# Leveraging Turbine-level Data for Improved Probabilistic Wind Power Forecasting

Ciaran Gilbert, *Student Member, IEEE*, Jethro Browell, *Member, IEEE*, and David McMillan, *Member, IEEE*

*Abstract*—This paper describes two methods for creating improved probabilistic wind power forecasts through the use of turbine-level data. The first is a feature engineering approach whereby deterministic power forecasts from the turbine-level are used as explanatory variables in a wind farm level forecasting model. The second is a novel bottom-up hierarchical approach where the wind farm forecast is inferred from the joint predictive distribution of the power output from individual turbines. Notably, the latter produces probabilistic forecasts that are coherent across both turbine and farm levels, which the former does not. The methods are tested at two utility scale wind farms and are shown to provide consistent improvements of up to 5% in terms of continuous ranked probability score compared to the best performing state-of-the-art benchmark model. The bottom-up hierarchical approach provides greater improvement at the site characterised by a complex layout and terrain, while both approaches perform similarly at the second location. We show that there is a clear benefit in leveraging readily available turbine-level information for wind power forecasting.

## I. INTRODUCTION

THE growth of weather dependent renewable energy sources is transforming power systems across the world with wide-ranging implications for system operation and market design. Energy forecasting is essential for reliable and economic power system operation due to the uncertain variation of supply and demand, which increases the difficulty of balancing the network and managing power flows [1].

Here we are concerned with short-term forecasting, where the prediction horizon is several hours to days ahead. For example, this type of forecast is used to inform trading strategies for participants in the day-ahead market and for power system operations. Numerical Weather Predictions (NWP) are key inputs into wind power forecasting models at such horizons [2]. Best practice in creating these models involves mapping the relationship between meteorological forecasts and corresponding wind farm power generation via a statistical learning technique. This allows for a fully data-driven statistical model, using inputs from the physics-based NWP, which implicitly accounts for complex physical processes influencing the wind to power conversion such as wake losses and any systematic bias in the weather forecasts [3].

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 [4]. However, there has been extensive research in the area of probabilistic forecasting, which is reviewed comprehensively in [5], driven by the economic value of quantifying uncertainty when making decisions [3]. Uncertainty

C. Gilbert, J. Browell and D. McMillan are with the University of Strathclyde, Glasgow, G1 1XQ, UK. email: ciaran.gilbert@strath.ac.uk

at a single point in the future is commonly quantified by producing a predictive probability distribution of future power production, called a density forecast. Density forecasts are central to probabilistic forecasting, and as such have received much attention from the research community. Non-parametric methods, where no particular distribution shape is assumed, have emerged as superior to estimating parametric probability distributions conditional on NWP and other inputs. Popular statistical methods for generating these forecasts emerged as additive quantile regression with splines [6], adapted resampling [7], and conditional Kernel Density Estimation [8].

More recently, a number of competitions have been run in order to compare forecasting methods on the same dataset and under controlled conditions [9]. The two winning teams from GEFCom (2012 and 2014) utilised Gradient Boosting regression Trees (GBT), the latter for quantile regression to produce density forecasts, with input features engineered from NWPs [10], [11]. Other entrants also employed GBTs but did not produce as skillful forecasts highlighting the importance of feature engineering in such methods. This approach was extended with spatio-temporal features engineered from a grid of NWP points to improve probabilistic forecast performance of both solar and wind power in [12]. Analog ensemble methods have also been successful in producing non-parametric density forecasts [13] where here the definition of the distance measure used to define the ensemble is critical and similar to feature engineering in regression.

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. Additionally, including coherency constraints in predictive models can improve performance at all levels of the hierarchy. Hierarchies can be both spatial and temporal in nature [14], [15]. 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 [16], [17]. As discussed in [18], 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, this is not the case for wind farms where each wind turbine provides a consistent weather dependent signal. In the wind power forecasting domain, [19] evaluates a method of deterministic forecast reconciliation via a generalised least squares method to generate coherent forecasts.

The concept of coherent *probabilistic* forecasts is explored in [18], [20] 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.

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 [21]. The most frequently used family is the Gaussian copula [22], [23], though temporal dependency has received more attention than spatial or spatial-temporal dependency. 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 [23], [24].

