The Journal of Engineering

The 7th International Conference on Renewable Power Generation (RPG 2018)

SCADA based nonparametric models for condition monitoring of a wind turbine

eISSN 2051-3305
Received on 26th October 2018
Accepted on 05th December 2018
doi: 10.1049/joe.2018.9284
www.ietdl.org

Ravi Kumar Pandit[1] ✉, *David Infield*[1]
[1]*University of Strathclyde, Glasgow, UK*
✉ *E-mail: ravi.pandit@strath.ac.uk*

Abstract: High operation and maintenance costs for offshore wind turbines push up the LCOE of offshore wind energy. Unscheduled maintenance due to unanticipated failures is the most prominent driver of the maintenance cost which reinforces the drive towards condition-based maintenance. SCADA based condition monitoring is a cost-effective approach where power curve used to assess the performance of a wind turbine. Such power curves are useful in identification of wind turbine abnormal behaviour. IEC standard 61400-12-1 outlines the guidelines for power curve modelling based on binning. However, establishing such a power curve takes considerable time and is far too slow to reflect changes in performance to be used directly for condition monitoring. To address this, data-driven, nonparametric models being used instead. Gaussian Process models and regression trees are commonly used nonlinear, nonparametric models useful in forecasting and prediction applications. In this paper, two nonparametric methods are proposed for power curve modelling. The Gaussian Process treated as the benchmark model, and a comparative analysis was undertaken using a Regression tree model; the advantages and limitations of each model will be outlined. The performance of these regression models is validated using readily available SCADA datasets from a healthy wind turbine operating under normal conditions.

1 Introduction

Due to the global energy crisis and thrust for clean energy, the use of wind energy has increased dramatically in recent times with both onshore and offshore wind turbines in wide-scale use. To sustain this growth, operation and maintenance (O&M) cost must be reduced. A wind turbine comprises expensive components such as a tower, blade, generator, etc. That makes replacement costly, and moreover, unexpected failures of these components lead to turbine downtime and thus increases the overall cost of energy. These unexpected failures are considered as the most significant drivers of O&M cost, particularly in case of offshore wind turbines due to their remote locations and associated logistical issues. Authors of [1] found that O&M costs make up 20–25% of the total lifetime costs of an offshore wind farm. Reducing this cost via condition monitoring is an essential target for research.

A Supervisory Control and Data Acquisition (SCADA) system can provide significant information about wind turbine operation. Condition monitoring based on SCADA data is a cost-effective approach and an effective way to monitor turbines and pinpoint potential failures and performance issues. SCADA data are retrospectively analysed aiming to detect failures in advance before they reach a catastrophic stage, [2]. For performance assessment of a wind turbine, the wind industry commonly employs the power curve. A wind turbine power curve is also useful in estimating power for given wind speed. Accurate power curve modelling is vital to the wind power provider's in the electricity market because they must bear the penalty for underestimation of day-ahead or hour-ahead energy generation [3]. The power curve role is also significant in the identifying of abnormal status and facilitates online condition monitoring which is vital for offshore wind farms because of accessibility and oversight issues, [2, 4]. The power curve provided by turbine manufacturers considers site-specific air density and wind speed as input parameters, but in reality local turbulence and the wear and tear of wind turbine components such as rotor and gearbox also affects the power curve, and this discrepancy between the empirical and the theoretical power curves have been found, resulting in inaccurate power estimation, [5]. Accurate of the power curve incorporating all influencing parameters is a current emerging research area.

Many papers have published that seek an alternative approach to power curve modelling, and these are fall broadly into two categories: parametric and non-parametric methods. The nonparametric approach often uses machine learning techniques and performs typically better than parametric methods, [6, 7]. Advanced algorithms like the Genetic algorithm (GA) and particle swarm optimisation are widely used parametric models, while neural networks, k nearest neighbour clustering (kNN), fuzzy c-means clustering and machine learning processes, such as Gaussian Processes, are now finding application for non-parametric approaches to wind engineering problems, as summarised in [8, 9].

