COMPARATIVE STUDY OF BINNING AND GAUSSIAN PROCESS BASED ROTOR CURVES OF A WIND TURBINE FOR THE PURPOSE OF CONDITION MONITORING

A. Ravi Kumar Pandit, *University of Strathclyde, UK*
B. David Infield, *University of Strathclyde, UK*

ABSTRACT

The wind turbines deteriorating performance cause failures, including catastrophic failures that lead to high operation and maintenance (O&M) cost. SCADA based continuous monitoring of wind turbines is a cost-effective approach and plays an essential role as turbine sizes increase, and they are placed in more remote locations, for example, offshore to improve the performance of wind turbines and reduces the O&M cost. Various studies suggest that the internal operation of the wind turbine depends on various variables, especially on rotor speed. The proper analysis of rotor speed can be useful for constructing effective SCADA based models for wind turbine condition monitoring. Gaussian Process (GP) is a nonparametric, stochastic process that's designed to solve regression and probabilistic classification problems. GP models powerful to solve nonlinear systems, however, its application are not much explored in wind turbine condition monitoring.
This paper describes a wind turbine condition monitoring Gaussian Process technique that uses the rotor speed to derive a rotor curve for wind turbine condition monitoring. Developed GP model, then compared with the conventional approach based on binned rotor curves together with individual bin probability distributions to identify operational anomalies. The proposed techniques have been validated experimentally using SCADA data sets obtained from operational turbines. Finally, comparative analysis of these techniques described outlining the strength and weakness of individual models.

1. INTRODUCTION

The growth of wind energy is fastest among all alternative forms of energy generation, and such a large scale of deployment of wind energy has brought challenges to assess the turbine performance. This challenges even more due to the fluctuating nature of wind speed and power production which makes turbines access for maintenance and repair work severe [1]. The wind turbines deteriorating performance causes failures, including catastrophic failures that lead to high operation and maintenance (O&M) cost. This is especially true in the case of offshore wind turbines, where access for routine maintenance or unscheduled repairs or inspections can be expensive and weather dependent. For instance, it suggested that O&M costs for offshore wind farms could account for up to 25- 30% of the energy costs [1]. Therefore, there is a need, to reduce O&M cost and SCADA based condition monitoring considered as an effective way to reduce the O&M cost of wind turbines [2].
Various statistical techniques used to monitor the performance of wind turbines, which classified into the parametric and nonparametric approach. Nonparametric models do not involve equations and very accurate to express the dynamic characteristics of wind turbines unlike parametric models [3]. Various nonparametric models used power curve as a critical indicator to assess the performance of wind turbines, see for examples [4,5]. However internal factors may not be analyzed it alone by power curve since the operational characteristics of turbines depend on variables such as rotor power, torque, and pitch angle. Continuous monitoring of the impact of these internal factor makes condition monitoring (CM) useful. The rotor curves useful for identifying the failures associated with wind turbines and have two types: i) A rotor speed curve represents the relationship between rotor speed and wind speed and ii) A rotor power curve defines the relationship between power output and rotor speed. Both these relationships are in nonlinear. The author of [6] constructed rotor curve and power curve to detect abnormal behavior of turbines and result concludes that rotor curve detected performance change due to down event, but it remains undetected in its corresponding power curve. Furthermore, A. Kusiak et al. [7], reference rotor speed curve developed based on multivariate outlier detection approach inspired by

k-means clustering and Mahalanobis distance. Using this reference rotor speed curve, the underperformance of a wind turbine determined by calculated values of kurtosis and skewness.

A Gaussian process (GP) is a nonparametric model where any finite number of random variables have a joint Gaussian distribution [8]. GP models can be expressed by the covariance function, which is used to define the similarity between inputs. GP models previously applied in wind power forecasting [9], solar power [10], and electricity price [11]. Despite GP models promising a result, its application to wind turbine condition monitoring not much explored.

In this paper, SCADA based Gaussian Process rotor curves presented which can be useful for the qualitative understanding of turbine health condition to detect failures at an early stage. The developed GP model is then compared with the binning technique, to identify the operational anomalies. Finally, the weaknesses and strengths of these techniques summarised.

The contents of this paper are as follows: Section 1 is the introduction. Section 2 describes the wind turbine rotor curves. Section 3 presents actual operational SCADA data sets and describes its pre-processing techniques. Section 4 synopsis the methodologies used to estimate the wind turbine blade rotor curves. Section 5 examines the comparative analysis of proposed techniques, and section 6 concludes the paper with intended future work.

