

Using Gaussian Process Theory for Wind Turbine Power Curve Analysis with Emphasis on the Confidence Intervals

Ravi Kumar Pandit, David Infield
University of Strathclyde, Glasgow, (United Kingdom)
ravi.pandit@strath.ac.uk

Abstract— High operation and maintenance (O&M) costs may affect the profitability and growth of wind turbine industries in long term, especially where offshore wind farms are concerned. With the increase in age of wind turbines and the expansion of offshore wind, the operation and maintenance (O&M) cost is expected to grow significantly which reinforces the drive towards condition based maintenance. Wind turbine power curves play a central role in the assessment of turbine operational health. Gaussian process theory is finding increasing application in this current emerging research area. This paper investigates the potential of Gaussian process models to improve the representation of wind turbine power curves and in particular the importance of confidence intervals as determined by such modeling.

Keywords -- Forecasting, Gaussian process, Operation & maintenance, power curve, Wind turbines.

I. INTRODUCTION

Because of the increasing demand for clean energy to combat climate change, wind energy is now playing a very important role worldwide. Recently installed wind turbines generally require less repair, and thus operation and maintenance (O&M) costs are lower. But in long term, as turbines get older and less reliable, condition monitoring can play a vital role in limiting O&M costs.

These older wind turbines now require more maintenance than those more recently installed. O&M costs constitute a significant percentage of the total annual costs of a wind turbine, as shown in Figure 2, taken from [2]. Because of the difficulties of access O&M costs are much higher for offshore wind turbines; in particular, those located far from the shore and exposed to more extreme winds and sea states. It is found that O&M costs make up 20-25% of the total lifetime costs of an offshore wind farm making this a target for research to reduce costs in the coming years, [2].

Condition monitoring is not a new concept but its application to wind turbine technology remains relatively unexplored. Different kinds of machine learning approach are presently gaining popularity in data analysis for the condition monitoring of wind turbines [3, 4]. There are two distinct types of approach to O&M; scheduled maintenance and unscheduled maintenance. Unscheduled maintenance due to unexpected failures of wind turbine components is considered to be a major reason for high maintenance costs and low availability. Scheduled

maintenance traditionally comprises routine maintenance at fixed time intervals, but condition based maintenance has the potential to remove unnecessary site visits, reduce down time and by avoiding catastrophic failure, reduce repair costs by anticipating faults.

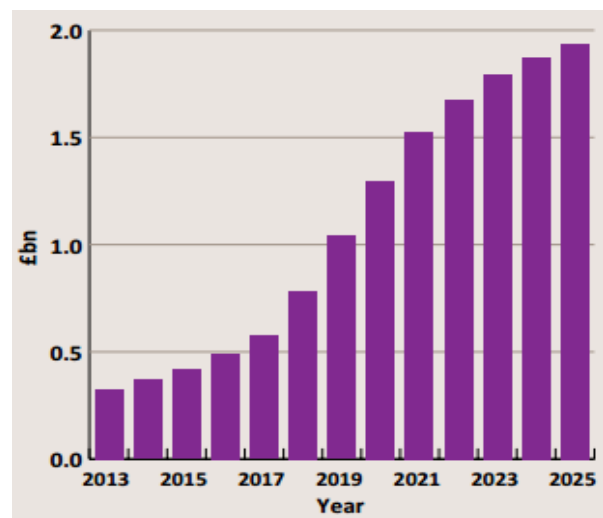


Fig 1: UK offshore wind O&M cost over the year's [5]

