This paper has been accepted for publication in the proceedings of the 14th International Conference on the European Electricity Market Conference 2017, Dresden, Germany. It should be cited as follows. A revised citation, including doi, will be placed here when available.

J. Browell and C. Gilbert, “Cluster-based regime-switching AR for the EEM 2017 Wind Power Forecasting Competition,” 14th International Conference on the European Energy Market (EEM), Dresden, 2017.

Cluster-based regime-switching AR for the EEM 2017 Wind Power Forecasting Competition

Jethro Browell and Ciaran Gilbert
University of Strathclyde, Glasgow, UK
{jethro.browell, ciaran.gilbert}@strath.ac.uk

***Abstract*—This paper describes the regime-switching auto-regressive models used to win the EEM 2017 Wind Power Forecasting Competition. The competition required participants to produce daily forecast wind power production for a portfolio of wind farms from 2 to 38 hours-ahead based on historic generation and numerical weather prediction analysis data only. The regimes used in the methodology presented are defined on the previous day's weather conditions using the k-medians clustering algorithm. Cross-validation is used to identify models with the best predictive power from a pool of candidate models. The final methodology produced a final weighted mean absolute error 4.5% lower than the second place team during the two-week competition period.**

***Index Terms*—Wind Power, Forecasting, Time Series, Clustering, Autoregression, Regime Switching**

I. Introduction

Wind power forecasting is an essential process in modern power system operation in networks with significant penetration of wind generation, and is central to successful electricity market participation in these regions. Short-term forecasts from hours- to days-ahead are used in generation and maintenance scheduling, and for trading in electricity markets forecasts; and very-short-term forecasts from minutes- to hours-ahead are used for participation in intra-day markets and by power system operators to balance supply and demand in real time [1], [2]. Due to the stochastic nature of the wind resource forecasts will always be required to inform decisions where future wind generation is a factor. Furthermore, the importance and value of high-quality forecasting will increase with the penetration of wind power. The growing demand for energy forecasts, and for improvement in forecast quality, has motivated a great deal of research and development, and also competitions to compare methodologies in a controlled environment, see [3], [4], for example.

This paper details team p9's winning approach to the problem set in the EEM 2017 Wind Power Forecasting Competition. The competition required participants to forecast the aggregated wind power generation from a portfolio of wind farms from 2 to 38 hours-ahead at 15 minute resolution on a daily basis for two weeks. Participants were provided with one year of historic power production and numerical weather prediction data (analysis only, no forecasts) to train their forecasting models, plus daily updates during the competition period. Team p9's solution was based on regime-switching auto-regressive models with regimes defined on the most recent day's weather. This approach won the competition with a 4.5% lower error score (the competition was scored on re-weighted mean absolute error) than the second place team.

Deterministic wind power forecasts, the focus for this competition, comprised of single-valued best estimates of future energy for a particular time-horizon are approaching technological maturity. A comprehensive review of the concerted research effort of the wider academic community can be found in [5], [6]. However, due to the stochastic nature of the wind there is a broad consensus in the academic community that forecasts should be probabilistic in order to quantify forecast uncertainty [7], [8]. Despite this, many forecast users still only utilise deterministic forecasts due to their interpretability and difficulties associated with incorporating complex probabilistic information into decision-making processes. Therefore, improving deterministic power forecasts is an important pursuit, and developing new methodologies is the focus of current research [9], [10]. Furthermore, improvements in deterministic forecasting will translate to improvements in probabilistic forecasting in many cases.

The methodology used in this paper has particular relevance to very-short-term wind power forecasting where it is typically assumed that statistical models based on time series analysis are superior to the those which rely on physical model outputs, i.e. Numerical Weather Prediction [6], [11]. The superiority of purely statistical models within this time horizon is due to a number of factors including: the most recent input measurements to a NWP may be several hours old by the time the forecast is issued, and errors introduced by the spatial interpolation process required to make predictions at a specific point of interest from gridded NWP output. The competition set-up did not include numerical weather prediction forecasts so only time series methods could be considered for all forecast horizons; however, we show that there is value in conditioning time series model on features derived NWP data.

