

# Comparison of advanced non-parametric models for wind turbine power curves

ISSN 1752-1416

Received on 4th September 2018

Revised 30th January 2019

Accepted on 19th March 2019

E-First on 10th April 2019

doi: 10.1049/iet-rpg.2018.5728

www.ietdl.org

Ravi Kumar Pandit<sup>1</sup> ✉, David Infield<sup>1</sup>, Athanasios Kolios<sup>2</sup>

<sup>1</sup>Department of Electronic & Electrical Engineering, University of Strathclyde, 16 Richmond St, Glasgow – G1 1XQ, Scotland, UK

<sup>2</sup>Department of Naval Architecture, Ocean & Marine Engineering, University of Strathclyde, 16 Richmond St, Glasgow – G1 1XQ, Scotland, UK

✉ E-mail: ravi.pandit@strath.ac.uk

**Abstract:** To continuously assess the performance of a wind turbine (WT), accurate power curve modelling is essential. Various statistical methods have been used to fit power curves to performance measurements; these are broadly classified into parametric and non-parametric methods. In this study, three advanced non-parametric approaches, namely: Gaussian Process (GP); Random Forest (RF); and Support Vector Machine (SVM) are assessed for WT power curve modelling. The modelled power curves are constructed using historical WT supervisory control and data acquisition, data obtained from operational three bladed pitch regulated WTs. The modelled power curve fitting performance is then compared using suitable performance, error metrics to identify the most accurate approach. It is found that a power curve based on a GP has the highest fitting accuracy, whereas the SVM approach gives poorer but acceptable results, over a restricted wind speed range. Power curves based on a GP or SVM provide smooth and continuous curves, whereas power curves based on the RF technique are neither smooth nor continuous. This study highlights the strengths and weaknesses of the proposed non-parametric techniques to construct a robust fault detection algorithm for WTs based on power curves.

## 1 Introduction

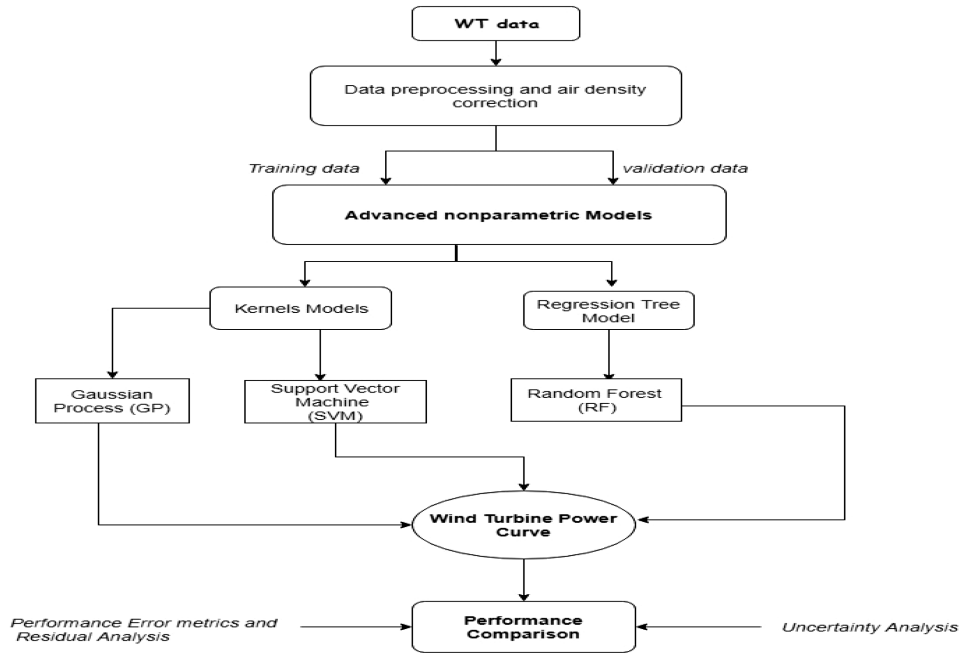
Unexpected failures of wind turbine (WT) components, in particular, the rotor, gearbox, and generator make operation and maintenance (O&M) more expensive and can add significantly to the overall cost of energy. Furthermore, offshore wind farm's O&M cost is higher due to transportation and logistics issues, and thus there is a pressing need to reduce the O&M cost by employing continuous condition monitoring and using predictive, and proactive maintenance strategies. Predictive maintenance can be useful in identifying failures at an early stage and preventing catastrophic damage. Bently and Hatch [1], define condition monitoring as the process of monitoring the condition of a machine through measurement of parameters such as vibration or temperature in such way that a significant change is indicative of a developing failure. Such methods are well described in [2, 3].

Supervisory control and data acquisition (SCADA)-based condition monitoring is considered to be financially savvy since the required information is accessible at no additional cost [4, 5]. Reflecting this advantage, some SCADA-based condition monitoring approaches have recently been proposed: vibration for rotor blades [6, 7]; advanced signal processing for gearboxes [8, 9] or bearings [10]; and for others [5, 11]. Among the various proposed SCADA-based models, the WT power curve is widely used since it reflects the turbine behaviour, which is helpful for fault diagnosis and condition monitoring.

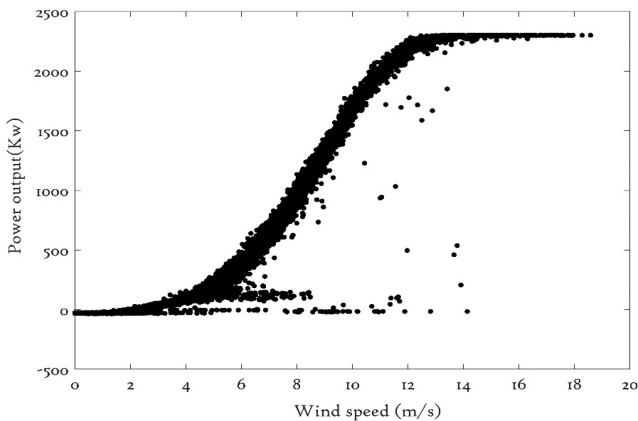
Power production is the key consideration when assessing a potential site for wind farm development. The power output of a WT is estimated from the power curve and wind speed profile for the site in question. The wind power generation affected by the site air density. The predicted long-term gross annual mean power output at a target site is calculated with the help of a WT power curve. Moreover, power curve models can be useful in forecasting and capacity factor estimation purposes. Numerous techniques have been introduced in the past to model WT power curves and these techniques generally divided into parametric and non-parametric. Parametric models are generally based on mathematical models that are often constructed from a family of functions with some variables that are fitted to correspond to the particular WT. Widely used parametric approaches are segmented

linear models [12], polynomial regression [13, 14], and models based on probabilistic distributions such as four- or five-parameter logistic distributions [15, 16]. In contrast to the parametric approach, non-parametric approaches do not enforce any pre-specified condition, and thus, the estimated power curve is as close as possible to the measured data subject to the smoothness of the fit. Owing to this, non-parametric models are able to model the power curve accurately over a wide range. Several studies have been conducted to develop an accurate power curve for performance evaluation and these methods include a copula power curve model [17], cubic spline interpolation [18], support vector machine (SVM) [19], neural networks [20, 21], and data-driven methods (e.g. Gaussian Process (GP) [22], Random Forest (RF) [23], and the k-nearest neighbour clustering [24]). A comprehensive review of the existing WT power curve monitoring techniques can be found in [16].