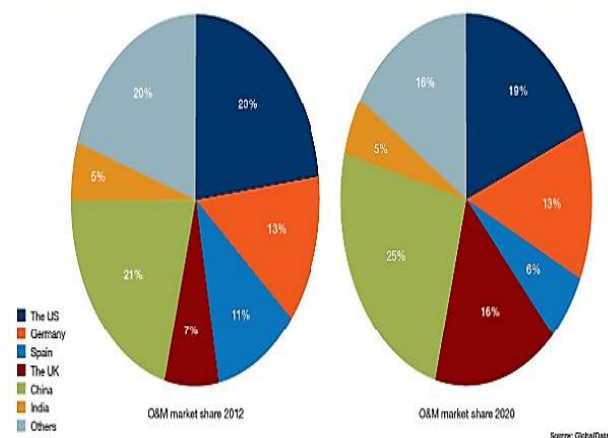


Fig 2: Operation & maintenance cost forecasting [16]

An essential tool when planning maintenance strategies is information about the probabilities of component failures and Gaussian process (GP) models can be an effective in this regard. GP analysis is a non-parametric machine

learning approach, [6], that is relatively simple conceptually and for this reason is finding favor in a number of application areas including wind farm O&M, see for example [7] and [8].

This paper is focused on modeling wind turbine power curves using a Gaussian process where the confidence interval identified by the GP can play a key role in turbine fault identification. Extensive wind farm operational data, as described below, has been used for model identification and exploration.

II. WIND FARM DESCRIPTION

Whitelee Wind Farm is large onshore wind farm situated on upland about 15 kms south of Glasgow, Scotland, and comprises 215 Siemens and Alstom wind turbines with a total capacity of 539 megawatts (MW), [9]. This wind farm is owned and operated by Scottish Power Renewables. The data used in this paper correspond to a full year of operation. All the SCADA data consists of 10 minute averages with maximum, minimum, standard deviation over the 10 minutes also being recorded. In addition, two met masts are located at the wind farm with instruments at a range of heights.

III. DATA PRE-PROCESSING

The wind turbine industry uses supervisory control and data acquisition (SCADA) data to monitor wind farm operation. As well assisting operators to spot severe performance deterioration resulting in shut-down it also has the potential to identify early indications of failure through subtler changes in performance, [10]. Using SCADA data for condition monitoring system can be very cost-effective as no additional hardware is required, but it must be used carefully as it is not immune from sensor error or possible malfunction of the data management system. There will inevitably be incidences of out of range values, missing data, unmatched timestamp values etc. For this reason, SCADA data must be carefully filtered to remove such problems; practical issues also include negative power measurements, turbine curtailment, and data affected by switching transitions within the 10 minute periods.

IV. INFLUENCE OF AIR DENSITY CORRECTION

Power curves are influenced by air density, since as is clear from Equation 1, from [11], power is proportional to air density.

$$P = 0.5 \cdot \rho \cdot v^3 \cdot A \cdot C_p \quad (1)$$

where, P = Power output (W)
 ρ = air density (kg/m^3)
 v = wind speed (m/sec^2)
 A = swept area (m^2)
 C_p = aerodynamic efficiency of a wind turbine.

Air density varies with wind turbine location and altitude and in particular with the air temperature. According to IEC Standard 61400-12-1, [12], an air density correction should be applied for modern variable speed turbines following Equation (2),

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

Here V_C and V_M are the corrected and measured wind speed respectively and the ambient air density, ρ can be obtained by equation (3),

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

where T is temperature in Kelvin and B is the barometric pressure in mbar.

V. WIND TURBINE PERFORMANCE CURVE

The power curve of the wind turbine gives the 10-minute averaged power generation as a function of 10-minute averaged wind speed measured at a suitable position upstream of the turbine rotor and at the hub height, and corrected for air density as explained. It is an effective indicator of the performance of a wind turbine. With the help of power curve any performance deterioration of individual wind turbines can be identified. Figure 3 shows a power curve based on raw unfiltered and uncorrected data, while Figure 4 shows a curve after air density correction, and more importantly after unwanted and potentially misleading data has been removed. In this work all such unwanted data has been removed at the outset, i.e. before plotting. If the IEC Standard is followed the scatter of data is reduced to a single line using a data reduction process known as binning. This approach will not be followed here; rather a GP will be used to fit a very general non-linear function to the data.

