

Comparative analysis of binning and support vector regression for wind turbine rotor speed based power curve use in condition monitoring

Ravi Pandit

Department of Electronic and Electrical Engineering
University of Strathclyde
Glasgow, United Kingdom
ravi.pandit@strath.ac.uk

David Infield

Department of Electronic and Electrical Engineering
University of Strathclyde
Glasgow, United Kingdom
david.infield@strath.ac.uk

Abstract—Unscheduled maintenance consumes a lot of time and effort and hence reduces the overall cost-effectiveness of wind turbines. Supervisory control and data acquisition (SCADA) based condition monitoring is a cost-effective approach to carry out diagnosis and prognosis of faults and to provide performance assessment of a wind turbine. The rotor speed based power curve, which describes the nonlinear relationship between wind turbine rotor speed and power output, is useful for performance appraisal of a wind turbine though limited work on this area has been undertaken to date. Support Vector Machine (SVM) is a data-driven, nonparametric approach used for both classification and regression problems developed initially from statistical learning theory (SLT) by Vapnik. SVM is useful in forecasting and prediction applications.

This paper deals with the application of support vector regression to estimate the rotor speed based power curve of a wind turbine and its usefulness in identifying potential faults. It is compared with a conventional approach based on a binned rotor speed power curve to identify operational anomalies. The comparative studies summarise the advantages and disadvantages of these techniques. SCADA data obtained from a healthy operational wind turbine is used to train and validate these methods.

Keywords—*anomaly detection, condition monitoring, support vector machine, nonparametric models, wind turbine.*

I. INTRODUCTION

Wind energy is a source of clean energy that shows a steady growth worldwide, for example, as per GWEC 2016 report [1], the total world installed capacity by the wind industry has changed from 1.29 GW in 1995 to 370 GW by the end of 2015. Such impressive growth increases the need to minimize the Cost of Energy (COE) of wind turbines. Operation and maintenance (O&M) costs represent a significant portion of the total annual

costs of a wind turbine and are higher for offshore wind turbines as compared to onshore wind turbines [2]. There is potential to improve the profitability of offshore wind by developing condition based maintenance.

SCADA based condition monitoring is a cost-effective approach that helps to ensure the security, reliability, and stability of wind turbines while in operation [3]. Continuous monitoring of the condition of wind turbines is a most effective way to reduce the O&M cost and prevent the fault from reaching a catastrophic stage through early detection [4,5]. Several techniques have been proposed to achieve this, and are described as follows.

Power curves are nonlinear relationships that define the relationship between power output and hub wind speeds and are widely used to analyze the operational health of a wind turbine using SCADA data sets. The International Standard IEC 61400-12-1, [6] outlines the guidelines for modeling an effective power curve using the ‘method of bins’ which is helpful for condition monitoring and measuring power performance characteristics though it is slow to respond and is not entirely accurate [6,7]. It is not always possible to analyze the performance of a wind turbine using a power curve. Other performance indicators such as rotor speed, blade pitch angle are also vital for active condition monitoring [8]. For example, Singh [9], used power and rotor speed curves to identify abnormal operation. The comparative studies of these two curves concluded that the rotor speed curve identified a performance change due to downtime event, but when reviewing its corresponding power curve, this performance change cannot be identified. Thus, it is desirable to consider other key performance indicators for active wind turbine condition monitoring [8,9].

Parametric and nonparametric approaches are widely used for data-driven models for wind turbine condition monitoring

[10,11]. Nonparametric models do not impose any pre-specified condition unlike parametric approaches and hence are best suited for describing the operational behavior of wind turbines. Artificial neural networks (ANNs) [12], Gaussian Process (GP) [13], kNN [14], support vector machine (SVM) [15] and Random forest [16] are the widely used nonparametric approaches for wind turbine condition monitoring purposes. Most of the nonparametric models used the power curve as a key performance indicator to assess the power performance of a wind turbine, see [11,17,18].

Support vector machine (SVM) is powerful data-driven, machine learning approach popularly used for classification and regression related problems first discovered by Vladimir Vapnik and his colleagues in 1992 [19]. SVM regression entirely relies on a kernel function; it is a nonparametric approach widely used in the wind turbine condition monitoring [20,21].