In this study, two methods are investigated to leverage turbine-level data and are compared to state-of-the-art benchmarks. The first is a feature engineering approach proposed in [25], 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 is not guaranteed. This work also expands on [25] by extending the case study to a second wind farm with different site characteristics and testing a second novel approach 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 forecasts reflect 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 paper are the proposed bottom-up hierarchical method and its evaluation, plus significantly expanding the evaluation of the feature engineering approach first introduced in [25]. We hypothesise that leveraging information from the turbine-level will enable us to improve 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. The advantages of the proposed hierarchical method are improved accuracy and coherency between turbine-level and wind farm total, however the nature of the wind farm (terrain, layout, size...) has a bearing on the extent of this improvement.

This paper is organised as follows: Section II details the forecasting methods and benchmark models, Section III describes the case study based on two utility scale wind farms in the UK, Section IV presents and evaluates the results, and conclusions are drawn in Section V. Supplementary information provides additional detail and results [26].

## II. 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 1, which details the training process, input data, and output forecast of each model. The turbine level feature engineering model 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 [25]. 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 [21], [22], [27]; vine copulas with a range of copula families are also considered [23], [24].

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 [11], [12]. Cubic spline basis functions are also included to capture diurnal bias in the NWP at the specific sites along the lines of [28]. Full details of all features are listed in the supplementary material [26].

### A. Gradient Boosting Trees

This section introduces the statistical learning technique used 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. The Gradient Boosting regression Tree algorithm (GBT) is an ensemble learning technique whereby powerful predictive models can be constructed by combining a number of individual regression trees, known as *weak learners* [29]. This technique can capture non-linear relationships such as the wind power curve, can be used with a variety of differentiable loss functions, and can intrinsically learn interactions between input features. GBTs are also naturally regularised by virtue of the way trees are constructed [30]. The use of GBTs for quantile regression here is motivated by their success in similar applications [10], [11] though this element could be substituted for other supervised learning algorithms. The gradient boosted tree $F_n(\mathbf{x}_t)$ is defined as the sum of $n$ regression trees

$$F_n(\mathbf{x}_t) = \sum_{i=1}^{n} f_i(\mathbf{x}_t) \tag{1}$$

where each $f_i(\mathbf{x}_t)$ is a regression tree. The ensemble of regression trees is constructed sequentially by estimating the new regression tree $f_{n+1}(\mathbf{x}_t)$ via

$$\underset{f_{n+1}}{\text{argmin}} \sum_t L\left(y_t, F_n(\mathbf{x}_t) + f_{n+1}(\mathbf{x}_t)\right) \tag{2}$$
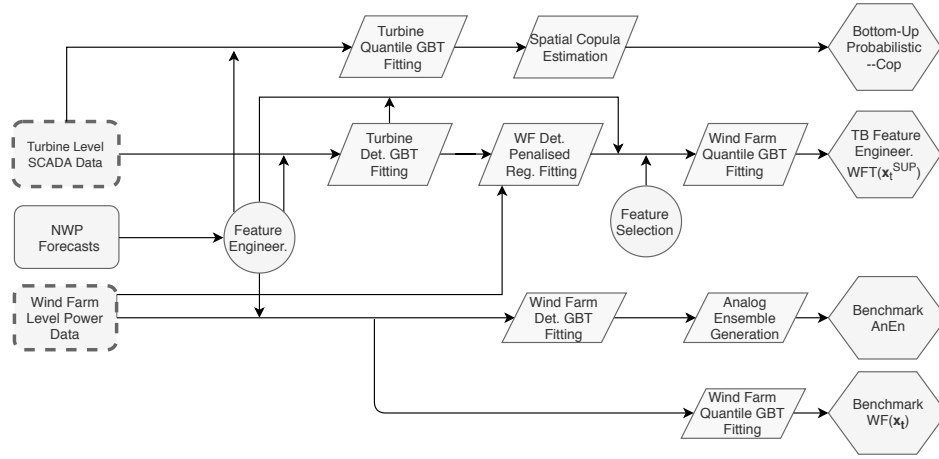
Fig. 1: Flowchart illustration of the entire forecasting model training methodology. Wind Farm Level Power Data and Turbine Level SCADA data are only required for model training and evaluation. The only inputs required to produce operational forecasts are Numerical Weather Predictions. TB indicates turbine, WF is wind farm, Det. is deterministic, and Reg. is regression

for some loss function $L(\cdot)$. Where $L(\cdot)$ is differentiable, this optimisation can be solved by steepest descent [30]. Turbine-level deterministic forecasts used as features in this study are produced by GBTs fit with a squared loss function, and density forecasts are produced using GBTs for multiple quantile regression (quantile loss function [1]) and then using spline interpolation, with knots at each predicted quantile and the boundaries 0 and nominal power, to estimate the predictive Cumulative Distribution Function (CDF).