2.ROTOR CURVES OF A WIND TURBINE

The rotor curves facilitated the relationship between the various wind turbines parameter and classified into two types. The rotor power curve signifies the nonlinear relationship between wind turbine rotor speed and power output while a rotor curve represents a mapping between rotor speed and wind speed, see figures 1 and 2. Unexpected failures impact rotor curves shape and can be useful for performance appraisal of a wind turbine.

Air density corrections: For accurate power curve modeling, IEC 61400-12-1 [12], 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,

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

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 varied 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 used for constructing error free rotor curves in upcoming sections.

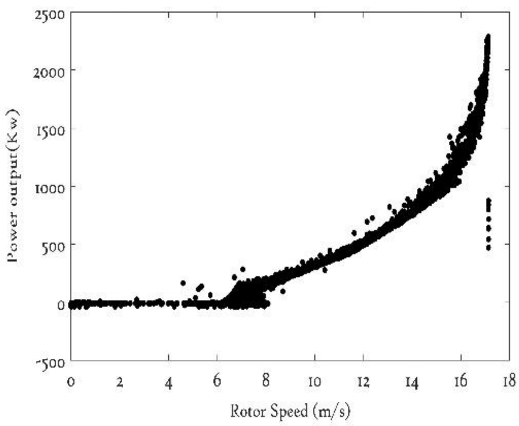

Figure 1: Rotor power curve

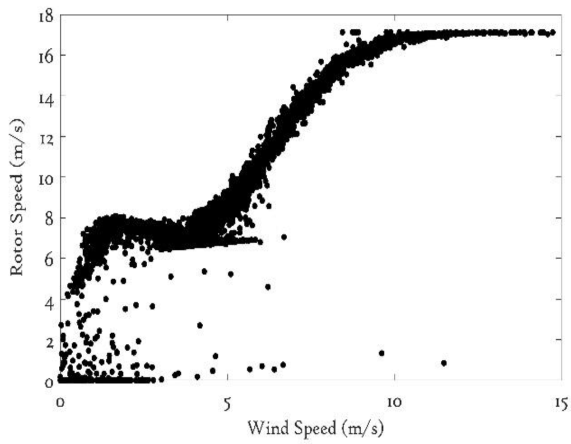

Figure 2: Rotor speed curve

3.SCADA DATA AND ITS PRE-PROCESSING

A Supervisory Control and Data Acquisition (SCADA) based condition monitoring is a cost-effective approach to observe the early warning of failures and performance issues [13]. SCADA system recorded the operational data which usually contain variables such as wind speed, wind direction, yaw angle, pitch angle, active power, reactive power, generator current, generator speed, gearbox temperature, generator winding temperature and ambient temperature. Proper analysis of these parameters over time can be useful to identify abnormal behavior related to developing failure. Typical SCADA datasets come with ten-minute intervals to reduce transmitted data bandwidth and memory storage. Continuous monitoring of stored SCADA data vital to improving the overall health of the turbine as well as its internal components. SCADA data sets used in this study are ten- minutes averaged data points covering parameters such as wind speed, power output, ambient temperature and these datasets is sorted by timestamp and These data contains an error caused by sensor failures, malfunctions of mechanical and data collection system. The impact of these error needs to minimize for fundamental analysis purposes. Hence pre-processing erroneous values is a first and significant step in the data-driven approach such as Gaussian Process. Following data filtering as outlined in [14], such as mismatch timestamp, negative power values, curtailments; SCADA data filtered and figure 3 and 4 are the pre-processed and air density corrected rotor curves of a typical wind turbine. These filtered and air density corrected rotor curves used in GP and binning approach.

