

A Hierarchical Approach to Probabilistic Wind Power Forecasting

Ciaran Gilbert, *Student Member, IEEE*, Jethro Browell, *Member, IEEE*, and David McMillan, *Member, IEEE*
University of Strathclyde, Glasgow, UK
{ciaran.gilbert, jethro.browell, d.mcmillan}@strath.ac.uk

Abstract—This paper describes a method to generate improved probabilistic wind farm power forecasts in a hierarchical framework with the incorporation of production data from individual wind turbines. Wind power forms a natural hierarchy as generated electricity is aggregated from the individual turbine, to farm, to the regional level and so on. To forecast the wind farm power generation, a layered approach is proposed whereby deterministic forecasts from the lower layer (turbine level) are used as input features to an upper-level (wind farm) probabilistic model. In a case study at a utility scale wind farm it is shown that improvements in probabilistic forecast skill (CRPS) of 1.24% and 2.39% are obtainable when compared to two very competitive benchmarks based on direct forecasting of the wind farm power using Gradient Boosting Trees and an Analog Ensemble, respectively.

Index Terms—Wind power, probabilistic forecasting, hierarchical forecasting, forecasting, wind power integration

I. INTRODUCTION

In certain electricity networks the paradigm shift from large centralized thermal power stations to decentralised and stochastic renewable energy sources has led energy forecasting to be essential for economic and secure grid operation. Significant uncertainties in the power network are now present on the supply-side which increases the difficulty of balancing the network [1]; to manage the influence of these renewable sources it is necessary to use predictions of future generation and, for risk minimisation, probabilistic forecasts [2].

Depending on the application, end-user requirements, and time horizon there are various ways of approaching the forecasting problem. A trader participating in the day-ahead markets will have very different forecast requirements to a network operator balancing supply and generation in real time. These are example applications of short-term forecasting (hours to days ahead) and very-short term forecasting (minutes to hours ahead) respectively and this paper will focus on the former.

Using the output of a Numerical Weather Prediction (NWP) a data-driven approach is possible for the short-term forecasting of a single wind farm. Typically, this involves mapping concurrent NWP input features to a measured power time series at the wind farm export cable using a statistical learning technique. This avoids any assumptions of the physical phenomena governing the wind to power process and is recommended best-practice because it implicitly accounts for wake-losses, terrain effects, turbine condition, and the systematic bias in weather forecasts [2]. It follows that there are two

sources of error in wind power forecasts: meteorological forecast errors and those from the wind-to-power conversion process, the latter being the focus of this paper.

Due to the stochastic nature of the wind resource, in academia the focus has been on developing probabilistic methods for forecasting, which is reviewed comprehensively in [3]. However, it has been noted that there is a disconnect between the research in academia and application in industry, which is mainly based around deterministic forecasting [2]. Deterministic wind power forecasting involves a *best-guess* estimate of future generation and is reviewed extensively in [4], [5].

An area of challenge for wind power prediction highlighted within the IEA Wind Task 36 [6] is the lack of widely available and publicised benchmark datasets to test various methodologies on. For this reason, forecasting competitions such as the 2012 and 2014 Global Energy Forecasting Competitions (GEFCom) [7], [8] are a valuable pursuit and provide learning for both forecast producers and users. A common theme amongst the two winning approaches to the two GEFCom wind forecasting competitions [9], [10] is the use of Gradient Boosting regression Trees (GBT) with a focus on feature engineering and cross validation to generate robust and powerful models. The five most successful models in the latest competition employed a fully non-parametric approach using techniques such as *k*-Nearest Neighbour (kNN), and Quantile Regression [8].

Large utility scale wind farms contain many turbines distributed over a wide geographical area with each turbine experiencing different conditions. The turbine level data enabling this study is also routinely measured and transmitted to operators making this data available for use in forecasting systems. Here we propose an extension to [11] to evaluate the use of turbine-level power data in a hierarchical probabilistic forecasting framework which has proved successful in other related applications [12], [13].