Studies to date have mostly focused on the power curve to assess turbine performance though it cannot alone describe the operational behavior of turbines because internal factors profoundly influence the turbine performance. Hence other key performance indicators are necessary for valid power performance assessments of a wind turbine. In this study, rotor speed curves that reveal the nonlinear relationship between rotor speed and power output are investigated.

In this paper, a novel support vector regression (SVR) approach for wind turbine rotor speed curve modeling and its application in anomaly detection for condition monitoring is proposed. The binning method is the benchmark for wind industries, but its application is limited to power curve only. It can be extended to calculate the rotor speed curve. This is then compared with SVR based rotor speed curve, and the advantages and limitations of these techniques are assessed.

This paper is outlined as follows: Section 1 is the introduction. Section 2 describes the wind turbine rotor speed curve modeling and air density correction. Section 3 provides the SCADA dataset description, and the pre-processing techniques applied. Section 4 describes the methodologies of support vector regression and binning. Section 5 described the comparative analysis of SVR and binning methods and Section 6 concludes the paper.

II. WIND TURBINE ROTOR CURVES

Rotor speed is one of the key performance indicators (KPIs) used to describe the critical relationship between power production and hub height wind speed. The nonlinear relationship between power output and rotor speed is called rotor speed power curve and is shown in figure 1. Another, rotor speed curve is defined as the relationship between rotor speed and wind speed (figure 2) where rotor speed increases with the increase of wind speed above cut-in wind speed [9]. Both rotor curves are useful for performance assessment and fault diagnosis purposes [8].

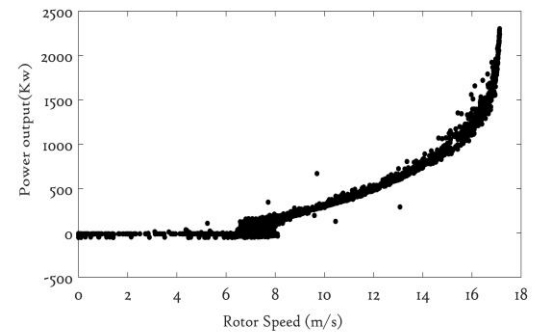


Fig. 1: Measured rotor power curve

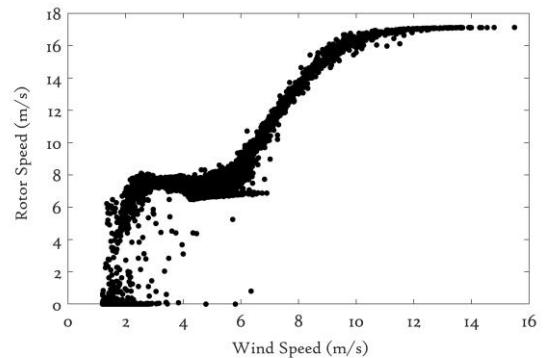


Fig. 2: Measured rotor speed curve

III. DATA DESCRIPTIONS

The SCADA data used in this study are from pitch controlled variable speed wind turbines located in Scotland, UK that comes with 10 min sampling interval resolution and contains more than 100 different signals ranging from timestamp, calculated values, set point, measurements of temperature, current, voltage, wind speed, power output, wind direction and so on. A full year of data starting from January 2012 to December 2012 is available. The SCADA data records the significant information such as operational and load history without any extra cost. Hence SCADA based condition monitoring is considered a cost-effective approach. However, data obtained from the SCADA system is not error free due to sensor failures and data malfunctions. To make sure that the SVR model accuracy is not affected by these errors initial data screening and filtering process needed. The criterion described in [22] including timestamp mismatches, out of range values, negative power values, and turbine power curtailments are applied to remove confusing and potentially misleading data.

The data used in this study is from a 2.3 MW Siemens turbine and records 4464 data points beginning with time stamp “1/7/2012 00:00 AM” and ending at time stamp “31/7/2012 23:50 PM”. Rotor scatters plots based on these data are shown in figures 1 and 2. These measured data points became 626 data points after pre-processing (table 1) and were used to model the rotor curves using SVR and binning methods in the next sections. The pre-processed and air density corrected rotor curves are shown in figure 4 and 5.