GP models are used extensively in the literature [25, 26] for a wide range of modelling applications; however, they have not much been used to explore issues related to WTs. SVM is another non-parametric method that has been introduced for WT power curve modelling [27, 28]. However, both methods suffer from a number of practical issues such as the cubic inversion issue associated with more massive datasets. Finally, the RF model is another non-parametric approach used to construct power curves. Unlike most of the non-parametric approaches, the RF does not need to be tuned or optimised, and it incorporates the prediction of several weak predictors [29]. As the name suggests, it is used to create a forest in some way and make it random while maintaining the direct relationship among the number of trees in the forest. Usually, a large number of trees indicates a more accurate result. It is worth noting that RF and decision tree techniques are not the same because with the RF random samples are used to obtain the root node and splitting the feature nodes, in contrast to decision trees [30]. These advanced non-parametric models are flexible and easy to implement, and computationally straightforward to implement.



**Fig. 1** Framework of the advanced non-parametric WT Power Curve models for performance comparison



**Fig. 2** Measured power curve of an industrial WT

## 2 Scientific contribution

As has already been mentioned, power curve model accuracy varies with the techniques and the particular data set used, and hence there is no one technique that performs best in all cases of observations obtained from different WTs. It is therefore essential to investigate the performance of different non-parametric techniques for power curve modelling to evaluate which technique is more accurate for a given dataset. Advanced non-parametric models such as the GP, SVM, and RF are gaining popularity because of their low computational cost and high accuracy. The direct comparison between these models can be useful in identifying the method that is more robust and computationally feasible. This paper aims to fill this gap.

The paper presents the implementation of three advanced non-parametric algorithms (GP, SVM, and RF) for modelling of WT power curves, and their accuracy has been compared using error performance metrics [root mean square error (RMSE),  $R^2$ , and mean absolute error (MAE)]. Comparison of the methods is made to identify the best approach taking into account the computational cost and required processing power. The SCADA dataset obtained from modern pitch-regulated WTs is used to train and validate the performance of the proposed non-parametric models. The outcomes should be useful in constructing power curve-based fault detection algorithms, where accuracy is paramount, for the purpose of condition monitoring. A framework for modelling WT power curves and its performance comparisons is presented in Fig. 1 and described as follows. The SCADA data extracted from operational

WTs, which is then filtered, and air density corrected. After this, datasets are divided into training and validation; training data points are used to train the models and validation data points are used to validate the performance of models. Performance error metrics, residuals analysis, and uncertainty analysis, are used to compare the performance of the models and based on this comparison, the best approach for WT power curve modelling is being suggested.

## 3 WT power curve modelling

The WT power curve describes the non-linear relationship between the power and hub height wind speed and is shown in Fig. 2. Primarily it is used to capture the WT performance. The electrical power output of the turbine is not only correlated with wind speed but also responds to turbulence intensity, wind direction, vertical, and horizontal shear, atmospheric stability, drive train temperature, yaw error and so on [31]. An accurate power curve is not only used to improve performance assessments but can play a significant role in identifying different WT fault types. Usually, an individual WT has unique power curve depending on the operating conditions for which it has been designed (e.g. wind speed range) and actively used for continuous monitoring the performance of a WT by differentiating between a normal and an abnormal state [31]. The WT power curve follows the sigmoid shape and any changes in its characteristic shape likely to indicate abnormal operation due to a fault. The wind speed between the cut-in and the cut-out speed ranges are considered significant because this operational region presents a significant opportunity to optimise the WT power generation process. The theoretical power obtained from a WT is given by

$$P = 0.5 \rho A C_p(\lambda, \beta) v^3 \quad (1)$$

where  $\rho$  is air density ( $\text{kg/m}^3$ ),  $A$  is swept area ( $\text{m}^2$ ),  $C_p$  is the power coefficient of the WT, and  $v$  is the hub height wind speed ( $\text{m/s}$ ). The power coefficient is a function of tip speed ratio ( $\lambda$ ) and pitch angle ( $\beta$ ). In addition to these two parameters, the power output of a WT affected by flow conditions, in particular, terrain, wind shear, turbulence intensity, and air density [31].

### 3.1 Air density correction

The IEC standard (61400-12-1) [32], recommend air density correction for accurate power curve measurement. The IEC standards suggest two approaches for air density correction based

on whether the turbine is a pitch or stall regulated. In this paper, SCADA datasets are from a pitch-regulated and hence corrected wind speed  $V_C$  is calculated using (2) and (3) as shown below

$$\rho = 1.225 \left[ \frac{288.15}{T} \right] \left[ \frac{B}{1013.3} \right] \quad (2)$$

and

$$V_C = V_M \left[ \frac{\rho}{1.225} \right]^{1/3} \quad (3)$$

where  $V_C$  and  $V_M$  are the corrected and measured wind speeds in m/s and the corrected air density is calculated by (7) where  $B$  is atmospheric pressure in mbar, and  $T$  the temperature in Kelvin in which 10 min average values obtained from SCADA data are used. The corrected wind speed ( $V_C$ ) from (3) is then used to calculate the power curve, normally by binning. The air density  $\rho$  is related to temperature by the gas law  $\rho = p/(R \cdot T)$ ; where  $p$  is absolute atmospheric pressure,  $R$  the gas constant, and  $T$  is the environmental temperature in Kelvin. The air density computed by using the gas law where air pressure and temperature measured from sonic anemometer at hub height. It is worth to note that because of the significant impact of temperature on air density, air density correction is sometimes referred to as temperature correction.

#### 4 SCADA data source and pre-processing

Original SCADA-based condition monitoring is a cost-effective approach which gathers the information provided by sensors without any extra cost. The WT SCADA system has more than 120 parameters such as the power output, hub height wind speed, ambient pressure, wind direction, vibrations, digital control signals, and ambient temperatures; comes with minimum, maximum, average, and standard deviation values. These SCADA datasets are essential in identifying early warning of failures and improving the performance of WTs. Despite such advantages, unexpected sensor or data collection malfunction can create errors in the SCADA datasets; if not dealt with, such errors will degrade the model accuracy. It is, therefore, necessary to filter the SCADA data in order to remove as far as possible any erroneous data before doing further analysis. The steps described in [33] have been applied; these include timestamp mismatches; negative power values; out of range values; abnormal wind speeds; and turbine power curtailment, all considered to be questionable and thus confusing data. However, it should be noted that despite such filtering,

SCADA data is not entirely free from error. Fig. 3 is a filtered, and air density corrected WT power curve.