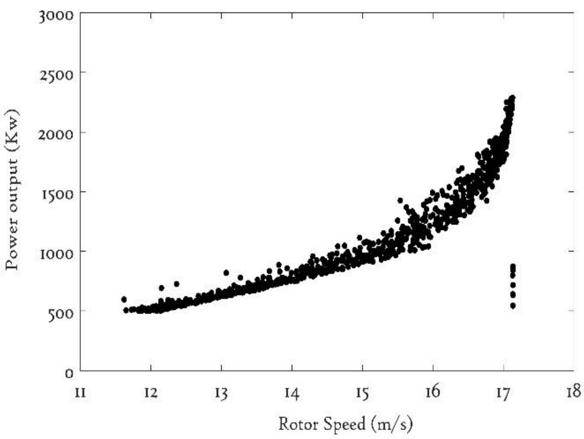

Figure 3: Filtered rotor power curve

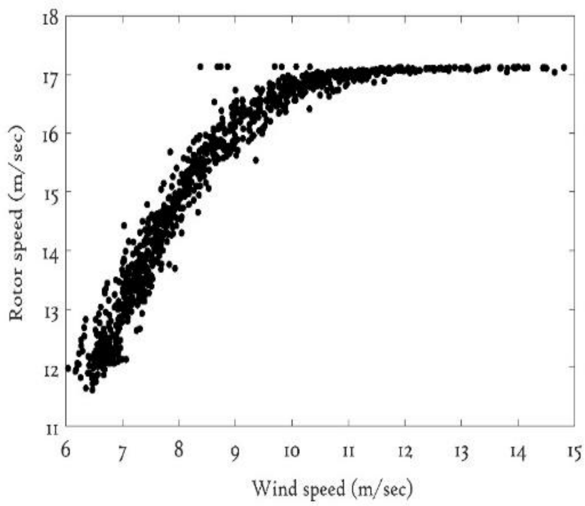

Figure 4: Filtered rotor speed curve

4.METHODOLOGIES

Pre-processed SCADA data sets are used to develop rotor curves of a wind turbine based on Gaussian Process and binning approach and described as follows,

4.1 ROTOR CURVES BASED ON GAUSSIAN PROCESS

Rasmussen and Williams [8] discussed a brief theoretical description of the Gaussian Process (GP) models and their applications. Here a brief review of the fundamentals is provided as follows.
Gaussian Process is a nonparametric, data-driven approach have state of the art performance in regression and classification problems. GP are a simple and general class of models of functions which is completely specified by its mean function and covariance function,

$$Y \sim GP(\mu, \Sigma) \quad (4)$$

where μ is the mean function, Σ is the covariance function that has an associated probability density function:

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

where $|\Sigma|$ is defined as a determinant of Σ, n is the dimension of random input vector x , and μ is mean vector of x . The term under exponential i.e.

$\frac{1}{2}(x-\mu)^T \sum^{-1}(x-\mu)$ is an illustration of a quadratic form.

GP models use a covariance function to define the prior covariance between any two function values where covariance function describes the dependency of two variables concerning each that represents the similarity between two points and hence determines closeness between two points. A brief description of covariance function discussed in [8] and its selection based on problem statement and nature of the data. The squared exponential covariance function is commonly applied and would be used in this paper. For any finite collection of inputs $\{x1, x2, \ldots\ldots, xn\}$, It is defined as:

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

SCADA dataset of the wind turbine comes with sensor errors, so it is desirable to add a noise term to the covariance function to improve the accuracy of the GP model. Hence equation (6) modified to be:

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

where σ_f^2 and l are known as the hyper-parameters. σ_f^2 Signifies the signal variance and l is a characteristic length scale which describes how quickly the covariance decreases with the distance between points. σ_n is the standard deviation of the noise fluctuation and gives information about model uncertainty. δ is the Kronecker delta, [8]. To accurately reflect the correlations presented in the processed rotor curves data (figures 3 and 4), the hyperparameter values of the squared exponential covariance function need to be optimised using method dthe escribed in [8] and the estimated rotor curve based on the GP model is shown in figures 5 and 6.

The estimated GP rotor curves come with confidence intervals (CIs) which is significant for estimated GP model uncertainty analysis. The confidence intervals contain the actual parameter value in some known proportion of repeated samples, on average. The width of confidence intervals is thought to index the precision of an estimate and CI of 95% is used in this study.

