This paper is a post-print of a paper submitted to and accepted for publication in Journal of Maintenance Engineering (JME), University of Manchester and is subject to *proceedings of 2nd International Conference in Maintenance Engineering (IncoME-II) and/or Journal of Maintenance Engineering (JME).*

Comparison of binned and Gaussian Process based wind turbine power curves for condition monitoring purposes

Ravi Kumar Pandit, David Infield

University of Strathclyde, Glasgow

United Kingdom (UK)

ravi.pandit@strath.ac.uk / david.infield@strath.ac.uk

ABSTRACT

Performance monitoring based on available SCADA data is a cost effective approach to wind turbine condition appraisal. A power curve of a wind turbine describes the relationship between power output and wind speed and is a key measure of wind turbine performance. The standard IEC method calculates a binned power curve from extensive measured data, however this approach requires an extended measurement period in order to limit the uncertainty associated with the calculated power curve, and is far too slow to be used directly for condition monitoring where any changes in operation need to be identified quickly. Hence an efficient approach needs to be developed to overcome this limitation and be able to detect anomalies quickly, thus detecting damage at an early stage so as to prevent catastrophic damage. A Gaussian Process (GP), which is a non-parametric machine learning approach, has the potential fit power curves quickly and effectively. This paper deals with the application of a Gaussian Process to power curve fitting and anomaly detection. This is compared with the conventional approach based on a binned power curve together with individual bin probability distributions to identify operational anomalies. The paper will outline the advantages and limitations of the Gaussian Process approach.

Keywords: Wind turbine, Gaussian Process, condition monitoring, power curve

This paper is a post-print of a paper submitted to and accepted for publication in Journal of Maintenance Engineering (JME), University of Manchester and is subject to *proceedings of 2nd International Conference in Maintenance Engineering (IncoME-II) and/or Journal of Maintenance Engineering (JME).*

1. INTRODUCTION

Wind energy is one of the viable alternatives to the coal or fossil based energy. A rapid increase in wind turbine installation is now replacing fossils sources to meet energy demand across the globe (Figure 1). As stated in [1], there has been a 29 % increase in this renewable source between 2014 and 2015 in UK alone and the trend is still increasing. Wind turbines currently contribute over 10% of the UK's total electricity needs, [2]. Compared to onshore wind turbine operation, maintenance costs are significantly higher in offshore. With the growth of offshore wind, particularly in the UK, there is an increase in operation and maintenance (O&M) costs which makes offshore turbine less profitable. As in report [3], it is found that O&M costs make up 20-25% of the total lifetime costs of an offshore wind farm. Reducing this cost via condition monitoring is an important target for research.