To analyse the performance of WTs, the SCADA data of 2.3 MW Siemens collected at a wind farm located in Scotland, UK has been used. SCADA contains 13,250 data point that beginning with time stamp '1/10/2012 00:00 AM' and ending at time stamp '31/12/2012 23:50 PM' and is sampled at 10 min average. These measured data points became 3960 data points after pre-processing (Table 1) and will be used for power curve modelling based on advanced non-parametric approaches in upcoming sections. In total, 13,250 SCADA data points were collected from 1 October 2012 to 31 December 2012 which are divided using 10-fold cross validation into training and validation datasets with a ratio of 70 and 30%, respectively. The training datasets would be used to train the models while validation datasets would be used to assess the performance of the models.

#### 5 WT power curve modelling

In this section, the algorithmic procedure of the proposed advanced non-parametric power curve modelling approaches is explained in detail. The three advanced non-parametric models, namely; GP, SVM, and RF were used to construct the power curve of a WT. Out of these three approaches, GP and SVM are powerful kernel-based methods while RF technique is based on a regression tree.

##### 5.1 Power curve model based on GP

GP model is a non-parametric, non-linear approach for accurate function approximation in high-dimensional space. GP models are very flexible, and its brief explanation can be found in [34]. Here, a brief description of GP for WTs power curve provided. GP models entirely specified by its mean function and covariance functions

$$f(x) \sim \text{GP}(\mu(x), k(x, x')) \quad (4)$$

where the mean function  $\mu(x)$  and the covariance function  $k(x, x')$  are defined by

$$\mu(x) = E[f(x)] \quad (5)$$

$$k(x, x') = E[(f(x) - \mu(x))(f(x') - \mu(x')))] \quad (6)$$

The GP models accuracy depends on its covariance function or kernels (a positive-definite function) and is describes the similarity between two points. The mean function  $\mu(x)$  for the simplicity often taken as zero because preferable to centre observed output to have a zero mean although it can be arbitrarily selected. The covariance function is the heart of the GP model and generally

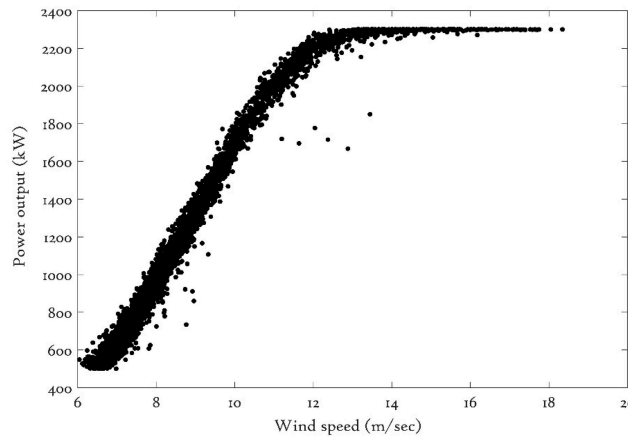
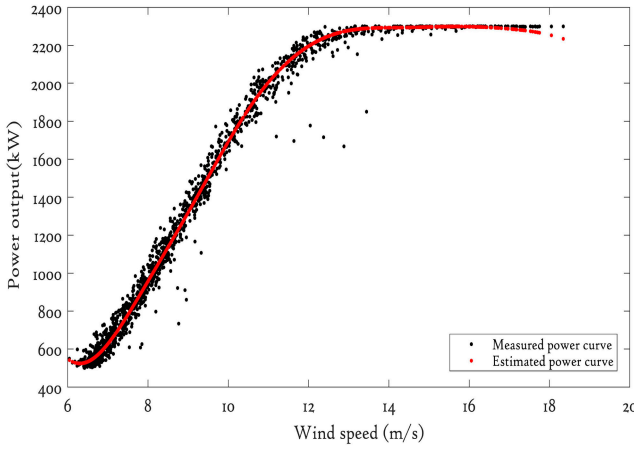


Fig. 3 Preprocessed and air density corrected power curve

Table 1 SCADA data description

Start timestamp	End timestamp	Measured data	Filtered data	Training data	Validation data
1/10/2012 00:00 AM	31/12/2012 23:50 PM	13,250	3960	2500	1460



**Fig. 4** GP power curve

divided into stationary and non-stationary functions and are explained briefly in [34]. The most widely used covariance function is squared exponential function is a stationary function and would be used in this paper as it is more suitable in expressing the non-linear relationship of wind speed and power output of a WT and is mathematically expressed as

$$k_{SE}(x, x') = \sigma_f^2 \exp\left(-\frac{(x - x')^2}{2l^2}\right) \quad (7)$$

The squared exponential covariance function ( $k_{SE}$ ) is a stationary covariance that calculates the covariance between any two points and is a function of Euclidean distance. The WTs SCADA datasets come with noise and measurement error which eventually will affect the GP model accuracy. Hence it is advisable to add a noise term added along with covariance function to minimise the impact of these errors and (7) further modified to

$$k_{SE}(x, x') = \sigma_f^2 \exp\left(-\frac{(x - x')^2}{2l^2}\right) + \sigma_n^2 \delta(x, x') \quad (8)$$

where  $\sigma_f^2$  and  $l$  are defined as the hyper-parameters.  $\sigma_f^2$  describes the signal variance and  $l$  is a characteristic length scale which signifies how quickly the covariance decreases with the distance between points.

To estimate the power curve of a WT for given SCADA datasets involves finding the most appropriate parameters and it is generally achieved by maximum log marginal likelihood or posterior estimation. Let us assume that we have a training set  $\mathcal{D}$  of  $n$  observations,  $\mathcal{D} = \{\mathbf{x}_i, y_i\}_{i=1}^n$ , where  $\mathbf{x}$  an input vector of  $D$  and  $y$  is a scalar output. The whole input dataset is represented by a  $\mathcal{D} \times n$  matrix since we have  $n$  cases of  $\mathbf{x}$  and with target values collected in a vector  $\mathbf{y}$ , can be written as  $\mathcal{D} = (\mathbf{X}, \mathbf{y})$ . Therefore, the function  $f(x_i)$  used to transform the input vector  $\mathbf{x}_i$  to the targeted value  $y_i$  using

$$y_i = f(\mathbf{x}_i) + \epsilon_i \quad (9)$$

where  $\epsilon_i$  is Gaussian noise with zero mean and variance  $\sigma_n^2$  such that  $\epsilon_i = N(0, \sigma_n^2)$ . The targeted value  $\mathbf{y}$  is a linear combination of Gaussian variables and hence is itself Gaussian [35]. To gather the information about an uncertain parameter, prior distribution used which can be either uninformative or informative. The obtain prior with probability distributions of new data points used to calculate the posterior distribution. The prior on  $\mathbf{y}$  becomes

$$E[\mathbf{y}] = E[f] + \epsilon = 0 \quad (10)$$

$$\text{cov}[\mathbf{y}] = K[\mathbf{X} \cdot \mathbf{X}] + \sigma_n^2 \mathbf{I} \quad (11)$$