Aspects of the electricity network provides a suitable framework when studying hierarchical forecasting; wind power behaviour is hierarchical by its nature, rising from the individual turbine, to farm, to regional level etc. At different levels of the hierarchy the behaviour and interactions between the measurements can pose different problems. At the lowest level the time-series is often characterized by intermittency and there is a tendency to observe noisy measurements, whereas at the top level these influences can be smoothed out. This has

been discussed extensively in [13]–[15] where it is emphasised that a simple bottom-up approach to forecasting the top level, by summing the constituent lower level forecasts, tend to deliver poor performance because of the low signal to noise ratio of the bottom hierarchy.

II. FORECASTING METHODOLOGY

In the composition of the hierarchical model, the first layer is composed of out-of-sample deterministic forecasts of each turbine and an aggregated deterministic farm level forecast. These are then used as input features to a top-level probabilistic model which provides the desired final forecast.

In the benchmark and individual turbine forecasting models the explanatory variables \mathbf{x}_t contains features derived from the NWP, namely forecast wind speed and direction at 10m and 100m above ground linearly interpolated to 30-minute resolution. Feature engineering is used to boost the predictive power of all models, this included temporal features (leading and lagging wind speed forecasts, averages, ratios, and gradients) [16] and time-of-day features to capture diurnal bias in the NWP which are modelled by cubic spline kernels [17].

In total three benchmark models are used. The deterministic and a probabilistic benchmark consist of power forecasts at the farm level produced using the GBT method with feature engineering inspired by the winning approach to the wind tracks of the 2012 and 2014 GEFCom competitions [9], [10]. The other probabilistic benchmark model uses the output of the GBT mean benchmark to generate a probabilistic forecast via the Analog Ensemble technique [18]. These robust and very competitive benchmarks are based on the conventional approach to forecasting, using solely the smoother wind farm level power data and are used to determine any improved skill of the hierarchical approach. A summary of the input features used in this study is shown in Table I for the GBT benchmarks and individual turbine models.

The Analog Ensemble (AnEns) consists of wind farm power observations which correspond to the most similar mean forecasts in the training dataset to the current mean forecast of interest. In this case the AnEns was conditioned by lead time and the algorithm used to rank past forecast similarity was k -nearest neighbours (kNN) with a Euclidean distance metric. The advantage of this method is that it is very computationally efficient although a large training period is necessary to produce competitive forecasts. A similar approach was used in the second place entry in the GEFCom2014 competition [19]. For a more detailed description of the AnEns method please refer to [18].

A. Gradient Boosting Trees

The Gradient Boosting Tree (GBT) framework is an ensemble learning algorithm that it is non-linear, non-parametric, and can be used in conjunction with various differentiable loss functions. This means that the same model framework can be used to fit the deterministic mean forecast, using the square loss function, and the probabilistic forecasts, using the quantile loss function.

TABLE I
LIST OF INPUT FEATURES BY MODEL - METEOROLOGICAL INPUTS AT 100M (UNLESS SPECIFIED)

| Model | Features | ID |
|---|---|--------------|
| GBT Bench- marks & Layer 1 Individual Turbines | 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. | |
| Layer 2 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-nl |
| | Average L1 aggregation (0 to 4 periods) | L1-agg-av-pl |
| | Average ws (0 to 4 periods) | ws-av-pl |

Individual trees can be fit very efficiently using the process of recursive partitioning but have limited predictive power and for this reason are often called *weak learners* [20]. Gradient boosting attempts to overcome this drawback by constructing a ‘stronger’ learner from an ensemble of weak learners. The gradient boosting tree $F_n(\mathbf{x}_t)$ is the sum of n regression trees

$$y_t = F_n(\mathbf{x}_t) + \epsilon_t \quad (1)$$

$$= \sum_{i=1}^n f_i(\mathbf{x}_t) + \epsilon_t \quad (2)$$

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

$$\operatorname{argmin}_{f_{n+1}} \sum_t L(y_t, F_n(\mathbf{x}_t) + f_{n+1}(\mathbf{x}_t)) \quad (3)$$