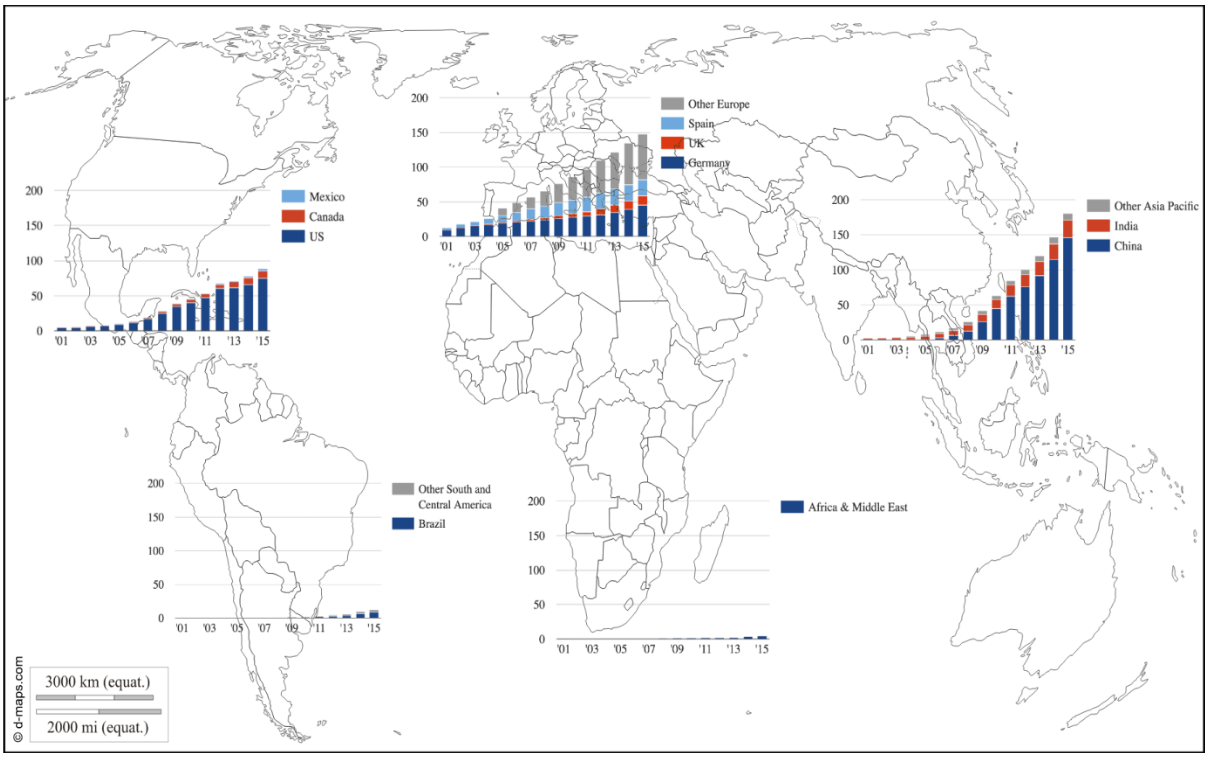


Figure 1: Global wind energy capacity by country and region between 2001 and 2015, gigawatts (GW), taken from [4]

Performance monitoring based on available SCADA data is a cost effective approach to turbine condition appraisal. Performance is conventionally assessed in terms of the wind turbine power curve that represents the relationship between the generated power and hub height wind speed,

[5]. Power curves also play a vital role in energy assessment, and performance and warranty formulations. Various methods have been used, for example in [6] enhanced parametric models (MHTan), for power curve prediction. Here, swarm optimization has been used, with Cuckoo search, and a backtracking search algorithm, with least square error (LSE) and maximum likelihood estimation (MLE) used in to compare the compare an existing parametric model with the proposed model. Laplacian Eigenmaps (LE) nonlinear dimensionality reduction and Linear Mixture Self-organizing Maps (LMSOM) classifiers are used in [7] for power curve based anomaly detection of a wind turbine. Any power curve comes with uncertainty and hence its analysis is important; it plays a key role in wind resource assessment. Reference [8] presents a deterministic method for uncertainty analysis in wind resource assessment and wind energy production estimation. Fontaine and Amstrong, [9], have experimented with the uncertainty analysis of a wind turbine located in Italy using the IEC approach and Monte Carlo analysis. However, regardless of the methods used for power prediction, it is worth noting that wind power prediction is not only dependent on external wind conditions but also on the structural and mechanical performance of the wind turbines systems, [10, 11]. Conventional power curves as defined in the IEC Standard are considered to be effective measure for accuracy and uncertainty. However, they take considerable time to establish and are thus far too slow to be used directly for condition monitoring, [12].

A Gaussian process (GP), [13], is a non-parametric machine learning method that is gaining in popularity in prediction and forecasting applications due to its simple concept and parsimony in terms of the assumptions required to construct a model as compared to other non-parametric methods (e.g. neural network or fuzzy network), [14]. Moreover, a GP provides a natural way to estimate the uncertainty associated with its predictions as compared to neural network models because of their complex model design.

This paper introduces the application of Gaussian Process to wind turbine power curve determination and anomaly detection. For any model dealing with condition monitoring of a wind turbine, uncertainty analysis is central. In a GP model, the confidence interval used to describe uncertainty around the predicted GP values is intrinsic to the model and can be directly used for anomaly detection. The IEC binned power curve is widely used for power performance and uncertainty analysis but, as already mentioned, is too slow for condition monitoring. It is however the benchmark of the wind sector. The paper presents a comparative analysis of the

IEC binned power curve and GP power curve in terms of uncertainty and identifies the advantages and disadvantages of the GP model.