A wide variety of well established and time series methods have been adapted for power forecasting including autoregressive [12] and autoregressive moving average [13] models, in addition to contemporary methods such as neural networks [14] and Markov chains [15]. Hybrid methods that combine several time series models have also been studied and shown to outperform individual methods in some cases [16]. Spatial models that consider multiple locations simultaneously have been developed and shown to improve forecast skill at all measurement locations [17], [18], and it has been shown that the spatial dependency structure itself is dynamic and exhibits seasonality and dependence wind direction, for example [19], [20].

Time series models may be conditioned on observed or unobserved regimes. Wind speed forecasting techniques based on switching between different models depending on wind direction is proposed in [19] with regimes selected via a cross-validation procedure. Hidden-Markov regime-switching methods have been developed to forecast offshore wind power with the number of regimes chosen to be three to reflect the three distinct regions of the wind farm power curve [21], [22]. More recently, cyclone detection has been used to predict periods of potentially large forecast error in day-ahead wind power forecasting [23] and atmospheric classification has been used to improved very-short-term spatio-temporal wind forecasting [24]. The large-scale meteorological situation has a clear bearing on forecast performance but it is often overlooked by studies, which restrict themselves to wind and power time series only. In this work, the mean wind vector for the 24 hours preceding the forecast issue time are used to define regimes on which simple time series models are conditioned.

While many approaches to wind forecasting have been proposed, it is often difficult to compare their performance since results will differ across datasets and implementing multiple sophisticated methods for comparison on the same dataset is challenging. For this reason, forecasting competitions are very valuable pursuit and provide valuable learning for both forecast producers and users.

This paper is organised as follows. Section II details the methodology used in the competition including the data-exploration used to inform model selection and Section III details the cross-validation and competition results. Section IV indications some proposed model improvements and finally, Section V presents the conclusions.

II. Methodology

The following section details the approach taken by team p9 from exploration of the competition training dataset to model fitting and evaluation.

A. Competition Framework

The competition was based on forecasting the 15 minute resolution aggregated wind power generation from a portfolio of wind farms from 2 to 38 hours-ahead on a daily basis. For each day's forecast the previous 24 hours of generation data and NWP analysis at 3 hour resolution data for 10 geographic locations is provided. A training dataset comprising 1 year of the same data is also provided for model fitting. The meteorological parameters provided are zonal and meridional wind speed at 2m, 80m and 100m above ground, temperature at 2m, and global surface radiation. Only wind speeds at 100m at one of the 10 locations and the most recent 1.5 hours of power data were inputs to p9's final model.

B. Data Exploration

The first stage in the process involves visually exploring the given data to identify the basic characteristics of the dataset.

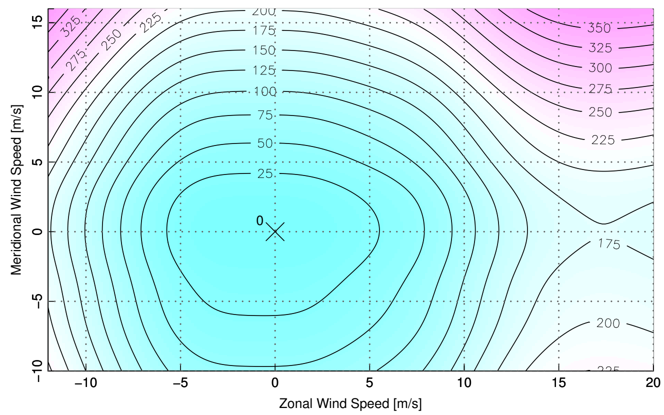

Fig. 1. Contour plot of the empirical power curve for the wind park portfolio. Power contours (labelled, MW) are plotted for zonal and meridional wind speed 100m above ground from 'location 4' in the competition training data. The dataset only spans the illustrated domain. There is a strong directional dependence on power performance and maximum wind speed.