Start timestamp	End timestamp	Measured data	Filtered data
1/7/2012 00:00 AM	31/7/2012 23:50 PM	4464	626

Table 1: SCADA dataset description

Air density corrections: For accurate power curve modeling, IEC 61400-12-1 [6], outlines the guidelines for an individual wind turbine. Since SCADA data used in this study are from pitch regulated wind turbines, so as per IEC standard, air density correction should be applied using the following equations

$$\rho = 1.225 \left[\frac{288.15}{T} \right] \left[\frac{B}{1013.3} \right] \quad (1)$$

$$\text{and, } V_C = V_M \left[\frac{\rho}{1.225} \right]^{\frac{1}{3}} \quad (2)$$

where, V_C and V_M are the corrected and measured wind speed in m/sec and the corrected air density is calculated by equation (1) where B is atmospheric pressure in mbar and T the temperature in Kelvin. The air density is affected by various factors such as location, altitude and ambient temperature. In equation (2), B and T records 10-minute average values obtained from SCADA datasets of an operational wind turbine. The calculated value of ρ then being used in equation (2) to calculate the corrected wind speed (V_C). This corrected wind speed is used for constructing rotor curves in upcoming sections.

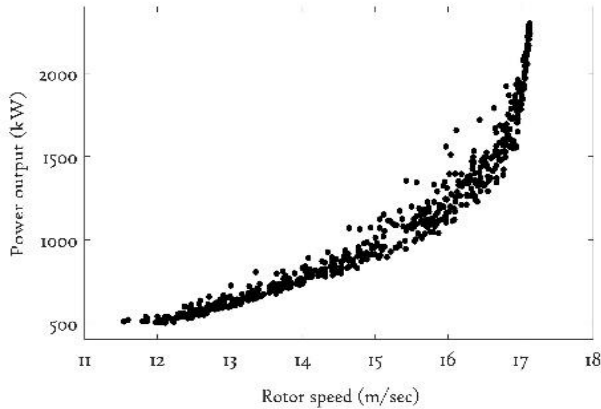


Fig. 3: Filtered rotor power curve

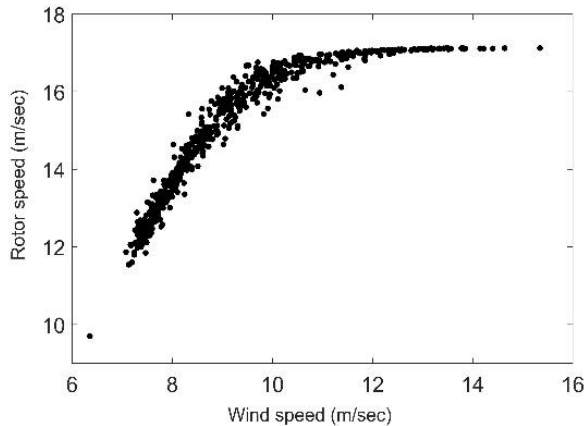


Fig. 4: Filtered rotor speed curve

IV. METHODOLOGIES

The two methods are proposed for rotor power curve estimation and are described as follows,

Support Vector Machine based Rotor Power Curve

A systematic and detailed explanation of Support Vector Machine (SVM) can be found in Ref. [19] and for more details see [15,20-21]. Because of SVM dependency on kernels, it is a nonparametric model and is useful in finding complex and nonlinear relationships between a high number of parameters. SVM is a particular class of algorithms widely used in classification and regression problems that are identified by the utilisation of kernels, absence of local minima, sparseness of the solution and capacity control obtained by acting on the margin, or on a number of support vectors [19]. Here we only provide a brief description of SVR technique for rotor curve modelling.

A Lagrangian function of the primal function calculated by inserting nonnegative multipliers α_n and α_n^* for calculating dual SVR formula (dual SVR formula perform faster than standard hence used in this study [23]) where the inner product of predictors $a_i' a_j$ is changed by its corresponding element from the Gram matrix t_{ij} for accurate support vector regression modeling. The Gram matrix is an n-by-n matrix that incorporates elements; $t_{ij} = T(a_i, a_j)$, where a_i, a_j are the pre-processed SCADA data points used for training, see section 3. Since SVM were first introduced with hard margin formulation [19], for which Lagrange dual formulation (because of the constraints) seems more natural for nonlinear systems [24]. Using this method, a nonlinear SVR calculates the optimal function $f(a)$ in the transformed predictor space where SVR search for the coefficient that minimize the Lagrangian function using dual formula [19,25],

$$L(\alpha) = 0.5 \sum_{i=1}^N \sum_{j=1}^N (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) T(a_i, a_j) + \varepsilon \sum_{i=1}^N (\alpha_i + \alpha_i^*) - \sum_{i=1}^N y_i (\alpha_i - \alpha_i^*) \quad (3)$$