The given training dataset  $(\mathbf{X}, \mathbf{y})$  used to train the GP model in order to estimate the targeted variable  $f$ , for a given new input  $\mathbf{x}_*$ . The distribution of with new input can be mathematically expressed as

$$\begin{bmatrix} \mathbf{y} \\ f_* \end{bmatrix} \sim N\left(0, \begin{bmatrix} K(\mathbf{X}, \mathbf{X}) + \sigma_n^2 \mathbf{I} & k(\mathbf{X}, \mathbf{x}_*) \\ k(\mathbf{x}_*, \mathbf{X}) & k(\mathbf{x}_*, \mathbf{x}_*) \end{bmatrix}\right) \quad (12)$$

$k(\mathbf{X}, \mathbf{x}_*) = [k(\mathbf{x}_*, \mathbf{X}_1)k(\mathbf{x}_*, \mathbf{X}_2)k(\mathbf{x}_*, \mathbf{X}_3)\dots k(\mathbf{x}_*, \mathbf{X}_n)]^T$  is the covariance between test and training data points in the form of a column vector and denoted by  $\mathbf{k}_{**}$  and,  $k(\mathbf{x}_*, \mathbf{x}_*)$  is the autocovariance function of the testing data points. It should be noted that the noise variance  $\sigma_n^2$  has been added to the diagonal of covariance of  $\mu$  in order to get the covariance for  $\mathbf{y}$ . The estimated output as per joint Gaussian distribution is given by

$$\bar{f}_* = \mathbf{k}_*^T (\mathbf{K} + \sigma_n^2 \mathbf{I})^{-1} \mathbf{y} \quad (13)$$

$$\text{Var}(f_*) = k(\mathbf{x}_*, \mathbf{x}_*) - \mathbf{k}_*^T (\mathbf{K} + \sigma_n^2 \mathbf{I})^{-1} \mathbf{k}_* \quad (14)$$

The obtained mean  $\bar{f}_*$  and variance  $\text{Var}(f_*)$  are the estimated values and its associated variance, respectively. It should be noted that the calculated mean  $\bar{f}_*$  is the continuous merger of the output  $\mathbf{y}$  while posterior variance  $\text{Var}(f_*)$  is a function of  $\mathbf{k}_*$ , hence it will be inversely proportional to the distance between test and training data points. Using the SCADA data of Table 1 power curve is estimated (in MATLAB) using (13) and (14) and is shown in Fig. 4. Fig. 4 suggest that GP power curve is smooth and continuous and, able to estimate measured power curve accurately.

It is worth noting, the posterior computation in GP regression is a trivial matter but suffers from two issues. The first is that computation mean and variance of (13) and (14) requires a matrix inversion and thus has asymptotic complexity called cubic inversion  $O(n)^3$  where  $n$  is the number of data points. The computational cost increases rapidly with  $n$  that it becomes challenging to fit GP models. The last issue is that if  $n$  becomes large, then the computation of the  $n \times n$  matrix becomes problematic and leads to GP model inaccuracy. These two limitations make GP less attractive. However, techniques based on state-space [36, 37] proposed to solve this issue but these methods require high processing power. Hence striking a balance between the data points is the key to constructing an effective GP algorithm for fault detection.

## 5.2 Power curve model based on SVM

The SVM is a non-parametric, machine learning technique which follows the principle of structural risk minimisation and widely used in solving a problem related to classification (called support vector classification) and regression (called support vector regression (SVR)). SVM based on statistical learning theory and [38, 39] provide a detailed explanation of SVM. A so-called dual SVR algorithm used where the inner product of predictors replaced by its corresponding element from the Gram matrix to build a power curve model. The Gram matrix is an  $n \times n$  matrix which contains elements  $g_{ij} = G(\mathbf{x}_i, \mathbf{x}_j)$ , where  $\mathbf{x}_i, \mathbf{x}_j$  are the training SCADA data points. Using non-negative multipliers, the Lagrangian function of the primal function constructed. This Lagrange dual formulation complements the non-linear system and hence taken into this study. A non-linear SVR calculates the optimal function  $f(\mathbf{x})$  in the transformed predictor space where the SVR search for the coefficient that minimises the Lagrangian function using the dual formula [38]

$$\begin{aligned} L(\alpha) = & 0.5 \sum_{i=1}^N \sum_{j=1}^N (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) G(\mathbf{x}_i, \mathbf{x}_j) \\ & + \epsilon \sum_{i=1}^N (\alpha_i + \alpha_i^*) - \sum_{i=1}^N y_i (\alpha_i - \alpha_i^*) \end{aligned} \quad (15)$$

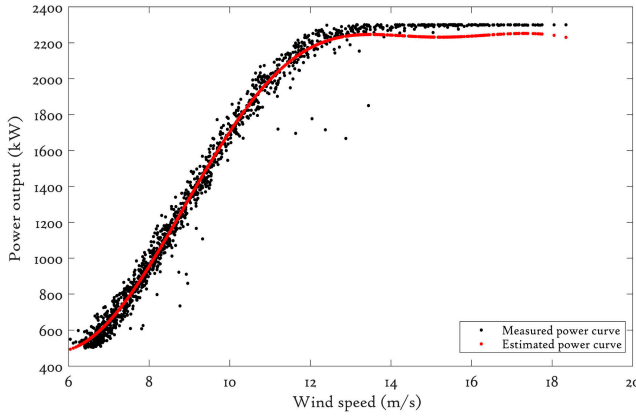


Fig. 5 SVR-based power curve

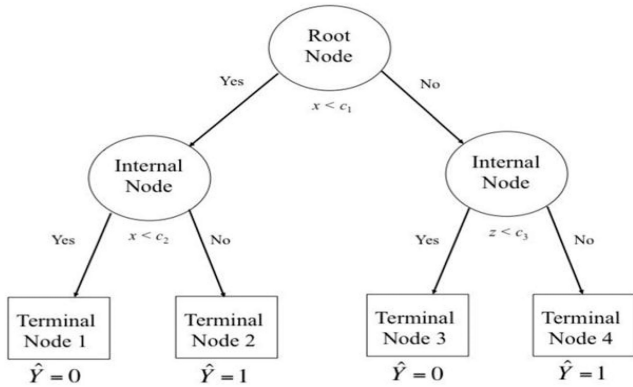


Fig. 6 Example tree diagram from CART analysis, [41]

Under the following constraints

$$\sum_{n=1}^N (\alpha_n - \alpha_n^*) = 0; \quad \forall n : 0 \leq \alpha_n \leq C; \quad \forall n : 0 \leq \alpha_n^* \leq C.$$

The function  $f(x)$  used to construct the SVR model for WT power curve and mathematically written as

$$f(x) = \sum_{n=1}^N (\alpha_n - \alpha_n^*) G(x_n, x) + b \quad (16)$$

This specific SVR called  $\varepsilon$ -SVR due to its scarcity representation capability [40]. The  $\varepsilon$ -insensitive loss function is used to build the objective function of the  $\varepsilon$ -SVR.