Inspection of the aggregate power versus wind speed and direction from the 10 NWP locations revealed that few of the NWP grid points provided were related to the location of the wind portfolio of interest. The portfolio power curve (with direction) at location 4 is illustrated in Figure 1. The familiar wind power curve is clearly visible and shows a significant directional dependence. It was also observed that the capacity of the portfolio appeared to increase over the course of the 1 year of training data; however, with the limited information available this was difficult to quantify and was not incorporated into our final model.

C. Benchmark Models

Initially, we implement a selection of standard benchmark models against which we can evaluate more complex approaches. These were: the 'mean forecast' where the forecast for every horizon is the mean of all historic data; persistence, where the forecasts for every horizon is equal to the most recent measurement; and an autoregressive (AR) model of order p, where the forecast is a weighted sum of p previous measurements (or forecasts in when making multiple set-ahead predictions). We also considered a generalised additive model based on a single lagged measurement and diurnal and annual seasonality; an AR(p) model with parameter estimation via the least absolute shrinkage and selection operator (LASSO); and a tree-based gradient boosting machine (GBM). Following evaluation of these models, further modifications are explored. All models are evaluated via k-fold cross-validation which allows for efficient out-of-sample testing over the relatively small training dataset to gives a representative measure predictive performance.

1) Autoregressive Model: Modelling the wind power time-series as an AR(p) process assumes that the wind power at time t is a weighted sum of p past measurements plus some error ϵ_t,

$$y_t = \phi_0 + \sum_{i=1}^{p} \phi_i y_{t-i} + \epsilon_t \quad , \tag{1}$$

where ϕ_0 is the model intercept and $\phi_i,\ i > 0$ are the autoregressive coefficients associated with the i^{th} lag. The model parameters are estimated by ordinary least squares (OLS), which produced the optimal linear unbiased predictor in the case where ϵ_t has constant variance and is serially uncorrelated. Although this condition is likely violated, the simplicity and robust performance of this model necessitate its inclusion. The order of the model p is chosen by examining the autocorrelation and partial correlation function of the wind power time series [25].

As the winds are effected by the daily heating and cooling of the Earth it is desirable to introduce the time-of-day as an exogenous dummy variable to capture diurnal profiles. An alternative approach would be to model the diurnal trend via some periodic function such as a Fourier series and/or de-trend the time series. The new model is denoted ARX(p) and written

$$y_t = \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{j=0}^{q-1} \eta_j D_{j,t} + \epsilon_t \quad , \tag{2}$$

where

$$D_{j,t} = \begin{cases} 1 & , \quad \lfloor \frac{t\%q}{r} \rfloor = j \\ 0 & , \quad \text{otherwise} \end{cases} \quad , \tag{3}$$

$t\ \%\ q$ denotes the remainder of t divided by q, and $\lfloor x \rfloor$ denotes the *floor* operator which returns the value of x rounded down to the nearest integer. In this work data are 15 minute resolution therefore $q = 96$. To obtain 96 dummy variables, one for each 15 minute period of the day $r = 1$, and to obtain hour-of-the-day dummy variables $r = \frac{96}{24} = 4$. The final value of r is chosen based on cross-validation. The parameters ϕ_i and η_j are estimated by OLS as for the AR(p) model. Note the intercept ϕ_0 is superseded by $\eta_j D_{j,t}$ which may be interpreted as a time-dependent intercept.

2) Generalized Additive Model: Generalised additive models may be used to model smooth non-linear responses explanatory variables, in contrast to the linear responses of the ARX models described above. This can be achieved by recasting the linear model as an additive model of smooth functions

$$y_t = \beta_1 f_1(y_{t-1}) + \beta_2 f_2(D_t) + \beta_3 f_3(A_t) + \epsilon_t \quad , \tag{4}$$

where $D_t = t\%q$, $A_t = t\%(q\times 365)$, and $f_i(\cdot),\ i = 1, 2, 3$ are smooth functions to be estimated. Here, we choose $f_1(\cdot)$ to be a cubic spline, and $f_2(\cdot)$ and $f_3(\cdot)$ to be cyclic cubic splines to capture smooth non-linear dependence on the first lagged measurement, time-of-day and day-of-year, respectively. The parameters $\beta_i,\ i = 1, 2, 3$ and those of the cubic splines are estimated by penalized least squares to control the 'wigglyness' of the cubic splines as described in [26].