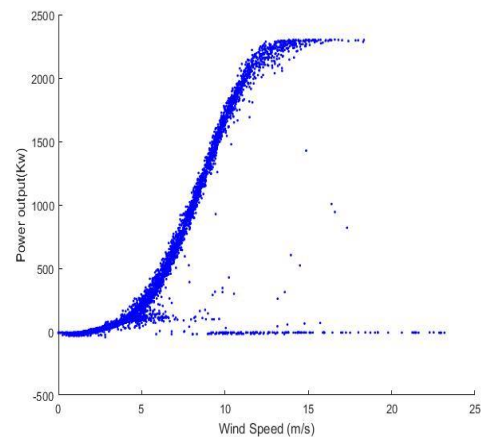


Fig 3: Monthly Power without data filtration

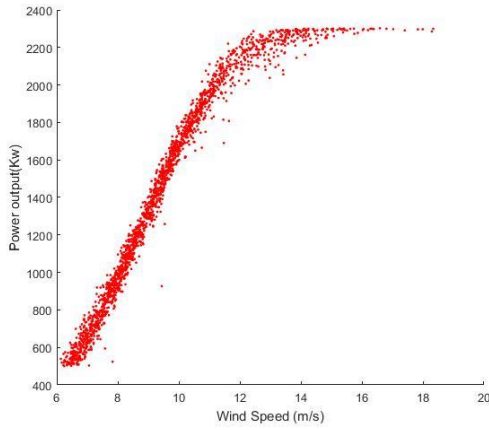


Fig 4: Monthly Power after data filtration

VI. BRIEF OUTLINE OF GAUSSIAN PROCESS THEORY AND APPLICATION TO POWER CURVES

A Gaussian process (GP) is a non-parametric machine learning approach for generating a function based on assigning probabilities that follow a multi-dimensional Gaussian probability distribution to individual data points. It has gained popularity because of its flexible non-parametric behavior, the use of a minimum number of assumptions and its conceptual simplicity. It is not only useful for point estimation but also provides valuable information about uncertainty which is informative for covariance function analysis of nonlinear training datasets, [6]. It is however computationally demanding as it involves the inversion of very large matrices.

As mentioned, all data points are treated as samples for random variables that individually have a Gaussian distribution. In the limit the function can be continuous with respect to time or space, [13]. The GP approach uses a so called lazy learning approach from machine learning, like the k nearest neighbor (kNN) approach. The quality and quantity of training datasets seriously affects the performance of the Gaussian process algorithm. From [9, 10], it is found that a GP may not give a good result when data sets are too small or too large. Theoretically, a GP model comprise a mean function and a covariance function as follows,

$$X \sim \text{GP}(m, K) \quad (4)$$

Where, m is the mean function; K the covariance function; and X a random function.

A GP assumes that the covariance between any set of data points relates to a multivariate Gaussian distribution. The covariance function (or Kernel) of a GP measures the similarity between nearby points. In this work the widely used squared exponential (SE) form of covariance function has been adopted, which is mathematically defined following [10], as:

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

The datasets generally contain noise and measurement errors so it is advisable to add a noise term to the covariance function in order to make covariance more representative. Hence equation (5) can be modified to become:

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

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 distance between points. One of the unique features of a GP is that it is determined solely by its mean and covariance functions and this property facilitates model fitting as only the first- and second-order moments of the process require specification. An example monthly power curve based on a GP is shown in figure 5 calculated from a single month of filtered SCADA data. In this paper, GP Matlab toolbox being used for modelling and analysis of power curve prediction [17]. Gaussian process estimation works well for power curve estimation as seen in figure 6 where the estimated power curve is compared with the filtered 10-minute data. The GP power curve has been used to estimate a time series of power values and these are compared with the measured data used to fit the GP in figure 7.

