Pandit, Ravi Kumar and Infield, David (2018) Comparative analysis of binning and Gaussian Process based blade pitch angle curve of a wind turbine for the purpose of condition monitoring. Journal of Physics: Conference Series, 1102 (1). ISSN 1742-6588 , http://dx.doi.org/10.1088/1742-6596/1102/1/012037

IOP Conf. Series: Journal of Physics: Conf. Series **1102** (2018) 012037 doi:10.1088/1742-6596/1102/1/012037

Comparative analysis of binning and Gaussian Process based blade pitch angle curve of a wind turbine for the purpose of condition monitoring

Ravi Kumar Pandit and David Infield

University of Strathclyde,16 Richmond St, Glasgow G1 1XQ, UK.
Corresponding author e-mail: ravi.pandit@strath.ac.uk

Abstract. Several studies have used the power curve as a critical indicator to assess the performance of wind turbines. However, the wind turbine internal operation is affected by various parameters, particularly by blade pitch angle. Continuous monitoring of blade pitch angle can be useful for power performance assessment of wind turbines. The blade pitch curve describes the nonlinear relationship between pitch angle and hub height wind speed which to date has been little explored for wind turbine condition monitoring. Gaussian Process models are nonlinear and nonparametric technique, based on Bayesian probability theory. Such models have the potential give results quickly and efficiently. In this paper, we propose a Gaussian Process model to predict blade pitch curve of a wind turbine for condition monitoring purposes. The obtained Gaussian Process based blade pitch curve is then compared with a conventional approach based on a binned blade pitch curve for identifying operational anomalies purposes. Finally, the weaknesses and strengths of these methods are summarised. SCADA data from healthy wind turbines are used to train and evaluate the performance of these techniques.

1. Introduction

The wind energy had witnessed exponential growth in past decades and became the fastest among the available alternative energy sources. Global Wind Energy (GWEC) on its Global Wind Report [1], found that more than 54 GW of a wind turbine was installed across the global market in 2016, which now includes more than 90 countries and cumulative capacity grew by 12.6% to reach a total of 486.8 GW [1]. However, expensive operation and maintenance (O&M) cost makes wind energy costly affairs and therefore, reducing O&M costs and improving the reliability has become an emerging research interest. Unexpected failures of the wind turbines components cause high repair costs and result in machine downtime and loss of revenue. K. Fischer and A.et al. [2], suggested that O&M cost is responsible for 20-30% of the life cycle cost of onshore and up to 30% of offshore wind farms. To protect turbine components from premature failures, improve reliability and reduces O&M costs, SCADA based condition monitoring is considered as an effective technique and are expected to grow in coming years. Several literatures are focused on the power curve to assess the performance of a wind turbine for the purpose of condition monitoring, see [3,4]. However, the internal operation of turbines is affected by various turbine parameters that make difficult to visualize the turbine operation with the help of power curve alone. The turbine performance parameter (e.g., blade pitch angle) can play an important role in analyzing the internal operations of wind turbines and described in the upcoming section.

The wind turbine blade pitch curve describes the nonlinear relationship between pitch angle and wind speed and can be useful for robust wind turbine condition monitoring. For example, Singh [5] used a misaligned wind vane as a case study for wind turbine performance assessments where both power curve and blade pitch angle were constructed to identify the performance change due to misaligned vane. The comparative studies of these curves suggest that the blade pitch curve detects performance change while it is unidentified by the power curve. Furthermore, in ref [6], the wind turbine blade curve used in which multivariate outlier detection approach based on k-means clustering and Mahalanobis distance are applied. Using this curves, kurtosis, and skewness calculated for identifying underperformance associated with wind turbines. Thus, pitch curve monitoring is significant for identifying the abnormal behavior due to failures (e.g., pitch failures).

There are several statistical methods mostly confined to power curve of a wind turbine fitting and are broadly divided into the parametric and nonparametric methods. The nonparametric approach does not impose any pre-specified model unlike parametric method and so nonparametric models performance is better as compared to parametric models [7]. Examples of the nonparametric techniques include artificial neural networks [8], Gaussian Process [9,10], and k-nearest neighbor [11]. A comprehensive review of parametric and nonparametric approach related to wind turbines can be found in [3,7].

A Gaussian Process (GP) is a nonparametric, data-driven approach and widely used in solving problems related to classification and regression. GP is defined as a collection of random variables, any finite number of which have a joint Gaussian distribution [12]. GP model is a robust approach for approximating the complex nonlinear model, and one of its advantages is that estimated values come with confidence intervals. The confidence intervals are calculated separately by estimating the magnitude of the associated uncertainty and play a vital role in uncertainty analysis [9].

The recent expansion of wind power system leads to the development of accurate methods for their parameter identification such as the wind turbine blade pitch curve. Recent wind turbine condition monitoring studies have mostly focused on the power curve for evaluating performance. However, this cannot reflect the complete turbine operation since the operational behavior of the wind turbines profoundly influence by a parameter such as rotor power, torque, and pitch angle. Valid assessments of these parameter improve the power performance of a wind turbine. In this study, blade pitch angle impact on wind turbine performance is analyzed using the blade pitch curve that reveals the nonlinear relationship between pitch angle and the hub height wind speed that can be useful for analyzing wind turbine performance and the detection of faults.