A Gaussian process (GP), [10], is a data-driven non-parametric machine learning method that is gaining in popularity in prediction and forecasting related applications due to its simple concept and parsimony in terms of the assumptions required to construct a model as compared to other non-parametric methods (e.g. neural network or fuzzy network), [11]. Moreover, a GP comes with intrinsic confidence intervals that provide a natural way to estimate the uncertainty associated with its estimations. Recent applications of GP models to wind turbines well explained in [12–14].

This paper proposes an intelligent SCADA data-driven, nonparametric approach to monitor the performance of turbine for active condition monitoring. Two nonparametric methods namely; Gaussian Process and Regression tree are used to estimate the power curve of a wind turbine; then the comparative analysis is undertaken to identify operational anomalies. GP and regression models developed using evolutionary strategy algorithms, and then the comparative analysis is undertaken regarding model fitting accuracy and distribution function analysis. The paper will outline the advantages and limitations of these techniques.

This paper organised as follows: Section 1 is the introduction. Section 2 describes the power curve modelling and the importance of air density corrections. Section 3 describes the SCADA dataset used and its pre-processing. Section 4 outlines the nonparametric models and this section further divided into two subsections explaining the Gaussian Process and Regression Tree models for power curve estimation. Section 5 describes the comparative analysis of the proposed models, and Section 6 concludes the paper.

2 Power curve of a wind turbine

The wind turbine power curve is widely used to assess the performance of wind turbines, and it describes the relationship

This is an open access article published by the IET under the Creative Commons Attribution License
(http://creativecommons.org/licenses/by/3.0/)

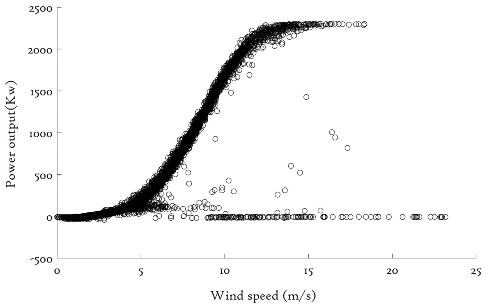

Fig. 1 *Measured power curve*

Table 1 SCADA dataset description

Start timestamp	End timestamp	Measured data	Filtered data
1/1/201200:00 AM	31/1/201223:50 PM	4464	1926

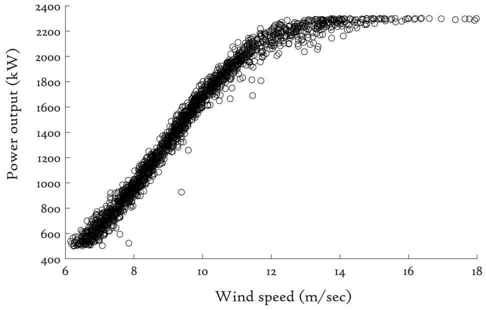

Fig. 2 *Pre-processed power curve*

between power production and hub height wind speed. Fig. 1 shows the raw data. This relationship is nonlinear and vital for identifying the abnormal status of wind turbine due to failures and or faults, [15]. With the help of a power curve, turbine power output and energy production estimated without knowing detailed knowledge of turbine operation and control. The performance of a wind turbine can be modelled on either the turbine power curve or the theoretical equation for power obtainable from the wind, mathematically expressed by:

$$P = 0.5\rho A C_p(\lambda, \beta)\nu^3 \tag{1}$$

where ρ is air density (kg/m^3), A is swept area (m^2), C_p is power coefficient of the wind turbine and ν is the hub wind speed (m/sec). The power coefficient is the function of tip speed ratio (λ) and pitch angle (β) and thus affects the power production of the wind turbine along with wind shear, turbulence, and inflow angle, [15, 16].