GBTs include several hyper-parameters which control both the tree fitting and boosting processes to optimise performance while preventing over-fitting [31]. The two key parameters tuned to minimise out-of-sample error via $k$-fold cross validation are the interaction depth and shrinkage. The interaction depth is the number of splits allowed to partition the input variable space per tree and the shrinkage or learning rate controls the weight of each tree in the ensemble.

### B. 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 the supplementary information [26]. The first benchmark is a wind farm-level GBT quantile regression model, WF($\mathbf{x}_t$), 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 [10], [11], [13] 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 GBT 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 [13]. For more information, the reader is referred to [32].

### C. 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 from related work by the authors [25]. 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

$$z_t = \sum_{i=1}^{D} \omega_i y_{i,t} + \epsilon_t \qquad (3)$$

which completes the constitution of the supplementary feature set $\mathbf{x}_\mathbf{t}^{\text{SUP}} = [\mathbf{x}_t, y_{1,t}, ..., y_{D,t}, z_t]$. The weights $\omega$ are estimated via elastic net regression motivated by the necessity to regularise turbine forecasts because they are highly correlated. The weights are calculated via

$$\boldsymbol{\omega} = \underset{\boldsymbol{\omega}}{\arg\min}\bigg\{ \frac{1}{2N}||\mathbf{Z} - \mathbf{Y}\boldsymbol{\omega}||_2^2 + \\ \lambda\bigg[(1-\alpha)\frac{1}{2}||\boldsymbol{\omega}||_2^2 + \alpha||\boldsymbol{\omega}||_1\bigg]\bigg\} \quad (4)$$

where $\alpha$ and $\lambda$ are hyper parameters requiring tuning, $\mathbf{Z}$ and $\mathbf{Y}$ are matrices of vertically stacked instances of $z_t$ and $\mathbf{y}_t$ [33]. 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 $\alpha$ and $\lambda$ are determined through grid search and $k$-fold cross validation.

The final wind farm level density forecast, WFT($\mathbf{x}_t^{\text{SUP}}$), 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 GBT model with all the available inputs from $\mathbf{x}_t^{\text{SUP}}$, then selecting and retaining only the features that have the greatest influence. 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 the supplementary information [26]. 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 [30].

### D. 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 $i^{\text{th}}$ turbine, and $y_i$ the corresponding realisation (time indices are dropped to avoid notational clutter). The predictive CDF of the $i^{\text{th}}$ turbine is

$$F_i(y_i) = P(Y_i \leq y_i) \tag{5}$$

for $i = 1, 2, ..D$ turbines. Sklar's theorem [34] 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

$$F(y_1, y_2, ..., y_D) = C\left(F_1(y_1), F_2(y_2), ..., F_D(y_D)\right) \quad, \tag{6}$$

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 6 can be alternatively written as

$$C(u_1, u_2, ..., u_D) = F\left(F_1^{-1}(u_1), F_2^{-1}(u_2), ..., F_D^{-1}(u_D)\right) \tag{7}$$

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 [21]. Next we introduce a range of options for the copula function.

*1) Gaussian Copula:* The Gaussian copula is given by

$$C(F_1(y_1), F_2(y_2), ..., F_D(y_D)) =$$
$$\Phi_\Sigma\left(\Phi^{-1}(F_1(y_1)), \Phi^{-1}(F_2(y_2)), ..., \Phi^{-1}(F_D(y_D))\right) \tag{8}$$

where $\Phi^{-1}(\cdot)$ indicates the inverse standard normal distribution function and $\Phi_\Sigma(\cdot)$ the $D$-dimensional normal distribution function with covariance matrix $\Sigma$ 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

$$v_i = \Phi^{-1}\left(F_i(y_i)\right) \tag{9}$$

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 $\hat{v}_i$. Each of the samples are back-transformed

$$\hat{u}_i = \Phi(\hat{v}_i) \tag{10}$$

and then transformed into the original power domain using the inverse CDF for the $i^{\text{th}}$ turbine

$$\hat{y}_i = F_i^{-1}(\hat{u}_i) \tag{11}$$

which are summed over the $D$-turbines to give a snapshot of the wind farm forecast generation $\hat{z}_j$ for a $j^{\text{th}}$ 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

$$\hat{F}(z) = \frac{1}{K} \sum_{j=1}^{K} \mathbf{1}(\hat{z}_j \leq z) \quad. \tag{12}$$

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 [21], [27], 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

$$\Sigma_{i,j} = \text{cov}(v_i, v_j) = \exp\left(-\frac{\Delta s}{\eta}\right) \tag{13}$$

where $\Delta s$ is the distance between turbines $i$ and $j$, and the parameter $\eta$ is fit using weighted least squares regression using empirical covariance and distance information.