under the following conditions,

$$\sum_{n=1}^N (\alpha_n - \alpha_n^*) = 0 ;$$

$$\forall n : 0 \leq \alpha_n \leq C ;$$

$$\forall n : 0 \leq \alpha_n^* \leq C.$$

The function $f(a)$ used to fit the SVR model for rotor curves is given by:

$$f(a) = \sum_{n=1}^N (\alpha_n - \alpha_n^*) T(x_n, x) + b \quad (4)$$

This specific SVR called ε -SVR and ε -insensitive loss function is used to build the objective function of the ε -SVR.

The Karush-Kuhn-Tucker (KKT) conditions are significant in dealing with constrained optimization and using KKT conditions [26] of quadratic programming, a limited number of the coefficients $(\alpha_n - \alpha_n^*)$ will assume nonzero values. The data points associated with nonzero coefficients having approximation errors equal to or larger than ε , are referred to as support vectors while the samples fall outside the ε -insensitive zone are not considered as support vectors and play no role in estimation. The KKT complementarity conditions are optimization constraints required to obtain optimal solutions and

for nonlinear SVM regression, these conditions are defined as follows,

$$\forall n : \alpha_n(\varepsilon + \xi_n - y_n + f(a_n)) = 0 ;$$

$$\forall n : \alpha_n(\varepsilon + \xi_n^* + y_n - f(a_n)) = 0 ;$$

$$\forall n : \xi_n(C - \alpha_n) = 0 ;$$

$$\forall n : \xi_n^*(C - \alpha_n^*) = 0 ;$$

The trade-off between the model complexity (flatness) and the degree to which larger deviations are tolerated in the optimization formulation is calculated by C [19] whereas ε manages the width of the ε -insensitive zone and is useful to train the data points and thus affects the number of support vectors. In summary, both C and ε -values affect SVR performance and, hence, it is important to calculate the optimal values for these parameters by using appropriate optimisation techniques. In addition, cross-validation folds (5 folds here) are used to prevent overfitting and an effective grid search [25,26]. The pre-processed data points (see section 3) were randomly shuffled and divided into training and testing data sets, for training and SVR model validation purposes. The Gaussian kernel or radial basis function (RBF) kernel is used to form qualified function and is mathematically defined as,

$$k(a, a') = \exp(-\gamma \|a - a'\|^2) \quad (5)$$

where γ is the kernel scale for a given points a and a' .

Using above described SVR theory, a rotor power curves is constructed and then compared with a measured rotor curve, see figure 5. Figure 5 suggests that the SVR based rotor power curve closely follows the measured curve. The SVR model suffers when dealing with massive numbers of data points due to cube inverse issue, see [24], so appropriately limiting the size of data sets is important for accurate SVR modeling.

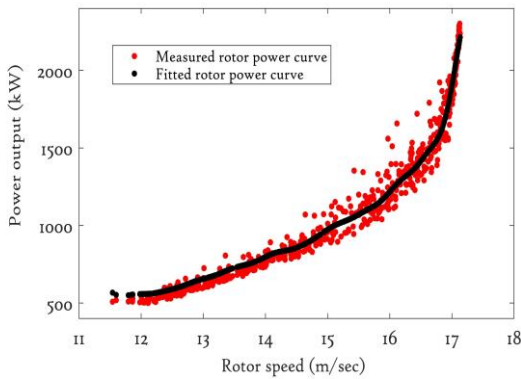


Fig 5: Estimated & measured rotor power curve

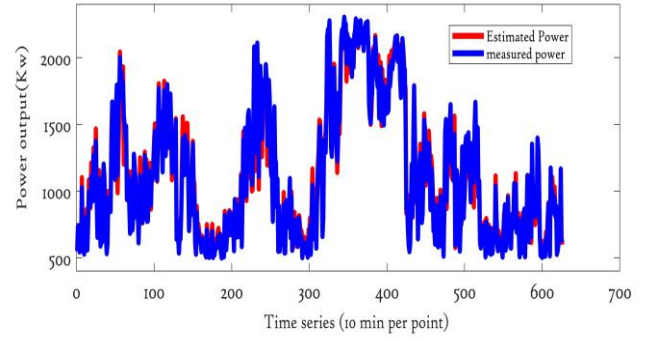


Fig 6: measured & estimated power in time series

SVR is non-parametric in nature, and hence residual analysis is essential. Residuals are defined as the difference between measured value and estimated values and are shown in figure 7. The frequency distribution of the residuals is shown in figure 8 together with a fitted Gaussian distribution; it is found that the distribution of residuals closely follows a Gaussian distribution.