The Karush–Kuhn–Tucker (KKT) conditions play an important role in dealing with constrained optimisation and using KKT conditions, [38] of the quadratic programming in which only a certain number of the coefficients ( $\alpha_n - \alpha_n^*$ ) will assume non-zero values. The datasets with non-zero coefficients, having approximation errors equal to or larger than  $\varepsilon$ , are referred to as support vectors. Other samples are deemed to be  $\varepsilon$ -insensitive are not support vectors and play no role in the model. Generally, the larger, the fewer the number of support vectors and the sparser the representation of the solutions. The KKT complementarity conditions are optimisation constraints required to obtain optimal solutions and for non-linear SVM regression these conditions are

$$\begin{aligned} \forall n : \alpha_n(\varepsilon + \xi_n - y_n + f(x_n)) &= 0; \\ \forall n : \alpha_n^*(\varepsilon + \xi_n^* + y_n - f(x_n)) &= 0; \\ \forall n : \xi_n(C - \alpha_n) &= 0; \\ \forall n : \xi_n^*(C - \alpha_n^*) &= 0; \end{aligned}$$

The trade-off between the model complexity (flatness) and the degree to which larger deviations tolerated in the optimisation

formulation is indicated by [38], whereas controls the width of the  $\varepsilon$ -insensitive zone and affects the number of support vectors.

The Gaussian kernel is also popularly known by radial basis function kernel because it makes computation faster and involves computations in higher dimensional space. In this research, the Gaussian kernel is used model SVR-based power curve and mathematically expressed as

$$k(x, x') = \exp(-\gamma \|x - x'\|^2) \quad (17)$$

where  $\gamma$  is the kernel scale for given points  $x$  and  $x'$ .

The training and validation SCADA datasets of Table 1 were used to train and test the SVR power curve model, and the result is shown in Fig. 5. The SVR-based power curve is continuous and accurately predicts the measured power curve. However, at above-rated wind speed, the SVR power curve accuracy deteriorates. The SVR model is optimised automatically to give the values of  $\gamma$ ,  $C$ , and  $\varepsilon$ . The optimised values are constant and independent of wind speed. In Fig. 5, the optimised values of these parameters are constant entire range of wind speed. At above-rated wind speed, the data points are relatively less dense but the calculated values of  $C$ ,  $\varepsilon$ , and  $\gamma$  are same irrespective of data quantities and are not optimised to reflect the availability of data in this wind speed region. As a result, the SVR algorithm accuracy suffers as shown in Fig. 5. Manually tuning these parameters can improve SVR model accuracy but that approach is time consuming, complex, and these manually calculated values will not be the same for all types of problem and varies with the types of datasets. The accuracy of the SVR model is affected by the quantity and quality of the data due to the cubic inversion issue, just like GP models (see Section 5.1).

### 5.3 Power curve model based on RF

The RF is a non-parametric, ensemble learning approach that uses a large number of individual, unpruned decision trees and merges together to get a more accurate and stable estimation. The systematic and detailed explanation of RF can be found in [23, 30]. Here, a brief description of RF would be provided. The RF is a collection of collection of Classification and Regression Trees (CARTs) in which CART splits the input space recursively, according to a predefined split criterion, to small rectangular regions and then fits a simple model, commonly a constant value, in each one of them, and this can be demonstrated by tree diagram, see Fig. 6.

In RF ensemble learning approach, a group of ‘weak learners’ used together to form a ‘strong learner’ to improve the performance. RF uses decision tree in which each tree is constructed from a bootstrap sample from the original dataset with an objective to increase diversity between members of the ensemble by restricting classifiers to work on different random subsets of the full feature space [42]. In the RF approach,  $k$  bootstrap sampled randomly and then a regression tree fit on each sample. After that, average values of  $k$  regression tree are taken in order to make an estimation.

In this paper, RF algorithm as per [43], used to estimate power curve of a WT where bootstrap samples are generated similar to bagging algorithm. The bootstrap aggregation produces non-correlated trees through different training samples, which gives immunity to noise. Nevertheless, instead of using all training data to fit the tree, only random predictor variables are used at each split. Splitting the decision improve the RF accuracy such that the reduction in the residual sum of squares is maximised [44]. Here, SCADA datasets divided into a training and validation datasets (Table 1) at the first node, all variables and values are considered, and after the split, further variable and the splitting condition are selected and this repeated again. While doing this, the same variable can be selected consecutively and hence this splitting technique called recursive binary splitting. To find the optimal values, randomly selected predictor ( $k_{tr}$ ) can vary. It should be noted that RF tried to search for the best split among the  $k_{tr}$  selected features and this selection is uniform. The randomly selected predictor ( $k_{tr}$ ) is same for all prediction trees and it is



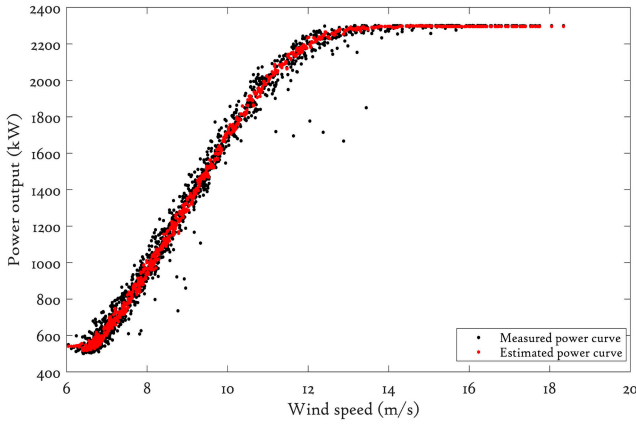


Fig. 7 RF-based power curve

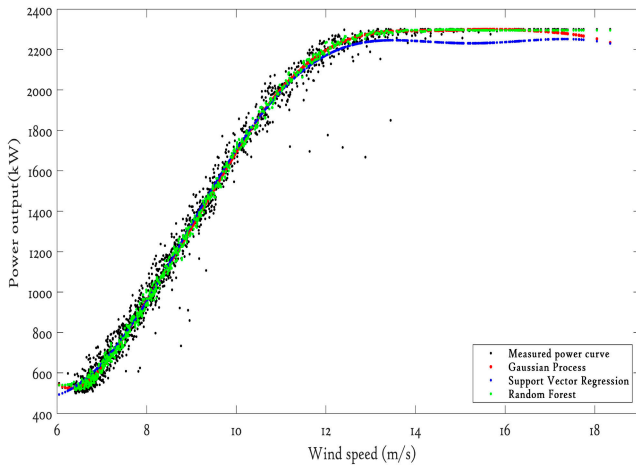


Fig. 8 Comparative analysis of non-parametric models

recommended to be the square root or one-third of the features number  $k$  as:  $k_{tr} = \sqrt{k}$  or,  $k_{tr} = k/3$ . After that, the RF algorithm is similar to the CART where by minimising the cost function, best split is obtained and repeat the procedure until full development of all trees. RF models are very good at capturing the non-linear relationship between features and the target and in minimising the overfitting issue. The out-of-bag error and the measure of variable importance are the two main important properties of RF models, see [23]. The spatial 10 min average training SCADA datasets (Table 1) are used to estimate the power curve based on RF and is shown in Fig. 7. Fig. 7 suggests that the RF-based power curve is accurately following measured power curve variance but neither it is continuous or smooth. It is worth to note that, the RF power curve is predictive model, not a descriptive model and hence it does not give a description of the relationship among the predictors. Moreover, confusing data makes RF inaccurate and confusing; hence, it is desirable to select appropriate and error free predictors that affect the target variable.