SCADA data used in this study is from a pitch-regulated wind turbine, and IEC standard 61400-12-1 recommends that air density correction should be applied to the SCADA datasets for accurate power curve modelling. The following formulas used for making air density corrections:

$$\rho = 1.225\left[\frac{288.15}{\mathrm{T}}\right]\left[\frac{\mathrm{B}}{1013.3}\right] \tag{2}$$

and,

$$V_C = \mathrm{V_M}\left[\frac{\rho}{1.225}\right]^{\frac{1}{3}} \tag{3}$$

where V_C and V_M are the corrected and measured wind speed in m/sec and the corrected air density is calculated by (2) where B is atmospheric pressure in mbar and T the temperature in Kelvin. It is worth highlighting that the wind site farms parameters: altitude and ambient temperature affect the air density. In (3), B and T records 10-minute average values obtained from SCADA datasets of an operational wind turbine. The calculated value of ρ is then used in (3) to calculate the corrected wind speed (V_C). This approach will be used in the next section for developing the correct and error-free power curve of a wind turbine using nonparametric models.

3 SCADA data pre-processing

A wind farm SCADA system provides valuable information, for example, load history and operation of individual turbines without additional cost. This makes SCADA based condition monitoring a cost-effective approach. SCADA based modelling can be useful in improving the overall health of a turbine as well as playing a significant role in the identification of components failures. The SCADA data used in this study are from an operational wind turbine located in Scotland, UK. More than 100 different signals included ranging from the timestamp, calculated values, set points, measurements of temperature, current, voltage, wind speed, power output, wind direction, and so on. The SCADA dataset used here consists of 10-minute averages with maximum, minimum, and standard deviation; the available data set corresponded to a full one month of operation and divided into operational data, status data and warning data.

SCADA data can include errors due to sensor failures and data collection faults. Model fitting degraded by missing, invalid and poorly processed SCADA data. Minimising such errors is an essential requirement for accurate analysis. Data acquisition errors and time-steps with missing or erroneous data steps excluded by pre-processing of the SCADA data. Furthermore, the criteria described in ref. [17] for example; such as timestamp mismatches, out of range values, negative power values, and turbine power curtailment can be used to remove misleading data. Despite following such procedures, the resulting SCADA data is not entirely free from error.

The data used in this paper is for a 2.3 MW Siemens turbine and contains 4464 data points, beginning with time stamp "1/1/2012 00:00 AM" and ending at time stamp "31/1/2012 23:50 PM". The data for this month of operation (unfiltered) shown in Fig. 1. These measured data points became 1926 data points after pre-processing (Table 1) and were used to develop nonparametric power curve models in subsequent sections. Fig. 2 shows the filtered and air density corrected (described in section 2) data.

4 Power curve estimation using nonparametric models

Nonparametric models are data-driven but not protected against overfitting, and hence this issue needs particular attention. Cross-validation is being used here to protect against overfitting by partitioning the data set into folds and estimating the accuracy for each of these. The primary objective of the cross-validation (CV) analysis is to determine whether the developed model for curve estimation is appropriate for power curve prediction independently of the data set. This helps to estimate how accurately an estimated model will perform for independent SCADA data sets. Here, the monthly SCADA data is partitioned into approximately 20% and 80% for training and validation respectively. Fivefold cross-validation was found to give satisfactory results. The approach used in this study is similar to leave-p-out CV approach. These techniques applied to both nonparametric models for better results. The algorithms for power curve fitting using Gaussian Process and Regression Tree described as follows.

4.1 Gaussian process

A Gaussian Process (GP) is a collection of random variables, any finite number of which have a joint Gaussian distribution; it is a nonlinear, nonparametric model useful in representing dependent data observed over the period. A GP defines a prior over functions, which converted into a posterior over functions and mathematically a GP is a distribution over the functions. GP entirely specified by a mean function and a covariance function, If $m(x)$ is the mean function and $k(x, x')$ is the covariance function of a real process, then the desired function $f(x)$ is defined as:

This is an open access article published by the IET under the Creative Commons Attribution License
(http://creativecommons.org/licenses/by/3.0/)

$$f(x) \sim \; GP(m(x), k(x, x')) \tag{4}$$

where k is the covariance function that has an associated probability density function:

$$P(x; m, k) = \frac{1}{(2\pi)^{\frac{n}{2}}\left|k\right|^{\frac{1}{2}}}\exp\left\{-\frac{1}{2}(x-m)^T k^{-1}(x-m)\right\} \tag{5}$$

where $|k|$ is defined as a determinant of k, n is the dimension of random input vector x, and m is mean vector of x. The term under exponential i.e. $1/2(x-m)^T k^{-1}(x-m)$ represents an example of a quadratic form.

The covariance function defines the covariance between random variables that relate to multivariate Gaussian distribution; its appropriate selection is vital for model accuracy. There are numerous possible covariance functions available, and suitability of these depends on the model application, see for example [10, 18]. For this paper, the squared exponential covariance function ($k_{SE)}$ is used which is technically a smooth sample function (ie infinitely differential) defined as:

$$k_{SE}(x, x') = \sigma_f^2 \exp\left(-\frac{(x, x')^2}{2l^2}\right) \tag{6}$$

As already described, the SCADA data contains noise and measurement error that affects the covariance function (k_{SE}) and hence it is wise to add a noise term to the covariance function in order to compensate for the effects of this additional uncertainty. To make, the covariance function more representative (5) is modified:

$$k_{\mathrm{SE}}(x, x') = \sigma_f^2 \exp\left(-\frac{(x, x')^2}{2l^2}\right) + \sigma_n^2 \delta(x, x') \tag{7}$$

where σ_f^2 and l are known as the hyper-parameters. σ_f^2 describe the signal variance and l is a characteristic length scale which describes how quickly the covariance decreases with the distance between points. Using the squared exponential covariance function, a monthly GP predicted power curve algorithm has been realised in MATLAB [19]. In Fig. 3, the comparison between measured and estimated power curve shows that the GP provides an effective estimation of the power curve.

Model residuals defined as the difference between measured and estimated values; its analysis is essential for understanding the behaviour of a GP model. A residual plot is shown in Fig. 4 as a time series and indicates that the estimated GP values are generally close to measured values. Theoretically, the residuals of a GP model should have a Gaussian distribution. To confirm this, the frequency distribution is shown in Fig. 5 together with a fitted Gaussian distribution and, as expected, the distribution of GP residuals is close to being Gaussian.

4.2 Regression tree

A regression tree/decision tree is a data-driven, nonparametric machine learning approach. It falls under the family of classification and regression trees (C&RT) and is a recursive partitioning method useful for estimating continuous dependent variables regression and categorical predictor variables for classification. Breiman, [20], developed the classic C&RT algorithm in 1984. The use of decision trees has grown due to its ease of implementation and interpretation as compared against alternative quantitative data-driven tools. It has proved useful in wind turbines condition monitoring applications such as detecting faults, errors, damage patterns, anomalies and abnormal operation. Regression trees are easy to interpret and gives fitting and estimation results quickly with low memory requirements. Basic Regression tree theory is described in refs. [20, 21] and used for power curve estimation as described below.

A Regression tree may be considered a variant of decision trees, designed to approximate real-valued functions and built through a process known as binary recursive partitioning, which is an iterative technique that splits the SCADA datasets into partitions or branches, and then continues splitting each partition into smaller groups as the method moves up each branch [20]. In this study, a decision tree with binary splits for regression adopted. Training data points initially grouped into the same partition and then regression tree algorithm begins allocating the data into the first two partitions or branches, using every possible binary split on every field. To minimise the squared deviations from the mean in the two partitions, the decision tree split, and then it applied to each of the new branches, and this process continues until it gives a satisfactory result as per described criterion in ref. [22].

