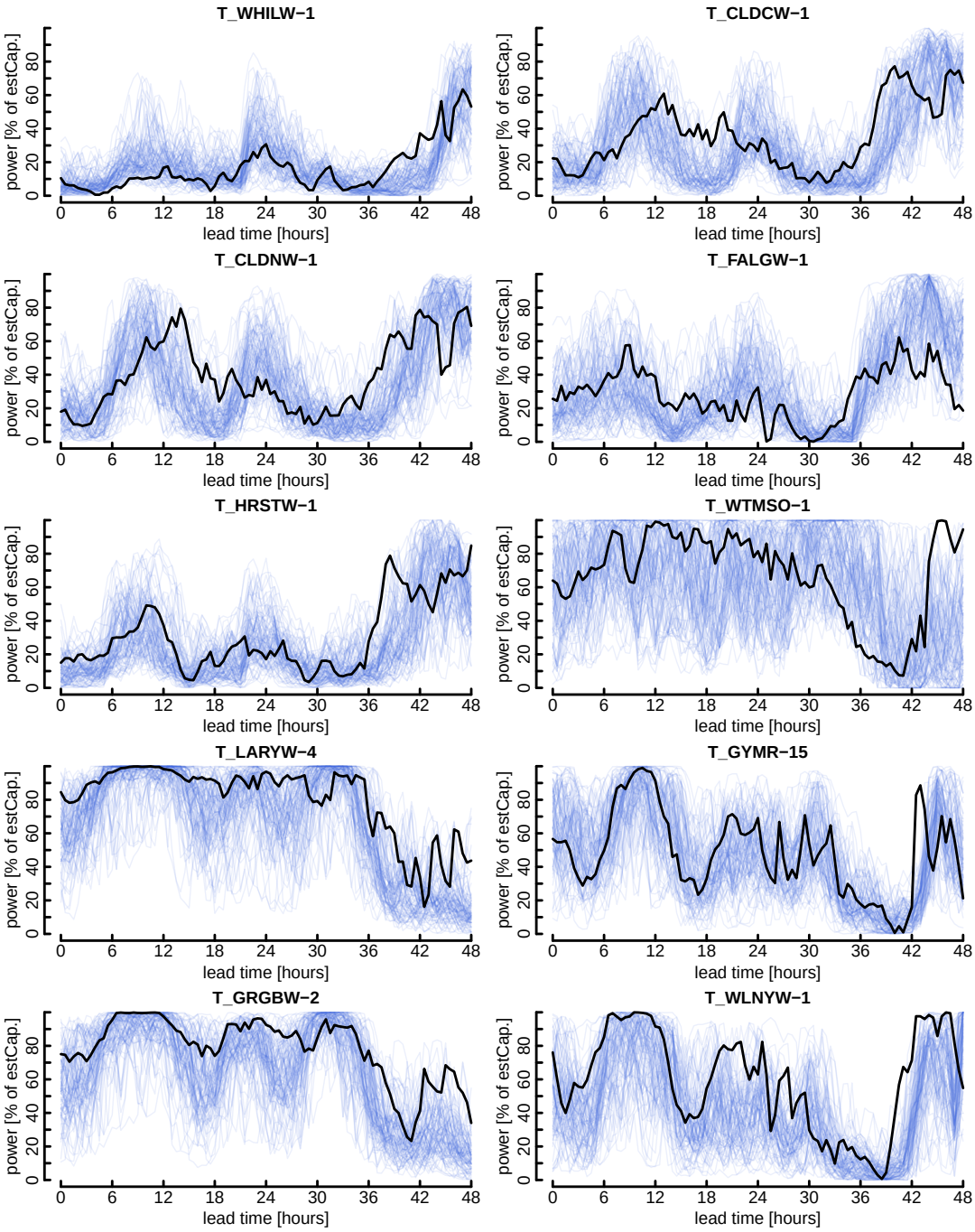


Figure 7.15: Multivariate scenario forecast as a spaghetti plot with 100 samples at the subset of BMU locations. The forecast is issued at 2018-02-25 12:00 UTC, and scenarios are blue and the corresponding measurement black

scores of the spatio-temporal covariances compared to the independence and purely temporal covariance cases, which demonstrates the importance of including the spatial information at these scales. The novelty in this chapter comes from the demonstration of the methodology at the space-time resolutions needed for future end-users of the forecasts, such as transmission system operators, and in explicitly quantifying the value of exploiting the spatial covariances in terms of strictly proper skill scores.

We also evaluate a regime-switching covariance model, where the covariance matrix is conditional on regimes based on clustered average day-ahead forecasts of the wind components. This gives two regimes based on prevailing easterly or westerly average wind direction. Although the results suggest a benefit, there is no statistically significant improvement against the typical spatiotemporal case. Importantly, the multivariate evaluation scores are the computational bottleneck in the process, as found elsewhere in the literature [191], which has two practical implications: the evaluation is based on a relatively small set of scenarios compared to the dimension of the problem, and there is no weighting applied to the variogram based score. Future work should consider the impact of both these limitations on the evaluation results, as well as a higher fidelity definition of the regimes. For example, principle component analysis, self organising maps, and more diverse meteorological data than simple wind speed forecasts at each wind farm, have proven successful for atmospheric regime modelling in related studies [96, 197]. Another interesting extension would be to evaluate the mean of the scenarios as a point forecast to gauge any improvement in forecasting skill against the median of the marginal density forecast, given the introduction of the spatio-temporal dependency structure in the multivariate forecast.

Chapter 8

Conclusions

Forecasting is now an essential component of power system operations. Participants in electricity markets, system operators, asset owners, and all who make decisions based on the future generation require these tools. Probabilistic forecasts quantify the uncertainty associated with a forecast, and are essential for both risk management and for participants in asymmetrically penalised electricity markets, regardless of their risk appetite. In this work, methodologies for generating forecasts of wind power and offshore access conditions are developed, linked by the growing challenge of modelling and/or leveraging high dimensional data.

It is shown that exploiting high dimensional data can improve forecast skill, provide opportunities for novel forecasting products, and that the high dimensional forecast outputs required at the system-wide scale are feasible. The latter point is especially relevant to utilities and transmission system operators, and will be of growing importance as the penetration of wind power increases on the network.

In Chapter 4 the value in post-processing ultra high resolution weather forecasts for power prediction is demonstrated, by improving the accuracy of point forecasts at Horns Rev I. The case study evaluates the performance of wind speed and power forecasting at a single turbine in the array in an exploratory analysis, and then investigates the performance of overall wind farm power prediction. The exploratory investigation uses single-input models to explicitly characterise the value in temporal smoothing of Numerical Weather Prediction (NWP) sources, which is shown to improve

performance. For wind farm power forecasting, using information from both standard and ultra-high resolution NWP sources proved to give significant increases in accuracy, especially with a feature selection stage, compared to using information content solely from either single source which are comparable to each other.

Turbine-level data provides valuable information about how a wind farm responds to different weather conditions, and the nature of forecast errors, which is not accessible when only considering wind farm total power production. In Chapter 5, two methods for improving wind power forecasting by leveraging data from individual wind turbines are evaluated. The first is a feature engineering approach whereby deterministic forecasts for individual turbines are aggregated and used as supplementary input variables to a conventional wind farm-level model. The second is a hierarchical bottom-up probabilistic approach which forecasts the joint predictive distribution of generation from all turbines in a copula framework, which is then used to produce a wind farm-level forecast. Both methods are shown to increase forecast skill compared to two highly competitive benchmarks, particularly at the site with complex terrain.

In Chapter 6, a novel forecasting tool for offshore operations is described. We forecast safety-critical conditions during transfer of personnel during offshore operations, and demonstrate its potential with a case study at an east coast wind farm in the UK. The proposed access forecasts predict vessel motion during transfer, accounting for weather uncertainty, up to 5-days ahead. Sharp and calibrated density forecasts of peak wave period and significant wave height are generated by post-processing NWP, with boosted generalised additive models for location, scale, and shape outperforming non-parametric methods. Scenario forecasts of these variables are then produced using the Gaussian copula to model temporal dependence and used as inputs to a data-driven vessel motion model. A method of visualisation of these forecasts is also suggested to best communicate the information content for end users.

Chapter 7 demonstrates a methodology for multivariate wind power forecasting, with a case study at a system-wide scale. Parametric models of covariance are used to model a very-high dimensional multivariate Gaussian distribution in a copula framework, enabling scenario forecasts for 92 wind balancing units on the GB network. There

is statistically significant improvement in the multivariate skill scores of the space-time covariance model compared to the independent and purely temporal covariance cases. A regime-switching approach is also proposed, where the covariance is conditional on regimes based on clustered average day-ahead wind component forecasts. This gives two regimes based on prevailing easterly or westerly average wind direction. Although the results are promising, there is no statistically significant improvement of the regime switching method against the spatiotemporal case; the simple regime definition and/or the computational compromises required for forecast evaluation, such as the small number of scenarios relative to the dimension of the problem, are possible reasons for this. The contribution of Chapter 7 is in the demonstration of the methodology at the space-time resolutions needed for future end-users of the forecasts, such as transmission system operators, and in explicitly quantifying the value of exploiting the spatial covariances in terms of strictly proper skill scores.

Future work related to the specific technical challenges of each chapter can be found in the relevant sections. However, a common theme is the use of dynamic dependency structures between forecast uncertainty in space and/or time, motivated by the idea that the nature of the dependency is not constant and related to large-scale meteorological processes. Although a simple approach is explored in Chapter 7, higher fidelity methods are clearly required.

More generally, although we focus on probabilistic forecasting methods, there is a lack of uptake of this uncertainty information in practice, even for low-dimensional problems; more work is needed in the communication of probabilistic forecasts to end-users, in terms of both the value and the underlying statistical content. We can begin to address this challenge via applied demonstration studies, the tailoring of forecasts to specific use-cases, and close collaboration between forecast users and forecasters.

Chapter 9

Published Work

Journal papers:

- C. Gilbert, J. Browell, and D. McMillan, "Probabilistic access forecasting for improved offshore operations", International Journal of Forecasting, vol. 37, no. 1, pp. 134-150, 2021 (open access)
- C. Gilbert, J.W. Messner, P. Pinson, P.J. Trombe, R. Verzijlbergh, P. van Dorp, and H. Jonker, "Statistical post processing of turbulence-resolving weather forecasts for offshore wind power forecasting", Wind Energy, vol. 23, no. 4, pp. 884-897, 2019
- C. Gilbert, J. Browell, and D. McMillan, "Leveraging turbine-level data for improved probabilistic wind power forecasting", IEEE Transactions on Sustainable Energy, vol. 11, no. 3, pp. 1152-1160, 2019, (open access)
- I. Würth, L. Valldecabres, E. Simon, C. Möhrlen, B. Uzunoglu, C. Gilbert, G. Giebel, D. Schlipf, and A. Kaifel, "Minute-scale forecasting of wind power — results from the collaborative workshop of IEA Wind task 32 and 36", Energies, Published Online, 2019, (open access)

Conference Papers:

- J. Browell, C. Gilbert, "ProbCast: Open-source production, evaluation and visualisation of probabilistic forecasts", IEEE PMAPS Conference, Liege, Belgium,

2020 (accepted)

- C. Gilbert, J. Browell and D. McMillan, "A data-driven vessel motion model for offshore access forecasting", IEEE OCEANS, Marseille, 2019

- C. Gilbert, J. Browell and D. McMillan, "Visualisation of probabilistic access forecasts for offshore operations", Journal of Physics: Conference Series, Wind Europe, Bilbao, 2019

- C. Gilbert, J. Browell and D. McMillan, "A Hierarchical Approach to Probabilistic Wind Power Forecasting", IEEE PMAPS Conference, Boise, USA, 2018

- J. Browell, C. Gilbert and D. McMillan, "Use of Turbine Level Data for Improved Wind Power Forecasting", IEEE PowerTech Conference, Manchester, UK, 2017

- J. Browell and C. Gilbert, "Cluster-based Regime-switching AR for the EEM Wind Power Forecasting Competition", 14th International Conference on the European Energy Market (EEM), Dresden, Germany, 2017. Invited Paper

Appendix A

Statistical Post-processing of Turbulence Resolving Weather Forecasts

This supplementary material contains extra results from the case study presented in Chapter 4. We present the full list of initial input features and the reduced set of selected features used for the wind farm level power prediction.

Table A.1: List of input features for wind farm level power prediction using either ECMWF or GRASP data, or both sources as in the combined case. Specific heights and locations are indicated in square brackets for before, after, and in the combined feature selection case. TB is short for turbine, RA means rolling average, and RV means rolling variance. The FFT features were calculated over a 400 minute window on the disk averaged wind speed and the features extracted from the high frequency band.

Source	Feature	Before	After	After (combined)
GRASP	Air density	[30m, 50m, 70m, 90m, 110m]		
	Temperature	[30m, 50m, 70m, 90m, 110m]		
	Wind Direction (wd)	[30m, 50m, 70m, 90m, 110m]		
	Wind Speed (ws)	[30m, 50m, 70m, 90m, 110m]	[70m, 90m]	[70m]
	Disk averaged ws	[TB 1, 2, ..., 79, all]		[TB 78]
	RA ws 300 min	[30m, 50m, 70m, 90m, 110m]	[70m]	
	RA ws 400 min	[30m, 50m, 70m, 90m, 110m]	[70m]	
	RA ws 500 min	[30m, 50m, 70m, 90m, 110m]		
	RV ws 300 min	[30m, 50m, 70m, 90m, 110m]		
	RV ws 400 min	[30m, 50m, 70m, 90m, 110m]		
	RV ws 500 min	[30m, 50m, 70m, 90m, 110m]		
	RA disk ws 300 min	[TB 1, 2, ..., 79, all]	[TB 77, 79]	[TB 79,]
	RA disk ws 400 min	[TB 1, 2, ..., 79, all]	[TB 78]	[TB 78, 79]
	RA disk ws 500 min	[TB 1, 2, ..., 79, all]	[TB 59, 67, 78, 79]	[TB 78,]
	RV disk ws 300 min	[TB 1, 2, ..., 79, all]		
	RV disk ws 400 min	[TB 1, 2, ..., 79, all]		
	RV disk ws 500 min	[TB 1, 2, ..., 79, all]		
	Rolling FFT mean	[TB 1, 2, ..., 79, all]		
	Rolling FFT sum	[TB 1, 2, ..., 79, all]		
ECMWF	Air density	[30m, 50m, 70m, 90m, 110m]		
	Temperature	[30m, 50m, 70m, 90m, 110m]		
	Wind Direction (wd)	[30m, 50m, 70m, 90m, 110m]		
	Wind Speed (ws)	[30m, 50m, 70m, 90m, 110m]		
	RA ws 300 min	[30m, 50m, 70m, 90m, 110m]	[70m, 90m]	[70m, 90m]
	RA ws 400 min	[30m, 50m, 70m, 90m, 110m]	[50m, 70m, 90m]	[50m, 70m]
	RA ws 500 min	[30m, 50m, 70m, 90m, 110m]	[50m]	[30m, 50m, 90m]
	RV ws 300 min	[30m, 50m, 70m, 90m, 110m]		
	RV ws 400 min	[30m, 50m, 70m, 90m, 110m]		
	RV ws 500 min	[30m, 50m, 70m, 90m, 110m]		

Appendix B

Leveraging Turbine-level Data for Probabilistic Wind Power Forecasting

This supplementary material contains information on the feature lists and variable importances at both wind farms, the mapped partial covariance at Wind Farm B, details of bivariate copula family selections from the vine copulas, and full deterministic results for both wind farms. Variable importance is a quantitative measure based on the number of times a variable has been selected for splitting in the GBM, with a weighting applied derived from the predictive improvement to the model from each split averaged over all the trees [199]; it highlights the most useful predictor variables.

B.1 Wind Farm B

The feature list at Wind Farm B for all models is contained in Table B.1. This includes the benchmark GBT models, the deterministic and probabilistic models used in the feature engineering method, and the probabilistic turbine models in the bottom-up probabilistic model.