3) LASSO: The least absolute shrinkage and selection operator [27] simultaneously performs linear regression and feature selection estimation by shrinking the absolute size of coefficients $\boldsymbol{\beta}$ by adding the ℓ_1 norm of $\boldsymbol{\beta}$ to the model cost function. For a set of T samples, $\{\mathbf{Y}, \mathbf{X}\}$, where $\mathbf{Y}$ and $\mathbf{X}$ are matrices of vertically stacked instances of y_t and $\mathbf{x}_t = [y_{t-1}, ...y_{t-p}]$, respectively, the LASSO cost function is given by

$$||\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}||_2^2 + \lambda||\boldsymbol{\beta}||_1 \quad . \tag{5}$$

The user-defined shrinkage parameter λ controls sparsity and is typically selected via a cross-validation procedure. Here, p is set by the largest lag determined to have statistical significance in the partial autocorrelation function of y_t with a significance level of 1%. The values of $\boldsymbol{\beta}$ and λ are estimated using the `R` package `glmnet` which minimises (5) by cyclical coordinate descent [28].

4) Gradient Boosting Machines: Gradient boosting constructs a powerful predictive model from an ensemble of *weak learners* where, in this case, each learner is a regression tree. The ensemble of regression trees is constructed sequentially by estimating a new tree according to some user-specified differentiable loss function. Importantly, the optimisation is solved by steepest descent [29]. 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 to control the learning rate of the fitting procedure and reduce the impact of individual trees in the ensemble. In this implementation, lagged measurements and time of day and year variables are used as inputs. For more information on this algorithm please refer to [29].

D. Cluster Based Regime-Switching

Motivated by knowledge that synoptic-scale meteorological conditions persists for several days (longer than the competition's 38-hour forecast horizon) and our observation that the production characteristics of the wind park portfolio vary significantly in different regions it's directional power curve, we develop a regime-switching approach in an attempt to model, and forecast, these distinct behaviours separately.

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 [19], [22], [24]. These models either utilise exogenous variable, such as wind direction [19] or atmospheric mode [24], or model some unobserved hidden-Markov process [12].

Here, we define a number of discrete regimes based on the clustering weather data available in the EEM competition paradigm. We define the mean wind vector $\theta_t = (\tilde{u}_t, \tilde{v}_t)$ where $\tilde{u}_t$ and $\tilde{v}_t$ are the mean zonal and meridional wind speeds over the 24 hours immediately preceding time t. The k-median algorithm [30] is then used to define k regimes. This algorithm generates disjoint regions R_k that collectively cover the input space spanned by θ_t.

A level plot, showing the history of the weather regime s_t during the training dataset is plotted in Figure 2 where $k = 5$

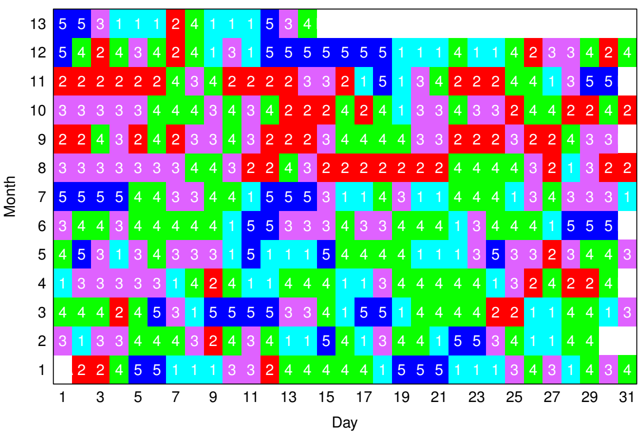

Fig. 2. Level plot of weather regime s_t for $k = 5$ clusters, overlaid with regime number. Month 13 consists of the competition period data.