for some loss function $L(\cdot)$. Where $L(\cdot)$ is differentiable, this optimisation can be solved by steepest descent written

$$g_n(\mathbf{x}_t) = \frac{\partial L(y_t, F_n(\mathbf{x}_t))}{\partial F_n} \quad (4)$$

$$f_{n+1}(\mathbf{x}_t) = -\rho_n g_n(\mathbf{x}_t) \quad (5)$$

where

$$\rho_n = \operatorname{argmin}_{\rho} \sum_t L(y_t, F_n(\mathbf{x}_t) - \rho g_n(\mathbf{x}_t)) \quad (6)$$

The user must specify the number of trees to fit, n , and the number of regions each tree divides the input space into. An additional shrinkage parameter may be included in (5) to

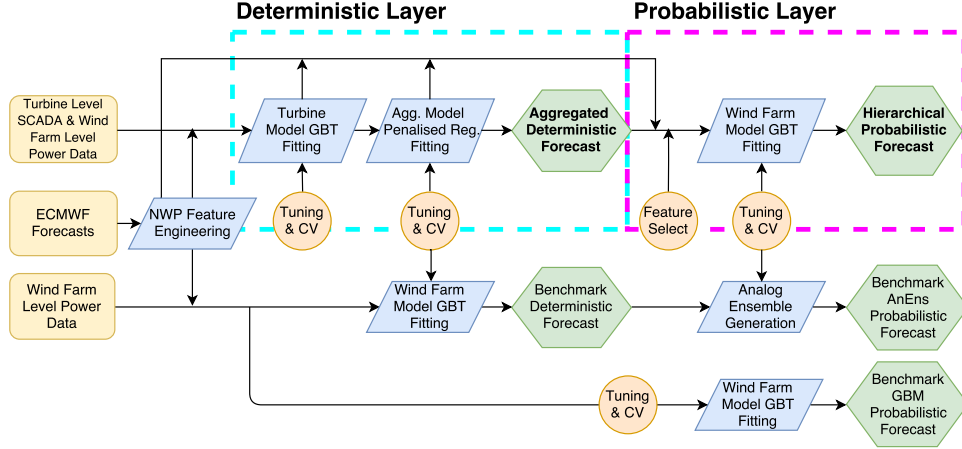


Fig. 1. Block-flow diagram illustration of the proposed layered methodology

control the learning rate of the fitting procedure and reduce the impact of individual trees. It is important to note that it is necessary to do extensive tuning of the hyper-parameters used for the boosting and tree-fitting processes [21].

B. Hierarchical Approach

The hierarchical methodology is effectively split into a two-layer approach consisting of a deterministic turbine layer and a probabilistic farm-level layer. Using a similar approach has proven successful in the context of incorporating off-site information to improve the accuracy of an individual wind farm [10]. In this case, it is proposed to improve accuracy at the wind farm level by using internal information from a lower hierarchy via deterministic mean forecasts for each wind turbine. A summary of this method is shown in Figure 1.

1) *Deterministic Layer*: The individual out-of-sample turbine mean forecasts $y_{i,t}$ are combined to produce a mean forecast for the wind farm z_t using a weighted sum

$$z_t = \sum_{i=1}^D \omega_i y_{i,t} + \epsilon_t \quad (7)$$

which is essentially a forecast calibration step whereby information from the wind farm smoother time-series is used to fit a weighted aggregation of the turbine forecasts. The two-way passage of information from the lower and higher levels of the hierarchy differentiate this method from a conventional bottom-up approach, the importance of which is emphasised in other hierarchical forecasting applications [15]. Weighting parameters are estimated by elastic net penalized regression [22]

$$\omega = \underset{\beta}{\operatorname{argmin}} \left\{ \frac{1}{2N} \|\mathbf{Z} - \mathbf{Y}\beta\|_2^2 + \lambda \left[(1 - \alpha) \frac{1}{2} \|\beta\|_2^2 + \alpha \|\beta\|_1 \right] \right\} \quad (8)$$

where λ is a tuning parameter, α is a compromise between ridge ($\alpha = 0$) and lasso ($\alpha = 1$) regression, \mathbf{Z} and \mathbf{Y} are matrices of vertically stacked instances of z_t and y_t .