Figures B.1 and B.2 illustrate the importance of selected variables across key quantiles in the probabilistic GBT benchmark and probabilistic feature engineering wind

Table B.1: Wind Farm B: list of input features by model - meteorological inputs at 100m (unless specified)

Model	Features	ID
All benchmarks & individual turbines both deterministic and probabilistic	Wind speed - 10m	ws-10m
	Wind speed	ws
	Wind direction - 10m	wd-10m
	Wind direction	-
	Leading ws (1 to 4 periods)	ws-pl-[1-4]
	Lagging ws (1 to 4 periods)	ws-nl-[1-4]
	Leading wd (-1 to -4 periods)	-
	Lagging wd (-1 to -4 periods)	-
	Average ws (0 to 4 periods)	ws-av-pl
	Average ws (0 to -4 periods)	ws-av-nl
	Average wd (0 to 4 periods)	-
	Average wd (0 to -4 periods)	wd-av-nl
	Standard deviation ws (0 to 4 periods)	-
	Standard deviation ws (0 to -4 periods)	-
	Standard deviation wd (0 to 4 periods)	-
	Standard deviation wd (0 to -4 periods)	-
	WS ratio (100m & 10m)	-
	WS gradient (-1 to 0 period)	-
	Direction difference (100m & 10m)	-
	Time of day splines (4)	-
	Persistence	pers.
Feature engineering wind farm probabilistic model	Layer 1 aggregation (mean)	L1-agg
	L1 turbine forecasts (2,13,11,17,23,25)	L1-T[#]
	L1 turbine forecast variance (all)	L1-var
	Leading L1 aggregation (1, 2, 3)	L1-agg-pl[#]
	Lagging L1 aggregation (-1)	L1-agg-nl1
	Average L1 aggregation (0 to 4 periods)	L1-agg-av-pl
	Average ws (0 to 4 periods)	ws-av-pl

farm model respectively. From the mapped parametric covariance of turbine 35 at Wind Farm B, shown in Figure B.3, the layout of the wind farm discussed in the results section of the main paper is shown.

Full deterministic results in terms of MAE and RMSE are presented in Tables B.2 and B.3 respectively. These are based on the median quantile (p50) in the final full density forecast and are therefore optimised for MAE.

Copula family selection results are displayed in Table B.4 for the copula vine method.

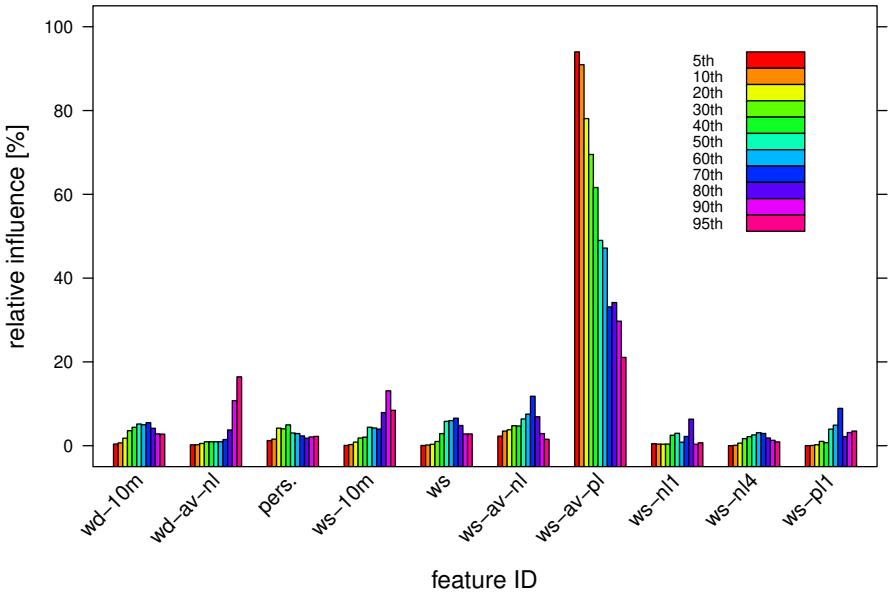


Figure B.1: Wind Farm B: relative influence of input features: WF($\mathbf{x}_t$) benchmark

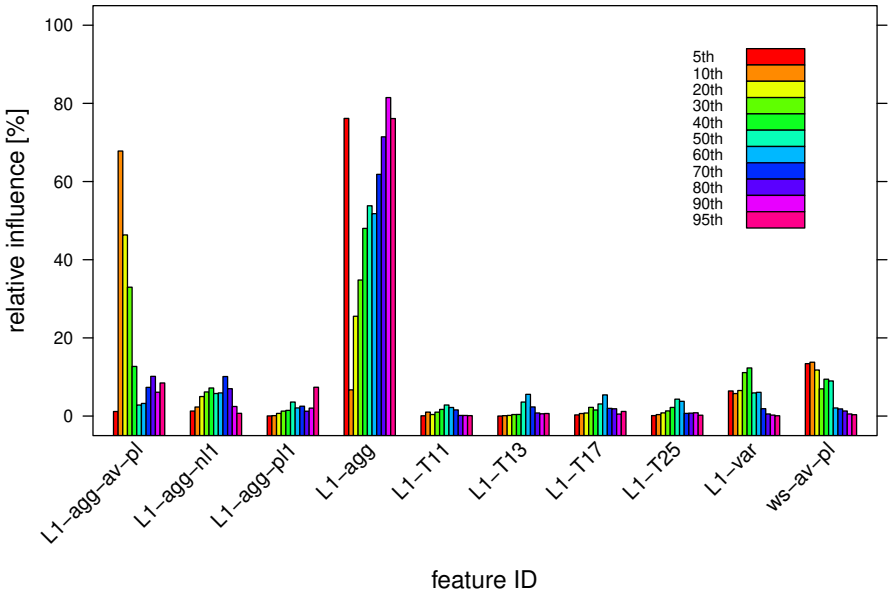


Figure B.2: Wind Farm B: relative influence of input features: WFT($\mathbf{x}_t^{\text{SUP}}$) feature engineering wind farm probabilistic model

Table B.2: P50 MAE Results at Wind Farm B - %Δ indicates improvement compared to specified benchmark above [MAE in % of max power]

Lead time	Benchmarks		Feature Engineering			Gaussian Copula, Empirical Cov.			Gaussian Copula, Parametric Cov.			Copula Vine		
	WF($\mathbf{x}_t$)	AnEn	WFT($\mathbf{x}_t^{\text{SUP}}$)	WF($\mathbf{x}_t$)	AnEn	EGCop	WF($\mathbf{x}_t$)	AnEn	PGCop	WF($\mathbf{x}_t$)	AnEn	VCop	WF($\mathbf{x}_t$)	AnEn
Hour	MAE	MAE	MAE	%Δ	%Δ	MAE	%Δ	%Δ	MAE	%Δ	%Δ	MAE	%Δ	%Δ
0-6	10.08	10.19	9.84	**2.37**	3.38	9.88	2.03	3.04	9.89	1.93	2.94	9.87	2.05	3.06
7-12	10.17	10.17	10.02	**1.50**	1.45	10.09	0.81	0.75	10.09	0.85	0.80	10.10	0.75	0.70
13-18	11.12	11.24	10.88	**2.13**	3.21	10.97	1.33	2.43	10.97	1.30	2.40	10.98	1.25	2.35
19-24	10.60	10.64	10.41	**1.77**	2.09	10.54	0.60	0.92	10.53	0.73	1.05	10.54	0.61	0.93
25-30	12.14	12.38	11.96	**1.47**	3.36	12.01	1.09	3.00	12.01	1.11	3.02	12.02	1.01	2.92
31-36	11.69	11.74	11.50	**1.60**	1.97	11.54	1.27	1.65	11.53	1.36	1.73	11.54	1.26	1.64
37-42	13.07	13.23	12.89	**1.42**	2.59	12.94	1.05	2.23	12.95	0.96	2.14	12.93	1.09	2.27
43-48	12.36	12.41	12.27	**0.78**	1.18	12.27	0.77	1.18	12.27	0.73	1.13	12.27	0.71	1.11
All	11.39	11.46	11.21	**1.61**	2.41	11.26	1.12	1.92	11.26	1.12	1.92	11.27	1.09	1.90

Table B.3: P50 RMSE Results at Wind Farm B - %Δ indicates improvement compared to specified benchmark above [RMSE in % of max power]

Lead time	Benchmarks		Feature Engineering			Gaussian Copula, Empirical Cov.			Gaussian Copula, Parametric Cov.			Copula Vine		
	WF($\mathbf{x}_t$)	AnEn	WFT($\mathbf{x}_t^{\text{SUP}}$)	WF($\mathbf{x}_t$)	AnEn	EGCop	WF($\mathbf{x}_t$)	AnEn	PGCop	WF($\mathbf{x}_t$)	AnEn	VCop	WF($\mathbf{x}_t$)	AnEn
Hour	RMSE	RMSE	RMSE	%Δ	%Δ	RMSE	%Δ	%Δ	RMSE	%Δ	%Δ	RMSE	%Δ	%Δ
0-6	14.69	14.96	**14.26**	2.95	4.70	14.28	2.82	4.57	14.34	2.39	4.15	14.27	2.87	4.63
7-12	14.75	15.02	**14.55**	1.38	3.16	14.57	1.20	2.99	14.63	0.83	2.62	14.56	1.29	3.07
13-18	16.27	16.58	**15.86**	2.51	4.34	15.93	2.09	3.93	15.98	1.74	3.59	15.92	2.13	3.97
19-24	15.40	15.62	**15.13**	1.78	3.18	15.23	1.13	2.53	15.27	0.85	2.26	15.21	1.27	2.67
25-30	17.75	18.13	**17.40**	1.97	4.01	**17.40**	1.99	4.03	17.46	1.68	3.73	**17.40**	2.01	4.05
31-36	17.16	17.24	16.83	1.92	2.39	16.83	1.91	2.38	16.88	1.60	2.07	**16.81**	2.02	2.49
37-42	19.42	19.69	19.11	1.60	2.99	19.10	1.65	3.04	19.17	1.29	2.69	**19.07**	1.79	3.18
43-48	18.68	18.63	18.48	1.06	0.78	18.48	1.08	0.80	18.55	0.67	0.39	**18.46**	1.18	0.90
All	16.83	17.04	**16.52**	1.85	3.09	16.54	1.71	2.96	16.60	1.36	2.61	**16.52**	1.80	3.05

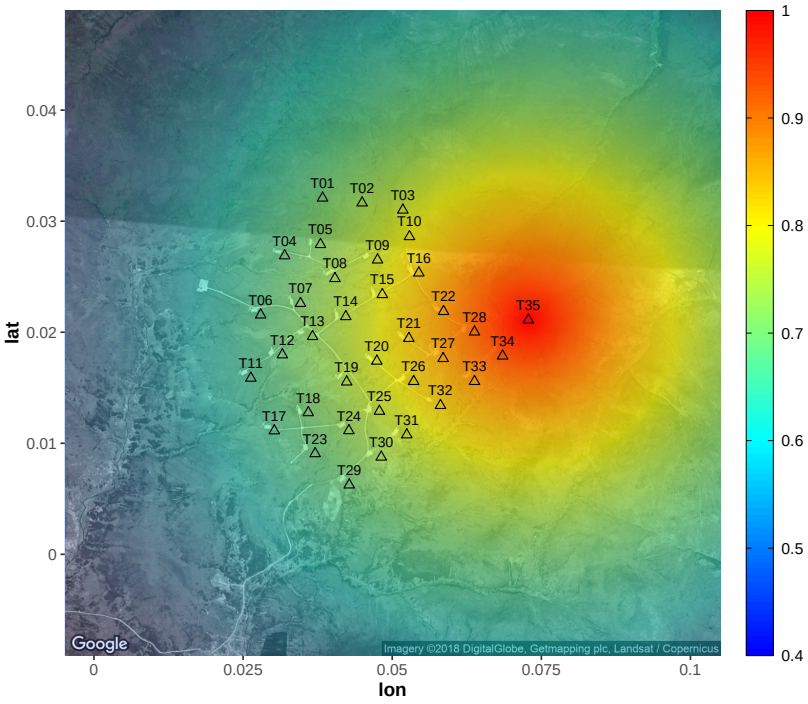


Figure B.3: Mapped parametric partial covariance at Wind Farm B of turbine 35 [2]. Latitude and longitude scales are indicative

Table B.4: Details of Copula Vine selections at Wind Farm B

Total Number	Family
6	Gaussian copula
355	Student t copula (t-copula)
2	Clayton copula
46	Gumbel copula
65	Frank copula
8	Joe copula
6	rotated Clayton copula (180 degrees; “survival Clayton”)
51	rotated Joe copula (180 degrees; “survival Joe”)
11	rotated Clayton copula (90 degrees)
9	rotated Gumbel copula (90 degrees)
3	rotated Joe copula (90 degrees)
9	rotated Clayton copula (270 degrees)
20	rotated Gumbel copula (270 degrees)
4	rotated Joe copula (270 degrees)

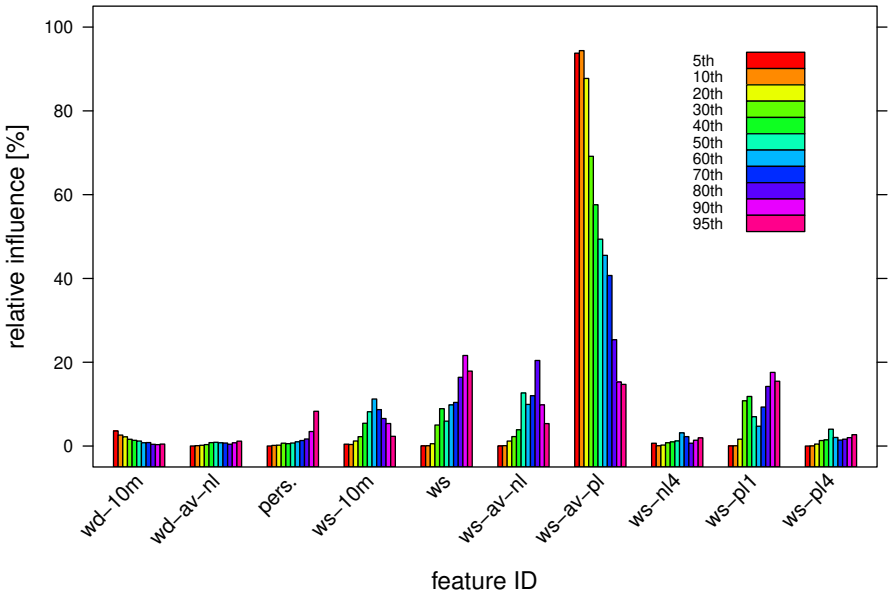


Figure B.4: Wind Farm A: relative influence of input features – WF($\mathbf{x}_t$) benchmark

B.2 Wind Farm A

The feature list at Wind Farm A for all models is contained in Table B.5. This includes the benchmark GBT models, the deterministic and probabilistic models used in the feature engineering method, and the probabilistic turbine models in the bottom-up probabilistic model. Figures B.4 and B.5 illustrate the importance of selected variables across key quantiles in the probabilistic GBT benchmark and probabilistic feature engineering wind farm model respectively.

Full deterministic results in terms of MAE and RMSE are presented in Tables B.6 and B.7 respectively. These are based on the median quantile (p50) in the final full density forecast and are therefore optimised for MAE.

Copula family selection results are displayed in Table B.8 for the copula vine method.