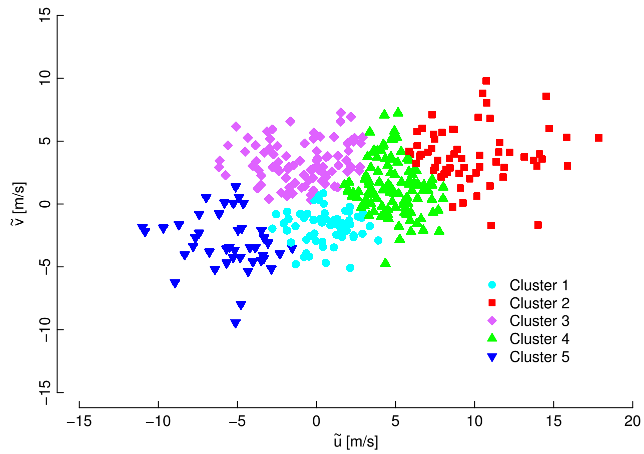

Fig. 3. Visualisation of k-median clusters for $k = 5$. The clusters pattern reflects the distinct regions of the directional power curve illustrated in Figure 1.

clusters. We observe that in general the regime persists for several days and therefore that it is reasonable to forecast up to 38-hours ahead assuming no change in regime. A possible extension would be to forecast the future regime, but that is beyond the scope of this work.

Separate AR(p) and ARX models are fit for each regime and forecasts are produced using the model corresponding to the regime at the forecast issue time. The number of clusters is selected via cross-validation.

The regime-switching ARX model is written

$$y_t = \sum_{i=1}^{p} \phi_i^{(s_t)} y_{t-i} + \sum_{j=0}^{q-1} \eta_j^{(s_t)} D_{j,t} + \epsilon_t \tag{6}$$

where

$$s_t = \begin{cases} 1 \text{ for } \theta_t \in R_1 \\ 2 \text{ for } \theta_t \in R_2 \\ \vdots \\ k \text{ for } \theta_t \in R_k \end{cases} . \tag{7}$$

Setting all $\eta_j = 0$ reduces (6) to the regime-switching AR model. The model above may be interpreted as a conditional AR/ARX model where the regression parameters and conditioned on the discrete switching variable s_t.

E. Forecast Blending

While it is expected that the performance of all models will deteriorate with forecast horizon, the best performing approach for a given horizon may not be the best for all others. Therefore, the final forecast we issue is taken from the model that has performed best in cross-validation for each specific horizon. The transition between models is blended using a the logistic function over one hour of the horizon purely for visual satisfaction with negligible impact on overall performance.

III. Results

For simplicity, 11-fold cross-validation is used with each month from February to December held out in turn. January is excluded for simplicity as lagged values are unavailable for the first day and also to reduce the influence of the apparently depreciated capacity observed for the first months of the year.

The competition ranking is based on mean absolute error (MAE), so that is the measure by which we evaluate candidate forecast models. The MAE is given by

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^{N} |y_i - \hat{y}_i| \tag{8}$$

where $\hat{\cdot}$ denotes a forecast and N is the total number of forecasts being evaluated. The competition ranking is based on the average of the MAE for the issue day (from 10:00 to 23:45, or 2 to 15.75 hours-ahead) and the MAE for the day-ahead (00:00 to 23:45, or 16 to 37.75 hours ahead). This has the effect of placing a greater weight on the earlier horizons, though our blending approach optimises for all horizons independently so no special action is required to optimise performance specifically for the competition.

The cross-validation results for the benchmark models for forecast horizons from 15 minutes to 38 hours ahead are plotted in Figure 4. Little separates the different forecasts for the shortest horizons, but the GAM and AR models are clearly superior to the others for horizons greater than 8 hours.