Conditioning the weighted sum on forecast wind direction was also tested by partitioning the data into a number of discrete directional bins, as described in [11], however in this study was shown to provide no value in improving the forecast via cross-validation. The optimal value of α and λ are determined through k -fold cross validation.

The tuning and regularisation strategy of the GBT models utilised k -fold cross validation, with 5 folds, via a grid-search of the interaction depth and shrinkage parameters. The parameters held constant were the number of trees (500), the minimum number of observations at each leaf node (30), and the bag fraction (75%) which is the fraction of the training set randomly chosen to fit the next tree in the ensemble. In the deterministic models the k -fold cross validation was used to minimise the Root Mean Squared Error (RMSE).

2) *Probabilistic Layer*: The desired final predictive distribution, determined using quantile regression (including the 5th, 10th, ..., 95th quantiles), is calculated using the NWP input features and the available deterministic forecasts generated by the lower hierarchy layer. This indicates that there are several possible configurations available to optimise. Feature selection is employed to determine the most valuable input features from a candidate pool that includes: 1) NWP features \mathbf{x}_t , 2) the mean forecast of the lower hierarchy z_t , 3) individual turbine mean forecasts $y_{i,t}$, and 4) temporal and spatial features derived from the lower hierarchy power forecasts. For this particular case study (see Section III) the final engineered feature list for the layered probabilistic model are summarised in Table I.

For the tuning of the quantile GBT models certain key quantiles (5th, 30th, 50th, 70th, 95th) in the distribution were tuned individually to minimise the pinball loss function [23]. For target quantile q the pinball loss function at time t is

$$\text{PB}_t^{(q)} = \begin{cases} (1 - q)(\hat{z}_t^{(q)} - z_t) & , \text{ if } z_t < \hat{z}_t^{(q)} \\ q(z_t - \hat{z}_t^{(q)}) & , \text{ if } z_t \geq \hat{z}_t^{(q)} \end{cases} \quad (9)$$

and spline interpolation was then used to determine the intervening quantile hyper-parameters. The same rigour of tuning

via cross validation is also applied to the benchmark models. For the AnEns benchmark, k fold cross validation is used to determine the optimum k neighbours which constitute the ensemble and it was computationally feasible to efficiently minimise the Continuous Rank Probability Score (CRPS).

III. CASE STUDY

The methodologies are tested at Gordonbush wind farm operated by SSE in the UK. This is a large utility scale park with 35 2MW turbines spread across an area of approximately 15km². 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. Training and testing data are partitioned with a 15 and 6 month length respectively. Numerical Weather Forecasts from the European Centre for Medium-Range Weather Forecasts (ECMWF) comprise of wind speed and direction at 10 and 100m above ground at the closest grid point to the wind farm, with 2 issue times per day. The methodologies described are implemented in R using the packages `glmnet`, `kknn` and `gbm` [24]–[27].

IV. RESULTS

To evaluate the mean forecast, the RMSE performance metric is used [28]. Probabilistic forecasts are evaluated in terms of their reliability (also called calibration) and sharpness. Reliability is a necessary condition for probabilistic forecasts to be useful in decision making, and reliable forecasts are differentiated between using the CRPS score. The calibration is calculated from an indicator variable $\eta_t^{(q)}$ for a given nominal quantile forecast

$$\eta_t^{(q)} = \begin{cases} 1 & , \text{ if } z_t < \hat{z}_t^{(q)} \\ 0 & , \text{ otherwise} \end{cases} \quad (10)$$