Table B.5: Wind Farm A: list of input features by model - meteorological inputs at 100m (unless specified)

Model	Features	ID
All benchmarks & individual turbines both deterministic and probabilistic	Wind speed - 10m	ws-10m
	Wind speed	ws
	Wind direction - 10m	wd-10m
	Wind direction	-
	Leading ws (1 to 4 periods)	ws-pl-[1-4]
	Lagging ws (1 to 4 periods)	ws-nl-[1-4]
	Leading wd (-1 to -4 periods)	-
	Lagging wd (-1 to -4 periods)	-
	Average ws (0 to 4 periods)	ws-av-pl
	Average ws (0 to -4 periods)	ws-av-nl
	Average wd (0 to 4 periods)	-
	Average wd (0 to -4 periods)	wd-av-nl
	Standard deviation ws (0 to 4 periods)	-
	Standard deviation ws (0 to -4 periods)	-
	Standard deviation wd (0 to 4 periods)	-
	Standard deviation wd (0 to -4 periods)	-
	WS ratio (100m & 10m)	-
	WS gradient (-1 to 0 period)	-
	Direction difference (100m & 10m)	-
	Time of day splines (4)	-
	Persistence	pers.
Feature engineering wind farm probabilistic model	Layer 1 aggregation (mean)	L1-agg
	L1 turbine forecasts (30,43,46)	L1-T[#]
	Leading L1 aggregation (1)	L1-agg-pl1
	Lagging L1 aggregation (-1)	L1-agg-nl1
	Average L1 aggregation (0 to 4 periods)	L1-agg-av-pl
	Average L1 aggregation (0 to -4 periods)	L1-agg-av-nl
	Average ws (0 to 4 periods)	ws-av-pl

Table B.6: P50 MAE Results at Wind Farm A- %Δ indicates improvement compared to specified benchmark above [MAE in % of max power]

Lead time	Benchmarks		Feature Engineering			Gaussian Copula, Empirical Cov.			Gaussian Copula, Parametric Cov.			Copula Vine		
	WF($\mathbf{x}_t$)	AnEn	WFT($\mathbf{x}_t^{\text{SUP}}$)	WF($\mathbf{x}_t$)	AnEn	EGCop	WF($\mathbf{x}_t$)	AnEn	PGCop	WF($\mathbf{x}_t$)	AnEn	VCop	WF($\mathbf{x}_t$)	AnEn
Hour	MAE	MAE	MAE	%Δ	%Δ	MAE	%Δ	%Δ	MAE	%Δ	%Δ	MAE	%Δ	%Δ
0-6	7.15	7.48	6.79	5.02	9.26	6.66	6.85	11.02	**6.61**	7.59	11.72	6.70	6.27	10.46
7-12	8.47	8.64	8.16	3.72	5.58	7.98	5.85	7.67	**7.92**	6.49	8.30	8.02	5.34	7.17
13-18	8.37	8.52	8.31	0.73	2.54	8.05	3.81	5.57	**8.01**	4.30	6.04	8.10	3.20	4.96
19-24	9.50	9.73	8.97	5.59	7.74	8.87	6.62	8.75	**8.82**	7.20	9.32	8.91	6.24	8.38
25-30	9.92	10.08	9.54	3.83	5.37	9.46	4.65	6.18	**9.42**	5.00	6.53	9.50	4.27	5.80
31-36	10.83	11.21	10.33	4.63	7.87	10.20	5.80	9.00	**10.16**	6.18	9.37	10.23	5.55	8.76
37-42	11.10	11.17	10.61	4.48	5.05	10.56	4.91	5.47	**10.52**	5.25	5.81	10.59	4.66	5.22
43-48	12.26	12.28	11.62	5.25	5.37	11.60	5.40	5.53	**11.57**	5.65	5.78	11.62	5.20	5.33
All	9.69	9.88	9.27	4.25	6.09	9.16	5.47	7.28	**9.11**	5.92	7.72	9.19	5.09	6.92

Table B.7: P50 RMSE Results at Wind Farm A - %Δ indicates improvement compared to specified benchmark above [RMSE in % of max power]

Lead time	Benchmarks		Feature Engineering			Gaussian Copula, Empirical Cov.			Gaussian Copula, Parametric Cov.			Copula Vine		
	WF($\mathbf{x}_t$)	AnEn	WFT($\mathbf{x}_t^{\text{SUP}}$)	WF($\mathbf{x}_t$)	AnEn	EGCop	WF($\mathbf{x}_t$)	AnEn	PGCop	WF($\mathbf{x}_t$)	AnEn	VCop	WF($\mathbf{x}_t$)	AnEn
Hour	RMSE	RMSE	RMSE	%Δ	%Δ	RMSE	%Δ	%Δ	RMSE	%Δ	%Δ	RMSE	%Δ	%Δ
0-6	11.02	11.28	10.51	4.56	6.79	10.33	6.23	8.42	10.37	5.85	8.05	**10.30**	6.49	8.67
7-12	13.04	13.24	12.66	2.93	4.38	12.44	4.57	5.99	12.50	4.11	5.54	**12.42**	4.74	6.17
13-18	12.72	12.91	12.67	0.39	1.80	12.37	2.80	4.17	12.43	2.32	3.69	**12.36**	2.85	4.21
19-24	14.86	15.13	14.22	4.28	5.97	14.18	4.59	6.27	14.25	4.12	5.81	**14.14**	4.82	6.50
25-30	15.38	15.34	**14.77**	4.01	3.75	14.79	3.87	3.61	14.88	3.31	3.05	**14.77**	4.01	3.76
31-36	16.65	17.22	16.09	3.39	6.54	16.00	3.89	7.03	16.09	3.40	6.56	15.96	**4.18**	7.31
37-42	17.12	17.17	16.63	2.88	3.17	16.61	2.95	3.25	16.70	2.47	2.77	16.57	**3.19**	3.49
43-48	19.02	18.96	**18.23**	4.17	3.87	18.34	3.55	3.24	18.45	2.99	2.68	18.29	3.82	3.52
All	15.16	15.33	15.64	3.44	4.49	14.57	3.89	4.95	14.64	3.39	4.46	14.54	**4.11**	5.17

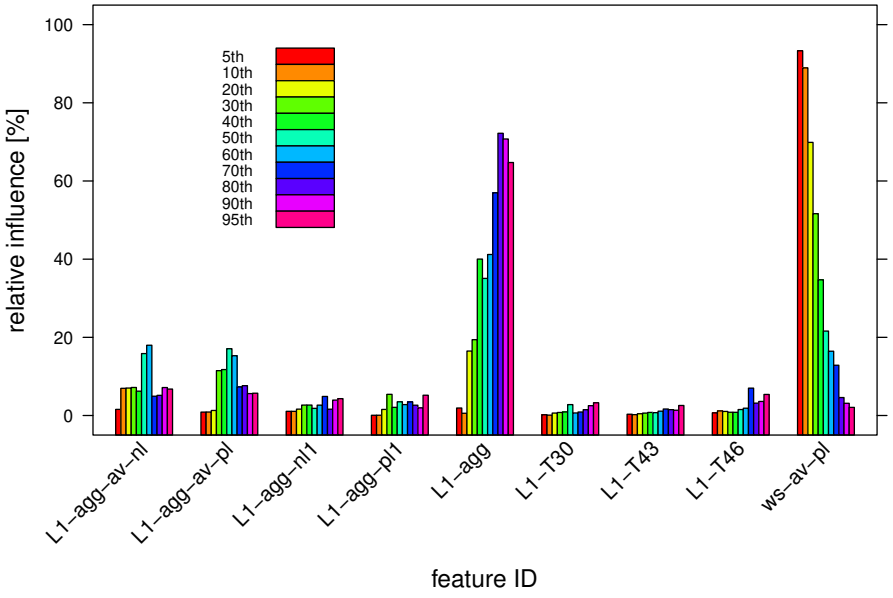


Figure B.5: Wind Farm A: relative influence of input features – WFT($\mathbf{x}_t^{\text{SUP}}$) feature engineering wind farm probabilistic model

Table B.8: Details of Copula Vine selections at Wind Farm A

Total Number	Family
41	Gaussian copula
628	Student t copula (t-copula)
5	Clayton copula
100	Gumbel copula
433	Frank copula
10	Joe copula
6	rotated Clayton copula (180 degrees; “survival Clayton”)
213	rotated Joe copula (180 degrees; “survival Joe”)
35	rotated Clayton copula (90 degrees)
2	rotated Gumbel copula (90 degrees)
37	rotated Clayton copula (270 degrees)
13	rotated Gumbel copula (270 degrees)

Appendix C

Probabilistic Access Forecasting for Improved Offshore Operations

This supplementary material contains information on the input features and formulation for every model tested in the forecasting analysis.

Table C.1: List of input features and base learners by model — significant wave height regression

Model	Features	Learner	Location, Scale, Shape*	Distribution
benchmark - glm	Significant height of combined wind waves and swell (swh)	linear	θ_1	Gamma
benchmark - gam	swh	p-spline	θ_1	Gamma
gbm	Wind direction — 10m Lead time Maximum individual wave height Mean square slope of waves Mean wave direction Mean wave period Peak wave period Significant height of total swell Significant height of wind waves swh Wind speed — 10m Average swh (-2 to 2 hours) Average swh (-3 to 3 hours) Average swh (-4 to 4 hours) Variance swh (-2 to 2 hours) Variance swh (-3 to 3 hours) Variance swh (-4 to 4 hours) Lagging swh (-1 to -4 periods) Leading swh (1 to 4 periods) Time-of-day Day-of-year Week-of-year Month-of-year Indicator of buoy location**	tree	N/A	N/A
gamlss	swh Lead time Mean wave direction Mean wave period*swh Time-of-day Month-of-year Indicator of buoy location**	p-spline p-spline cyclical p-spline p-spline cyclical p-spline cyclical p-spline linear	θ_1,θ_2 θ_1 θ_1 θ_1 θ_1 θ_1 θ_1	Generalised Beta Prime
gamboostLSS	Wind direction — 10m Lead time Maximum individual wave height Mean square slope of waves Mean wave direction Mean wave period Peak wave period Significant height of total swell Significant height of wind waves swh Wind speed — 10m Average swh (-2 to 2 hours) Average swh (-3 to 3 hours) Average swh (-4 to 4 hours) Variance swh (-2 to 2 hours) Variance swh (-3 to 3 hours) Variance swh (-4 to 4 hours) Lagging swh (-1 to -4 periods) Leading swh (1 to 4 periods) Time-of-day, Day-of-year Indicator of buoy location**	cyclical p-spline p-spline p-spline p-spline cyclical p-spline p-spline p-spline p-spline p-spline p-spline p-spline p-spline p-spline p-spline p-spline p-spline p-spline p-splines p-splines bi-variate (tensor) cyclical p-spline linear	θ_1,θ_2,θ_3,θ_4	Generalised Beta Prime

* For parametric regression models, learners applied to the location (θ_1), scale (θ_2), or shape parameters (θ_3, θ_4)

** At the site the buoy was moved midway through data-set

Table C.2: List of input features and base learners by model — peak wave period regression

Model	Features	Learner	Location, Scale, Shape*	Distribution
benchmark - glm	Peak wave period (pwp)	linear	θ_1	Gamma
benchmark - gam	pwp	p-spline	θ_1	Gamma
gbm	Wind direction — 10m	tree	N/A	N/A
	Lead time			
	Maximum individual wave height			
	Mean square slope of waves			
	Mean wave direction			
	Mean wave period			
	pwp			
	Significant height of total swell			
	Significant height of wind waves			
	Significant height of combined wind waves and swell			
	Wind speed — 10m			
	Average pwp (-2 to 2 hours)			
	Average pwp (-3 to 3 hours)			
	Average pwp (-4 to 4 hours)			
	Variance pwp (-2 to 2 hours)			
	Variance pwp (-3 to 3 hours)			
	Variance pwp (-4 to 4 hours)			
	Lagging pwp (-1 to -4 periods)			
	Leading pwp (1 to 4 periods)			
	Time-of-day			
	Day-of-year			
	Week-of-year			
	Month-of-year			
	Indicator of buoy location**			
gamlss	pwp	p-spline	θ_1,θ_2	Generalised Beta Prime
	Significant height of combined wind waves and swell	p-spline	θ_1	
	Lead time	p-spline	θ_1	
	Mean wave direction	cyclical p-spline	θ_1	
	pwp*Significant height of combined wind waves and swell	p-spline	θ_1	
	Time-of-day	cyclical p-spline	θ_1	
	Month-of-year	cyclical p-spline	θ_1	
	Indicator of buoy location**	linear	θ_1	
gamboostLSS	Wind direction — 10m	cyclical p-spline	θ_1,θ_2,θ_3,θ_4	Generalised Beta Prime
	Lead time	p-spline		
	Maximum individual wave height	p-spline		
	Mean square slope of waves	p-spline		
	Mean wave direction	cyclical p-spline		
	Mean wave period	p-spline		
	pwp	p-spline		
	Significant height of total swell	p-spline		
	Significant height of wind waves	p-spline		
	Significant height of combined wind waves and swell	p-spline		
	Wind speed — 10m	p-spline		
	Average pwp (-2 to 2 hours)	p-spline		
	Average pwp (-3 to 3 hours)	p-spline		
	Average pwp (-4 to 4 hours)	p-spline		
	Variance pwp (-2 to 2 hours)	p-spline		
	Variance pwp (-3 to 3 hours)	p-spline		
	Variance pwp (-4 to 4 hours)	p-spline		
	Lagging pwp (-1 to -4 periods)	p-splines		
	Leading pwp (1 to 4 periods)	p-splines		
	Time-of-day, Day-of-year	bi-variate (tensor) cyclical p-spline		
	Indicator of buoy location**	linear		

* For parametric regression models, learners applied to the location (θ_1), scale (θ_2), or shape parameters (θ_3,θ_4)
** At the site the buoy was moved midway through data-set

Table C.3: List of input features and base learners — clustered peak wave direction logistic regression

Model	Features	Learner	Location, Scale, Shape*	Distribution
gbm	Wind direction — 10m (wind-d) Lead time Maximum individual wave height Mean square slope of waves Mean wave direction (mwd) Mean wave period Peak wave period Significant height of total swell Significant height of wind waves Significant height of combined wind waves and swell Wind speed — 10m Average wind-d (-2 to 2 hours) Average wind-d (-3 to 3 hours) Average wind-d (-4 to 4 hours) Variance wind-d (-2 to 2 hours) Variance wind-d (-3 to 3 hours) Variance wind-d (-4 to 4 hours) Lagging wind-d (-1 to -4 periods) Leading wind-d (1 to 4 periods) Average mwd (-2 to 2 hours) Average mwd (-3 to 3 hours) Average mwd (-4 to 4 hours) Variance mwd (-2 to 2 hours) Variance mwd (-3 to 3 hours) Variance mwd (-4 to 4 hours) Lagging mwd (-1 to -4 periods) Leading mwd (1 to 4 periods) direction difference between mwd & wind-d Time-of-day Day-of-year Week-of-year Month-of-year Indicator of buoy location**	tree	N/A	N/A

* For parametric regression models, learners applied to the location (θ_1), scale (θ_2), or shape parameters (θ_3, θ_4)
** At the site the buoy was moved midway through data-set

Appendix D

Regime Switching Multivariate Wind Power Forecasting

This supplementary material contains extra results from the case study presented in Chapter 7. We present the full list of BMUs considered, a comparison between the non-separable and separable parametric covariance fits, more pinball loss results, the PIT histograms for each BMU, and Diebold-Marioano results over the test dataset.