2. WIND TURBINE DESCRIPTION AND DATA PRE-PROCESSING

Located about 15 kms south of Glasgow, Scotland, Whitelee Wind Farm (owned and operated by Scottish Power Renewables) is large onshore wind farm a comprising 215 Siemens and Alstom wind turbines with a total capacity of 539 megawatts (MW), [15]. SCADA data covering a full year of operation and consisting of 10 minute averages of wind speed and power, with maximum, minimum, and standard deviations is available for analysis. SCADA data provides operational data for individual wind turbines at effectively zero cost and can help to provide an efficient and cost-effective way to identify early warning of failures and related issues. SCADA data is not perfect; it comes with errors due to sensor and data logging faults, and thus needs careful pre-processing. A range of filtering criteria can be applied to eradicate timestamp mismatches, out of range values, negative power values, and turbine power curtailment, see for example [16]. Figure 1 shows unfiltered power curve data for a particular turbine, whilst Figure 2 shows the impact of data filtering. Note that a 500 kw power generation threshold has been applied in this instance.

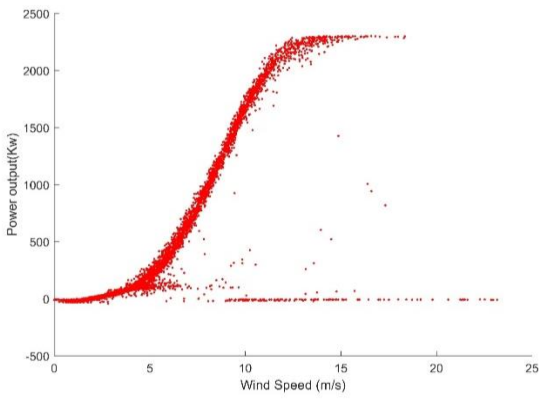


Figure 1: Measured power curve (raw data)

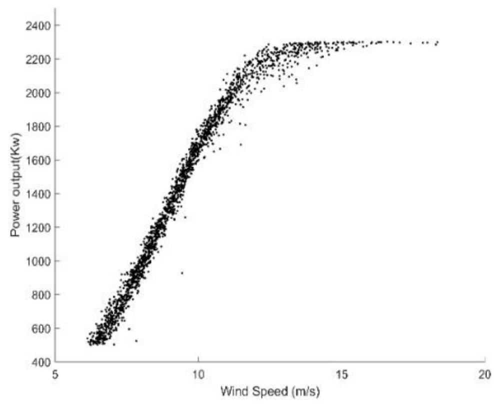


Figure 2: Measured power curve (filtered data)

3. POWER CURVE OF A WIND TURBINE

Available the 10-minute average SCADA values of power and hub wind speeds describe the nonlinear relationship between them and represented by power curve or performance curve of a wind turbine and mathematically expressed as follows,

$$P = \frac{1}{2}\rho\pi r^2 v^3 C_p \qquad (1)$$

where C_p is the power coefficient defined as the ratio of power extracted by wind turbine to the total power contained; r is the radius of the turbine rotor; v is the wind velocity; ρ is the air density and P is the turbine power output.

Early detection of wind turbine performance degradation has been used for condition monitoring of a wind farm. For example, [17] used power curve analysis to monitor several wind turbines using a data transformation approach based on 3^{rd} and 4^{th} polynomials. Furthermore, Yan et al., [18], introduced an inverse data transformation method to identify changes in wind turbine performance.

Measurements collected by the SCADA system comprise 10-minute averaged values of nacelle anemometer wind speed, turbine power output, ambient temperature and atmospheric pressure.

The air density ρ is calculated from temperature and atmospheric pressure. Before any power curve based analysis is undertaken, air density correction needs to be applied to the data because power capture by turbine rotor is directly proportional to the air density passing through it. In the IEC standard, [12], the air density correction done with regards to wind speed only and using equation (2):

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

where V_C and V_M are the corrected wind speed and measured wind speed respectively and ambient air density, ρ can be calculated by equation (3) as:

$$\rho = 1.225 \left[\frac{288.15}{T}\right]\left[\frac{B}{1013.3}\right] \qquad (3)$$

where, T is ambient temperature in absolute degree and B is the barometric pressure in mbar. Both are 10-minute average values taken from the SCADA data.