## 6 Performance comparison

The advanced non-parametric models discussed in Section 5 are used for comparative analysis in order to find out which of the proposed advanced non-parametric models is appropriate to represent power curves accurately and what are advantageous and disadvantageous of proposed advanced non-parametric models. The residual analysis, performance error metrics are used for comparative studies of proposed models. The advanced non-parametric models result presented in Section 5 were compared in Fig. 8 together with the measured power curve. The GP-based power curve is relatively more accurate and has a continuous and smooth fitting which closely following the expected variance at all wind speed range while RF-based power curve is neither continuous nor smooth because it is built on CART theory [23, 30] but closely matching the measured power curve. The performance

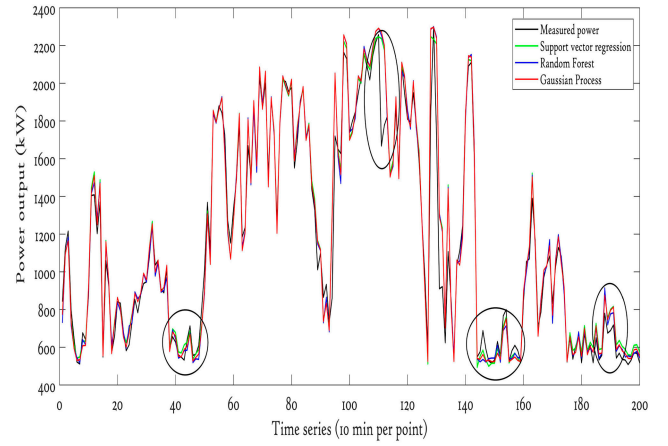


Fig. 9 Comparative analysis of non-parametric models in terms of time series

Table 2 Evolution of non-parametric models using performance metrics

MODELS	RMSE	$R^2$	MAE	RANK
GP	62.690	0.990	39.806	1
SVM	65.086	0.989	46.226	2
RF	65.444	0.989	42.568	3

of the power curve based on SVM deteriorates after rated wind speed because of the unavailability of reasonable numbers of SCADA data points as shown in Fig. 8. Fig. 9 shows the estimated power values in the time series of proposed advanced non-parametric models.

### 6.1 Using performance error metrics

There are several statistical performance metrics that can be used to measure performance of the estimated power curves such as the RMSE, normalised mean absolute percentage error (NMAPE), symmetric MAPE, the MAE, and the coefficient of determination ( $R^2$ ), [21]. In this paper, we use three goodness-of-fit indicators, namely the MAE, the RMSE, and the  $R^2$  to evaluate the goodness-of-fit statistics of the advanced non-parametric power curve models which are mathematically expressed as

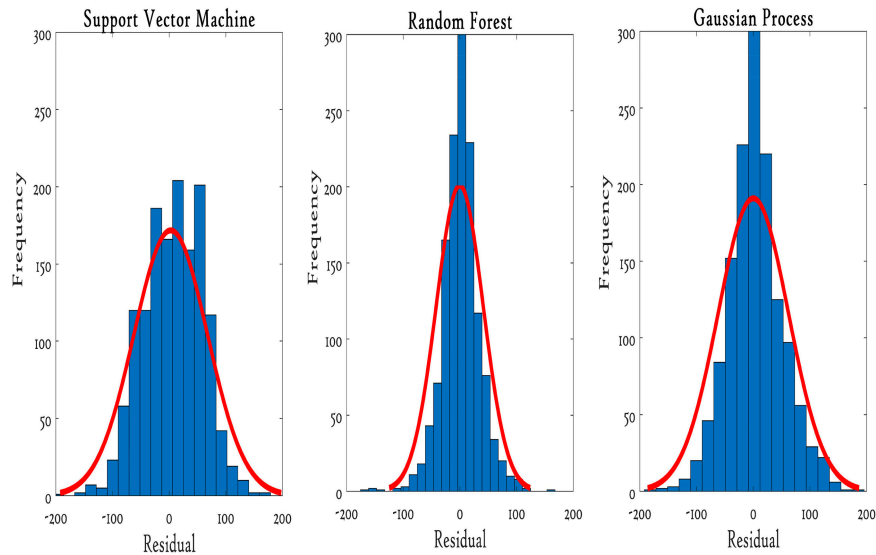
$$MAE = \frac{\sum_{i=1}^n \text{abs}(X'_i - X_i)}{n} \quad (18)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (X'_i - X_i)^2}{n}} \quad (19)$$

$$R^2 = 1 - \frac{SSE}{TSS} \quad (20)$$

where  $X'$  are the predicated values for  $n$  different predictions, and  $X$  are the measured values. SSE is the sum of squared errors and TSS is the total sum of squares.

The RMSE is the square root of the mean of the squared difference between the measured and predicted values of power. The MAE is the mean of the absolute values of the differences between the measured and predicted values of power. The coefficient of determination ( $R^2$ ), describes how close the data are to the fitted non-parametric models and is calculated as the square of the correlation between estimated output and measured values using (20). The RMSE, MAE, and  $R^2$  values of the discussed algorithms have been tabulated in Table 2. The lower values of RMSE or MAE suggest better estimation of power curve while a higher value of  $R^2$  indicates a better coincidence of observed and estimated results. Based on these three performance metrics, GP algorithm rank 1 and gives the most accurate power curve while



**Fig. 10** Comparative studies of histogram fitting of non-parametric models

RF-based power curve ranks 3 and relatively gives inaccurate power curve.

### 6.2 Using models residuals analysis

The GP, RF, and SVM models are data-driven, non-linear techniques whose residual distribution needs to be analysed. Residuals are the difference between the measured values and estimated values and can be useful in identifying the deviation between the data and the regression model, which widely used to measure the variability in the response variable. The frequency distribution of the calculated residuals of advanced non-parametric models is shown in Fig. 10 together with a fitted Gaussian distribution and found that distribution of GP residuals is close to Gaussian as compare to other non-parametric models.