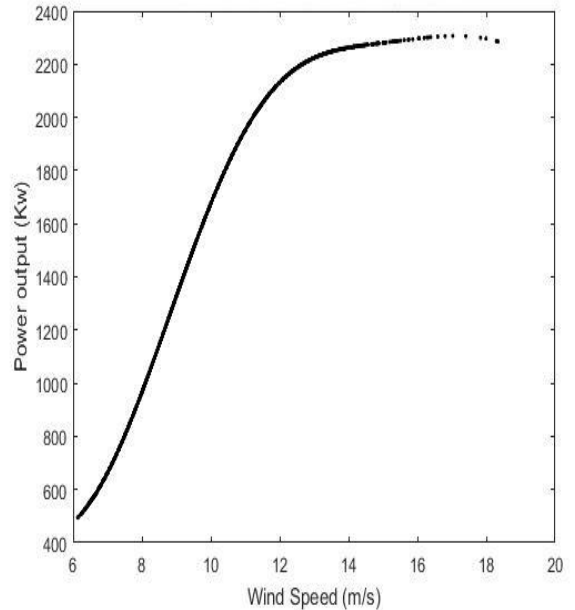


Fig 5: Predicted GP power curve

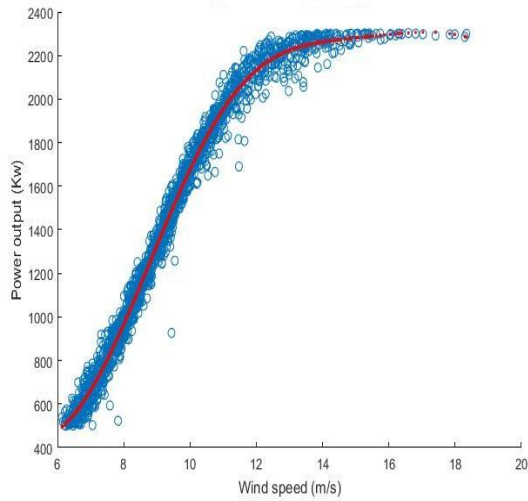


Fig 6: Predicted PC & measured PC

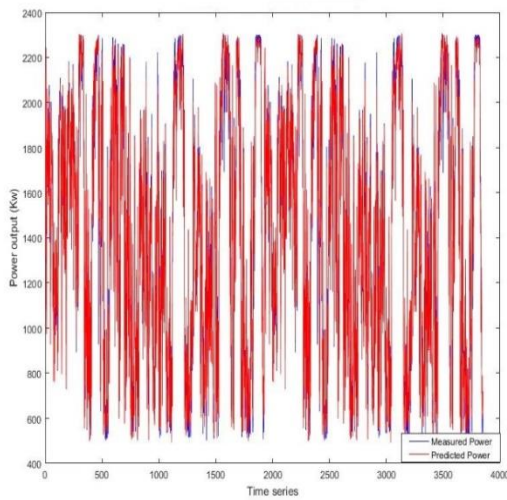


Fig 7: measured power and predicted power

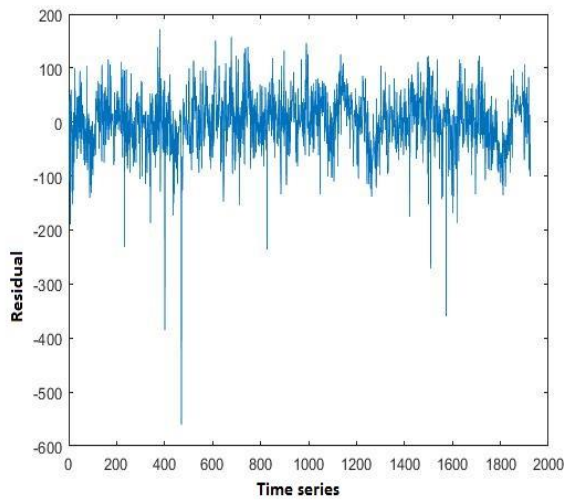


Fig 8: Gaussian process Residual

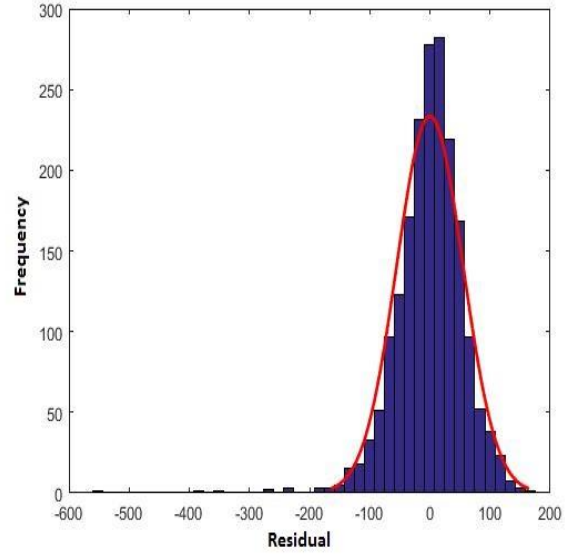


Fig 9: Residual Histogram with distribution fit

GP is a nonlinear, non-parametric regression model hence residual analysis is important, [14]. Residuals are defined as the difference between observed value and predicted values as given below:

$$e = y - y' \quad (7)$$

where, e is the residual, y the observed value and y' the value predicted by the model. The measured and predicted power shown in figure 7 and in good agreement as indicated by the generally small residuals, plotted in figure 8. The frequency distribution of the residuals is shown in figure 9 together with a fitted Gaussian distribution. As hoped for the distribution of residuals is close to being Gaussian. This provides further justification of the GP model.

VII. CONFIDENCE INTERVAL ESTIMATION FOR A GAUSSIAN PROCESS

Another parameter which plays vital role in interpretation of wind turbine power curves is the 'confidence interval'. Mathematically, the confidence interval gives valuable information about the uncertainty surrounding an estimate, [15]. The confidence interval itself is an estimate. Confidence intervals are intended to reflect the unknown sample population. They can also provide a threshold beyond which data is likely to reflect a malfunction and thus assist in wind turbine anomaly detection. GPs, by representing both model errors, and input measurement noise, allow straightforward estimation of confidence limits.

Figure 10 shows the calculated Gaussian process power curve together with the estimated 95% confidence limits based solely on model errors (but not measurement noise). These confidence limits can be used to identify unexpected data that are most likely to be the result of operational faults.