*2) 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 [23], [24]. 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 [23] and for more detail please refer to [35].

## III. 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 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 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` [36]–[40].

## IV. RESULTS

The skill of probabilistic forecasts is evaluated using proper scoring rules and according the principle that it is desirable for density forecasts to be as sharp as possible subject to calibration [41]. Sharpness is a measure of the spread of the distribution and calibration (or reliability) is the property that the forecast spread matches that of the observations. Calibration of individual quantiles $q$ is calculated as

$$\hat{a}^{(q)} = \frac{1}{N} \sum_{t=1}^{N} \mathbf{1}(z_t < \hat{z}_t^{(q)}) \quad , \qquad (14)$$

where $\mathbf{1}(\cdot)$ is the indicator function. If the forecast is calibrated, the empirical coverage should satisfy $\hat{a}^{(q)} \approx q$ for all $q$ [1]. Calibration is visualised using reliability diagrams and quantile bias $b^{(q)} = q - \hat{a}^{(q)}$ [42]. The sharpness and
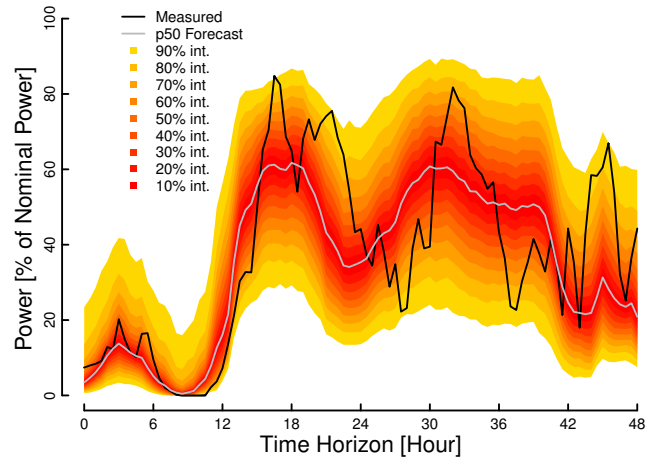


Fig. 2: Example density forecast using the parametric Gaussian copula approach at Wind Farm A

calibration can be both quantified via the Continuous Ranked Probability Score (CRPS) [41]

$$\text{CRPS} = \frac{1}{N} \sum_{t=1}^{N} \int_{-\infty}^{\infty} \{F_t(z) - \mathbf{1}(z \geq z_t)\}^2 dz \qquad (15)$$

which compares the predictive forecast distribution $F_t$ to observation $z_t$ and rewards both sharpness and reliability.

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. 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 the supplementary information [26].

At Wind Farm A, all of the proposed methods show improvements over the two benchmarks across the whole forecast horizon. The CRPS and improvement over benchmark metrics at Wind Farm A are detailed in Table I. The feature engineering method reduces CRPS by 3.95% and 5.46% compared to direct wind farm-level forecasting using

TABLE I: Results at Wind Farm A. $\%\Delta$ indicates improvement compared to specified benchmark [CRPS in % of max power]

| Horizon | Benchmarks | | Feature Engineering | | | Gaussian Copula, Empirical Cov. | | | Gaussian Copula, Parametric Cov. | | | Copula Vine | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | WF($x_t$) [11] | AnEn [32] | WFT($x_t^{SUP}$) | WF($x_t$) | AnEn | EGCop | WF($x_t$) | AnEn | PGCop | WF($x_t$) | AnEn | VCop | WF($x_t$) | AnEn |
| Hour | CRPS | CRPS | CRPS | $\%\Delta$ | $\%\Delta$ | CRPS | $\%\Delta$ | $\%\Delta$ | CRPS | $\%\Delta$ | $\%\Delta$ | CRPS | $\%\Delta$ | $\%\Delta$ |
| 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 |