from which the mean of the series of outcomes is calculated over the entire testing dataset to give the actual coverage $\hat{a}^{(q)}$ of the quantile

$$\hat{a}^{(q)} = \frac{1}{N} \sum_{t=1}^N \eta_t^{(q)} \quad (11)$$

the coverage is then determined over the distribution quantiles and is illustrated via a reliability diagram [1]. An extension to this measure includes deviations from nominal proportions to highlight the sometimes subtle differences between model calibration. This is known as the bias $b^{(q)}$ of the quantile forecast [29]

$$b^{(q)} = q - \hat{a}^{(q)} \quad (12)$$

The sharpness and calibration can be both quantified via the CRPS [30]

$$\text{CRPS} = \frac{1}{N} \sum_{t=1}^N \int_{-\infty}^{\infty} \{F_t(z) - \mathbf{1}(z \geq z_t)\}^2 dz \quad (13)$$

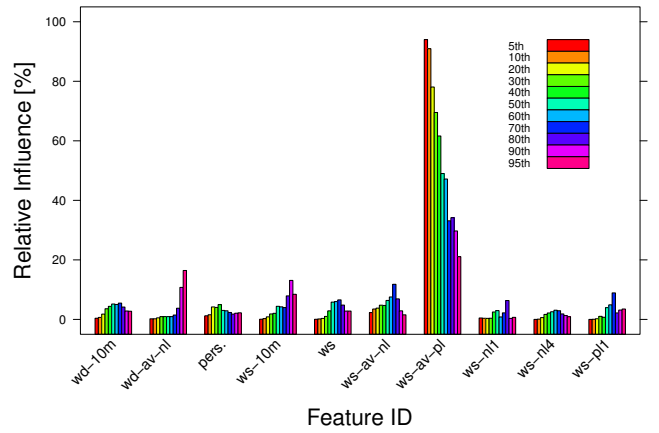


Fig. 2. Relative influence of input features: GBT Benchmark

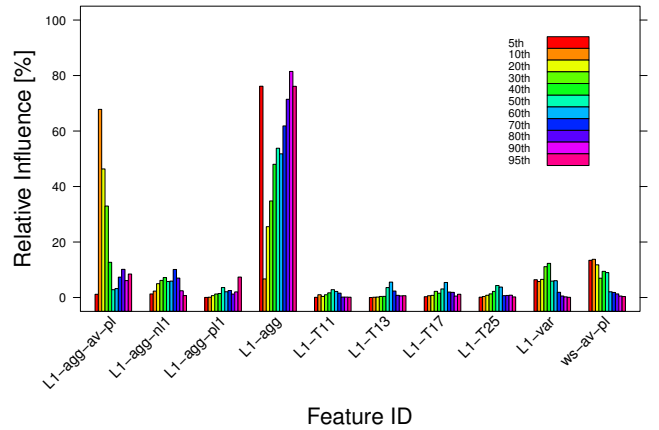


Fig. 3. Relative influence of input features: Hierarchical Model

where $\mathbf{1}(\cdot)$ is the indicator function. This score essentially compares the predictive forecast distribution F_t against the empirical distribution of the observation z_t .

For ensemble decision trees it is important to evaluate variable importance via the measure "relative influence" which allows the user to look under the hood; this is a quantitative measure based on the number of times a variable has been selected for splitting, with a weighting applied derived from the predictive improvement to the model from each split averaged over all the trees [31]. Figures 2 and 3 illustrate the importance of selected variables across key quantiles in the probabilistic GBT benchmark and probabilistic hierarchical model.

For the benchmark model the engineered feature giving the average of the leads of wind speed shows the highest influence in the model, especially at the lower tail of the distribution. At the upper tail of the distribution a greater mix of features contribute to the predictive power of the model, including the circular average of wind direction from the negative lags. This indicates that at the higher levels of power production, wind direction and other features have more influence on the power forecast, compared to the lower power generation levels which are simply dominated by average wind speed. The influence of

the engineered temporal features across all quantiles indicates the importance of what is essentially smoothing of the deterministic meteorological forecast by using average features and leading/lagging inputs.