4. GAUSSIAN PROCESS THOERY FOR POWER CURVE MODELING

Unexpected failures and associated unscheduled maintenance is a major reason for high wind turbine maintenance costs. However, by developing an efficient automated fault detection algorithm based on real time monitoring of performance it is hoped that most failures can be anticipated and maintenance thus planned around this, improving the reliability of wind turbine and ultimately reducing the cost of maintenance. A Gaussian Process (GP) is a non-linear machine learning approach able to support wind turbine condition monitoring. The GP is used to describe a distribution over functions and it is a collection of random variables any finite number of which have a joint Gaussian distribution, [13]. The overall model of GP is defined by its mean and covariance functions. If $m(x)$ is the mean function and $k(x, x')$ is the covariance function of a real process, then the desired function $f(x)$ is defined as:

$$f(x) \sim GP\big(m(x), k(x, x')\big) \qquad (4)$$

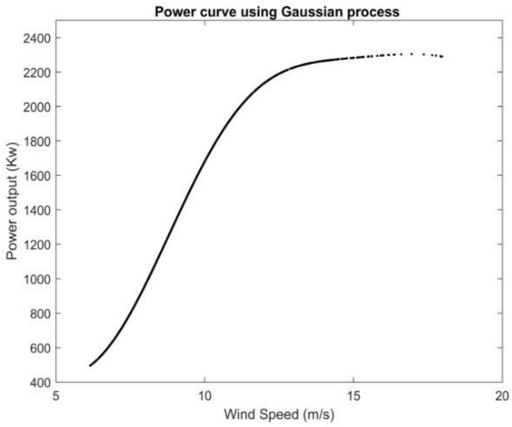
Covariance between any set of data points relates to a multivariate Gaussian distribution as per the GP assumption. The covariance function (or kernel) characterizes correlations between different values and selection of suitable covariance function for a GP model is considered to be vital in terms of model accuracy. Various covariance functions are available depending upon the need and application and well describe in [13]. In this paper, the squared exponential covariance function (k_{SE}) is used, defined as:

$$k_{SE}\,(x, x') = \sigma_f^2 \exp\left(-\frac{(x-x')^2}{2l^2}\right) \qquad (5)$$

The SCADA data comes with noise and measurement errors which affect the covariance function, k_{SE} hence it is generally recommended to add a noise term into the covariance function in order to make covariance more feasible and representative and hence equation (5) is modified:

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

where σ_f^2 and l are defined as the hyper-parameters. σ_f^2 describe the signal variance and l is a characteristic length scale which describes how quickly the covariance decreases with distance between points. The squared exponential covariance function is technically a smooth sample function and is infinitely differential. Using the squared exponential covariance function, a monthly GP predicted power curve algorithm has been developed in MATLAB and is shown in Figure 3. Prediction of power capture by a wind turbine at future time is significant for unit commitment and maintenance scheduling. The GP model is able to model the power curve accurately as shown in Figures 4 and 5.



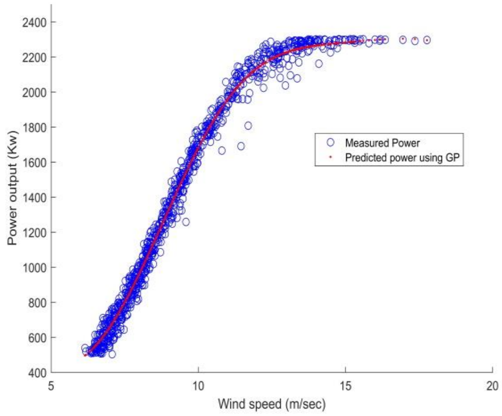


Figure 3: Predicted GP power curve   Figure 4: Comparison of GP power curve with data