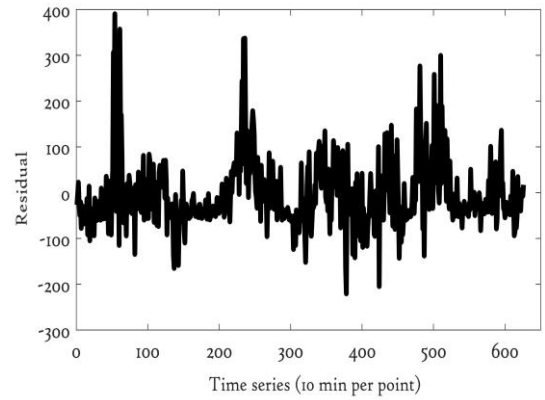


Fig. 7: SVR Residual

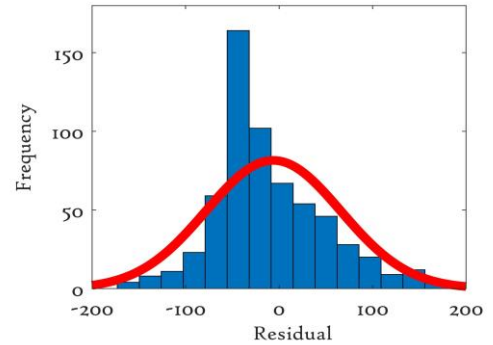


Fig. 8: Residual Histogram with distribution fit

Binned based Rotor Power Curve

The IEC 61400-12 proposed data reduction technique to calculate the power curve of a wind turbine is called 'binning.' The binning technique consists of grouping output power measurements into wind speed bins so that average output power obtained for each bin. The bin width is 0.5 m/sec. Adding more data points gives more certainty to the average value. It is worth

noting that wind speed is the most significant source of uncertainty in power performance assessment and methods such as meteorological Mast (IEC 61400-12-1 (2005), Ground-based LiDAR / SoDAR (IEC 61400-12-1), Nacelle-mounted LiDAR, Nacelle-mounted anemometry (IEC 61400-12-2 (2013)) and Existing nacelle-mounted anemometry (SCADA) are commonly used to measure wind speed for performance assessment [7]. Here, the binning method is applied to calculate rotor power curve using following equations,

$$Vr_i = \frac{1}{N_i} \sum_{j=1}^{N_i} Vr_{n,i,j}$$

$$P_i = \frac{1}{N_i} \sum_{j=1}^{N_i} P_{n,i,j}$$

where, Vr_i = normalised and averaged rotor speed in bin i ;

$Vr_{n,i,j}$ = normalised rotor speed of data sets j in bin i .

P_i = normalised and averaged power in bin i ;

$P_{n,i,j}$ = normalised power of data set j in bin i .

N_i = number of 10 min average data sets in bin i .

Using binning, a rotor power curve constructed together with error bars is compared with a measured rotor power curve in figure 9. Two standard deviations (i.e., 95% confidence intervals) of measured power values are used to calculate the error bars which reflect the uncertainty associated with each bin of the rotor power curve. However, the binning method accuracy is compromised by choosing the bin width of 0.5 m/sec because within each bin the measured power will depend strongly and non-linearly on wind speed; a wide bin would result in a systematic bias, and there is need in practice to get sufficient data points in each bin to be of statistical significance [26].