Table D.1: List of wind balancing mechanisms units (BMUs) used in the case study, with associated details from the UK Government renewable energy planning database [3]. GC means generating capacity, which for the BMUs is subject to change. NC is the site nameplate capacity

BMU id	Site Name	Country	Lat [deg]	Lon [deg]	BMU GC [MW]	Site Pn [MW]
E_ABRTW-1	Auchrobert	Scotland	55.624	-3.984	35.988	36.000
E_AIRSW-1	Airies	Scotland	54.802	-4.658	38.000	35.000
E_ASHWW-1	Andershaw	Scotland	55.509	-3.846	35.000	36.000
E_ASLVW-1	Assel Valley	Scotland	55.226	-4.779	25.314	25.000
E_BABAW-1	Baillie	Scotland	58.568	-3.677	52.500	52.500
E_BETHW-1	Beinn Tharsuinn	Scotland	57.805	-4.329	29.750	29.800
E_BRDUW-1	Braes O'Doune	Scotland	56.270	-4.059	74.000	72.000
E_BRYBW-1	Berry Burn	Scotland	57.447	-3.475	66.700	66.700
E_BTUIW-2	Beinn an Tuirc 2	Scotland	55.573	-5.587	43.700	43.700

Table D.1: List of wind balancing mechanisms units (BMUs) used in the case study, with associated details from the UK Government renewable energy planning database [3]. GC means generating capacity, which for the BMUs is subject to change. NC is the site nameplate capacity. (Continued)

BMU id	Site Name	Country	Lat [deg]	Lon [deg]	BMU GC [MW]	Site Pn [MW]
E_CLDRW-1	Clashindarroch 2	Scotland	57.366	-2.866	37.627	37.000
E_CLFLW-1	Clachan Flats	Scotland	56.288	-4.943	15.000	15.000
E_DALSW-1	Dalswinton	Scotland	55.184	-3.651	30.000	30.000
E_GDSTW-1	Gordonstown Hill	Scotland	57.462	-2.483	12.977	12.500
E_GFLDW-1	Goole fields	England	53.666	-0.889	32.684	32.800
E_GLCHW-1	Glenchamber	Scotland	54.963	-4.754	26.000	27.500
E_GLOFW-1	Glens of Foudland	Scotland	57.425	-2.636	26.700	26.000
E_GNFSW-3	Gunfleet sands - demo	England	51.703	1.192	11.700	12.000
E_HBHDW-1	Harburnhead	Scotland	55.809	-3.532	51.700	51.700
E_HLTWW-1	Hill of Towie	Scotland	57.623	-3.341	48.000	42.000
E_HRHLW-1	Hare Hill ext.	Scotland	55.363	-4.126	30.000	29.800
E_MINSW-1	Minsca	Scotland	55.122	-3.353	36.800	36.800
E_TULWW-2	Tullo South ext.	Scotland	56.827	-2.405	25.000	10.000
T_ACHRW-1	A'Chruach 1	Scotland	56.153	-5.310	42.600	48.300
T_ANSUW-1	An Suidhe	Scotland	56.219	-5.219	19.354	19.300
T_ARCHW-1	Arecleoch	Scotland	55.069	-4.797	114.000	120.000
T_BEINW-1	Beinneun	Scotland	57.096	-4.965	108.800	85.000
T_BLLA-1	Black Law	Scotland	55.762	-3.763	119.355	124.000
T_BLLA-2	Black Law ext.	Scotland	55.762	-3.763	69.000	48.400
T_BOWLW-1	Barrow	England	53.991	-3.298	90.000	90.000
T_BRBEO-1	Burbo Bank 2	England	53.480	-3.270	254.200	258.000
T_CLDCW-1	Clyde	Scotland	55.442	-3.543	225.400	350.000
T_CLDNW-1	Clyde ext.	Scotland	55.504	-3.548	197.800	172.800
T_CLDSW-1	Clyde	Scotland	55.442	-3.543	197.800	350.000
T_COUWW-1	Cour	Scotland	55.681	-5.485	20.500	20.500
T_CRGHW-1	Carraig Gheal	Scotland	56.339	-5.295	46.000	46.000
T_CRMLW-1	Corriemoillie	Scotland	57.669	-4.773	48.000	48.500
T_CRYRW-2	Crystal Rig 2a	Scotland	55.899	-2.529	138.000	138.000

Table D.1: List of wind balancing mechanisms units (BMUs) used in the case study, with associated details from the UK Government renewable energy planning database [3]. GC means generating capacity, which for the BMUs is subject to change. NC is the site nameplate capacity. (Continued)

BMU id	Site Name	Country	Lat [deg]	Lon [deg]	BMU GC [MW]	Site Pn [MW]
T_CRYRW-3	Crystal Rig ext. II	Scotland	55.913	-2.545	13.800	20.700
T_DDGNO-2	Dudgeon East	England	53.117	0.613	90.000	402.000
T_DNLWW-1	Dun Law	Scotland	55.806	-2.846	29.750	17.200
T_DRSLW-1	Dersalloch	Scotland	55.304	-4.488	69.000	69.000
T_DUNGW-1	Dunmaglass	Scotland	57.251	-4.256	188.000	94.000
T_EDINW-1	Edinbane	Scotland	57.469	-6.432	41.400	41.400
T_EWHLW-1	Ewe Hill ext.	Scotland	55.107	-3.334	37.000	36.800
T_FALGW-1	Fallago Rig	Scotland	55.831	-2.688	144.000	144.000
T_FARR-1	Farr	Scotland	57.334	-4.103	92.000	92.000
T_FARR-2	Farr	Scotland	57.334	-4.103	92.000	92.000
T_FSDLW-1	Freasdail	Scotland	55.777	-5.471	22.200	22.500
T_GLWSW-1	Galawhistle	Scotland	55.540	-3.975	55.200	66.000
T_GNAPW-1	Glen App	Scotland	55.001	-4.990	22.000	22.000
T_GNFSW-1	Gunfleet Sands	England	51.731	1.218	108.000	108.000
T_GNFSW-2	Gunfleet Sands II	England	51.727	1.246	64.000	64.800
T_GORDW-1	Gordonbush	Scotland	58.059	-3.959	70.000	70.000
T_GRGBW-1	Greater Gabbard	England	51.918	1.928	167.000	504.000
T_GRGBW-2	Greater Gabbard	England	51.918	1.928	167.000	504.000
T_GRGBW-3	Greater Gabbard	England	51.918	1.928	167.000	504.000
T_GRIFW-1	Griffin	Scotland	56.581	-3.734	102.000	156.000
T_GRIFW-2	Griffin	Scotland	56.581	-3.734	102.000	156.000
T_GYMR-15	Gwynt y Mor	Wales	53.454	-3.627	148.000	576.000
T_GYMR-17	Gwynt y Mor	Wales	53.454	-3.627	148.000	576.000
T_GYMR-26	Gwynt y Mor	Wales	53.454	-3.627	148.000	576.000
T_GYMR-28	Gwynt y Mor	Wales	53.454	-3.627	148.000	576.000
T_HADHW-1	Hadyard Hill	Scotland	55.257	-4.732	130.000	120.000
T_HMGTO-1	Humber Gateway A	England	53.387	0.501	108.000	219.000
T_HMGTO-2	Humber Gateway A	England	53.387	0.501	111.000	219.000

Table D.1: List of wind balancing mechanisms units (BMUs) used in the case study, with associated details from the UK Government renewable energy planning database [3]. GC means generating capacity, which for the BMUs is subject to change. NC is the site nameplate capacity. (Continued)

BMU id	Site Name	Country	Lat [deg]	Lon [deg]	BMU GC [MW]	Site Pn [MW]
T_HRSTW-1	Harestanes	Scotland	55.239	-3.574	142.300	136.000
T_KILBW-1	Kilbraur	Scotland	58.040	-4.059	67.500	47.500
T_LARYW-3	London Array	England	51.622	1.496	171.000	630.000
T_LARYW-4	London Array	England	51.622	1.496	171.000	630.000
T_LCLTW-1	lochluichart	Scotland	57.655	-4.817	69.000	51.000
T_LNCSW-1	Centrica (Lincs)	England	53.184	0.498	150.000	270.000
T_LNCSW-2	Centrica (Lincs)	England	53.184	0.498	150.000	270.000
T_MILWW-1	Millennium	Scotland	57.155	-4.884	65.000	40.000
T_MKHLW-1	Mark Hill	Scotland	55.128	-4.746	53.000	56.000
T_OMNDW-1	Ormonde	England	54.089	-3.439	150.750	150.000
T_PNYCW-1	Pen Y Cymoedd	Wales	51.656	-3.696	228.000	228.000
T_RCBKO-2	Race Bank	England	53.136	0.589	290.000	286.500
T_RREW-1	Robin Rigg East	Scotland	54.764	-3.696	90.000	90.000
T_RRWW-1	Robin Rigg West	Scotland	54.747	-3.729	90.000	84.000
T_SHRSW-1	Sheringham Shoal	England	53.135	1.148	158.400	317.000
T_SHRSW-2	Sheringham Shoal	England	53.135	1.148	158.400	317.000
T_STRNW-1	Strathy North	Scotland	58.490	-4.034	68.000	67.700
T_TDBNW-1	Toddleburn	Scotland	55.769	-2.877	55.200	27.600
T_THNTO-1	Thanet	England	51.430	1.633	150.000	300.000
T_THNTO-2	Thanet	England	51.430	1.633	150.000	300.000
T_WDNSO-1	West of Duddon Sands	England	53.985	-3.462	191.000	389.000
T_WDNSO-2	West of Duddon Sands	England	53.985	-3.462	191.000	389.000
T_WHILW-1	Whitelee	Scotland	55.681	-4.279	305.000	322.000
T_WHILW-2	Whitelee ext.	Scotland	55.677	-4.287	206.000	108.000
T_WLNYO-2	Walney 2	England	54.081	-3.609	183.600	183.600
T_WLNYW-1	Walney 1	England	54.039	-3.516	182.000	183.600
T_WTMSO-1	Westermost Rough	England	53.098	0.561	205.000	210.000

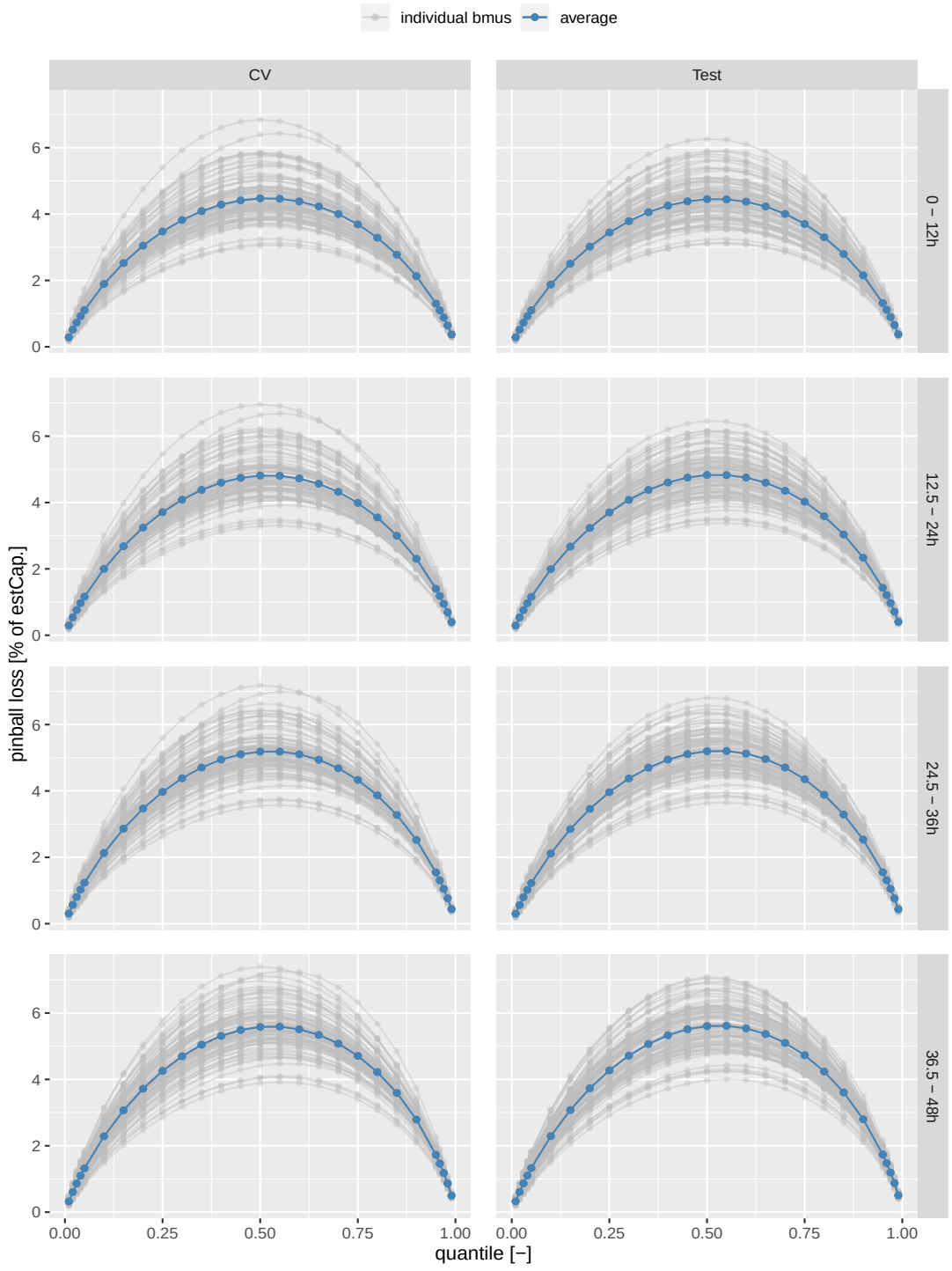
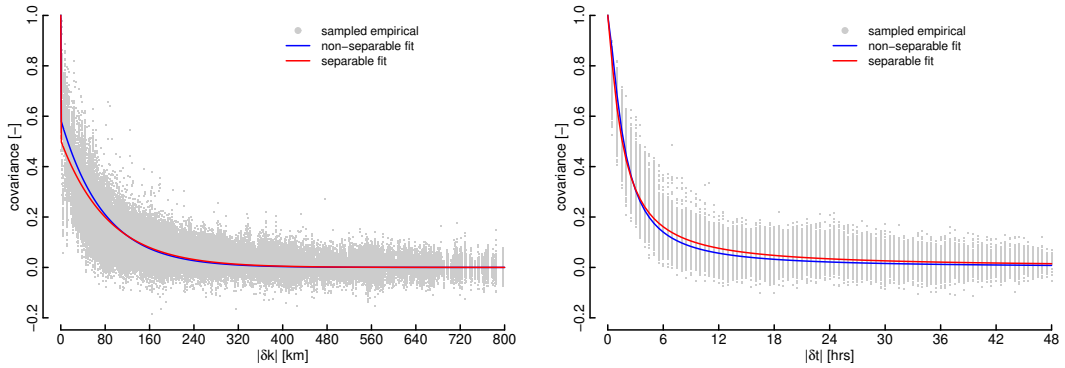


Figure D.1: Pinball loss in lead time groups during both cross validation (CV) and testing datasets

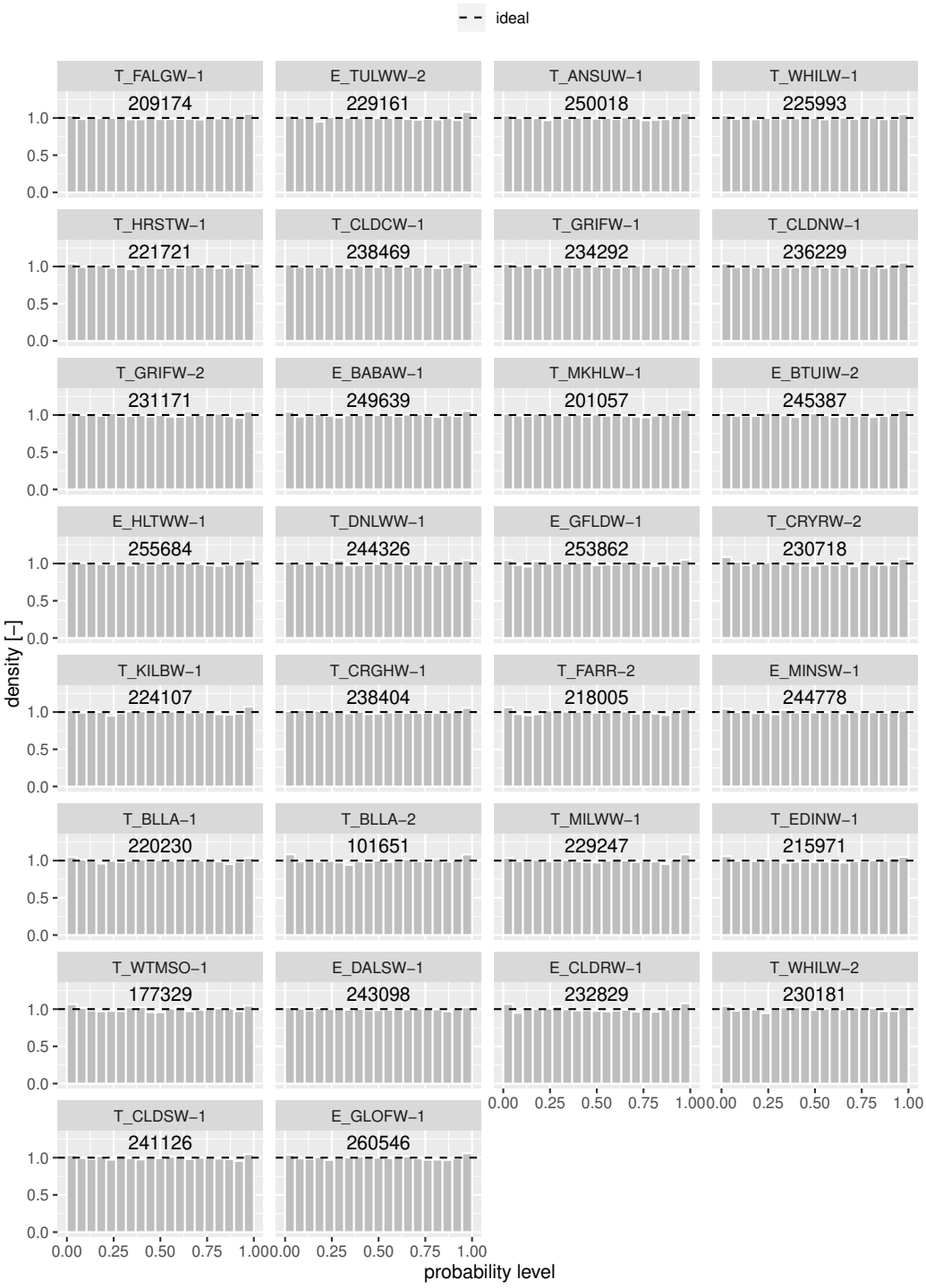


(a) spatial covariance, i.e $|\delta t| = 0$ (b) temporal covariance , i.e $|\delta k| = 0$