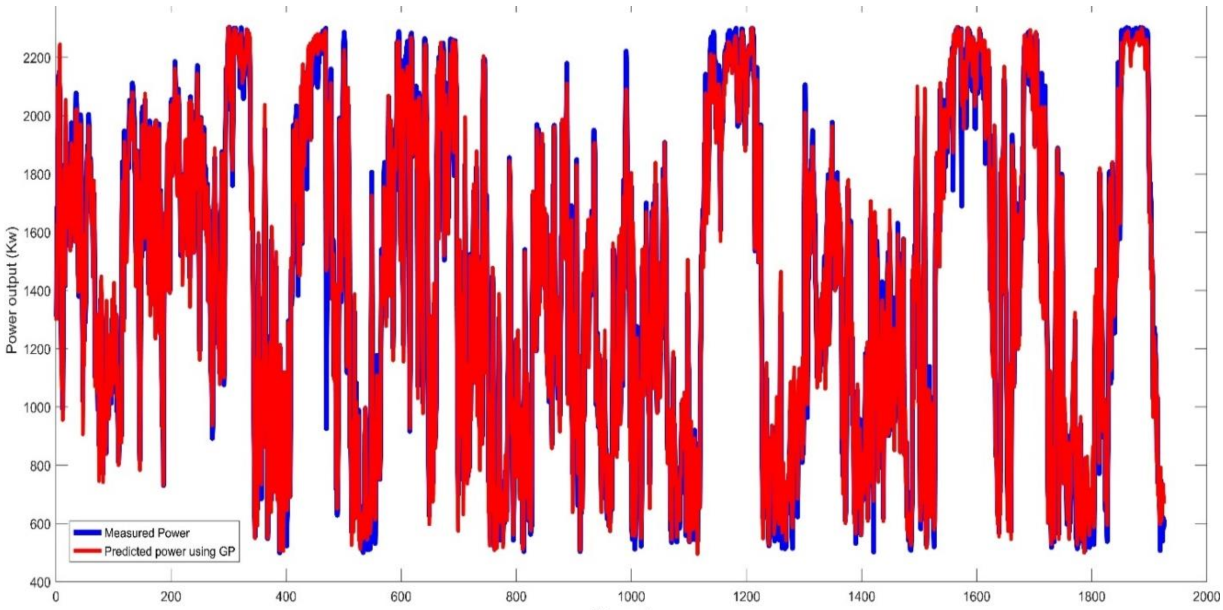
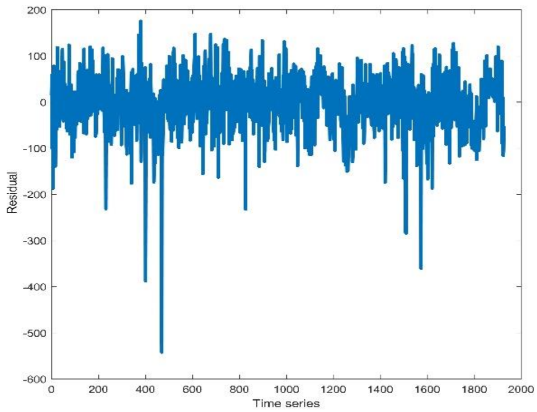


Figure 5: Comparison between measured power and predicted power from the GP model

The difference between measured and predicted values are called residuals, [19]. Due to the non-parametric, non-linear behaviour of a GP model, residual analysis is important. The residual plot (Figure 6) indicates predicted GP values close to measured values and hence model errors are small. Theoretically, residuals of a GP model should be Gaussian. The frequency distribution of the residuals is shown in Figure 7 together with a fitted Gaussian distribution and, as expected, the distribution of residual is close to being Gaussian.