For the proposed model, the out-of-sample aggregated mean forecast from the turbine level (layer 1) has most of the influence across the quantiles, along with the average leads of this feature. The only meteorological feature selected at this higher layer of the model is the average wind speed of the positive leads, which shows the importance of this explanatory variable at both levels of the hierarchy.

The case study results are given in Table II showing the forecast performance of the proposed method versus time horizon of the lower hierarchy (deterministic) and the upper hierarchy (probabilistic) models. These are compared with the benchmark models to give a percentage change in forecast skill. The deterministic results show that averaged across all the horizons it is possible to gain a 1.61% reduction in RMSE using a hierarchical approach.

The probabilistic hierarchical approach led to an average improvement in CRPS of 1.24% over the GBT benchmark, which is in-line with the deterministic improvement. It should be emphasised that this benchmark method is state-of-the-art and based on the winning method from the GEFCom2014, competition with additional focus on feature engineering. In the context of simply better utilisation of existing data, with no necessary large investments in new equipment, this modest improvement could have an accumulated large impact across a portfolio with no necessary cost overheads. In terms of forecast reliability, as shown in Figure 4, overall the proposed model is well calibrated and gives improvements over the GBT benchmark (Figure 4 (a)), at the upper end of the distribution with some slightly diminished performance at the lower end of the distribution.

The AnEns CRPS is marginally greater than the GBT models which, given the computational advantages and relative simplicity of this approach, is notable. However the reliability reveals that the calibration of this forecast is poor, especially in the lower half of the distribution. The hierarchical method offers a 2.39% improvement in CRPS when compared to the AnEns benchmark. Overall, using a hierarchical probabilistic method resulted in improvement in forecast CRPS of 1.24% and 2.39% averaged over the forecast horizon, compared to very competitive benchmark models based on direct forecasting wind farm power via GBT regression and the AnEns technique respectively. Figure 5 gives an example of the resulting density forecast from the hierarchical approach.

Future work should consider the utility of this approach in combination with downscaled NWP and if similar improvements are possible on very-short time scales where statistical methods typically outperform those based on NWP. It would be prudent to test the approach on wind farms with different characteristics, such as offshore and very-large onshore farms to test if the approach generalises. Additionally, comparisons could be made to bottom-up hierarchical forecasting approaches such as estimating the joint distribution of the