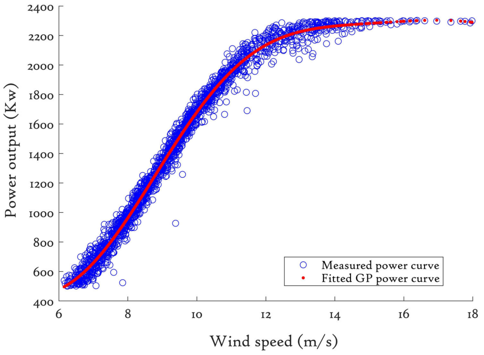

Fig. 3 *Measured and GP estimated power curve comparison*

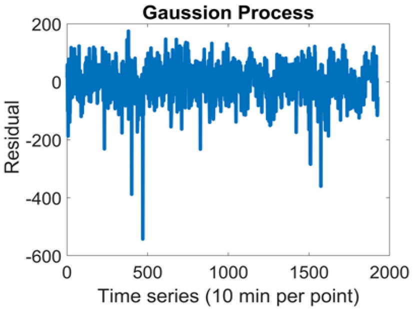

Fig. 4 *GP residuals time series plot*

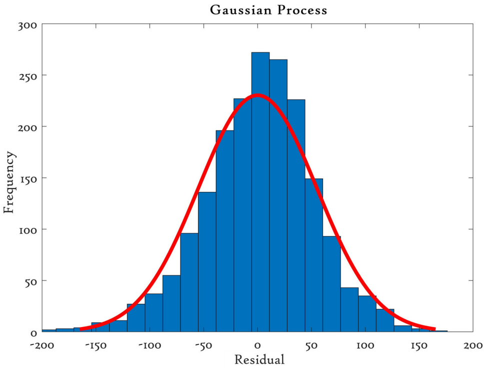

Fig. 5 *GP residuals histogram fitting*

To estimate the power curve, the root node down to a leaf node methodology outlined above used. The leaf node contains the estimated value, and here a minimum leaf size of 30 has been used to prevent overfitting while delivering an accurate estimate of the power curve. Using training datasets, the estimated value of each leaf node is calculated using [20, 23] and the outcome compared with the measured power curve. The algorithm realised in MATLAB; Fig. 6 shows the result. Since creating a decision tree is very complicated, especially for large datasets like the one used in this study. It needs many branches and makes the overall model complex and time consuming to compute. Moreover, the estimated decision tree based power curve reliability depends on feeding the exact internal and external information at the onset. Hence the effect of a small change in input data can cause a substantial change in the tree and makes the modelling process potentially

This is an open access article published by the IET under the Creative Commons Attribution License
(http://creativecommons.org/licenses/by/3.0/)