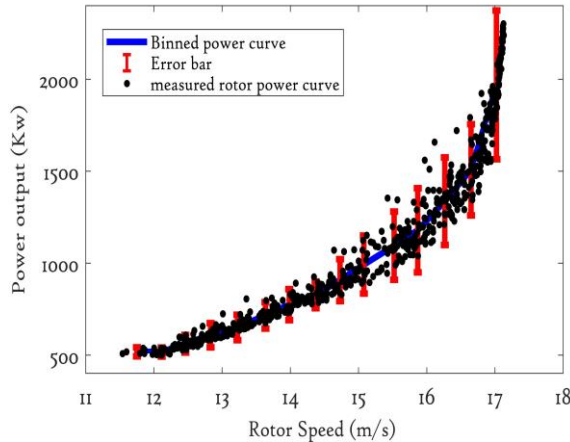


Fig. 9: Wind turbine rotor power curve with error bars

V. COMPARATIVE ANALYSIS

In this section, a comparative analysis of binned and Support Vector Regression (SVR) based rotor power curve is presented. The SVR based rotor power curve closely follows the binned

and measured rotor power curves, see figure 10. Though the accuracy of the SVR depends upon the quantity and quality of the data, as well as the appropriate method used. However, it should be noted that binning is not the most accurate approach and also it is too slow to respond to changes in conditions in time which are critical to early fault detection. These two limitations make binning approaches challenging in wind turbine condition monitoring. SVR is fast to respond due to its nonparametric nature. Also, the comparative analysis suggests that SVR is effective in rotor curve estimation, however, the accuracy of SVM regression depends upon the quantity and quality of the data, as well as the method used for fitting.

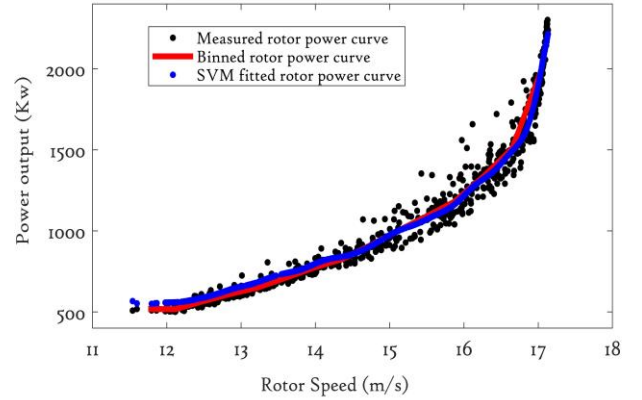


Fig 10: Comparative analysis of rotor power curve based on binned and SVR models

VI. CONCLUSION

This paper has proposed a support vector regression for estimating the rotor power curves of a wind turbine. Historical SCADA data sets obtained from operational wind turbines are used to train and test the models. The quality and quantity of data affect SVR model accuracy, related to cubic inversion issues. Nevertheless, with limited data sets, SVR gives better accuracy. The SVR based rotor power curve has been compared with the conventional approach based on a binned rotor curve together with individual bin probability distributions to identify operational anomalies. The comparative studies suggest that SVR model is superior.

Future work will develop an uncertainty analysis for SVR rotor power curve and then apply this to the development of an effective fault detection algorithm.

Acknowledgment

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.

REFERENCES

- [1] GWEC. Global wind report 2015. Gwec. Wind energy Technol., 75 (2016).
- [2] Rene van de Pieterman. Cost Modelling for Offshore Wind Operations and Maintenance. April 10, 2013.
- [3] P. Dao, W. Staszewski, T. Barszce, T. Uhl. Condition monitoring and fault detection in wind turbines based on cointegration analysis of SCADA data. Renew. Energy, 116 (Part B) (2018), pp. 107-122.