Figure D.2: Comparison between separable and non-separable covariance model fits for the purely temporal and purely spatial cases

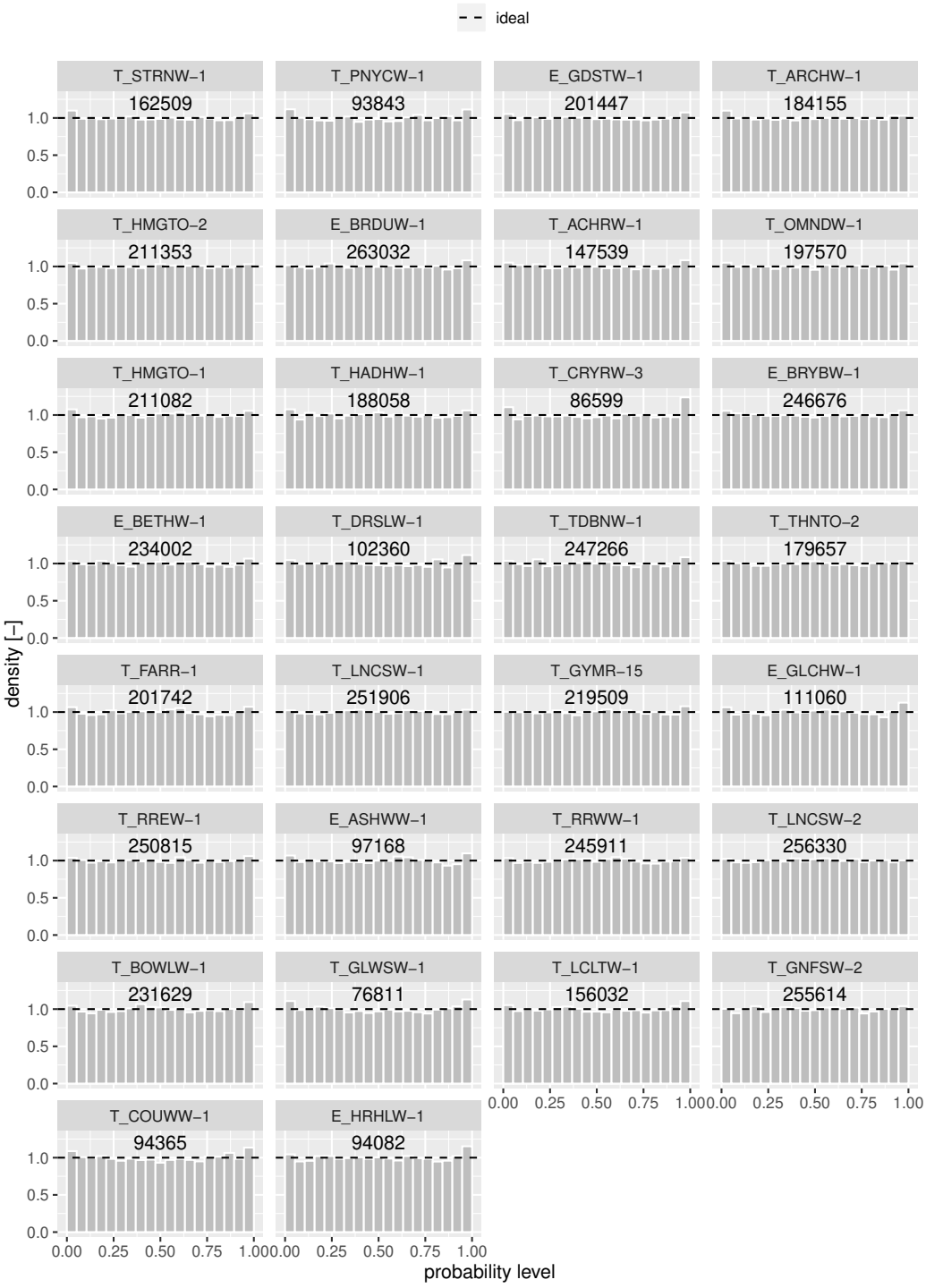
Table D.2: Diebold-Mariano test statistics based on the energy score differential on the testing data, where p values are shown in bold

	independence	temporal	spatiotemporal	RS-spatiotemporal
independence	-	3.150 **0.002**	7.175 **<0.001**	6.093 **<0.001**
temporal	-3.150 **0.002**	-	4.418 **<0.001**	4.764 **<0.001**
spatiotemporal	-7.175 **<0.001**	-4.418 **<0.001**	-	0.788 **0.431**
RS-spatiotemporal	-6.093 **<0.001**	-4.764 **<0.001**	-0.788 **0.431**	-



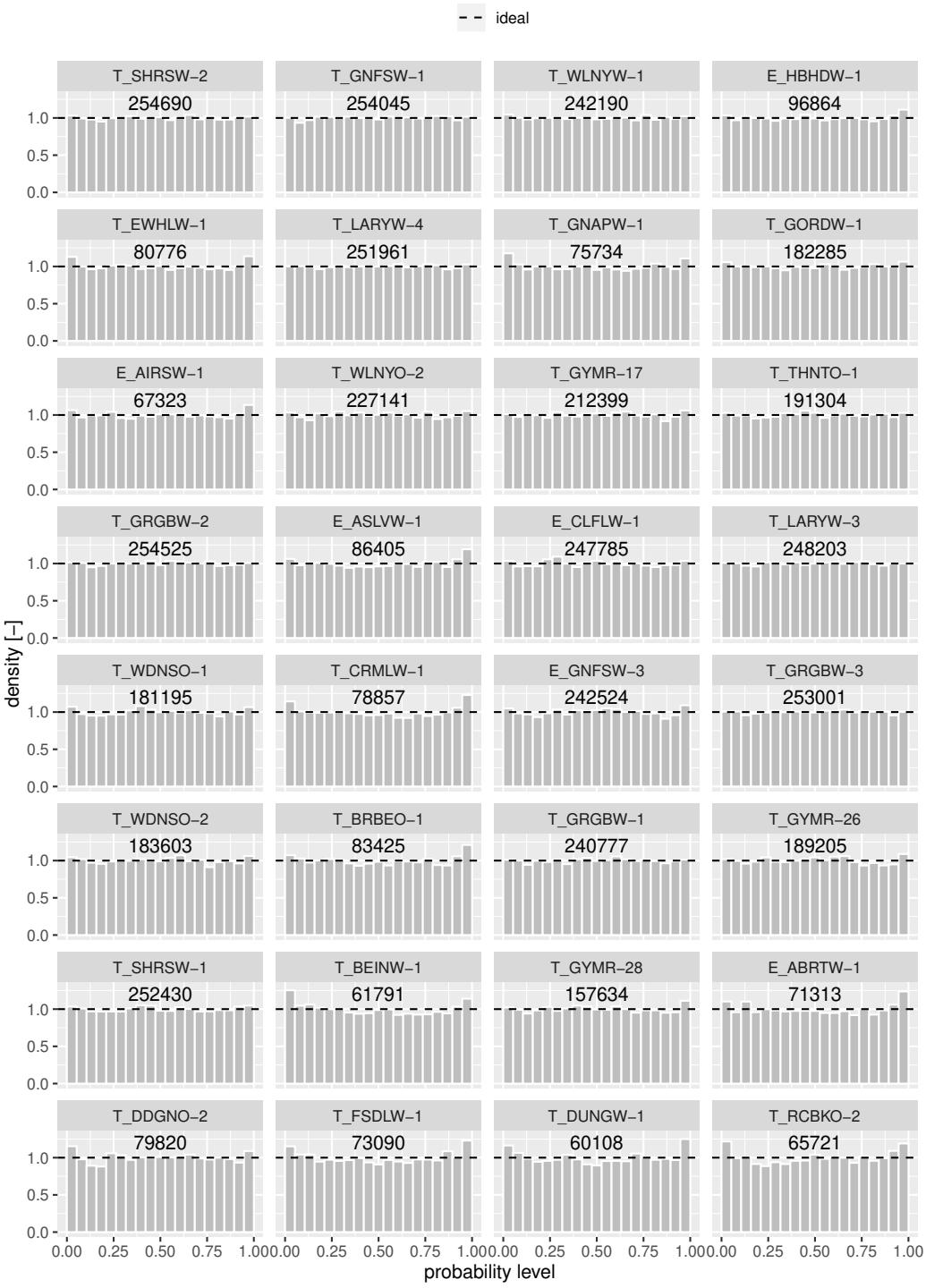
(a) part 1

Figure D.3: PIT histograms of all BMUs considered during cross validation, the number of observations available are overlaid on the plots



(b) part 2

Figure D.3: PIT histograms of all BMUs considered during cross validation, the number of observations available are overlaid on the plots



(c) part 3

Figure D.3: PIT histograms of all BMUs considered during cross validation, the number of observations available are overlaid on the plots

Bibliography

[1] BEIS, "Digest of UK Energy Statistics (DUKES): renewable sources of energy," UK Government - Department for Business, Energy & Industrial Strategy, Tech. Rep., 2019. [Online]. Available: https://www.gov.uk/government/collections/digest-of-uk-energy-statistics-dukes

[2] D. Kahle and H. Wickham, "ggmap: Spatial visualization with ggplot2," *R Journal*, vol. 5, no. 1, pp. 144–161, 2013. [Online]. Available: https://journal.r-project.org/archive/2013-1/kahle-wickham.pdf

[3] UK Government, "Renewable Energy Planning Data - GOV.UK." [Online]. Available: https://www.gov.uk/government/collections/renewable-energy-planning-data

[4] GWEC, "Global Wind Report 2019," Global Wind Energy Council, Tech. Rep., 2019. [Online]. Available: https://gwec.net/global-wind-report-2019/

[5] WindEurope, "Wind energy in Europe in 2019," WindEurope, Tech. Rep., 2019. [Online]. Available: https://windeurope.org/about-wind/statistics/european/wind-energy-in-europe-in-2019/

[6] BEIS, "Contracts for Difference (CfD): Allocation Round 3," UK Government - Department for Business, Energy & Industrial Strategy, Tech. Rep., 2019. [Online]. Available: https://www.gov.uk/government/collections/contracts-for-difference-cfd-third-allocation-round

[7] Ofgem, "State of the Energy Market 2019," Office of Gas and Electricity Markets - National Regulatory Authority, Tech. Rep., 2019. [Online]. Available: https://www.ofgem.gov.uk/publications-and-updates/state-energy-market-2019

[8] BBC, "SSE Renewables to build first subsidy-free wind farm - BBC News," 2020. [Online]. Available: https://www.bbc.co.uk/news/uk-scotland-highlands-islands-51325543

[9] M. Shahidehpour, H. Yamin, and Z. Li, *Market Operations in Electric Power Systems*. New York, USA: John Wiley & Sons, Inc., 4 2002. [Online]. Available: http://doi.wiley.com/10.1002/047122412X

[10] J. M. Morales, A. J. Conejo, H. Madsen, P. Pinson, and M. Zugno, *Integrating Renewables in Electricity Markets: Operational Problems*, ser. International Series in Operations Research & Management Science. Boston, MA: Springer US, 2014, vol. 205. [Online]. Available: http://link.springer.com/10.1007/978-1-4614-9411-9

[11] D. M. Newbery, "Electricity liberalization in Britain: The quest for a satisfactory wholesale market design," *The Energy Journal*, vol. 26, no. 01, pp. 43–70, 9 2005. [Online]. Available: http://www.iaee.org/en/publications/ejarticle.aspx?id=2091

[12] National Grid ESO, "ESO innovation strategy," National Grid ESO, Tech. Rep., 2019. [Online]. Available: https://www.nationalgrideso.com/innovation/strategy

[13] N. Mazzi and P. Pinson, "Wind power in electricity markets and the value of forecasting," in *Renewable Energy Forecasting*, ser. Woodhead Publishing Series in Energy, G. Kariniotakis, Ed. Elsevier, 2017, pp. 259–278. [Online]. Available: http://www.sciencedirect.com/science/article/pii/B978008100504000010X

[14] J. Sanz Rodrigo, L. Frías Paredes, R. Girard, G. Kariniotakis, K. Laquaine, N. Stoffels, and L. von Bremen, "The role of predictability in the investment phase of wind farms," in *Renewable Energy Forecasting*, ser. Woodhead Publishing

Series in Energy, G. Kariniotakis, Ed. Elsevier, 2017, pp. 341–357. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/B9780081005040000147

[15] J. F. Manwell, J. G. McGowan, and A. L. Rogers, *Wind Energy Explained*. Chichester, UK: John Wiley & Sons, Ltd, 12 2009.

[16] T. Burton, N. Jenkins, D. Sharpe, and E. Bossanyi, *Wind Energy Handbook*, 2nd ed. Wiley, 2011.

[17] C. J. Crabtree, D. Zappalá, and S. I. Hogg, "Wind energy: UK experiences and offshore operational challenges," *Proceedings of the Institution of Mechanical Engineers, Part A: Journal of Power and Energy*, vol. 229, no. 7, pp. 727–746, 11 2015. [Online]. Available: http://journals.sagepub.com/doi/10.1177/0957650915597560

[18] P. Bauer, A. Thorpe, and G. Brunet, "The quiet revolution of numerical weather prediction," *Nature*, vol. 525, no. 7567, pp. 47–55, 9 2015. [Online]. Available: http://www.nature.com/doifinder/10.1038/nature14956

[19] F. Ziel and R. Weron, "Day-ahead electricity price forecasting with high-dimensional structures: Univariate vs. multivariate modeling frameworks," *Energy Economics*, vol. 70, pp. 396–420, 2 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S014098831730436X

[20] J. W. Messner and P. Pinson, "Online adaptive lasso estimation in vector autoregressive models for high dimensional wind power forecasting," *International Journal of Forecasting*, vol. 35, no. 4, pp. 1485–1498, 10 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0169207018300347

[21] S. B. Taieb, J. W. Taylor, and R. J. Hyndman, "Hierarchical Probabilistic Forecasting of Electricity Demand With Smart Meter Data," *Journal of the American Statistical Association*, pp. 1–17, 3 2020. [Online]. Available: https://www.tandfonline.com/doi/full/10.1080/01621459.2020.1736081

[22] S. Ma, R. Fildes, and T. Huang, "Demand forecasting with high dimensional data: The case of SKU retail sales forecasting with intra- and inter-category promotional information," *European Journal of Operational Research*, vol. 249, no. 1, pp. 245–257, 2 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0377221715007845

[23] Spyros Makridakis, "The M5 Competition – MOFC," 2020. [Online]. Available: https://mofc.unic.ac.cy/m5-competition/

[24] T. Hastie, R. Tibshirani, and M. Wainwright, *Statistical learning with sparsity: The lasso and generalizations*. CRC Press, 1 2015.