The regime-switching AR/ARX approaches were implemented for 2 to 6 cluster numbers and 5 was found to be optimal with hour-of-day dummy variables $r = 4$ for the ARX model. Performance of these models is illustrated in Figure 5. A regime-switching GAM was also implemented but performed worse than the original GAM, perhaps due to the higher burden of training data required for reliable parameter estimation.

Importantly, this result reveals that at horizons up to 12 hours ahead the regime-switching AR model out-performs the

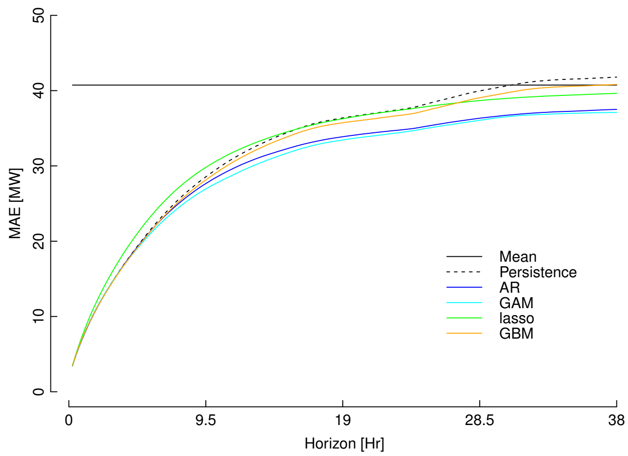

Fig. 4. Results of 11-fold cross-validation for benchmark models.

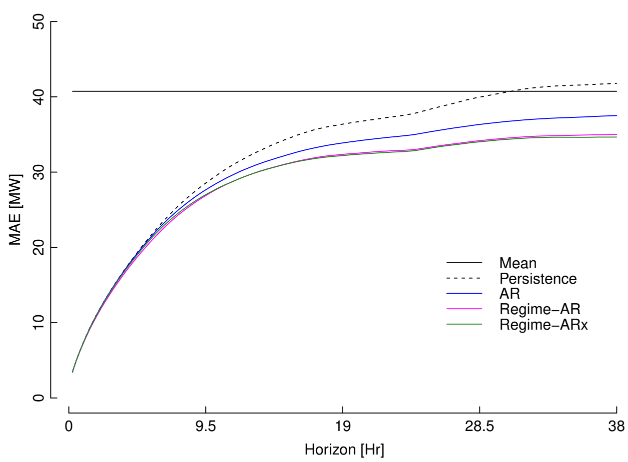

Fig. 5. Cross validation results for AR and cluster-based regime switching AR and ARX models used in competition entries.

regime ARX, but at horizons greater than 12 hours the regime-switching ARX is marginally better. Therefore, forecasts from these two models are blended around the 12 hour-ahead horizon to produce our final forecasts for submission during the competition period.

The poor performance of the regime-switching ARX model at the very short-term time-scale could be due to the unsophisticated nature of the engineered diurnal features. At instances where the wind speed is low and the last measured power value is near the minimum, the regime-switching ARX can give spurious results in the very short term due to the dummy variable multipliers. However, as the AR forecast develops through the horizon it will tend to the mean and the indicator multipliers will give a more meaningful forecast.

A. Competition Forecasts

The competition period entailed submitting 14 forecasts on a daily rolling basis for the quarter hourly generation from 2 to 38 hours ahead. Team p9's entries and the code used to produce them may be downloaded here [31]. Forecast blending was only introduced from day 5 onwards.

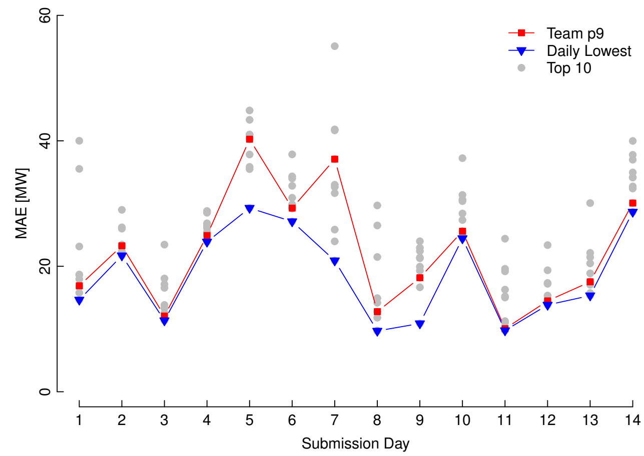

Fig. 6. Daily performance of p9 and other teams during the competition.