- [4] Amirat Y, Benbouzid M, Al-Ahmar E, Bensaker B and Turri S 2009. A brief status on condition monitoring and fault diagnosis in wind energy conversion systems *Renew. Sust. Energ. Rev.* 13(9) 2629-36.
- [5] Gustavo de Novaes PiresLeite, Alex MaurícioAraújo and Pedro André Carvalho Rosas. Prognostic techniques applied to maintenance of wind turbines: a concise and specific review. *Renewable and Sustainable Energy Reviews*. Volume 81, Part 2, January 2018, Pages 1917-1925.
- [6] Wind Turbines—Part 12-1: Power Performance Measurements of Electricity Producing Wind Turbines, British Standard, IEC 61400-12-1, 2006.
- [7] R. K. Pandit and D. Infield, "Comparison of binned and Gaussian Process based wind turbine power curve for condition monitoring purposes," 2nd International Conference on Maintenance Engineering (InCoME), 2017, University of Manchester, UK.
- [8] A. Kusiak and A. Verma. Monitoring Wind Farms with Performance Curves. *IEEE Transactions on Sustainable Energy*, vol. 4, no. 1, pp. 192-199, Jan. 2013. doi: 10.1109/TSTE.2012.2212470.
- [9] Singh, P. (2013), thesis. Analytical Techniques of SCADA Data to Assess Operational Wind Turbine. Performance. Glasgow: University of Strathclyde.
- [10] N. Yampikulsakul, E. Byon, S. Huang, S. Sheng and M. You. Condition Monitoring of Wind Power System with Nonparametric Regression Analysis. *IEEE Transactions on Energy Conversion*, vol. 29, no. 2, pp. 288-299, June 2014. doi: 10.1109/TEC.2013.2295301.
- [11] S. Shokrzadeh, M. Jafari Jozani and E. Bibeau. Wind Turbine Power Curve Modeling Using Advanced Parametric and Nonparametric Methods. *IEEE Transactions on Sustainable Energy*, vol. 5, no. 4, pp. 1262-1269, Oct. 2014. doi: 10.1109/TSTE.2014.2345059.
- [12] F. Pelletier, C. Masson, A. Tahan. Wind turbine power curve modelling using artificial neural network. *Renew Energy*, 89 (2016), pp. 207-214.
- [13] R. K. Pandit and D. Infield. Using Gaussian process theory for wind turbine power curve analysis with emphasis on the confidence intervals. 6th International Conference on Clean Electrical Power (ICCEP), Santa Margherita Ligure, 2017, pp. 744-749. doi: 524 10.1109/ICCEP.2017.8004774.
- [14] Kusiak A, Zheng H, Song Z. On-line monitoring of power curves. *Renewable Energy* 2009; 34: 1487–1493.
- [15] Nassim Laouti Nida Sheibat-Othman. Support Vector Machines for Fault Detection in Wind Turbines. *IFAC Proceedings Volumes Volume 44, Issue 1, January 2011, Pages 7067-7072.*
- [16] Y. Lin, U. Kruger, J. Zhang, Q. Wang, L. Lamont and L. E. Chaar. Seasonal Analysis and Prediction of Wind Energy Using Random Forests and ARX Model Structures. *IEEE Transactions on Control Systems Technology*, vol. 23, no. 5, pp. 1994-2002, Sept. 2015. doi: 10.1109/TCST.2015.2389031.
- [17] T. Ouyanga, A. Kusiak, Y. Heb. Modeling wind-turbine power curve: a data partitioning and mining approach. *Renew. Energy*, 102 (A) (2017), pp. 1-8.
- [18] M. Lydia, S.S. Kumar, A.I. Selvakumar, G.E.P. Kumar. A comprehensive review on wind turbine power curve modeling techniques. *Renewable and Sustainable Energy Reviews*, 30 (2014), pp. 452-460.
- [19] Vapnik, V. *The Nature of Statistical Learning Theory*. Springer, New York, 1995.
- [20] N. Laouti, N. Sheibat-Othman, S. Othman. Support vector machines for fault detection in wind turbines. *IFAC Proc Vol*, 44 (1) (2011), pp. 7067-7072.
- [21] L.B. Jack, A.K. Nandi. Fault detection using support vector machines and artificial neural network, augmented by genetic algorithms. *Mechanical System and Signal Processing*, 16 (2002), pp. 373-390.
- [22] M. Schlechtingen and I. F. Santos. Comparative analysis of neural network and regression based condition monitoring approaches for wind turbine fault detection. *Mech. Syst. Signal Process.*, vol. 25, no. 5, pp. 1849–1875, 2011.
- [23] Vladimir Cherkassky, Yunqian Ma. Practical selection of SVM parameters and noise estimation for SVM regression. *Neural Networks* 17 (2004) 113–126. *Neural Networks*.
- [24] Olivier Chapelle. Training a Support Vector Machine in the Primal. http://www.kyb.mpg.de/fileadmin/user_upload/files/publications/attachments/primal_%5b0%5d.pdf.
- [25] Support vector machine toolbox, MATLAB 2017b.
- [26] K.R. Miller and V. Vapnik. 1999. Using Support Vector Machine for Time Series Prediction. Cambridge: MIT, p. 243-253.
- [27] Breeze. Quantifying Change in Power Performance using SCADA Data Methods of measuring power performance and a case study to apply side-by-side testing. Part 3 of 3: How to quantify change in power performance. Available online at http://www.greenbyte.com/fileadmin/documents/Quantifying_Change_in_Power_Performance_using_SCADA_Data.pdf.