(a) Quantile bias          (b) Reliability of selected methods
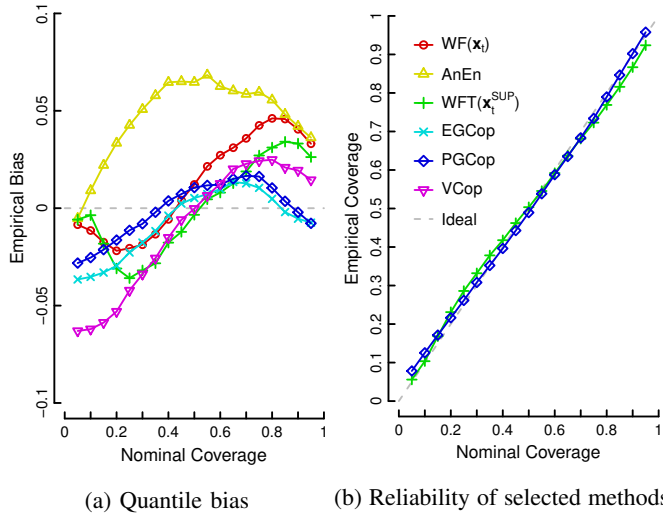
Fig. 3: Wind Farm A calibration plots



(a) Quantile bias          (b) Reliability of selected methods
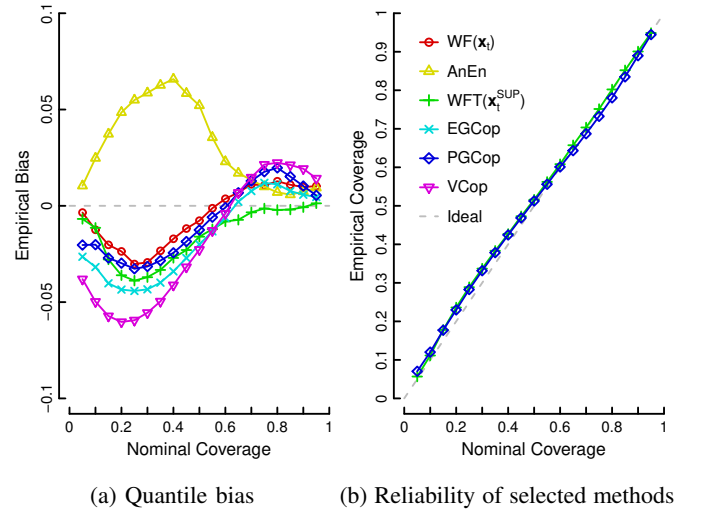
Fig. 4: Wind Farm B calibration plots

WF($x_t$) and AnEn respectively. The only difference between the WF($x_t$) 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($x_t$) and AnEn respectively. The calibration plots in Figure 3 reveal that the turbine-level feature engineering and copula methods also marginally improve the reliability of the forecast compared to the WF($x_t$) 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 II, all proposed methods outperform the benchmarks, though to a lesser extent than Wind Farm A. Unlike Wind Farm A, the feature engineering approach provides the greatest improvement reducing CRPS by 1.24% and 2.39% compared to the WF($x_t$) and AnEn benchmarks respectively. This improvement is also consistent across lead-times. The quantile bias plots, shown in Figure 4a, illustrate that the model calibration is slightly diminished when compared to the WF($x_t$) benchmark from the 15th-60th percentile, but otherwise provides improvement outside this range. The reliability diagram in Figure 4b reveals that the proposed models are well calibrated and that variations between the models are small.

Bootstrapping [43] 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 I and II. The results of this process are presented via boxplots in Figure 5 and show that improvement in CRPS compared to benchmarks is pronounced at both sites.

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($x_t$) benchmark.

The regular layout of turbines at Wind Farm B is evident in the covariance matrix for that wind farm, shown in Figure 6 and the layout of the farm can be found in the supplementary material [26]. 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 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 7,

TABLE II: Results for Wind Farm B. $\%\Delta$ indicates improvement compared to specified benchmark [CRPS in % of max power]