This paper is a post-print of a paper submitted to and accepted for publication in Journal of Maintenance Engineering (JME), University of Manchester and is subject to *proceedings of 2nd International Conference in Maintenance Engineering (IncoME-II) and/or Journal of Maintenance Engineering (JME).*



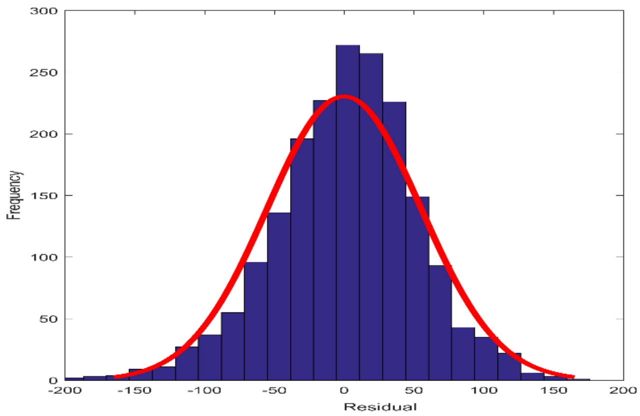


Figure 6: Residual time series plot

Figure 7: Residual histogram fitting

In order to assess the GP power curve accuracy, confidence intervals can be used. Confidence intervals are a useful measure of uncertainty and the precision of model estimates. Confidence intervals provide significant information about the uncertainty surrounding an estimation, but are themselves model based estimates [20]. Confidence interval estimation was used in [21] for wind forecasting to better control wind power generators used with batteries. This control requires accurate wind power forecasts and their confidence intervals. GP confidence intervals for annual wind power production are presented in [22]. GP power curves with estimated 95% confidence intervals have been used in this paper, as shown in Figure 8.

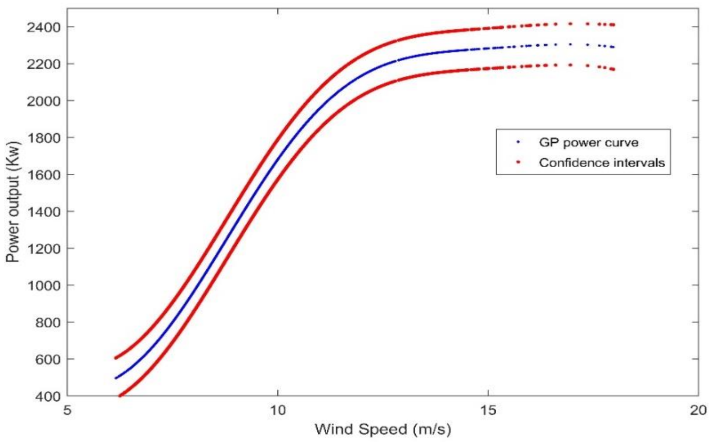


Figure 8: GP power curve with confidence interval

5. COMPARATIVE ANALYSIS OF BINNED POWER CURVE AND GP POWER CURVE

The IEC binned power curve is considered to be an accurate approach to power curve determination and uncertainty analysis and thus has been selected as the benchmark against which to assess GP models. The error bars are a graphical representation of the variability of wind data and spread of data around the mean value, [23]. They are an effective tool to represent the error or uncertainty in the power curve of a wind turbine. Following the IEC, the binned power curve is shown with error bars (95% confidence interval) in Figure 9. It is found that the spread of data is high (due to high error bars values) between cut in and rated wind speed, and small above rated, reflecting in the main the manner in which wind turbine power is controlled.

Conventional IEC Standard binned power curves described above are compared with a GP power curve shown in Figure 10. The GP power curve model closely follows the IEC standard power curve. Above rated wind speed, there is less SCADA data available and as a result the GP curve is less well determined with some mismatch with the binned power curve. This observation confirms that a GP model based on too little data can be inaccurate. 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 $O(n^3)$ issue associated with matrix inversion described in [24]. Hence an optimum size of dataset is a necessary and prerequisite for accurate GP modelling.