[25] P. Bühlmann and S. van de Geer, "Introduction," in *Statistics for high-dimensional data: methods, theory and applications*. Springer, Berlin, Heidelberg, 2011, pp. 1–6. [Online]. Available: http://link.springer.com/10.1007/978-3-642-20192-9_1

[26] G. Giebel and G. Kariniotakis, "Wind power forecasting—a review of the state of the art," in *Renewable Energy Forecasting*, ser. Woodhead Publishing Series in Energy, G. Kariniotakis, Ed. Elsevier, 2017, pp. 59–109. [Online]. Available: http://www.sciencedirect.com/science/article/pii/B9780081005040000032

[27] G. Giebel, J. Badger, I. Martí Perez, P. Louka, G. Kallos, A. M. Palomares, C. Lac, and G. Descombes, "Short-term forecasting using advanced physical modelling - The results of the anemos project: Results from mesoscale, microscale and CFD modelling," in *European Wind Energy Conference and Exhibition 2006, EWEC 2006*, vol. 3. European Wind Energy Association (EWEA), 4 2006, pp. 2431–2459.

[28] C. P. Sweeney, P. Lynch, and P. Nolan, "Reducing errors of wind speed forecasts by an optimal combination of post-processing methods," *Meteorological Applications*, vol. 20, no. 1, pp. 32–40, 3 2013. [Online]. Available: http://doi.wiley.com/10.1002/met.294

[29] C. Sweeney, R. J. Bessa, J. Browell, and P. Pinson, "The future of forecasting for renewable energy," *WIREs Energy and Environment*, vol. 9, no. 2, 3 2020. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/wene.365

[30] M. Landry, T. P. Erlinger, D. Patschke, and C. Varrichio, "Probabilistic gradient boosting machines for GEFCom2014 wind forecasting," *International Journal of Forecasting*, vol. 32, no. 3, pp. 1061–1066, 7 2016. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0169207016000145

[31] J. R. Andrade and R. J. Bessa, "Improving Renewable Energy Forecasting with a Grid of Numerical Weather Predictions," *IEEE Transactions on Sustainable Energy*, vol. 8, no. 4, pp. 1571–1580, 4 2017. [Online]. Available: http://ieeexplore.ieee.org/document/7903735/

[32] L. Silva, "A feature engineering approach to wind power forecasting: GEFCom 2012," *International Journal of Forecasting*, vol. 30, no. 2, pp. 395–401, 4 2014. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0169207013000836

[33] T. Hong, P. Pinson, S. Fan, H. Zareipour, A. Troccoli, and R. J. Hyndman, "Probabilistic energy forecasting: Global Energy Forecasting Competition 2014 and beyond," *International Journal of Forecasting*, vol. 32, no. 3, pp. 896–913, 7 2016. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0169207016000133

[34] G. Reikard, P. Pinson, and J.-R. Bidlot, "Forecasting ocean wave energy: The ECMWF wave model and time series methods," *Ocean Engineering*, vol. 38, no. 10, pp. 1089–1099, 7 2011. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0029801811000837

[35] C. Monteiro, R. Bessa, V. Miranda, A. Botterud, J. Wang, G. Conzelmann, and INESC Porto, "Wind power forecasting : state-of-the-art 2009." Argonne National Laboratory (ANL), Argonne, IL (United States), Tech. Rep., 11 2009. [Online]. Available: http://www.osti.gov/servlets/purl/968212-IMKXCO/

[36] R. Bessa, C. Möhrlen, V. Fundel, M. Siefert, J. Browell, S. Haglund El Gaidi, B.-M. Hodge, U. Cali, and G. Kariniotakis, "Towards Improved Understanding of the Applicability of Uncertainty Forecasts in the Electric Power Industry," *Energies*, vol. 10, no. 9, p. 1402, 9 2017. [Online]. Available: http://www.mdpi.com/1996-1073/10/9/1402

[37] E. N. Lorenz, "Deterministic Nonperiodic Flow," *Journal of the Atmospheric Sciences*, vol. 20, no. 2, pp. 130–141, 3 1963. [Online]. Available: http://journals.ametsoc.org/doi/abs/10.1175/1520-0469%281963%29020%3C0130%3ADNF%3E2.0.CO%3B2

[38] M. Calaf, C. Meneveau, and J. Meyers, "Large eddy simulation study of fully developed wind-turbine array boundary layers," *Physics of Fluids*, vol. 22, no. 1, p. 015110, 1 2010. [Online]. Available: http://aip.scitation.org/doi/10.1063/1.3291077

[39] Y.-T. Wu and F. Porté-Agel, "Modeling turbine wakes and power losses within a wind farm using LES: An application to the Horns Rev offshore wind farm," *Renewable Energy*, vol. 75, pp. 945–955, 3 2015. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0960148114003590

[40] D. Mehta, A. van Zuijlen, B. Koren, J. Holierhoek, and H. Bijl, "Large Eddy Simulation of wind farm aerodynamics: A review," *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 133, pp. 1–17, 10 2014. [Online]. Available: http://dx.doi.org/10.1016/j.jweia.2014.07.002

[41] J. Schalkwijk, E. J. Griffith, F. H. Post, and H. J. J. Jonker, "High-Performance Simulations of Turbulent Clouds on a Desktop PC: Exploiting the GPU," *Bulletin of the American Meteorological Society*, vol. 93, no. 3, pp. 307–314, 3 2012. [Online]. Available: http://journals.ametsoc.org/doi/abs/10.1175/BAMS-D-11-00059.1

[42] J. Schalkwijk, H. J. J. Jonker, A. P. Siebesma, and F. C. Bosveld, "A Year-Long Large-Eddy Simulation of the Weather over Cabauw: An Overview," *Monthly*

Weather Review, vol. 143, no. 3, pp. 828–844, 3 2015. [Online]. Available: http://journals.ametsoc.org/doi/10.1175/MWR-D-14-00293.1

[43] D. S. Wilks, "Statistical Forecasting," in *Statistical Methods in the Atmospheric Sciences*. Elsevier, 1 2019, ch. 7, pp. 235–312. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780128158234000079

[44] H. R. Glahn and D. A. Lowry, "The Use of Model Output Statistics (MOS) in Objective Weather Forecasting," *Journal of Applied Meteorology*, vol. 11, no. 8, pp. 1203–1211, 12 1972. [Online]. Available: http://journals.ametsoc.org/doi/abs/10.1175/1520-0450%281972%29011%3C1203%3ATUOMOS%3E2.0.CO%3B2

[45] M. P. Clark and L. E. Hay, "Use of Medium-Range Numerical Weather Prediction Model Output to Produce Forecasts of Streamflow," *Journal of Hydrometeorology*, vol. 5, no. 1, pp. 15–32, 2 2004. [Online]. Available: http://journals.ametsoc.org/doi/abs/10.1175/1525-7541%282004%29005%3C0015%3AUOMNWP%3E2.0.CO%3B2

[46] Y. Mao and A. Monahan, "Linear and nonlinear regression prediction of surface wind components," *Climate Dynamics*, vol. 51, no. 9-10, pp. 3291–3309, 11 2018. [Online]. Available: https://doi.org/10.1007/s00382-018-4079-5

[47] D. S. Wilks and T. M. Hamill, "Comparison of Ensemble-MOS Methods Using GFS Reforecasts," *Monthly Weather Review*, vol. 135, no. 6, pp. 2379–2390, 6 2007. [Online]. Available: http://journals.ametsoc.org/doi/10.1175/MWR3402.1

[48] J. W. Messner, G. J. Mayr, A. Zeileis, and D. S. Wilks, "Heteroscedastic Extended Logistic Regression for Postprocessing of Ensemble Guidance," *Monthly Weather Review*, vol. 142, no. 1, pp. 448–456, 1 2014. [Online]. Available: http://journals.ametsoc.org/doi/10.1175/MWR-D-13-00271.1

[49] H. A. Nielsen, T. S. Nielsen, H. Madsen, M. J. S. I. Pindado, and I. Marti, "Optimal combination of wind power forecasts," *Wind Energy*, vol. 10, no. 5, pp. 471–482, 9 2007. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/we.237

[50] J. W. Messner, A. Zeileis, J. Broecker, and G. J. Mayr, "Probabilistic wind power forecasts with an inverse power curve transformation and censored regression," *Wind Energy*, vol. 17, no. 11, pp. 1753–1766, 11 2014. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/we.1666

[51] J. W. Messner, P. Pinson, J. Browell, M. B. Bjerregård, and I. Schicker, "Evaluation of wind power forecasts—An up-to-date view," *Wind Energy*, p. we.2497, 3 2020. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/we.2497

[52] Y. Zhang, J. Wang, and X. Wang, "Review on probabilistic forecasting of wind power generation," *Renewable and Sustainable Energy Reviews*, vol. 32, no. 0, pp. 255–270, 4 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1364032114000446

[53] 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*, vol. 3, no. 4, pp. 827–836, 10 2012. [Online]. Available: http://ieeexplore.ieee.org/document/6209459/

[54] P. Pinson, C. Chevallier, and G. N. Kariniotakis, "Trading Wind Generation From Short-Term Probabilistic Forecasts of Wind Power," *IEEE Transactions on Power Systems*, vol. 22, no. 3, pp. 1148–1156, 8 2007. [Online]. Available: http://ieeexplore.ieee.org/document/4282048/

[55] 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, 8 2008. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1364032107000354

[56] 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," *Technical Report, ANEMOS.plus*, 2011.

[57] L. Landberg, "Short-term prediction of local wind conditions," *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 89, no. 3-4, pp. 235–245, 3 2001. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0167610500000799

[58] 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, 12 2000. [Online]. Available: http://doi.wiley.com/10.1002/1099-1115%28200012%2914%3A8%3C813%3A%3AAID-ACS622%3E3.0.CO%3B2-6

[59] 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, no. 2, pp. 177–184, 5 2005. [Online]. Available: https://asmedigitalcollection.asme.org/solarenergyengineering/article/127/2/177/461531/On-the-Uncertainty-of-Wind-Power

[60] H. Bludszuweit, J. Dominguez-Navarro, and A. Llombart, "Statistical Analysis of Wind Power Forecast Error," *IEEE Transactions on Power Systems*, vol. 23, no. 3, pp. 983–991, 8 2008. [Online]. Available: http://ieeexplore.ieee.org/document/4530750/

[61] 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, 8 2012. [Online]. Available: http://doi.wiley.com/10.1111/j.1467-9876.2011.01026.x

[62] R. Buizza, "Ensemble Forecasting and the Need for Calibration," in *Statistical Postprocessing of Ensemble Forecasts*. Elsevier, 1 2018, ch. 2, pp.

15–48. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780128123720000029

[63] S. Alessandrini, S. Sperati, and P. Pinson, "A comparison between the ECMWF and COSMO Ensemble Prediction Systems applied to short-term wind power forecasting on real data," *Applied Energy*, vol. 107, pp. 271–280, 7 2013. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0306261913001499

[64] P. Pinson and H. Madsen, "Ensemble-based probabilistic forecasting at Horns Rev," *Wind Energy*, vol. 12, no. 2, pp. 137–155, 3 2009. [Online]. Available: http://doi.wiley.com/10.1002/we.309

[65] P. Pinson and W. Messner, Jakob, "Application of Postprocessing for Renewable Energy," in *Statistical Postprocessing of Ensemble Forecasts*. Elsevier, 1 2018, ch. 9, pp. 241–266. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780128123720000091

[66] J. Lujano-Rojas, G. Osório, and J. Catalão, "New probabilistic method for solving economic dispatch and unit commitment problems incorporating uncertainty due to renewable energy integration," *International Journal of Electrical Power & Energy Systems*, vol. 78, pp. 61–71, 6 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0142061515004950

[67] C. Uckun, A. Botterud, and J. R. Birge, "An Improved Stochastic Unit Commitment Formulation to Accommodate Wind Uncertainty," *IEEE Transactions on Power Systems*, vol. 31, no. 4, pp. 2507–2517, 7 2016. [Online]. Available: http://ieeexplore.ieee.org/document/7272778/

[68] Z. Zhou, A. Botterud, J. Wang, R. Bessa, H. Keko, J. Sumaili, and V. Miranda, "Application of probabilistic wind power forecasting in electricity markets," *Wind Energy*, vol. 16, no. 3, pp. 321–338, 4 2013. [Online]. Available: http://doi.wiley.com/10.1002/we.1496

[69] 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, 1 2009. [Online]. Available: http://doi.wiley.com/10.1002/we.284

[70] 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, 1 2011. [Online]. Available: http://doi.wiley.com/10.1002/we.401

[71] R. J. Bessa, "From marginal to simultaneous prediction intervals of wind power," in *2015 18th International Conference on Intelligent System Application to Power Systems (ISAP)*. IEEE, 9 2015, pp. 1–6. [Online]. Available: http://ieeexplore.ieee.org/document/7325536/

[72] P. Pinson and R. Girard, "Evaluating the quality of scenarios of short-term wind power generation," *Applied Energy*, vol. 96, pp. 12–20, 8 2012. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0306261911006994

[73] E. B. Iversen, J. M. Morales, J. K. Møller, P.-J. Trombe, and H. Madsen, "Leveraging stochastic differential equations for probabilistic forecasting of wind power using a dynamic power curve," *Wind Energy*, vol. 20, no. 1, pp. 33–44, 1 2017. [Online]. Available: http://doi.wiley.com/10.1002/we.1988

[74] M. Clark, S. Gangopadhyay, L. Hay, B. Rajagopalan, R. Wilby, M. Clark, S. Gangopadhyay, L. Hay, B. Rajagopalan, and R. Wilby, "The Schaake Shuffle: A Method for Reconstructing Space–Time Variability in Forecasted Precipitation and Temperature Fields," *Journal of Hydrometeorology*, vol. 5, no. 1, pp. 243–262, 2 2004. [Online]. Available: http://journals.ametsoc.org/doi/abs/10.1175/1525-7541%282004%29005%3C0243%3ATSSAMF%3E2.0.CO%3B2

[75] G. Giebel, J. Cline, H. Frank, W. Shaw, P. Pinson, B.-M. Hodge, G. Kariniotakis, J. Madsen, and C. Möhrlen, "Wind power forecasting: IEA Wind Task 36 & future research issues," *Journal of Physics:*

Conference Series, vol. 753, no. 3, 9 2016. [Online]. Available: https://iopscience.iop.org/article/10.1088/1742-6596/753/3/032042

[76] T. Hong, P. Pinson, and S. Fan, "Global Energy Forecasting Competition 2012," *International Journal of Forecasting*, vol. 30, no. 2, pp. 357–363, 4 2014. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0169207013000745

[77] E. Mangalova and O. Shesterneva, "K-nearest neighbors for GEFCom2014 probabilistic wind power forecasting," *International Journal of Forecasting*, vol. 32, no. 3, pp. 1067–1073, 7 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0169207015001429

[78] L. Delle Monache, F. A. Eckel, D. L. Rife, B. Nagarajan, and K. Searight, "Probabilistic Weather Prediction with an Analog Ensemble," *Monthly Weather Review*, vol. 141, no. 10, pp. 3498–3516, 10 2013. [Online]. Available: http://journals.ametsoc.org/doi/10.1175/MWR-D-12-00281.1

[79] S. Alessandrini, L. Delle Monache, S. Sperati, and J. Nissen, "A novel application of an analog ensemble for short-term wind power forecasting," *Renewable Energy*, vol. 76, pp. 768–781, 4 2015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0960148114007915

[80] S. Alessandrini, L. Delle Monache, S. Sperati, and G. Cervone, "An analog ensemble for short-term probabilistic solar power forecast," *Applied Energy*, vol. 157, pp. 95–110, 2015. [Online]. Available: http://dx.doi.org/10.1016/j.apenergy.2015.08.011

[81] F. Ziel, C. Croonenbroeck, and D. Ambach, "Forecasting wind power – Modeling periodic and non-linear effects under conditional heteroscedasticity," *Applied Energy*, vol. 177, pp. 285–297, 9 2016. [Online]. Available: http://arxiv.org/abs/1606.00546

[82] G. Athanasopoulos, R. A. Ahmed, and R. J. Hyndman, "Hierarchical forecasts for Australian domestic tourism," *International Journal of Forecasting*, vol. 25,

no. 1, pp. 146–166, 1 2009. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0169207008000691

[83] G. Athanasopoulos, R. J. Hyndman, N. Kourentzes, and F. Petropoulos, "Forecasting with temporal hierarchies," *European Journal of Operational Research*, vol. 262, no. 1, pp. 60–74, 10 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0377221717301911