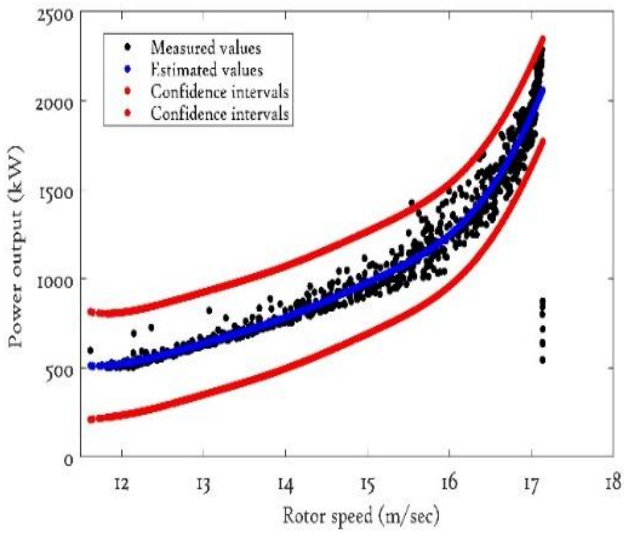

Figure 5: GP rotor power curve

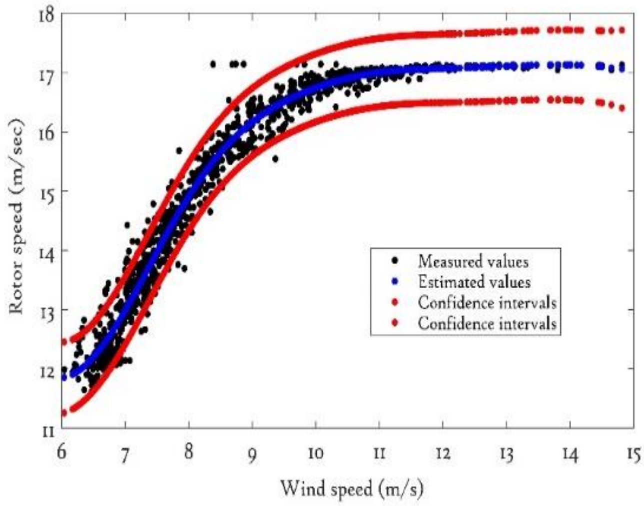

Figure 6: GP Rotor speed curve

The estimated GP power curve result suggest that the uncertainty is high between 12 m/sec to16 m/sec rotor speed range but after that its uncertainty started decreasing while in case of predicted GP rotor speed curve, uncertainty started increasing after rated wind speed, see figures 5 and 6. Due to the presence of a low number of SCADA data at different rotor speed and wind speed range caused high uncertainty of the estimated GP rotor curves. This is because the GP confidence intervals itself is an estimate and its accuracy highly depends on amount and quality of data.

4.3 ROTOR CURVES BASED ON BINNING

The IEC 61400-12 [12], uses a data reduction approach, called 'method of bin' to calculate the power curve of a wind turbine where ten-minute averaged SCADA data grouped into wind speed interval of 0.5 m/sec. The IEC standard uses nacelle wind speed to calculate the power curve. In binning, the average power output for each bin is obtained by grouping power measurements into wind speed bins. The bin width kept at 0.5 m/sec wide wind speed interval. Using large data points gives more certainty for average value in the power curve using 'method of bin'. In this study, the binning method applied to calculate rotor curves using the following equations,

$$V_i = \frac{1}{N_i} \sum_{j=1}^{Ni} V_{n,i,j}$$

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

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

where,
V_i = normalised and averaged wind speed in bin i.
$V_{n,i,j}$ = normalised wind speed of data sets j in bin i.
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.

Wind turbine rotor curves constructed together with error bars and compared with measured rotor curves shown in figure 7 and 8. The uncertainty of rotor curves based on binning assess via its error bars. The two standard deviations (i.e., 95% confidence intervals) of measured power values and measured rotor speed are used to calculate the error bars of rotor power curve and rotor speed curve respectively. This obtained error bar used to measure the uncertainty associated with a bin of the respective rotor curves. But, the accuracy of 'method of bin' is weaken due to the selection of bin width of 0.5 m/sec because within each bin the output (Power, rotor speed) will depend strongly and non-linearly on input (wind speed, rotor speed) and a wide bin would result in a systematic bias, and the need in practice to get sufficient data points in each bin to be of statistical significance. The binned rotor curves as expected following the measured rotor curves of a wind turbine. The uncertainty of binned rotor power curve is started increasing after 16 m/s rotor speed (figure 7) while between cut in and rated wind speed range, the uncertainty of binned rotor speed curve is high, see figure 8.