### 6.3 Using models uncertainty

The WT power curve vastly used by wind industries to identify the failures that cause the turbine to underperform and do preventive maintenance in order to prevent downtime and catastrophic stage. Uncertainty analysis is significant for constructing robust fault detection algorithm based on WT power curve. The GP estimate confidence intervals (CIs) along with the mean function that makes uncertainty analysis is simple and straightforward. The GP CIs is calculated by the standard deviation of the variance of the estimated function (14) using

$$\text{CIs} = \bar{f}_* \pm 2\sqrt{\text{Var}[f_*]} \quad (21)$$

Equation (21) conclude that GP CIs is the pointwise mean plus and minus two times the standard deviation for given input data points and it corresponds to 95% confidence region which defines the significance level of 0.05 for the calculated prior and posterior, respectively. The  $\pm$  of the CIs used to represent the upper and lower CIs of the GP model, respectively. The data points lie outside these CIs likely to suggest turbine underperformance and hence vital for fault detection algorithms based on the GP model. However, in the case of RF and SVM models, uncertainty analysis is complicated due to the extra mathematical challenges associated with it. Some authors proposed techniques to calculate the CIs for RF [45] and SVM [46] models but that requires high power processing, and computational cost and consequently makes the O&M cost higher.

## 7 Conclusions

The performance of WT can be described by power curve, and therefore, accurate modelling of the power curve can be useful for assessment and monitoring of the turbine's performance, energy

warranty formulations, and power forecasting, as well as sizing the storage capacity for wind power integration. Advanced non-parametric models (GP, SVM, and RF) for estimating WT power curves based on SCADA datasets obtained from operational WTs are presented in this paper. GP and SVM are kernel methods while RF is a regression tree method inspired by the CART principle. The computational results have demonstrated that the GP has the highest fitting accuracy, and can reflect the dynamic properties of a power curve whose distribution function is close to the Gaussian distribution. The accuracy of each method is evaluated using performance error metrics (RMSE, MAE, and  $R^2$ ) and based on this the GP model ranks 1 (see Table 2). The accuracy of SVM model suffers at above rated wind speed because of a lower density of SCADA data points in that region, whilst with same SCADA data points, the GP model performs better across entire wind speed range (including above rated wind speed). The uncertainty analysis in GP model is undertaken using CIs which is simple and straightforward and which makes fault detection using the GP algorithm robust and free from further complex mathematical calculations unlike RF and SVM methods where the uncertainty analysis leads to extra mathematical computations and cost. This makes both RF and SVM are less attractive for WT condition monitoring from an economic as well as a technical point of view.

Future work will use these proposed models to construct WT fault detection algorithms and so identify which model is most useful for identifying failures without the generation of unwanted false positives.

## 8 Acknowledgments

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

## 9 References

- [1] Bently, D., Hatch, C.: 'Fundamentals of rotating machinery diagnostics' (ASME publishers, New York, 2003)
- [2] Qiao, W., Lu, D.: 'A survey on wind turbine condition monitoring and fault diagnosis – part II: signals and signal processing methods', *IEEE Trans. Ind. Electron.*, 2015, **62**, (10), pp. 6546–6557
- [3] Márquez, F., Tobias, A., Pérez, J., *et al.*: 'Condition monitoring of wind turbines: techniques and methods', *Renew. Energy*, 2012, **46**, pp. 169–178. doi: 10.1016/j.renene.2012.03.003
- [4] Dai, J., Yang, W., Cao, J., *et al.*: 'Long ageing assessment of a wind turbine over time by interpreting wind farm SCADA data', *Renew. Energy*, 2018, **116**, (Part B), pp. 199–208. doi: 10.1016/j.renene.2017.03.097
- [5] Dao, P., Staszewski, W., Barszcz, T., *et al.*: 'Condition monitoring and fault detection in wind turbines based on cointegration analysis of SCADA data', *Renew. Energy*, 2018, **16**, (Part B), pp. 107–122. doi: 10.1016/j.renene.2017.06.089
- [6] Yang, W., Lang, Z., Tian, W.: 'Condition monitoring and damage location of wind turbine blades by frequency response transmissibility analysis. *IEEE*