TABLE I
FINAL RESULTS TABLE - TOP 10

Ranking	Team	Weighted MAE (MW)
1	p9	19.5906
2	4C	20.5070
3	p5	20.5590
4	Zephyr	21.6042
5	dmlab	22.1197
6	Keanu	22.5210
7	return42	22.6332
8	p25	22.8069
9	DSAP_XXQ	23.3469
10	DSAP_group1	23.4810

The results for each day of the competition period are shown in Figure 6 which illustrates that (with the exception of day 5 and 7) the p9's final forecast model closely tracks the best performing team on each day.

Performance of the top teams 10 as published by the competition organisers is reproduced in Table I. Team p9's final error score is 4.5% lower compared to the second place team.

IV. Discussion

Some refinements to the models presented here may yield further improvement in forecast performance. As mentioned in Section II-B, the capacity of the wind farm portfolio of interest appeared to increase over the training period. Parametrising this capacity may have improved final forecast performance, but we were unable to reliably estimate the portfolio's capacity with the data available. In addition, given more training data it may have been possible identify a greater number or more precise weather regimes based on a greater number of weather variables, which may also have improved forecast performance.

The EEM Wind Power Forecasting Competition has provided a platform for researchers and forecast producers to compare different methodologies in a controlled environment;

however, since NWP forecasts were not available for participants in the competition the usefulness of the winning methodologies is limited. Day-ahead forecasts driven by NWP typically exhibit 40%-60% improvement compared to persistence [6], while our winning entry achieved less than 20% without NWP.

V. Conclusions

This paper details team p9's approach to the problem set in the EEM 2017 Wind Power Forecasting Competition. The solution was based on a blend of regime-switching auto-regressive models with regimes defined on the previous day's wind conditions identified using k-medians clustering. This approach won the competition with a weighted mean absolute error 4.5% lower than the second-place finishers.

Rigorous evaluation of multiple candidate models, and the blending of the best performing models for specific forecast horizons was key to the success of team p9, as was a combination of wind energy and meteorology domain knowledge which informed modelling decisions.

Acknowledgements

We would like to thank Dirk Hladik, Hannes Hobbie, the competition sponsor EnBW, and the EEM Conference for organising the EEM Wind Power Forecasting Competition. Jethro Browell is supported by the University of Strathclyde's EPSRC Doctoral Prize, grant number EP/M508159/1, and Ciaran Gilbert by the University of Strathclyde's EPSRC Centre for Doctoral Training in Wind and Marine Energy Systems, grant number EP/L016680/1.

Data Statement: Data and code produced by this work are freely available to download from [31], DOI: 10.15129/01f10e89-3a33-4437-8b2e-9b53c70dad4f. At the time of writing, wind power generation data may not be shared due to legal restrictions.

References

[1] T. Ackermann, Ed., *Wind power in power systems*, 2nd ed. John Wiley & Sons: New York, 2012.

[2] J. Morales, A. Conejo, H. Madsen, P. Pinson, and M. Zugno, *Integrating Renewable in Electricity Markets*. Springer, 2014.

[3] T. Hong, P. Pinson, and S. Fan, "Global energy forecasting competition 2012," *International Journal of Forecasting*, vol. 30, pp. 357–363, 2014.

[4] 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, in press.

[5] 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.

[6] 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, project funded by the European Commission under the 6th Framework Program, Priority 6.1: Sustainable Energy Systems.

[7] P. Pinson, "Wind energy: Forecasting challenges for its operational management," *Statistical Science*, vol. 28, no. 4, pp. 564–585, 2013.

[8] 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.

[9] 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, pp. 1–1, 2017.