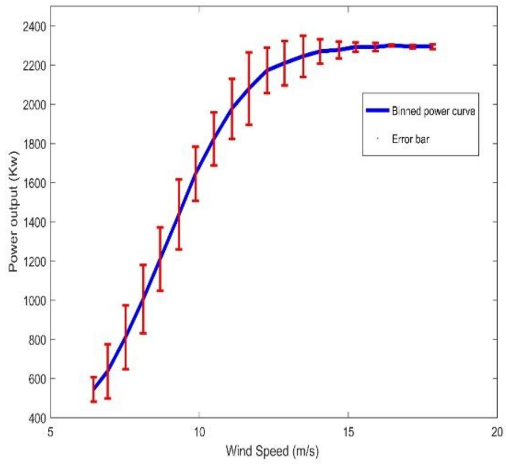


Figure 9: Binned power curve with error bars

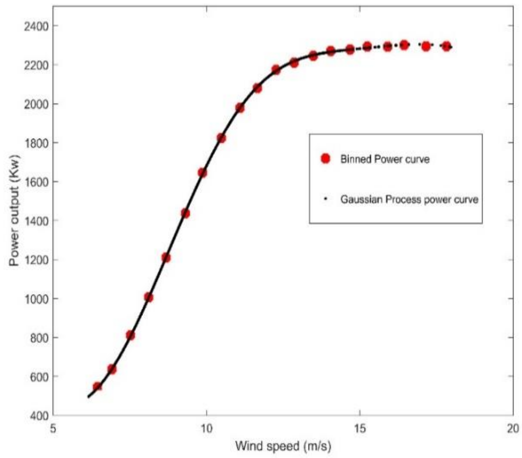


Figure 10: Binned and GP power curve comparison

The GP predicted power curve is compared with the IEC binned power curve (with 95% confidence interval of error bars) in terms of uncertainty in Figure 11. The part of the power curve between cut in and rated wind speeds, where maximum power tracking takes place, is the most critical for condition monitoring purposes. This is because small changes in turbine efficiency (perhaps due to blade damage or excessive drive train losses) can be readily detected, whereas above rated wind speed where power is limited by the control system, this loss of efficiency will be masked by greater wind power input. The smaller confidence intervals for the GP model compared to the binned power curve between cut in and rated wind speed indicate that GP model is more accurate for this critical range. Since a GP model is a form of interpolation, its uncertainty increases towards the two ends of the data set (i.e. low and high winds).

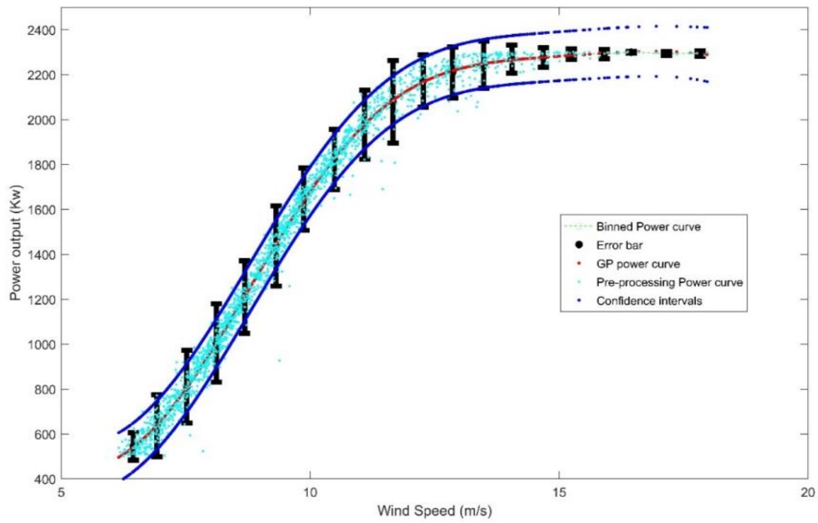


Figure 11: Overview of GP power curve and binned power curve uncertainties

This can be seen in Figure 11, but is not of concern for condition monitoring applications for the reasons explained above. In short, the confidence interval of the GP power curve, being smaller for the critical wind speed range, is better able to reject the unhealthy or faulty data than the binned power curve. For accurate and early anomaly detection, the smallest possible confidence intervals are required. The IEC binned power curve is relatively slow to respond whilst the GP model can be established from limited data with good accuracy as seen in Figures 8,10 and 11. It should be noted that, incoming data can be considered as noisy (due to measurement error) in GP, [13,20]. This issue has not been included here and will be addressed in future research.

6. DISCUSSION AND CONCLUSION

Gaussian Processes provide an attractive data analysis framework and the basis of a potentially effective automated fault detection system for the wind sector. The accuracy of a GP power curve depends upon the quantity and quality of the data, but can yield accurate results based on limited data. A low number of power-wind speed pairs may not give a smooth power curve while a high number of data is also not desirable because of large matrix inversion associated with GP. The confidence limits of a Gaussian Process are key to effective anomaly detection. For a binned power curve, the data variation is highest on the rising section of the power curve. By comparing a binned power curve with a GP power curve it is found that the latter is more accurate over the rising section of the power curve.