- Trans. Ind. Electron.*, 2015, **62**, (10), pp. 6558–6564. doi: 10.1109/TIE.2015.2418738
- [7] Tang, J., Soua, S., Mares, C., *et al.*: ‘An experimental study of acoustic emission methodology for in service condition monitoring of wind turbine blades’, *Renew. Energy*, 2016, **99**, pp. 170–179. doi: 10.1016/j.renene.2016.06.048
- [8] Ifigeneia, A.: ‘Accounting for nonstationarity in the condition monitoring of wind turbine gearboxes’, PhD thesis, University of Sheffield, 2013
- [9] Lei, Y., Lin, J., He, Z., *et al.*: ‘A review on empirical mode decomposition in fault diagnosis of rotating machinery’, *Mech. Syst. Signal Process.*, 2013, **35**, (1), pp. 108–126. doi: 10.1016/j.ymssp.2012.09.015
- [10] Yang, D., Li, H., Hu, Y., *et al.*: ‘Vibration condition monitoring system for wind turbine bearings based on noise suppression with multipoint data fusion’, *Renew. Energy*, 2016, **92**, pp. 104–116. doi: 10.1016/j.renene.2016.01.099
- [11] Castellani, F., Astolfi, D., Sdringola, P., *et al.*: ‘Analyzing wind turbine directional behavior: SCADA data mining techniques for efficiency and power assessment’, *Appl. Energy*, 2015, **185**, (Part 2), pp. 1076–1086. doi: 10.1016/j.apenergy.2015.12.049
- [12] Khalfallah, M., Koliub, A.: ‘Suggestions for improving wind turbines power curves’, *Desalination*, 2007, **209**, (3), pp. 221–229. doi: 10.1016/j.desal.2007.04.031
- [13] Giorsetto, P., Utsurogi, K.: ‘Development of a new procedure for reliability modeling of wind turbine generators’, *IEEE Trans. Power Appar. Syst.*, 1983, **PAS-102**, (1), pp. 134–143. doi: 10.1109/TPAS.1983.318006
- [14] Chedid, R., Akiki, H., Rahman, S.: ‘A decision support technique for the design of hybrid solar-wind power systems’, *IEEE Trans. Energy Convers.*, 1998, **13**, (1), pp. 76–83
- [15] Kusiak, A., Zheng, H., Song, Z.: ‘On-line monitoring of power curves’, *Renew. Energy*, 2009, **34**, (6), pp. 1487–1493. doi: 10.1016/j.renene.2008.10.022
- [16] Lydia, M., Kumar, S., Selvakumar, A., *et al.*: ‘A comprehensive review on wind turbine power curve modeling techniques’, *Renew Sustain. Energy Rev.*, 2014, **30**, pp. 452–460. doi: 10.1016/j.rser.2013.10.030
- [17] Gill, S., Stephen, B., Galloway, S.: ‘Wind turbine condition assessment through power curve copula modeling’, *IEEE Trans. Sustain. Energy*, 2012, **3**, (1), pp. 94–101. doi: 10.1109/TSTE.2011.2167164
- [18] Shokrzadeh, S., Jozani, M., Bibeau, E.: ‘Wind turbine power curve modeling using advanced parametric and nonparametric methods’, *IEEE Trans. Sustain. Energy*, 2014, **5**, (4), pp. 1262–1269. doi: 10.1109/TSTE.2014.2345059
- [19] Díaz, S., Cartaa, J., Matias, J.: ‘Performance assessment of five MCP models proposed for the estimation of long-term wind turbine power outputs at a target site using three machine learning techniques’, *Appl. Energy*, 2018, **209**, pp. 455–477. doi: 10.1016/j.apenergy.2017.11.007
- [20] Jung, S., Kwon, S.: ‘Weighted error functions in artificial neural networks for improved wind energy potential estimation’, *Appl. Energy*, 2013, **111**, pp. 778–790. doi: 10.1016/j.apenergy.2013.05.060
- [21] Marvuglia, A., Messineo, A.: ‘Monitoring of wind farms’ power curves using machine learning techniques’, *Appl. Energy*, 2012, **98**, pp. 574–583. doi: 10.1016/j.apenergy.2012.04.037
- [22] Pandit, R., Infield, D., Carroll, J.: ‘Incorporating air density into a Gaussian process wind turbine power curve model for improving fitting accuracy’, *Wind Energy*, 2019, **22**, pp. 302–315. <https://doi.org/10.1002/we.2285>
- [23] Lahouar, A., Slama, J.: ‘Hour-ahead wind power forecast based on random forests’, *Renew. Energy*, 2017, **109**, pp. 529–541. doi: 10.1016/j.renene.2017.03.064
- [24] Zhang, Y., Wang, J.: ‘K-nearest neighbors and a kernel density estimator for GEFCom2014 probabilistic wind power forecasting’, *Int. J. Forecast.*, 2016, **32**, (3), pp. 1074–1080. doi: 10.1016/j.ijforecast.2015.11.006
- [25] Meer, D., Widén, J., Munkhammar, J.: ‘Review on probabilistic forecasting of photovoltaic power production and electricity consumption’, *Renew. Sustain. Energy Rev.*, 2018, **81**, (Part 1), pp. 1484–1512. doi: 10.1016/j.rser.2017.05.212
- [26] Hong, T., Fan, S.: ‘Probabilistic electric load forecasting: a tutorial review’, *Int. J. Forecast.*, 2016, **32**, (3), pp. 914–938. doi: 10.1016/j.ijforecast.2015.11.011
- [27] Cortes, C., Vapnik, V.: ‘Support-vector networks’, *Mach. Learn.*, 1995, **20**, (3), pp. 273–297. doi: 10.1007/BF00994018
- [28] Baccarini, L., Silva, V., Menezes, B., *et al.*: ‘SVM practical industrial application for mechanical faults diagnostic’, *Expert Syst. Appl.*, 2011, **38**, (6), pp. 6980–6984. doi: 10.1016/j.eswa.2010.12.017
- [29] Lin, Y., Kruger, U., Zhang, J., *et al.*: ‘Seasonal analysis and prediction of wind energy using random forests and arx model structures’, *IEEE Trans. Control Syst. Technol.*, 2015, **23**, (5), pp. 1994–2002. doi: 10.1109/TCST.2015.2389031
- [30] Freund, Y.: ‘Boosting a weak learning algorithm by majority’, *Inf. Comput.*, 1995, **121**, (2), pp. 256–285. doi: 10.1006/inco.1995.1136
- [31] Paiva, L., Rodrigues, C., Palma, J.: ‘Determining wind turbine power curves based on operating conditions’, *Wind Energy*, 2014, **17**, pp. 1563–1575. doi: 10.1002/we.1651
- [32] IEC 61400-12-1: ‘Wind turbines – part 12-1: power performance measurements of electricity producing wind turbines’, 2006
- [33] Schlechtingen, M., Santos, I.: ‘Comparative analysis of neural network and regression based condition monitoring approaches for wind turbine fault detection’, *Mech. Syst. Signal Process.*, 2011, **25**, (5), pp. 1849–1875. doi: 10.1016/j.ymssp.2010.12.007
- [34] Rasmussen, C., Williams, C.: ‘Gaussian processes for machine learning’ (The MIT Press, Cambridge, MA, USA, 2006), ISBN 026218253X
- [35] Chatfield, C., Collins, A.: ‘Introduction to multivariate analysis’ (Springer, New York, NY, USA, 1980)
- [36] Hartikainen, J., Särkkä, S.: ‘Kalman filtering and smoothing solutions to temporal Gaussian process regression models’. IEEE Int. Workshop on Machine Learning for Signal Processing, Kittila, 2010, pp. 379–384. doi: 10.1109/MLSP.2010.5589113
- [37] Sarkka, S., Solin, A., Hartikainen, J.: ‘Spatiotemporal learning via infinite-dimensional Bayesian filtering and smoothing’, *IEEE Signal Process. Mag.*, 1992, **30**, (4), pp. 51–61. doi: 10.1109/MSP.2013.2246292
- [38] Boser, B.E., Guyon, I.M., Vapnik, V.: ‘A training algorithm for optimal margin classifiers’. Proc. fifth annual workshop on Computational Learning Theory – COLT, 1992, p. 144. ISBN 089791497X. doi: 10.1145/130385.130401
- [39] Fan, R.E., Chen, P., Lin, C.: ‘A study on SMO-type decomposition methods for support vector machines’, *IEEE Trans. Neural Netw.*, 2006, **17**, (4), pp. 893–908. doi: 10.1109/TNN.2006.875973
- [40] Cherkassky, V., Ma, Y.: ‘Practical selection of SVM parameters and noise estimation for SVM regression’, *Neural Netw.*, 2004, **17**, (1), pp. 113–126. doi: 10.1016/S0893-6080(03)00169-2
- [41] Berk, R.: ‘Statistical learning from a regression perspective’ (Springer, New York, 2009)
- [42] Sammut, C., Webb, G.: ‘Encyclopedia of machine learning and data mining’ (Springer, Boston, MA, 2017)
- [43] Breiman, L.: ‘Random forests’, *Mach. Learn.*, 2001, **45**, (1), pp. 5–32
- [44] Breiman, L., Friedman, J., Olshen, R., *et al.*: ‘Classification and regression trees’ (Chapman and Hall, New York, 1984)
- [45] Wager, S., Hastie, T., Efron, B.: ‘Confidence intervals for random forests: the jackknife and the infinitesimal jackknife’, *J. Mach. Learn. Res.*, 2014, **15**, pp. 625–1651. Available at <http://www.jmlr.org/papers/volume15/wager14a/wager14a.pdf>, accessed on 20 July 2018
- [46] Jiang, B., Zhang, X., Cai, T.: ‘Estimating the confidence interval for prediction errors of support vector machine classifiers’, *J. Mach. Learn. Res.*, 2008, **9**, pp. 521–540. Available at <http://www.jmlr.org/papers/volume9/jiang08a/jiang08a.pdf>, accessed on 20 July 2018