| Horizon | Benchmarks | | Feature Engineering | | | Gaussian Copula, Empirical Cov. | | | Gaussian Copula, Parametric Cov. | | | Copula Vine | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | WF($\mathbf{x}_t$) [11] | AnEn [32] | WFT($\mathbf{x}_t^{SUP}$) | 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 | $\%\Delta$ | $\%\Delta$ | CRPS | $\%\Delta$ | $\%\Delta$ | CRPS | $\%\Delta$ | $\%\Delta$ | CRPS | $\%\Delta$ | $\%\Delta$ |
| 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) Wind Farm A      (b) Wind Farm B
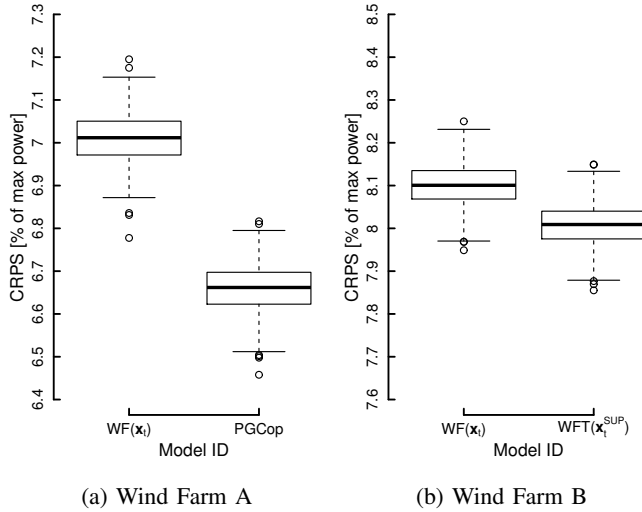
Fig. 5: Boxplots showing the bootstrap sample distributions of mean CRPS for the best benchmark and proposed model at both wind farms



Fig. 6: Parametric covariance matrix at Wind Farm B

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 III. The median (p50 in Figure 2) of each predictive distribution is taken as the deterministic forecast and evaluated in terms of Mean Absolute Error (MAE) [44]. As expected, the behaviour of the results is very similar to the probabilistic case. Performance evaluations separated by forecast horizon and in terms of Root Mean Square Error are available in the supplementary information [26].

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.
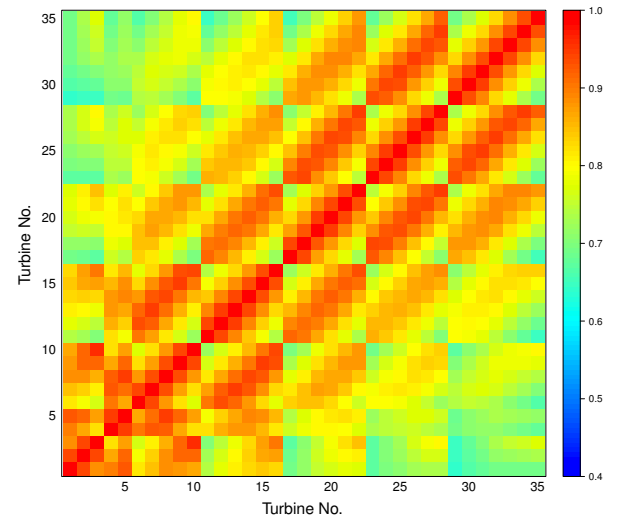
TABLE III: Deterministic forecast performance based on the median (p50) of each predictive distribution — $\%\Delta$ indicates improvement compared to specified benchmark [MAE as % of nominal power]

| Model | Wind Farm A | | | Wind Farm B | | |
|---|---|---|---|---|---|---|
| | MAE | $\%\Delta$WF($\mathbf{x}_t$) | $\%\Delta$AnEn | MAE | $\%\Delta$WF($\mathbf{x}_t$) | $\%\Delta$AnEn |
| WF($\mathbf{x}_t$) [11] | 9.69 | – | – | 11.39 | – | – |
| AnEn [32] | 9.88 | – | – | 11.49 | – | – |
| WFT($\mathbf{x}_t^{SUP}$) | 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 |

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 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