[10] J. Browell, C. Gilbert, and D. McMillan, "Use of turbine level data for improved wind power forecasting," in *IEEE PowerTech Conference*, Manchester, UK, 06 2017.

[11] X. Zhu and M. G. Genton, "Short-term wind speed forecasting for power system operations," *International Statistical Review*, vol. 80, no. 1, pp. 2–23, 2012.

[12] 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)*, pp. 555–576, 2012.

[13] E. Erdem and J. Shi, "{ARMA} based approaches for forecasting the tuple of wind speed and direction," *Applied Energy*, vol. 88, no. 4, pp. 1405 – 1414, 2011.

[14] G. Li and J. Shi, "On comparing three artificial neural networks for wind speed forecasting," *Applied Energy*, vol. 87, no. 7, pp. 2313–2320, 2010, cited By 204.

[15] M. Yoder, A. S. Hering, W. C. Navidi, and K. Larson, "Short-term forecasting of categorical changes in wind power with markov chain models," *Wind Energy*, pp. n/a–n/a, 2013.

[16] C. Feng, M. Cui, B.-M. Hodge, and J. Zhang, "A data-driven multi-model methodology with deep feature selection for short-term wind forecasting," *Applied Energy*, vol. 190, pp. 1245 – 1257, 2017.

[17] 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, 2006.

[18] D. C. Hill, D. McMillan, K. R. W. Bell, and D. Infield, "Application of auto-regressive models to UK wind speed data for power system impact studies," *IEEE Transactions on Sustainable Energy*, vol. 3, no. 1, pp. 134–141, 2012.

[19] A. S. Hering and M. G. Genton, "Powering up with space-time wind forecasting," *Journal of American Statistical Association*, vol. 105, pp. 92–104, 2010.

[20] J. Dowell, S. Weiss, D. Hill, and D. Infield, "Short-term spatio-temporal prediction of wind speed and direction," *Wind Energy*, vol. 17, no. 12, pp. 1945–1955, 2014.

[21] P. Pinson, L. E. A. Christensen, H. Madsen, P. E. Sørensen, M. H. Donovan, and L. E. Jensen, "Regime-switching modelling of the fluctuations of offshore wind generation," *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 96, no. 12, pp. 2327–2347, 2008.

[22] P. Pinson and H. Madsen, "Adaptive modelling and forecasting of offshore wind power fluctuations with Markov-switching autoregressive models," *Journal of Forecasting*, vol. 31, no. 4, pp. 281–313, 2012.

[23] A. Steiner, C. Köhler, I. Metzinger, A. Braun, M. Zirkelbach, D. Ernst, P. Tran, and B. Ritter, "Critical weather situations for renewable energies — Part A: Cyclone detection for wind power," *Renewable Energy*, vol. 101, pp. 41–50, 2017.

[24] J. Browell, D. R. Drew, and K. Philippopoulos, "Improved very-short-term wind forecasting using atmospheric classification," *arXiv:1705.07158*, 2017.

[25] G. R. P. Box and G. M. Jenkins, *Time Series Analysis: Forecasting and Control*, revised ed., E. Robinson, Ed. Oakland, California: Holden Day, 1976.

[26] S. N. Wood, *Generalized Additive Models: an introduction with R*. Chapman and Hall/CRC, 2006.

[27] T. Hastie, R. Tibshirani, and M. Wainwright, *Statistical Learning with Sparsity: the Lasso and Generalizations*, ser. Monographs on Statistics & Applied Probability. Chapman and Hall/CRC, 2015.

[28] 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.

[29] J. Friedman, "Greedy function approximation: a gradient boosting machine," *Annals of Statistics*, vol. 29, no. 5, pp. 1189–1232, 2001.

[30] F. Leisch, "A toolbox for k-centroids cluster analysis," *Computational Statistics and Data Analysis*, vol. 51, no. 2, pp. 526–544, 2006.

[31] J. Browell and C. P. Gilbert. (2017) EEM 2017 wind power forecasting competition: Competition data and code used by the winning team.