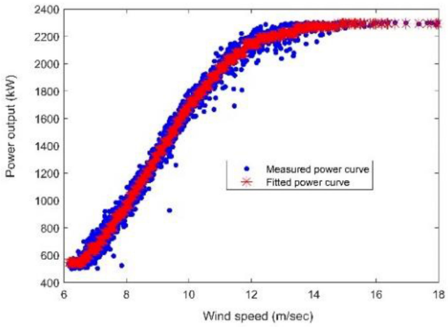

Fig. 6 *Measured and Regression tree estimated power curve comparisons*

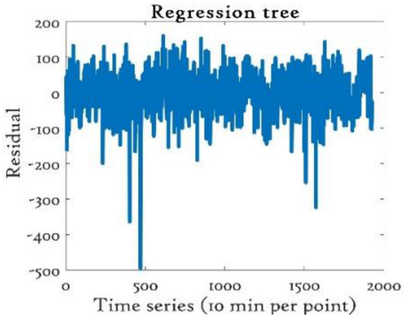

Fig. 7 *Residual plot in time series*

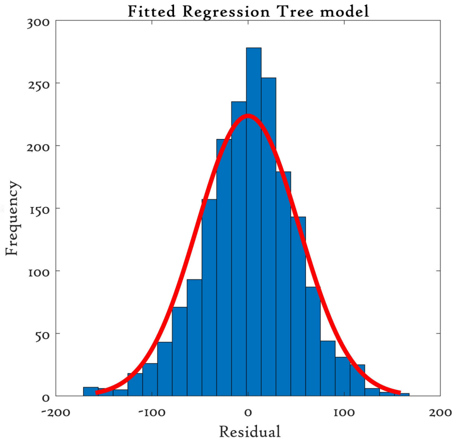

Fig. 8 *Residual plot in time series*

unstable. Even though decision trees follow natural relationships between events, it is difficult to plan for contingencies that arise from a decision process, and such oversights can lead to severe decisions. To solve this issue, Random Forest techniques have proposed that combine many decision trees based on slightly different versions of the dataset, see [21]. Random Forest modelling is out of the scope of this study but would address in future work.

The obtained residuals of the decision/regression tree model shown as a time series in Fig. 7. The corresponding distribution function showed in Fig. 8 and, like the GP model roughly has the form of a Gaussian distribution.

5 Manuscript preparation comparative analysis of the two models

Nonparametric models are The GP and regression tree (RT) based power curves compared to assess their respective fitting accuracy. Fig. 9 and 10 compare the power curve estimates. The two approaches give very similar results, and statistical analysis is required to differentiate them.

Statistical performance indicators used to evaluate the goodness-of-fit of the models described as follows. The root mean square error (RMSE) has been used as a standard statistical metric

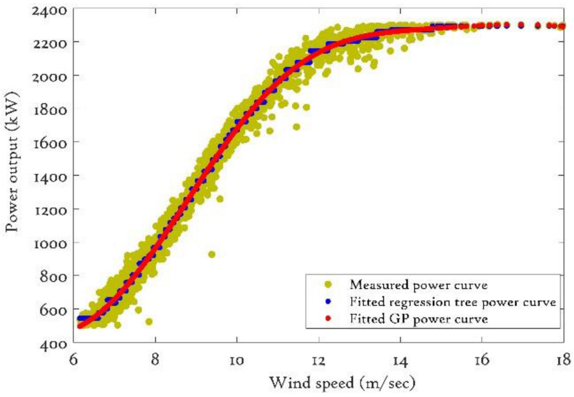

Fig. 9 *Nonparametric Power Curve models*

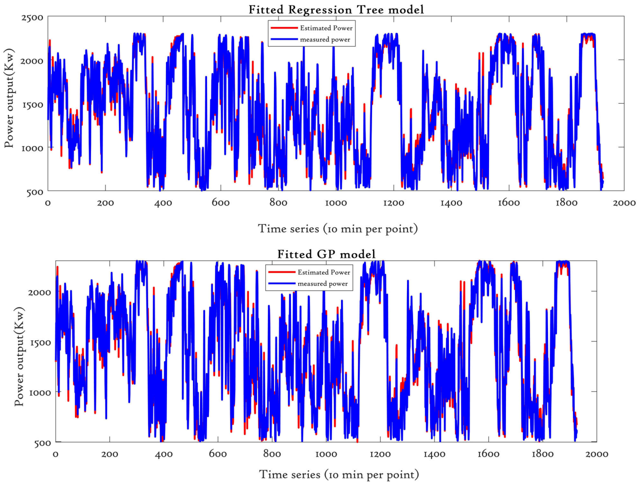

Fig. 10 *Estimated power comparison in time series*

to measure model performance, [24], and widely used to quantify the magnitude of the residuals; it defined as:

$$\mathrm{RMSE} = \sqrt{\frac{\Sigma_{i=1}^{n}(y_i' - y_i)^2}{n}} \tag{8}$$

The mean absolute error (MAE) is widely used as an error assessment indicator to facilitate comparison with existing models and signifies the closeness of estimated results with observed values and mathematically expressed as:

$$\mathrm{MAE} = \frac{\Sigma_{i=1}^{n}\mathrm{abs}(y_i' - y_i)}{n} \tag{9}$$

where y' are the estimated values for n different predictions, and y are the measured values.