Future work will use this result in anomaly detection by designing appropriate GP models and comparing them with available binned power curve based anomaly detection methods to judge how much earlier a GP model can detect anomalous wind turbine operation.

ACKNOWLEDGEMENTS

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] UK government report on "Renewable sources of energy" . chapter 6, renewable energy sources.

[2] Wind Energy in the UK State of the Industry Report 2015

[3] Cost Modelling for Offshore Wind Operations and Maintenance by Rene van de Pieterman, April 10, 2013

[4] Global Wind Report 2016 and the BP Statistical Review of World Energy 2015.

[5] 'Wind Energy Handbook' by Tony Burton, David Sharpe, Nick Jenkins, Ervin Bossanyi, DOI: 10.1002/0470846062

[6] 'Development of an enhanced parametric model for wind turbine power curve' by Ehsan Taslimi-Renani, Mostafa Modiri-Delshad, http://dx.doi.org/10.1016/j.apenergy.2016.05.124

[7] 'Wind turbines abnormality detection through analysis of wind farm power curves' by shuangyuan Wang, Yixiang Huang, Lin Li, Chengliang Liu, https://doi.org/10.1016/j.measurement.2016.07.006

[8] Lackner MA, Rogers AL, Manwell JF. Uncertainty analysis in wind resource assessment and wind energy production estimation. 45th AIAA Aerospace Sciences Meeting and Exhibit, Reno; 2007.

[9] Fontaine A, Armstrong P. Uncertainty analysis in energy yield assessment. EWEC2007, Milan; 2007.

[10] Tavner P, Edwards C, Brinkman A, Spinato F. Influence of wind speed on wind turbine reliability. Wind Eng 2006;30(1):55–72.

[11] Duenas-Osorio L, Basu B. Unavailability of wind turbine due to wind-induced accelerations. Eng Struct 2008;30(4):885–93.

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

[13] C. E. Rasmussen & C. K. I. Williams, Gaussian Processes for Machine Learning, the MIT Press, 2006, ISBN 026218253X.

[14] Elizabeta Lazarevska. A neuro-fuzzy model for wind speed prediction based of statistical learning theory. Journal of Electrical Engineering and Information Technologies, Vol. 1, No. 1–2, pp. 45–55 (2016)

[15] Whitelee Windfarm official home page, https://www.scottishpower.co.uk/whitelee/

[16] 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.

[17] Kusiak A, Verma A. Monitoring wind farms with performance curves. IEEE transactions on Sustainable Energy 2013;4(1):192e9

[18] Yan Y, Osadciw LA, Benson G, White E. Inverse data transformation for change detection in wind turbine diagnostics. In: Proceedings of Canadian conference on electrical and computer engineering 2009. p. 944e9.

[19] Morris H. DeGroot, Mark J. Schervish, ’Probability and Statistics’, 4th Edition.

[20] Neyman, J. (1937). "Outline of a Theory of Statistical Estimation Based on the Classical Theory of Probability". Philosophical Transactions of the Royal Society A. 236: 333 380. doi:10.1098/rsta.1937.0005.

[21] Tatsuya Iizaka, Toru Jintsugawa, Hideyuki Kondo, "A Wind Power Forecasting Method and Its Confidence Interval Estimation ". doi: 10.1002/eej.22326.

[22] Alain Bensoussan. ‘Confidence intervals for annual wind power production’. ESAIM: PROCEEDINGS, January 2014, Vol. 44, p. 150-158

[23] Sarkar, A; Blackwell, A; Jamnik, M; Spott, M (2015). "Interaction with uncertainty in visualisations" doi:10.2312/eurovisshort.20151138

[24] Joseph Gonzalez. “Linear-Time Inverse Covariance Matrix Estimation in Gaussian Processes”. Available online at http://www.cs.cmu.edu/~sahong/docs/jegonzal_sahong_final.pdf