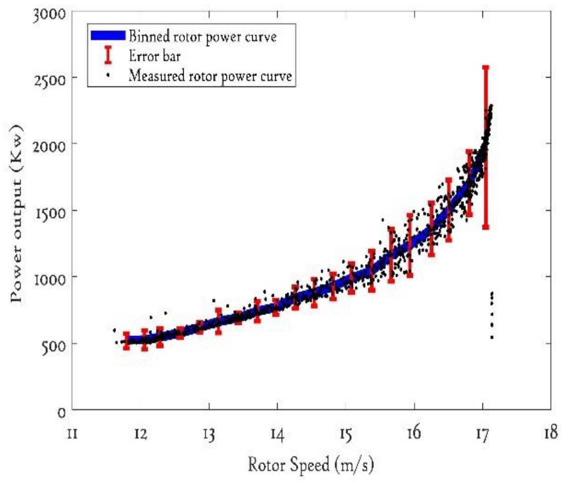

Figure 7: Binned rotor power curve

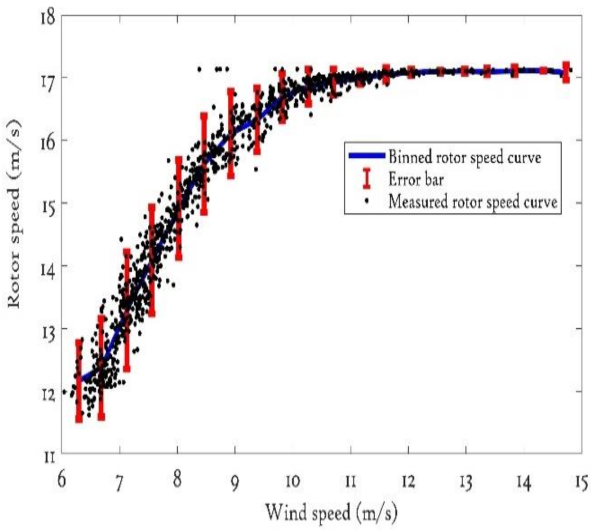

Figure 8: Binned Rotor speed curve

5.COMPARATIVE STUDIES

The Brief comparative studies of rotor curves based on Gaussian Process and binning presented in this section.
Gaussian Process based rotor power curve able to predict expected variance accurately which is then compared with binned rotor power curve and shown in figure 9. However, the GP model uncertainty as compared to binned approach

relatively high across the rotor speed region. This is because the accuracy of GP models depends upon the quantity and quality of the data, as well as the suitable method used.

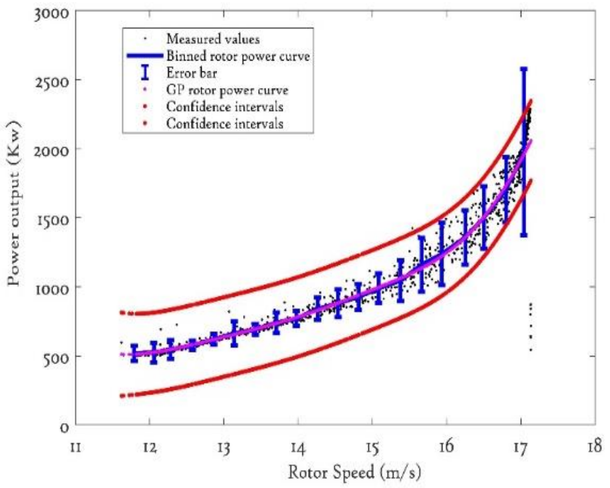

Figure 9: Comparative analysis of rotor power curve based on GP and binning

Rotor speed curve based on Gaussian Process and binning presented and their comparative analysis conclude that GP closely following the binned and measured rotor power curve as shown in figure 10. The uncertainty of GP model is low between cut in and rated wind speed range as compare to binned approach. Above rated wind speed, there are fewer SCADA data available, and as a result, the GP curve less well determined with some mismatch with the binned rotor power curve. This observation concludes that a GP model based on too little data can leads to inaccurate results. On the other hand, a large dataset leads to high complexity, high processing costs and potentially inaccurate results due to the mathematical challenges posed by the $O(n^3)$ an issue associated with matrix inversion described in [15]. Hence an optimum size of the dataset is a necessary and prerequisite for accurate GP modeling.