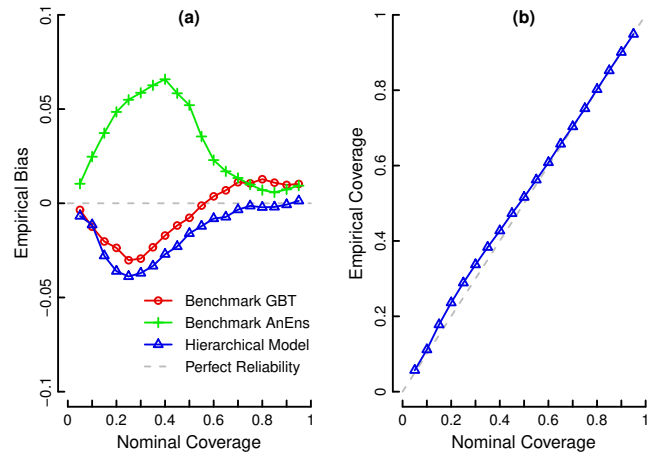


Fig. 4. (a) Quantile Bias Plots (b) Hierarchical Model Reliability Diagram

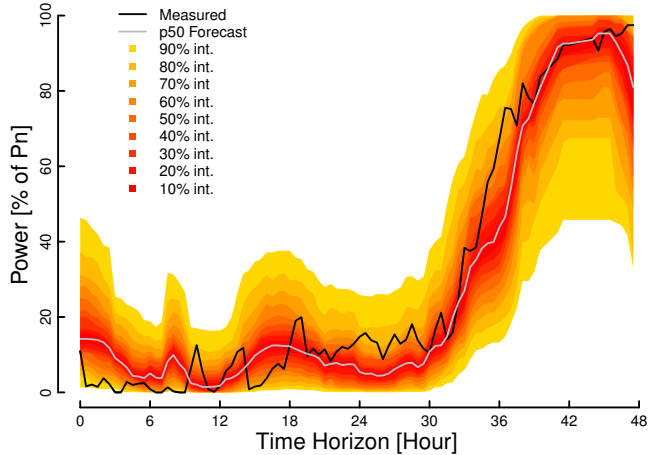


Fig. 5. Example forecast from final layered model - 01/05/2015

lower level turbine forecasts via a copula and the marginal distributions of the individual turbines.

V. CONCLUSIONS

This paper investigates a novel method for probabilistic hierarchical forecasting of wind farm power, incorporating turbine-level data using a layered approach where generated deterministic forecasts from the lower hierarchy (turbine level) are used as input features to a higher-level wind farm probabilistic model. On a case study at a utility scale wind farm in the UK it is shown that average improvements in CRPS of 1.24% and 2.39% are obtainable when compared to two very competitive benchmark models based on GBT regression and the AnEns technique respectively.

ACKNOWLEDGEMENTS

The authors thanks SSE plc for the supply of wind farm data, ECMWF for the supply of meteorological data, Sam Hawkins at Vattenfall UK for discussions on forecasting the methodology, and the reviewers for their advice. Ciaran Gilbert is supported by the University of Strathclyde's EPSRC Centre for Doctoral Training in Wind and Marine Energy Systems,

TABLE II
CASE STUDY RESULTS BY HORIZON RANGE [% OF PN]

| Horizon (Hr) | Deterministic - RMSE | | | Probabilistic - CRPS | | | | |
|--------------|----------------------|--------------|-------|----------------------|--------------|--------------|-----------|-------------|
| | Bench. | Hierarchical | %Imp. | Bench. GBT | Bench. AnEns | Hierarchical | %Imp. GBT | %Imp. AnEns |
| 0-6 | 14.32 | 13.94 | 2.68 | 7.14 | 7.23 | 7.01 | 1.77 | 3.06 |
| 7-12 | 14.48 | 14.41 | 0.49 | 7.13 | 7.17 | 7.05 | 1.13 | 1.74 |
| 13-18 | 15.99 | 15.55 | 2.77 | 7.91 | 8.06 | 7.78 | 1.71 | 3.47 |
| 19-24 | 15.18 | 15.02 | 1.08 | 7.48 | 7.55 | 7.38 | 1.26 | 2.23 |
| 25-30 | 17.48 | 17.04 | 2.53 | 8.67 | 8.80 | 8.56 | 1.32 | 2.82 |
| 31-36 | 16.79 | 16.58 | 1.26 | 8.30 | 8.37 | 8.19 | 1.35 | 2.14 |
| 37-42 | 18.96 | 18.65 | 1.66 | 9.38 | 9.50 | 9.28 | 1.06 | 2.32 |
| 43-48 | 18.16 | 18.07 | 0.49 | 8.90 | 8.99 | 8.86 | 0.42 | 1.41 |
| All | 16.48 | 16.21 | 1.61 | 8.10 | 8.20 | 8.00 | 1.24 | 2.39 |

grant number EP/L016680/1. **Data Statement:** Due to confidentiality agreements with research collaborators, access to wind power data is restricted. Numerical weather predictions are from ECMWF and are available at www.ecmwf.int for qualifying research institutions or commercial customers.