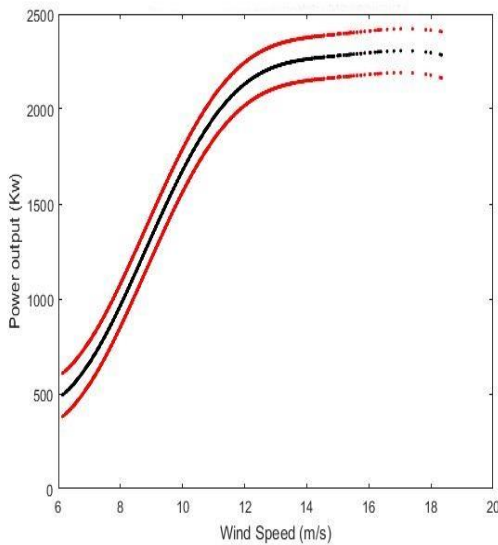


Fig 10: Predicated PC with confidence interval

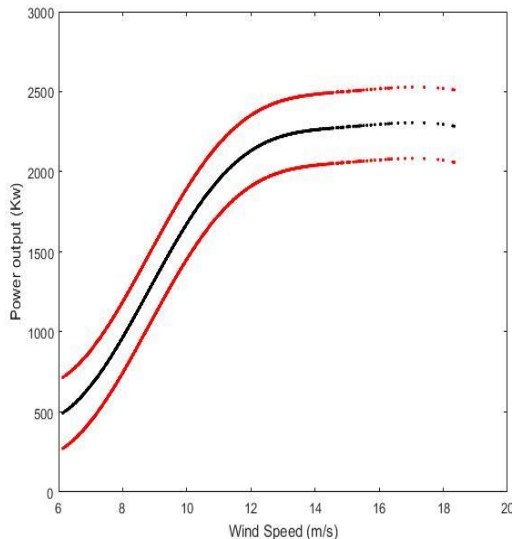


Fig 11: PC with modified confidence limit

In a Gaussian process, the incoming data can be considered as noisy due to measurement error. The GP's intrinsic estimates of confidence do not include this additional uncertainty, but the model does separately estimate the magnitude of the associated uncertainty, in this case treated as constant across the wind speed range. It is desirable to modify the confidence intervals to deal with incoming data measurement noise so that the fault identification process is at its most effective, [6,15]. The modified confidence limits relate to a higher overall standard deviation of the GP model. The modified confidence interval together with the GP power curve shown in figure 11.

VIII. CONCLUSION AND DISCUSSION

Wind turbine power curves are useful for monitoring the power production and performance of wind turbines and important in their condition monitoring. Because of its flexibility and the limited number of model parameters required, GP models are attractive and straightforward to implement. The application of GPs to power curve modelling has been shown to be effective in fitting the curve and in the identification of confidence limits. These results will be used in future work on anomaly detection for condition monitoring purposes, and their efficiency compared with other available methods.

ACKNOWLEDGMENT

The authors would like to thank CDT, Strathclyde staff and Sgurr energy, Glasgow for their continuous support.

REFERENCES

- [1] Operational and Maintenance Costs for Wind Turbines, <http://www.windmeasurementinternational.com/wind-turbines/om-turbines.php>
- [2] Cost Modelling for Offshore Wind Operations and Maintenance By Rene van de Pieterman , April 10, 2013
- [3] Models for monitoring wind farm power, Andrew Kusiak, Haiyang Zheng, Zhe Song, Renewable Energy, science direct, March 2009.
- [4] Modeling of Wind Turbine Power Curve Based on Gaussian Process, JIN ZHOU, PENG GUO, XUE-RU WANG, Proceedings of the 2014 International Conference on Machine Learning and Cybernetics, Lanzhou, 13-16 July, 2014
- [5] A Guide to UK Offshore Wind Operations and Maintenance, by GL Garrad Hassan
- [6] C. E. Rasmussen & C. K. I. Williams, Gaussian Processes for Machine Learning, the MIT Press, 2006, ISBN 026218253X
- [7] Xueru Wang, Jin Zhou, Peng Guo, 'Wind turbine gearbox forecast using Gaussian process model', Published in: Control and Decision Conference (2014 CCDC), The 26th Chinese.
- [8] Niya Chen, Zheng Qian, Xiaofeng Meng Ian T. Nabney, Short-Term Wind Power Forecasting Using Gaussian Processes, Twenty-Third International Joint Conference on Artificial Intelligence.
- [9] http://www.whiteleewindfarm.co.uk/about_windfarm?nav, Whitelee windfarm.
- [10] Zaher, A, McArthur, S.D.J., Infield, D.G., Patel, Y. "Online Wind Turbine Fault Detection Through Automated SCADA Data Analysis," Wind Energy, 2009, Published online in Wiley Interscience (www.interscience.wiley.com) DOI: 10.1002/we.319
- [11] Modelling of the Variation of Air Density with Altitude through Pressure, Humidity and Temperature, WindPRO / ENERGY, EMD International A/S
- [12] Wind Turbines—Part 12-1: Power Performance Measurements of Electricity Producing Wind Turbines, British Standard, IEC 61400-12-1, 2006
- [13] Ping Li, Songcan Chen, A review on Gaussian Process Latent Variable Models, CAAI Transactions on Intelligence Technology xx (2016)

- [14] Morris H. DeGroot, Mark J. Schervish, 'Probability and Statistics', 4th Edition.
- [15] 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.
- [16] Keep on turning: the cost of O&M by world wind technology.
- [17] Gaussian Process Regression Models, Matlab toolbox.