Another commonly used statistical measure is the coefficient of determination (R^2) which describes how close the data are to the fitted regression, [24], it is defined as: $R^2 = 1 - (\mathrm{SSE}/\mathrm{TSS})$; where SSE is the sum of squared errors, and TSS is the total sum of squares. (Table 2 summarises these error statistics for the two models. It is found that GP model performs better than the regression tree, although it is more time consuming to calculate.

6 Conclusion and future work

In this paper, two SCADA based nonparametric models for the wind turbine power curve proposed. Power curve modelling using a Gaussian Process model with squared exponential covariance function is simple and straightforward. Though GP power curve model accuracy can degrade with a large number of data points, and a low number of data points may yield poor estimation, a useful compromise has found here. The GP power curve has been compared with a regression tree/decision tree model and found to

This is an open access article published by the IET under the Creative Commons Attribution License (http://creativecommons.org/licenses/by/3.0/)

Table 2 *Error statistics for the two models*

Models	RMSE	R^2	MAE	Prediction speed	Training speed	Remarks
RT	59.089	0. 99	43.96	~290000 obs/sec	0.58 sec	Not continuous response and Poor smoothnes s
GP	55.223	0. 99	40.71	~61000 obs/sec	45.71 sec	Strong smoothnes s

be superior as given by the statistical performance indicators summarised in Table 1. Even though the time taken to run and evaluate the power curve algorithm is faster for the regression tree, it suffers from overfitting while the GP model strikes a right balance between algorithm smoothness and model optimisation time. In both nonparametric models, a cross-validation analysis was performed for accurate wind power curve prediction and to prevent overfitting. The regression tree model comes with limitations such as complexity, high computational cost and instability that summarised in [20, 25]. To overcome these issues, the tree-ensemble method can be used and analysed. Furthermore, the regression tree estimated output is not continuous (see Figs. 6 and 9) though it can minimise by using gradient boosting and boosted regression trees (BRT), see [26], but not discussed in this paper. Uncertainty analysis measures the degree of 'wrongness' of nonparametric models and traditionally described by a loss and cost function which is useful for developing efficient anomaly detection algorithms for wind turbines. The Gaussian Process models come with intrinsic confidence intervals that define the uncertainty of the models while this is difficult to estimate in regression tree approach. Future work will focus on comparative performance validations of the tree-ensemble technique, Gaussian Process models, and also other advanced nonparametric models, for condition monitoring of a wind turbine.

7 Acknowledgments

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 642108.

8 References

[1] van de Pieterman, R.: 'Cost Modelling for Offshore Wind Operations and Maintenance', April 2013.

[2] Joon-Young, P., Jae-Kyung, L., Ki-Yong, O.*, et al.*: 'Development of a novel power curve monitoring method for wind turbines and its field tests', *IEEE Trans. Energy Convers.*, 2014, **29**, pp. 119–128

[3] Panapakidis, I.P., Dagoumas, A.S.: 'Day-ahead electricity price forecasting via the application of artificial neural network based models', *Appl. Energy*, 2016, **172**, pp. 132–151

[4] Kusiak, A., Verma, A.: 'Monitoring wind farms with performance curves', *IEEE Trans Sustain Energy*, 2013, **4**, pp. 192–199

[5] Spinato, F., Tavner, P., Van Bussel, G.*, et al.*: 'Reliability of wind turbine subassemblies', *IET Renew. Power Gener.*, 2009, **3**, pp. 387–401

[6] Lydiaa, M., Suresh Kumarb, S., Immanuel Selvakumara, A.*, et al.*: 'A comprehensive review on wind turbine power curve modeling techniques', *Renewable Sustainable Energy Rev.*, 2014, **30**, pp. 452–460

[7] Sohoni, V., Gupta, S.C., Nema, R.K.: 'A Critical Review on Wind Turbine Power Curve Modelling Techniques and Their Applications in Wind Based Energy Systems'. doi: 10.1155/2016/8519785b.