(a) Covariance matrix

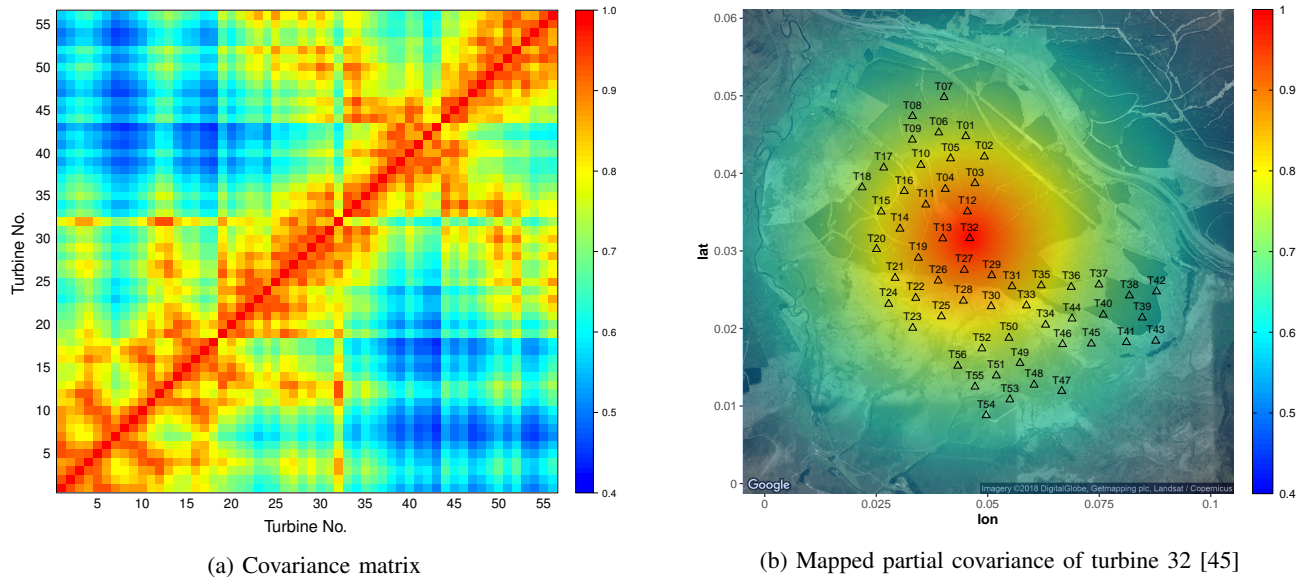

(b) Mapped partial covariance of turbine 32 [45]

Fig. 7: Parametric covariance plots at Wind Farm A. Note that the latitude and longitude scales of (b) are indicative

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.

## V. CONCLUSIONS

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 [25]. 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, [22] 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.

## ACKNOWLEDGEMENTS

## REFERENCES

[1] J. M. Morales, A. J. Conejo, H. Madsen, P. Pinson, and M. Zugno, *Integrating Renewables in Electricity Markets*, 2014, vol. 205.

[2] G. Giebel and G. Kariniotakis, "Wind power forecasting – a review of the state of the art," in *Renewable Energy Forecasting*, G. Kariniotakis, Ed., 2017, pp. 59 – 109.

[3] 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, 2017.

[4] A. Costa, A. Crespo, J. Navarro, G. Lizcano, H. Madsen, and E. Feitosa, "A review on the young history of the wind power short-term prediction," *Renewable and Sustainable Energy Reviews*, vol. 12, no. 6, pp. 1725–1744, 2008.

[5] 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, 2014.

[6] H. A. Nielsen, H. Madsen, and T. S. Nielsen, "Using quantile regression to extend an existing wind power forecasting system with probabilistic forecasts," *Wind Energy*, vol. 9, no. 1-2, pp. 95–108, 2006.

[7] P. Pinson and G. Kariniotakis, "Conditional Prediction Intervals of Wind Power Generation," *IEEE Transactions on Power Systems*, vol. 25, no. 4, pp. 1845–1856, 2010.

[8] 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, 2012.

[9] 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, 2016.

[10] L. Silva, "A feature engineering approach to wind power forecasting: GEFCom 2012," *International Journal of Forecasting*, vol. 30, no. 2, pp. 395–401, 2014.

[11] 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, 2016.

[12] J. R. Andrade and R. J. Bessa, "Improving Renewable Energy Forecasting with a Grid of Numerical Weather Predictions," *IEEE Transactions on Sustainable Energy*, vol. PP, no. 99, p. 1, 2017.

[13] 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.

[14] 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.

[15] 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.

[16] B. J. Dangerfield and J. S. Morris, "Top-down or bottom-up: Aggregate versus disaggregate extrapolations," *International Journal of Forecasting*, vol. 8, pp. 233–241, 1992.

[17] 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, 2011.

[18] S. B. Taieb, J. W. Taylor, and R. J. Hyndman, "Hierarchical Probabilistic Forecasting of Electricity Demand with Smart Meter Data," 2017, accessed 2019-01-05. [Online]. Available: https://robjhyndman.com/papers/HPFelectricity.pdf

[19] Y. Zhang and J. Dong, "Least squares-based optimal reconciliation method for hierarchical forecasts of wind power generation," *IEEE Transactions on Power Systems*, pp. 1–1, 2018.

[20] S. B. Taieb, J. W. Taylor, and R. J. Hyndman, "Coherent Probabilistic Forecasts for Hierarchical Time Series," *Proceedings of the 34th International Conference on Machine Learning*, vol. 70, pp. 3348–3357, 2017.