REFERENCES

- [1] J. M. Morales, A. J. Conejo, H. Madsen, P. Pinson, and M. Zugno, *Integrating Renewables in Electricity Markets*, 2014, vol. 205. [Online]. Available: <http://link.springer.com/10.1007/978-1-4614-9411-9>
- [2] R. Bessa, C. Möhrlein, 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. [Online]. Available: <http://www.mdpi.com/1996-1073/10/9/1402>
- [3] 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. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1364032114000446>
- [4] C. Monteiro, R. Bessa, V. Miranda, A. Botterud, J. Wang, and G. Conzelmann, "Wind Power Forecasting: State-of-the-Art 2009," Argonne National Laboratory ANL/DIS-10-1, Tech. Rep., 2009.
- [5] G. Giebel, R. Brownsword, G. Kariniotakis, M. Denhard, and C. Draxl, *The State-of-the-Art in Short-Term Prediction of Wind Power*. ANEMOS.plus, 2011.
- [6] G. Giebel, J. Cline, H. Frank, W. Shaw, P. Pinson, B.-M. Hodge, G. Kariniotakis, J. Madsen, and C. Möhrlein, "Wind Power Forecasting: IEA Wind Task 36 and Future Research Issues," *Journal of Physics Conference Series*, vol. 753, no. January 2017, 2016. [Online]. Available: <http://www.nrel.gov/docs/fy17osti/67542.pdf>
- [7] T. Hong, P. Pinson, and S. Fan, "Global Energy Forecasting Competition 2012," *International Journal of Forecasting*, vol. 30, pp. 357–363, 2014.
- [8] 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*, 2016.
- [9] L. Silva, "A feature engineering approach to wind power forecasting: {GEFCOM} 2012," *International Journal of Forecasting*, vol. 30, no. 2, pp. 395–401, 2014.
- [10] 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.
- [11] J. Browell, C. Gilbert, and D. McMillan, "Use of Turbine Level Data for Improved Wind Power Forecasting," in *IEEE PowerTech Conference*, Manchester, UK, 2017.
- [12] 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.
- [13] 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.
- [14] 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. [Online]. Available: <http://proceedings.mlr.press/v70/taieb17a.html>
- [15] —, "Hierarchical Probabilistic Forecasting of Electricity Demand with Smart Meter Data," pp. 1–30, 2017.
- [16] 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.
- [17] 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.
- [18] 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/abs/10.1175/MWR-D-12-00281.1>
- [19] 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>
- [20] L. Breiman, J. Friedman, C. J. Stone, and R. A. Olshen, *Classification and Regression Trees*. Chapman & Hall/CRC, 1984.
- [21] A. Natekin and A. Knoll, "Gradient boosting machines, a tutorial," *Frontiers in Neuroinformatics*, vol. 7, no. DEC, 2013.
- [22] 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>
- [23] J. Morales, A. Conejo, H. Madsen, P. Pinson, and M. Zugno, *International Series in Operations Research & Management Science*, 2011.
- [24] R Core Team, "R: A Language and Environment for Statistical Computing," Vienna, Austria, 2016. [Online]. Available: <https://www.r-project.org/>
- [25] 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.
- [26] K. Hechenbichler and K. Schliep, "Weighted k-Nearest-Neighbor Techniques and Ordinal Classification," Collaborative Research Center 386, Tech. Rep., 2004. [Online]. Available: <https://epub.uni-muenchen.de/1769/>
- [27] G. Ridgeway and with contributions from others, "gbm: Generalized Boosted Regression Models," 2015. [Online]. Available: <https://cran.r-project.org/package=gbm>
- [28] 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.
- [29] P. Pinson, G. Kariniotakis, H. A. Nielsen, T. S. Nielsen, and H. Madsen, "Properties of quantile and interval forecasts of wind generation and their evaluation," *Proceedings of the European Wind Energy Conference & Exhibition*, no. October 2015, pp. 2–6, 2006.
- [30] 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.
- [31] J. Elith, J. R. Leathwick, and T. Hastie, "A working guide to boosted regression trees," *Journal of Animal Ecology*, vol. 77, pp. 802–813, 2008. [Online]. Available: <http://avesbiodiv.mncn.csic.es/estadistica/bt1.pdf>