[8] Panahi, D., Deilami, S., Masoum, M.A.S.: 'Evaluation of parametric and non-parametric methods for power curve modelling of wind turbines', 2015 9th Int. Conf. on Electrical and Electronics Engineering (ELECO), Bursa, Turkey, 2015, pp. 996–1000. doi: 10.1109/ELECO.2015.7394497.

[9] Yampikulsakul, N., Byon, E., Huang, S.*, et al.*: 'Condition monitoring of wind power system with nonparametric regression analysis', *IEEE Trans. Energy Convers.*, 2014, **29**, pp. 288–299, idoi: 10.1109/TEC.2013.2295301

[10] Rasmussen, C.E., Williams, C.K.I.: '*Gaussian processes for machine learning*', (the MIT Press, Cambridge, UK, 2006) ISBN 026218253X

[11] Chen, N., Qian, Z., Nabney, I.T.*, et al.*: 'Wind power forecasts using Gaussian processes and numerical weather prediction', *IEEE Trans. Power Syst.*, March 2014, **29**, pp. 656–665, idoi:10.1109/TPWRS.2013.2282366

[12] Bockhorst, J., Barbe, C.: 'Gaussian Processes for Short-Horizon Wind Power Forecasting', http://www.cs.uwm.edu/~joebock/papers/maics10.pdf

[13] Pandit, R.K., Infield, D.: 'Comparison of binned and Gaussian Process based wind turbine power curves for condition monitoring purposes', *J. Maint. Eng.*, 2017, **25**, ISBN no :978-1-912505-25-8

[14] Bulaevskaya, V., Wharton, S., Clifton, A.*, et al.*: 'Wind power curve modeling in simple and complex terrain using statistical models', *J. Renewable Sustainable Energy*, 2015, **7**

[15] Wind Turbines—Part 12-1: Power Performance Measurements of Electricity Producing Wind Turbines, British Standard, IEC 61400-12-1, 2006

[16] Wang, Y., Infield, D.G.: 'Power curve based online condition monitoring for wind turbines', COMDEM 2013 conference. doi: 10.13140/2.1.4492.9928

[17] Schlechtingen, M., Santos, I.F.: 'Comparative analysis of neural network and regression based condition monitoring approaches for wind turbine fault detection', *Mech. Syst. Signal Process*, 2016, **25**, pp. 1849–1875

[18] Kronberger, G., Kommenda, M.: 'Evolution of Covariance Functions for Gaussian Process Regression using Genetic Programming'

[19] Gaussian Process Regression Models, MATLAB toolbox version 2017b

[20] Breiman, L., Friedman, J.H., Olshen, R.A.*, et al.*: '*Classification and regression trees*', (Wadsworth & Brooks/Cole Advanced Books & Software, Madrid, Spain, 1984). ISBN 978-0-412-04841-8.

[21] Sutton, C.D.: 'Classification and regression trees, bagging, and boosting', *Handb. Stat.*, 2005, **24**

[22] Salehi-Moghaddami, N., Yazdi, H.S., Poostchi, H.: 'Correlation based splitting criterion in the multi-branch decision tree', *Cent. Eur. J. Comput. Sci.*, 2011, **1**, pp. 205–220, doi:10.2478/s13537-011-0017-x

[23] Sørensen, J.D.: *'Optimal. 'risk-based operation and maintenance planning for offshore wind turbines*', (Aalborg, Denmark)

[24] Glantz, S.A., Slinker, B.K.: '*Primer of applied regression and analysis of variance*', (McGraw-Hill, New York, NY, USA, 1990). ISBN 0-07-023407-8.

[25] Van Putten, W.: '*CART: stata module to perform classification and regression tree analysis*', (Statistical Software Components, Boston, MA, USA, 2006).

[26] Elith, J., Leathwick, J.R., Hastie, T.: 'A working guide to boosted regression trees', *J. Anim. Ecol.*, 2008, **77**, pp. 802–813

This is an open access article published by the IET under the Creative Commons Attribution License
(http://creativecommons.org/licenses/by/3.0/)