[21] J. Tastu, P. Pinson, and H. Madsen, "Space-time trajectories of wind power generation: Parametrized precision matrices under a Gaussian copula approach," *Lecture Notes in Statistics*, vol. 217, pp. 267–296, 2015.

[22] P. Pinson, H. Madsen, G. Papaefthymiou, and B. Klöckl, "From Probabilistic Forecasts to Statistical Scenarios of Short-term Wind Power Production," *Wind Energy*, vol. 12, no. 1, pp. 51–62, 2009.

[23] R. J. Bessa, "On the quality of the Gaussian copula for multi-temporal decision-making problems," *19th Power Systems Computation Conference, PSCC 2016*, 2016.

[24] 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, Jan 2018.

[25] C. Gilbert, J. Browell, and D. McMillan, "A hierarchical approach to probabilistic wind power forecasting," in *2018 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*, Boise, USA, 2018.

[26] ——, "Supplementary material: Leveraging turbine-level data for improved probabilistic wind power forecasting," University of Strathclyde, Tech. Rep., 2019, https://doi.org/10.15129/6b949176-4608-430c-b13c-b201afb921c4.

[27] P. Pinson and R. Girard, "Evaluating the quality of scenarios of short-term wind power generation," *Applied Energy*, vol. 96, pp. 12–20, 2012.

[28] 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, 2016.

[29] L. Breiman, J. Friedman, C. J. Stone, and R. A. Olshen, *Classification and Regression Trees*. Chapman & Hall/CRC, 1984.

[30] T. Hastie, R. Tibshirani, and J. H. Friedman, *The elements of statistical learning: data mining, inference, and prediction*, 2nd ed. Springer, 2009.

[31] A. Natekin and A. Knoll, "Gradient boosting machines, a tutorial," *Frontiers in Neurorobotics*, vol. 7, 2013.

[32] 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.

[33] 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.

[34] A. Sklar, "Fonctions de répartition à n dimensions et leurs marges," *Publ. Inst. Statist. Univ. Paris*, 1959.

[35] 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.

[36] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2018. [Online]. Available: https://www.R-project.org/

[37] J. Friedman, T. Hastie, and R. Tibshirani, "Regularization Paths for Generalized Linear Models via Coordinate Descent," *Journal of Statistical Software*, vol. 2010, no. 1, pp. 1–22, 33.

[38] U. Schepsmeier, J. Stoeber, E. C. Brechmann, B. Graeler, T. Nagler, and T. Erhardt, "VineCopula: statistical inference of vine copulas," 2018. [Online]. Available: https://cran.r-project.org/web/packages/VineCopula

[39] 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/

[40] G. Ridgeway and with contributions from others, "gbm: Generalized Boosted Regression Models," 2015. [Online]. Available: https://cran.r-project.org/package=gbm

[41] 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, pp. 243–268, 2007.

[42] P. Pinson, H. A. Nielsen, J. K. Møller, H. Madsen, and G. N. Kariniotakis, "Non-parametric probabilistic forecasts of wind power: required properties and evaluation," *Wind Energy*, vol. 10, no. 6, pp. 497–516, 2007.

[43] B. Efron and R. J. Tibshirani, *An introduction to the bootstrap*. CRC press, 1994.

[44] 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, 2005.

[45] D. Kahle and H. Wickham, "ggmap: Spatial Visualization with ggplot2," *The R Journal*, vol. 5, pp. 144–161, 2013.

**Ciaran Gilbert** received the MSc in Sustainable Engineering with a speciality in offshore renewable energy in 2015. He is currently a PhD student at the Wind & Marine Energy Systems Centre for Doctoral Training at the University of Strathclyde, Glasgow, Scotland. His research interests are in energy, offshore access, and electricity price forecasting.

**Jethro Browell** (S15–M'16) received the M.Phys. degree in Mathematics and Theoretical Physics from the University of St Andrews, UK, in 2011 and a Ph.D. in wind energy systems from The University of Strathclyde, UK, in 2015. He is a Research Fellow at the University of Strathclyde's Institute for Energy and Environment where his research interests span all aspects of energy forecasting and associated decision-making in electricity markets and power system operation. He is an EPSRC Innovation Fellow and a member of the IEEE Power & Energy Society.

**David McMillan** received the PhD in wind turbine engineering in 2008, and holds the position of Senior Lecturer at University of Strathclyde Wind Energy & Control Centre (WECC). He has been principal investigator on grants totalling over 1m, with a focus on cost reduction for wind energy in the operational phase.