[84] B. J. Dangerfield and J. S. Morris, "Top-down or bottom-up: Aggregate versus disaggregate extrapolations," *International Journal of Forecasting*, vol. 8, no. 2, pp. 233–241, 10 1992. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/016920709290121O

[85] R. J. Hyndman, R. A. Ahmed, G. Athanasopoulos, and H. L. Shang, "Optimal combination forecasts for hierarchical time series," *Computational Statistics & Data Analysis*, vol. 55, no. 9, pp. 2579–2589, 9 2011. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0167947311000971

[86] S. B. Taieb, J. W. Taylor, and R. J. Hyndman, "Coherent probabilistic forecasts for hierarchical time series," in *34th International Conference on Machine Learning, ICML 2017*, vol. 7, 2017, pp. 5143–5155. [Online]. Available: http://proceedings.mlr.press/v70/taieb17a.html

[87] Y. Zhang and J. Dong, "Least Squares-based Optimal Reconciliation Method for Hierarchical Forecasts of Wind Power Generation," *IEEE Transactions on Power Systems*, vol. PP, no. c, pp. 1–1, 2018. [Online]. Available: https://ieeexplore.ieee.org/document/8453006/

[88] C. D. Modica, P. Pinson, and S. B. Taieb, "Online Forecast Reconciliation in Wind Power Prediction," in *21st Power Systems Computation Conference*, Porto, Portugal, 2020. [Online]. Available: http://pierrepinson.com/wp-content/uploads/2020/04/DiModicaetal2020-revised.pdf

[89] L. Bai and P. Pinson, "Distributed Reconciliation in Day-Ahead Wind Power Forecasting," *Energies*, vol. 12, no. 6, p. 1112, 3 2019. [Online]. Available: https://www.mdpi.com/1996-1073/12/6/1112

[90] J. Jeon, A. Panagiotelis, and F. Petropoulos, "Probabilistic forecast reconciliation with applications to wind power and electric load," *European Journal of Operational Research*, vol. 279, no. 2, pp. 364–379, 12 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0377221719304242

[91] J. Tastu, P. Pinson, and H. Madsen, "Space-Time Trajectories of Wind Power Generation: Parametrized Precision Matrices Under a Gaussian Copula Approach," in *Modeling and Stochastic Learning for Forecasting in High Dimensions*. Springer International Publishing, 2015, pp. 267–296. [Online]. Available: http://link.springer.com/10.1007/978-3-319-18732-7_14

[92] R. J. Bessa, "On the quality of the Gaussian copula for multi-temporal decision-making problems," in *2016 Power Systems Computation Conference (PSCC)*. IEEE, 6 2016, pp. 1–7. [Online]. Available: http://ieeexplore.ieee.org/document/7541001/

[93] T. Gneiting, "Nonseparable, Stationary Covariance Functions for Space–Time Data," *Journal of the American Statistical Association*, vol. 97, no. 458, pp. 590–600, 6 2002. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1198/016214502760047113

[94] P. Ailliot and V. Monbet, "Markov-switching autoregressive models for wind time series," *Environmental Modelling & Software*, vol. 30, pp. 92–101, 4 2012. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1364815211002222

[95] 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, 7 2012. [Online]. Available: http://dx.doi.org/10.1002/for.1194

[96] J. Browell, D. R. Drew, and K. Philippopoulos, "Improved very short-term spatio-temporal wind forecasting using atmospheric regimes," *Wind Energy*, vol. 21, no. 11, pp. 968–979, 11 2018. [Online]. Available: http://doi.wiley.com/10.1002/we.2207

[97] T. Gneiting, K. Larson, K. Westrick, M. G. 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, 9 2006. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1198/016214506000000456

[98] 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, 3 2010. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1198/jasa.2009.ap08117

[99] Z. Wang, W. Wang, C. Liu, Z. Wang, and Y. Hou, "Probabilistic Forecast for Multiple Wind Farms Based on Regular Vine Copulas," *IEEE Transactions on Power Systems*, vol. 33, no. 1, pp. 578–589, 1 2018. [Online]. Available: http://ieeexplore.ieee.org/document/7909035/

[100] F. Golestaneh, P. Pinson, R. Azizipanah-Abarghooee, and H. B. Gooi, "Ellipsoidal Prediction Regions for Multivariate Uncertainty Characterization," *IEEE Transactions on Power Systems*, vol. 33, no. 4, pp. 4519–4530, 7 2018. [Online]. Available: https://ieeexplore.ieee.org/document/8253871/

[101] F. Golestaneh, P. Pinson, and H. B. Gooi, "Polyhedral Predictive Regions for Power System Applications," *IEEE Transactions on Power Systems*, vol. 34, no. 1, pp. 693–704, 1 2019. [Online]. Available: https://ieeexplore.ieee.org/document/8423709/

[102] J. K. Møller, M. Zugno, and H. Madsen, "Probabilistic Forecasts of Wind Power Generation by Stochastic Differential Equation Models," *Journal of Forecasting*, vol. 35, no. 3, pp. 189–205, 4 2016. [Online]. Available: http://doi.wiley.com/10.1002/for.2367

[103] A. Staid, J.-P. Watson, R. J. Wets, and D. L. Woodruff, "Generating short-term probabilistic wind power scenarios via nonparametric forecast error density estimators," *Wind Energy*, vol. 20, no. 12, pp. 1911–1925, 12 2017. [Online]. Available: http://doi.wiley.com/10.1002/we.2129

[104] R. P. Worsnop, M. Scheuerer, T. M. Hamill, and J. K. Lundquist, "Generating wind power scenarios for probabilistic ramp event prediction using multivariate statistical post-processing," *Wind Energy Science*, vol. 3, no. 1, pp. 371–393, 6 2018. [Online]. Available: https://www.wind-energ-sci.net/3/371/2018/

[105] R. Schefzik, T. L. Thorarinsdottir, and T. Gneiting, "Uncertainty Quantification in Complex Simulation Models Using Ensemble Copula Coupling," *Statistical Science*, vol. 28, no. 4, pp. 616–640, 11 2013. [Online]. Available: http://projecteuclid.org/euclid.ss/1386078881

[106] R. Schefzik and A. Möller, "Ensemble Postprocessing Methods Incorporating Dependence Structures," in *Statistical Postprocessing of Ensemble Forecasts*. Elsevier, 1 2018, ch. 4, pp. 91–125. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780128123720000042

[107] Z. Ben Bouallègue, T. Heppelmann, S. E. Theis, and P. Pinson, "Generation of Scenarios from Calibrated Ensemble Forecasts with a Dual-Ensemble Copula-Coupling Approach," *Monthly Weather Review*, vol. 144, no. 12, pp. 4737–4750, 12 2016. [Online]. Available: http://arxiv.org/abs/1511.05877

[108] M. Shafiee, "Maintenance logistics organization for offshore wind energy: Current progress and future perspectives," *Renewable Energy*, vol. 77, no. 1, pp. 182–193, 5 2015. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0960148114007605

[109] C. Stock-Williams and S. K. Swamy, "Automated daily maintenance planning for offshore wind farms," *Renewable Energy*, vol. 133, pp. 1393–1403, 4 2019. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0960148118310620

[110] C. A. Irawan, D. Ouelhadj, D. Jones, M. Stålhane, and I. B. Sperstad, "Optimisation of maintenance routing and scheduling for offshore wind farms," *European Journal of Operational Research*, vol. 256, no. 1, pp. 76–89, 1 2017. [Online]. Available: http://dx.doi.org/10.1016/j.ejor.2016.05.059

[111] A. Kovács, G. Erdős, Z. J. Viharos, and L. Monostori, "A system for the detailed scheduling of wind farm maintenance," *CIRP Annals*, vol. 60, no. 1, pp. 497–501, 2011. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0007850611000503

[112] R. Dawid, D. McMillan, and M. Revie, "Decision Support Tool for Offshore Wind Farm Vessel Routing under Uncertainty," *Energies*, vol. 11, no. 9, p. 2190, 8 2018. [Online]. Available: http://www.mdpi.com/1996-1073/11/9/2190

[113] M. Stålhane, H. Vefsnmo, E. E. Halvorsen-Weare, L. M. Hvattum, and L. M. Nonås, "Vessel Fleet Optimization for Maintenance Operations at Offshore Wind Farms Under Uncertainty," *Energy Procedia*, vol. 94, no. 1876, pp. 357–366, 9 2016. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1876610216308591

[114] F. Besnard, M. Patriksson, A. Stromberg, A. Wojciechowski, K. Fischer, and L. Bertling, "A stochastic model for opportunistic maintenance planning of offshore wind farms," in *2011 IEEE Trondheim PowerTech*. Trondheim: IEEE, 6 2011, pp. 1–8. [Online]. Available: http://ieeexplore.ieee.org/document/6019376/

[115] N. Raknes, K. Ødeskaug, M. Stålhane, and L. Hvattum, "Scheduling of Maintenance Tasks and Routing of a Joint Vessel Fleet for Multiple Offshore Wind Farms," *Journal of Marine Science and Engineering*, vol. 5, no. 1, p. 11, 2 2017. [Online]. Available: http://www.mdpi.com/2077-1312/5/1/11

[116] Y. Dalgic, I. Lazakis, I. Dinwoodie, D. McMillan, and M. Revie, "Advanced logistics planning for offshore wind farm operation and maintenance activities," *Ocean Engineering*, vol. 101, pp. 211–226, 6 2015. [Online]. Available: http://dx.doi.org/10.1016/j.oceaneng.2015.04.040

[117] T. Gintautas and J. D. Sørensen, "Improved Methodology of Weather Window Prediction for Offshore Operations Based on Probabilities of Operation Failure," *Journal of Marine Science and Engineering*, vol. 5, no. 2, p. 20, 5 2017. [Online]. Available: http://www.mdpi.com/2077-1312/5/2/20

[118] J. W. Taylor and J. Jeon, "Probabilistic forecasting of wave height for offshore wind turbine maintenance," *European Journal of Operational Research*, vol. 267, no. 3, pp. 877–890, 6 2018. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0377221717311402

[119] V. M. Catterson, D. McMillan, I. Dinwoodie, M. Revie, J. Dowell, J. Quigley, and K. Wilson, "An economic impact metric for evaluating wave height forecasters for offshore wind maintenance access," *Wind Energy*, vol. 19, no. 2, pp. 199–212, 2 2016. [Online]. Available: http://doi.wiley.com/10.1002/we.1826

[120] M. O'Connor, T. Lewis, and G. Dalton, "Weather window analysis of Irish west coast wave data with relevance to operations & maintenance of marine renewables," *Renewable Energy*, vol. 52, pp. 57–66, 4 2013. [Online]. Available: http://dx.doi.org/10.1016/j.renene.2012.10.021

[121] R. T. Walker, J. van Nieuwkoop-McCall, L. Johanning, and R. J. Parkinson, "Calculating weather windows: Application to transit, installation and the implications on deployment success," *Ocean Engineering*, vol. 68, pp. 88–101, 8 2013. [Online]. Available: http://dx.doi.org/10.1016/j.oceaneng.2013.04.015

[122] J. Browell, I. Dinwoodie, and D. McMillan, "Forecasting for day-ahead offshore maintenance scheduling under uncertainty," in *Risk, Reliability and Safety: Innovating Theory and Practice - Proceedings of the 26th European Safety and Reliability Conference, ESREL 2016.* Taylor & Francis Group, 6000 Broken Sound Parkway NW, Suite 300, Boca Raton, FL 33487-2742: CRC Press, 9 2017, p. 182. [Online]. Available: http://www.crcnetbase.com/doi/10.1201/9781315374987-171

[123] T. Fazeres-Ferradosa, F. Taveira-Pinto, E. Vanem, M. T. Reis, and L. d. Neves, "Asymmetric copula–based distribution models for met-ocean data in offshore wind engineering applications," *Wind Engineering*, vol. 42, no. 4, pp. 304–334, 8 2018. [Online]. Available: http://journals.sagepub.com/doi/10.1177/0309524X18777323

[124] G. Leontaris, O. Morales-Nápoles, and A. Wolfert, "Probabilistic scheduling of offshore operations using copula based environmental time series – An application for cable installation management for offshore wind farms," *Ocean Engineering*, vol. 125, pp. 328–341, 10 2016. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0029801816303614

[125] R. J. Bessa, V. Miranda, A. Botterud, and J. Wang, "'Good' or 'bad' wind power forecasts: a relative concept," *Wind Energy*, vol. 14, no. 5, pp. 625–636, 7 2011. [Online]. Available: http://doi.wiley.com/10.1002/we.444

[126] H. Madsen, P. Pinson, G. Kariniotakis, H. A. Nielsen, and T. S. Nielsen, "Standardizing the Performance Evaluation of Short-Term Wind Power Prediction Models," *Wind Engineering*, vol. 29, no. 6, pp. 475–489, 12 2005. [Online]. Available: http://journals.sagepub.com/doi/10.1260/030952405776234599

[127] 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.

[128] 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, 11 2007. [Online]. Available: http://doi.wiley.com/10.1002/we.230

[129] H. Hersbach, "Decomposition of the Continuous Ranked Probability Score for Ensemble Prediction Systems," *Weather and Forecasting*, vol. 15, no. 5, pp.

559–570, 10 2000. [Online]. Available: http://journals.ametsoc.org/doi/abs/10.1175/1520-0434%282000%29015%3C0559%3ADOTCRP%3E2.0.CO%3B2

[130] S. Bentzien and P. Friederichs, "Decomposition and graphical portrayal of the quantile score," *Quarterly Journal of the Royal Meteorological Society*, vol. 140, no. 683, pp. 1924–1934, 7 2014. [Online]. Available: http://doi.wiley.com/10.1002/qj.2284

[131] 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, 3 2007. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1198/016214506000001437

[132] D. S. Wilks, "Forecast Verification," in *Statistical Methods in the Atmospheric Sciences*. Elsevier, 2019, pp. 369–483. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/B9780128158234000092

[133] T. Gneiting, A. E. Raftery, A. H. Westveld, and T. Goldman, "Calibrated Probabilistic Forecasting Using Ensemble Model Output Statistics and Minimum CRPS Estimation," *Monthly Weather Review*, vol. 133, no. 5, pp. 1098–1118, 5 2005. [Online]. Available: http://journals.ametsoc.org/doi/abs/10.1175/MWR2904.1

[134] M. Matos, R. Bessa, A. Botterud, and Z. Zhou, "Forecasting and setting power system operating reserves," in *Renewable Energy Forecasting*. Elsevier, 2017, pp. 279–308. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/B9780081005040000111

[135] T. Gneiting and R. Ranjan, "Comparing Density Forecasts Using Threshold- and Quantile-Weighted Scoring Rules," *Journal of Business & Economic Statistics*, vol. 29, no. 3, pp. 411–422, 7 2011. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1198/jbes.2010.08110

[136] T. L. Thorarinsdottir and N. Schuhen, "Verification: Assessment of Calibration and Accuracy," in *Statistical Postprocessing of Ensemble Forecasts*. Elsevier, 4

2018, pp. 155–186. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780128123720000066

[137] M. Scheuerer and T. M. Hamill, "Variogram-Based Proper Scoring Rules for Probabilistic Forecasts of Multivariate Quantities," *Monthly Weather Review*, vol. 143, no. 4, pp. 1321–1334, 4 2015. [Online]. Available: http://journals.ametsoc.org/doi/10.1175/MWR-D-14-00269.1

[138] F. Ziel and K. Berk, "Multivariate Forecasting Evaluation: On Sensitive and Strictly Proper Scoring Rules," *[preprint]*, 10 2019. [Online]. Available: http://arxiv.org/abs/1910.07325

[139] B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap*. Boston, MA: Springer US, 1993. [Online]. Available: http://link.springer.com/10.1007/978-1-4899-4541-9

[140] D. N. Politis, "The Impact of Bootstrap Methods on Time Series Analysis," *Statistical Science*, vol. 18, no. 2, pp. 219–230, 5 2003. [Online]. Available: http://projecteuclid.org/euclid.ss/1063994977

[141] C. Roach, "Reconciled boosted models for GEFCom2017 hierarchical probabilistic load forecasting," *International Journal of Forecasting*, vol. 35, no. 4, pp. 1439–1450, 10 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0169207018301791

[142] R. J. Hyndman and S. Fan, "Density Forecasting for Long-Term Peak Electricity Demand," *IEEE Transactions on Power Systems*, vol. 25, no. 2, pp. 1142–1153, 5 2010. [Online]. Available: http://ieeexplore.ieee.org/document/5345698/

[143] J. Xie and T. Hong, "Temperature Scenario Generation for Probabilistic Load Forecasting," *IEEE Transactions on Smart Grid*, pp. 1–1, 2016. [Online]. Available: http://ieeexplore.ieee.org/document/7529083/

[144] F. X. Diebold and R. S. Mariano, "Comparing Predictive Accuracy," *Journal of Business & Economic Statistics*, vol. 13, no. 3, pp. 253–263, 7 1995.

[Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/07350015.1995.10524599

[145] F. X. Diebold, "Comparing Predictive Accuracy, Twenty Years Later: A Personal Perspective on the Use and Abuse of Diebold–Mariano Tests," *Journal of Business & Economic Statistics*, vol. 33, no. 1, pp. 1–1, 1 2015. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/07350015.2014.983236

[146] S. Drechsel, G. J. Mayr, J. W. Messner, and R. Stauffer, "Wind speeds at heights crucial for wind energy: Measurements and verification of forecasts," *Journal of Applied Meteorology and Climatology*, vol. 51, no. 9, pp. 1602–1617, 2012.

[147] T. Heus, C. C. van Heerwaarden, H. J. J. Jonker, A. Pier Siebesma, S. Axelsen, K. van den Dries, O. Geoffroy, A. F. Moene, D. Pino, S. R. de Roode, and J. Vilà-Guerau de Arellano, "Formulation of the Dutch Atmospheric Large-Eddy Simulation (DALES) and overview of its applications," *Geoscientific Model Development*, vol. 3, no. 2, pp. 415–444, 9 2010. [Online]. Available: https://www.geosci-model-dev.net/3/415/2010/

[148] J. Schalkwijk, H. J. J. Jonker, A. P. Siebesma, and E. Van Meijgaard, "Weather Forecasting Using GPU-Based Large-Eddy Simulations," *Bulletin of the American Meteorological Society*, vol. 96, no. 5, pp. 715–723, 5 2015. [Online]. Available: http://journals.ametsoc.org/doi/10.1175/BAMS-D-14-00114.1

[149] R. J. Stevens, J. Graham, and C. Meneveau, "A concurrent precursor inflow method for Large Eddy Simulations and applications to finite length wind farms," *Renewable Energy*, vol. 68, pp. 46–50, 8 2014. [Online]. Available: http://dx.doi.org/10.1016/j.renene.2014.01.024

[150] J. Meyers and C. Meneveau, "Large Eddy Simulations of Large Wind-Turbine Arrays in the Atmospheric Boundary Layer," in *48th AIAA Aerospace Sciences Meeting Including the New Horizons Forum and Aerospace Exposition*, no. January. Reston, Virigina: American Institute of Aeronautics and Astronautics, 1 2010, pp. 1–10. [Online]. Available: http://arc.aiaa.org/doi/10.2514/6.2010-827

[151] M. L. Aitken, R. M. Banta, Y. L. Pichugina, and J. K. Lundquist, "Quantifying wind turbine wake characteristics from scanning remote sensor data," *Journal of Atmospheric and Oceanic Technology*, vol. 31, no. 4, pp. 765–787, 4 2014. [Online]. Available: http://journals.ametsoc.org/doi/10.1175/JTECH-D-13-00104.1

[152] T. L. Thorarinsdottir and T. Gneiting, "Probabilistic forecasts of wind speed: ensemble model output statistics by using heteroscedastic censored regression," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, vol. 173, no. 2, pp. 371–388, 4 2010. [Online]. Available: http://doi.wiley.com/10.1111/j.1467-985X.2009.00616.x

[153] W. Messner, Jakob, J. Mayr, Georg, and A. Zeileis, "Heteroscedastic Censored and Truncated Regression with crch," *The R Journal*, vol. 8, no. 1, p. 173, 2016. [Online]. Available: https://journal.r-project.org/archive/2016-1/messner-mayr-zeileis.pdf

[154] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *Annals of Statistics*, vol. 29, no. 5, pp. 1189–1232, 2001.

[155] A. Natekin and A. Knoll, "Gradient boosting machines, a tutorial," *Frontiers in Neurorobotics*, vol. 7, no. DEC, p. 21, 2013. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/24409142

[156] T. Hastie, R. Tibshirani, and J. H. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. Springer, 2009.

[157] J. H. Friedman, "Stochastic gradient boosting," *Computational Statistics & Data Analysis*, vol. 38, no. 4, pp. 367–378, 2 2002. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167947301000652

[158] The H2O.ai team, "h2o: R Interface for H2O," 2018. [Online]. Available: https://github.com/h2oai/h2o-3

[159] C. Vincent, G. Giebel, P. Pinson, and H. Madsen, "Resolving Nonstationary Spectral Information in Wind Speed Time Series Using the Hilbert–Huang

Transform," *Journal of Applied Meteorology and Climatology*, vol. 49, no. 2, pp. 253–267, 2 2010. [Online]. Available: http://journals.ametsoc.org/doi/10.1175/2009JAMC2058.1

[160] P. A. Jiménez, J. Navarro, A. M. Palomares, and J. Dudhia, "Mesoscale modeling of offshore wind turbine wakes at the wind farm resolving scale: a composite-based analysis with the Weather Research and Forecasting model over Horns Rev," *Wind Energy*, vol. 18, no. 3, pp. 559–566, 3 2015. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/we.1708

[161] J. Sanz Rodrigo, R. A. Chávez Arroyo, P. Moriarty, M. Churchfield, B. Kosović, P.-E. Réthoré, K. S. Hansen, A. Hahmann, J. D. Mirocha, and D. Rife, "Mesoscale to microscale wind farm flow modeling and evaluation," *Wiley Interdisciplinary Reviews: Energy and Environment*, vol. 6, no. 2, p. e214, 3 2017. [Online]. Available: http://doi.wiley.com/10.1002/wene.214

[162] A. K. Joensen, G. Giebel, L. Landberg, H. Madsen, and H. A. Nielsen, "Model output statistics applied to wind power prediction," in *European Wind Energy Conference*, Nice, France, 1999, pp. 1177–1180.

[163] C. Gilbert, J. Browell, and D. McMillan, "A Hierarchical Approach to Probabilistic Wind Power Forecasting," in *2018 IEEE International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*. Boise: IEEE, 6 2018, pp. 1–6. [Online]. Available: https://ieeexplore.ieee.org/document/8440571/

[164] G. Anastasiades and P. McSharry, "Quantile forecasting ofwind power using variability indices," *Energies*, vol. 6, no. 2, pp. 662–695, 2 2013. [Online]. Available: www.mdpi.com/journal/energiesArticle

[165] H.-z. Wang, G.-q. Li, G.-b. Wang, J.-c. Peng, H. Jiang, and Y.-t. Liu, "Deep learning based ensemble approach for probabilistic wind power forecasting," *Applied Energy*, vol. 188, pp. 56–70, 2 2017. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0306261916317421

[166] J. Catalão, H. Pousinho, and V. Mendes, "Short-term wind power forecasting in Portugal by neural networks and wavelet transform," *Renewable Energy*, vol. 36, no. 4, pp. 1245–1251, 4 2011. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0960148110004477

[167] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 67, no. 2, pp. 301–320, 4 2005. [Online]. Available: http://doi.wiley.com/10.1111/j.1467-9868.2005.00503.x

[168] R. B. Nelsen, *An Introduction to Copulas*, ser. Springer Series in Statistics. New York, NY: Springer New York, 2006. [Online]. Available: http://link.springer.com/10.1007/0-387-28678-0

[169] K. Aas, C. Czado, A. Frigessi, and H. Bakken, "Pair-copula constructions of multiple dependence," *Insurance: Mathematics and Economics*, vol. 44, no. 2, pp. 182–198, 4 2009. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167668707000194

[170] R Development Core Team 3.0.1., "A Language and Environment for Statistical Computing," *R Foundation for Statistical Computing*, vol. 2, pp. https://www.R–project.org, 2013. [Online]. Available: https://www.r-project.org/

[171] J. Friedman, T. Hastie, and R. Tibshirani, "Regularization Paths for Generalized Linear Models via Coordinate Descent," *Journal of Statistical Software*, vol. 33, no. 1, pp. 1–22, 2010. [Online]. Available: http://www.jstatsoft.org/v33/i01/

[172] K. Hechenbichler and K. Schliep, "Weighted k-Nearest-Neighbor Techniques and Ordinal Classification," Collaborative Research Center 386, Tech. Rep., 2004. [Online]. Available: https://epub.ub.uni-muenchen.de/1769/

[173] G. Ridgeway and H. Southworth, "gbm: Generalized boosted regression models. {R} package version 2.1," 2014. [Online]. Available: http://cran.r-project.org/package=gbm

[174] R. J. Bessa, V. Miranda, A. Botterud, J. Wang, and E. M. Constantinescu, "Time Adaptive Conditional Kernel Density Estimation for Wind Power Forecasting," *IEEE Transactions on Sustainable Energy*, vol. 3, no. 4, pp. 660–669, 10 2012. [Online]. Available: http://ieeexplore.ieee.org/document/6220264/

[175] R. Tawn, J. Browell, and I. Dinwoodie, "Missing data in wind farm time series: Properties and effect on forecasts," *Electric Power Systems Research*, vol. 189, 2020.

[176] B. Maples, G. Saur, M. Hand, R. Van De Pietermen, and T. Obdam, "Installation, Operation, and Maintenance Strategies to Reduce the Cost of Offshore Wind Energy," National Renewable Energy Laboratory (NREL), Tech. Rep., 2013.

[177] F. N. Fritsch and R. E. Carlson, "Monotone Piecewise Cubic Interpolation," *SIAM Journal on Numerical Analysis*, vol. 17, no. 2, pp. 238–246, 4 1980. [Online]. Available: http://epubs.siam.org/doi/10.1137/0717021

[178] C. Gonçalves, L. Cavalcante, M. Brito, R. J. Bessa, and J. Gama, "Forecasting conditional extreme quantiles for wind energy," *Electric Power Systems Research*, vol. 190, 2021.

[179] R. A. Rigby and D. M. Stasinopoulos, "Generalized additive models for location, scale and shape (with discussion)," *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, vol. 54, no. 3, pp. 507–554, 6 2005. [Online]. Available: http://doi.wiley.com/10.1111/j.1467-9876.2005.00510.x

[180] B. Hofner, A. Mayr, and M. Schmid, "gamboostLSS : An R Package for Model Building and Variable Selection in the GAMLSS Framework," *Journal of Statistical Software*, vol. 74, no. 1, pp. 1–31, 2016. [Online]. Available: http://www.jstatsoft.org/v74/i01/

[181] A. Mayr, N. Fenske, B. Hofner, T. Kneib, and M. Schmid, "Generalized additive models for location, scale and shape for high dimensional data-a flexible approach based on boosting," *Journal of the Royal Statistical Society: Series*

C (Applied Statistics), vol. 61, no. 3, pp. 403–427, 5 2012. [Online]. Available: http://doi.wiley.com/10.1111/j.1467-9876.2011.01033.x

[182] P. H. C. Eilers and B. D. Marx, "Flexible smoothing with B -splines and penalties," *Statistical Science*, vol. 11, no. 2, pp. 89–121, 5 1996. [Online]. Available: http://projecteuclid.org/euclid.ss/1038425655

[183] B. Hofner, T. Kneib, and T. Hothorn, "A unified framework of constrained regression," *Statistics and Computing*, vol. 26, no. 1-2, pp. 1–14, 1 2016. [Online]. Available: http://link.springer.com/10.1007/s11222-014-9520-y

[184] A. Perperoglou, W. Sauerbrei, M. Abrahamowicz, and M. Schmid, "A review of spline function procedures in R," *BMC Medical Research Methodology*, vol. 19, no. 1, p. 46, 12 2019. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/30841848

[185] J. Dowell, A. Zitrou, L. Walls, T. Bedford, and D. Infield, "Analysis of wind and wave data to assess maintenance access to offshore wind farms," in *Safety, Reliability and Risk Analysis: Beyond the Horizon - Proceedings of the European Safety and Reliability Conference, ESREL 2013.* CRC Press, 9 2014, pp. 743–750. [Online]. Available: http://www.crcnetbase.com/doi/abs/10.1201/b15938-114

[186] C. Gilbert, J. Browell, and D. McMillan, "Code for: "Probabilistic Access Forecasting for Improved Offshore Operations" — University of Strathclyde," 2020. [Online]. Available: https://doi.org/10.15129/d4055a23-f2f7-482d-9cb2-4e15960b17f6

[187] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861–874, 6 2006. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S016786550500303X

[188] D. M. Stasinopoulos and R. A. Rigby, "Generalized Additive Models for Location Scale and Shape (GAMLSS) in R," *Journal of Statistical Software*, vol. 23, no. 7, pp. 1–46, 4 2007. [Online]. Available: http://www.jstatsoft.org/v23/i07/

[189] G. Will, "Visualizing and Clustering Data that Includes Circular Variables," Montana State University, Tech. Rep., 2016. [Online]. Available: http://www.math.montana.edu/graduate/writing-projects/2016/16will.pdf

[190] M. A. Matos, R. J. Bessa, C. Goncalves, L. Cavalcante, V. Miranda, N. Machado, P. Marques, and F. Matos, "Setting the maximum import net transfer capacity under extreme RES integration scenarios," in *2016 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*. IEEE, 10 2016, pp. 1–7. [Online]. Available: https://ieeexplore.ieee.org/document/7764145/

[191] J. Tastu, H. Madsen, and P. Pinson, "Short-term wind power forecasting: probabilistic and space-time aspects," Technical University of Denmark, Tech. Rep., 2013.

[192] T. Gneiting, M. Genton, and P. Guttorp, "Geostatistical Space-Time Models, Stationarity, Separability, and Full Symmetry," in *Statistical Methods for Spatio-Temporal Systems*, B. Finkenstaedt, L. Held, and V. Isham, Eds. Statistical Methods for Spatio-Temporal Systems, 10 2006, pp. 151–175. [Online]. Available: http://www.crcnetbase.com/doi/abs/10.1201/9781420011050.ch4

[193] O. Ledoit and M. Wolf, "A well-conditioned estimator for large-dimensional covariance matrices," *Journal of Multivariate Analysis*, vol. 88, no. 2, pp. 365–411, 2 2004. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0047259X03000964

[194] N. J. Higham, "Computing the nearest correlation matrix–a problem from finance," *IMA Journal of Numerical Analysis*, vol. 22, no. 3, pp. 329–343, 7 2002.

[195] J. Friedman, T. Hastie, and R. Tibshirani, "Sparse inverse covariance estimation with the graphical lasso," *Biostatistics*, vol. 9, no. 3, pp. 432–441, 7 2008. [Online]. Available: http://arxiv.org/abs/0708.3517

[196] A. Couto, P. Costa, L. Rodrigues, V. V. Lopes, and A. Estanqueiro, "Impact of Weather Regimes on the Wind Power Ramp Forecast in Portugal," *IEEE*

Transactions on Sustainable Energy, vol. 6, no. 3, pp. 934–942, 7 2015. [Online]. Available: http://ieeexplore.ieee.org/document/6863716/

[197] R. Neal, D. Fereday, R. Crocker, and R. E. Comer, "A flexible approach to defining weather patterns and their application in weather forecasting over Europe," *Meteorological Applications*, vol. 23, no. 3, pp. 389–400, 7 2016. [Online]. Available: http://doi.wiley.com/10.1002/met.1563

[198] J. Browell and C. Gilbert, "ProbCast: Open-source production, evaluation and visualisation of probabilistic forecasts," in *PMAPS 2020 - the 16th International Conference on Probabilistic Methods Applied to Power Systems*, Liege, Belgium, 2020. [Online]. Available: https://pureportal.strath.ac.uk/en/publications/probcast-open-source-production-evaluation-and-visualisation-of-p

[199] J. Elith, J. R. Leathwick, and T. Hastie, "A working guide to boosted regression trees," *Journal of Animal Ecology*, vol. 77, no. 4, pp. 802–813, 7 2008. [Online]. Available: http://avesbiodiv.mncn.csic.es/estadistica/bt1.pdf