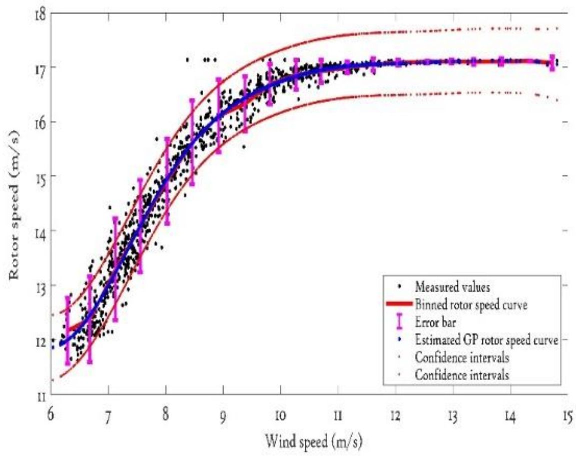

Figure 10: Comparative analysis of rotor speed curve based on GP and binning

6. CONCLUSION AND DISCUSSION

Gaussian Process is a powerful nonlinear approach that provides an attractive data analysis framework. The GP comes with a confidence interval that plays an essential role in identifying the early warning of failures and performance issues a wind turbine.

In this Study, the comparative analysis of rotor curves of a wind turbine based on Gaussian Process and binning techniques presented. With reasonable data points, GP model gives relatively better results. However, GP rotor curves models uncertainty suffers due to lack of a reasonable number of the data point, but in a binned rotor power curve, the data variation is highest on the rising section of the power curve hence perform relatively better.

Future work will use this result to construct effective GP based fault detection algorithm for wind turbines condition monitoring.

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.

REFERENCES

[1] Rene van de Pieterman (2013). Cost Modelling for Offshore Wind Operations and Maintenance. Renewable energy world.

[2] P. Dao, W. Staszewski, T. Barszce, T. Uhl (2018). Condition monitoring and fault detection in wind turbines based on cointegration analysis of

SCADA data. Renew. Energy, 116 (Part B), pp. 107-122.
[3] M. Lydia, S.S. Kumar, A.I. Selvakumar, G.E. Prem Kumar (2014). A comprehensive review on wind turbine power curve modeling techniques. Renew. Sustain. Energy Rev., 30, pp. 452-460.
[4] J. Gottschall, J. Peinke (2008). How to improve the estimation of power curves for wind turbines. Environ. Res. Lett., 3, Article 015005.
[5] S. Shokrzadeh, M. Jafari Jozani and E. Bibeau (2014). Wind Turbine Power Curve Modeling Using Advanced Parametric and Nonparametric Methods. IEEE Transactions on Sustainable Energy, vol. 5, no. 4, pp. 1262-1269, doi: 10.1109/TSTE.2014.2345059.
[6] Singh, P. (2013). Analytical Techniques of SCADA Data to Assess Operational Wind Turbine Performance. Glasgow: University of Strathclyde.
[7] A. Kusiak and A. Verma (2013). Monitoring Wind Farms with Performance Curves. IEEE Transactions on Sustainable Energy, vol. 4, no. 1, pp.192-199, doi: 10.1109/TSTE.2012.2212470.
[8] C.E. Rasmussen, C.K.I. Williams (2006). Gaussian Processes for machine learning. The MIT Press (2006).ISBN:026218253X. <http://www.gaussianprocess.org/gpml/chapters/RW.pdf>.
[9] Yan J, Li K, Bai EW (2013). Prediction error adjusted Gaussian Process for short-term wind power forecasting. In: 2013 IEEE int work intel energy syst; pp.173–8.
doi: https://doi.org/10.1109/IWIES.2013.6698581. ISBN 978-1-4799-1135-6.
[10] Sheng, S. Member, J. Xiao, Y. Cheng, S. Member, Q.Ni (2017). Short-term solar power forecasting based on weighted Gaussian process regression. IEEE Trans Ind Electron, 0046 (c), pp. 1-10, 10.1109/TIE.2017.2714127.
[11] Mori H, Nakano K (2015). Development of advanced Gaussian process for LMP forecasting. 18th International Conference on Intelligent System Application to Power Systems (ISAP); doi: https://doi.org/10.1109/ISAP.2015.7325553 [ISBN 9781509001903].
[12] Wind Turbines—Part 12-1: Power Performance Measurements of Electricity Producing Wind Turbines, British Standard, IEC 61400-12-1, 2006.
[13] A. Zaher, S. D. J. McArthur, D. G. Infield, and Y. Patel (2009). Online Wind Turbine Fault Detection Through Automated SCADA Data Analysis, Wind Energy, vol.12, no.6, pp.574593.
[14] M. Schlechtingen and I. F. Santos (2011). 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.
[15] Joseph Gonzalez. Linear-Time Inverse Covariance Matrix Estimation in Gaussian Processes .http://www.cs.cmu.edu/~sahong/docs/jegonzal_sahong_final.pdf