In this paper, we propose a Gaussian Process approach to estimate the blade pitch curve of a wind turbine for application in identifying potential faults. The obtained GP based blade pitch curve is then compared with a conventional approach based on a binned blade pitch curve together with individual bin probability distributions to identify operational anomalies. Finally, the weaknesses and strengths of these methods are summarised. SCADA data from healthy wind turbines are used to train and evaluate the performance of these techniques.

This paper is structured as follows: Section 1 is the introduction. Section 2 describes the wind turbine blade pitch curve. Section 3 describes the SCADA dataset and its pre-processing. Section 4 outlines the methodologies and this section further divided into subsections explaining Gaussian Process (GP) and the binning approach to estimate the wind turbine blade pitch curve. Section 5 discusses the comparative analysis of proposed techniques, and Section 6 concludes the paper with intended future work.

2. Operational Curves of a Wind Turbine

The power curve widely used by wind operators to assess power performance, warranty formulations, energy assessment, and fault diagnosis applications. Power curve shows the power output strong dependence on wind speed (Figure 1) and mathematically expressed by following cubic relation [13],

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

where ρ is air density (kg/m^3), A is swept area (m^2) , C_p is the power coefficient of the wind turbine and v is the hub height wind speed (m/sec). The shape of the power curve governed by the cubic relation of these parameters.

A typical blade pitch curve described the nonlinear relationship between turbine pitch angle and wind speed and shown in Figure 2. Blade pitch angle customarily set to maximize the power production at below rated wind speed and at above-rated wind speed, pitch angle continuously adjusts to restrict power production to rated wind speed.

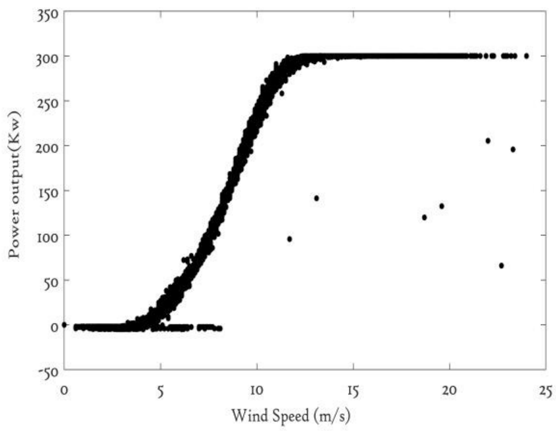

Figure 1: Measured power curve

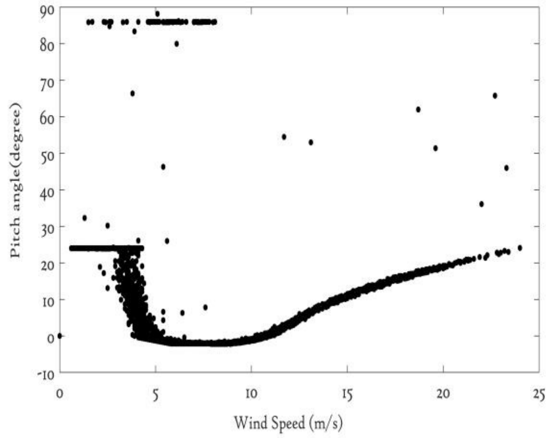

Figure 2: Measured pitch curve

The measured power curve and pitch curve (Figure 1 and 2) uses 10 minute average SCADA data obtained from a pitch controlled wind turbine. The IEC standard (61400-12-1) [14], advocate air density correction before any further analysis in which a corrected wind speed V_C is calculated using equations (2) and (3) as shown below,

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

and,

$$V_C = V_M \left[\frac{\rho}{1.225}\right]^{\frac{1}{3}} \tag{3}$$

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 (2) where B is atmospheric pressure in mbar and T the temperature in Kelvin for which 10 minute average values obtained from SCADA data are used. The corrected wind speed (V_C) from equation (3) will be used for constructing blade pitch curve based on Gaussian Process and binning methods in upcoming sections.

3. SCADA data description and pre-processing

The wind farms linked with SCADA data that contains enormous data sets which is useful for continuous condition monitoring of a wind turbine. SCADA data may be used efficiently to 'tune' a wind farm operation which is useful in identifying early warnings of possible failures and improving power production across many turbines in all conditions. SCADA can be used with any resolution, but 10-minute intervals generally recommended to reduce transmitted data bandwidth. SCADA data records historical and operational data of individual turbine where each data set contains average values of more than 100 parameters such as torque, wind speed, wind direction, drive train acceleration, and tower acceleration and so on. These parameters materialized with maximum, minimum, standard deviation. The datasets used in this study comes from a wind farm located in southern Europe. Monthly data used for model training and testing purposes contains 4464 data points beginning with time stamp ''1/08/2010 00:00 AM'' and ending at time stamp ''31/08/2010 23:50''. Using these data points, measured power and blade pitch curves are shown in Figure 1 and 2 respectively. It is worth to note that the monthly SCADA data used in this study is having an average monthly temperature of 29.779°C. However, the malfunction of sensors, mechanical systems, and the data collection system makes SCADA data subject to error and this affects the overall SCADA based condition monitoring of a wind turbine. Hence, pre-processing of SCADA is a must for minimizing the impact of these errors in order to extract useful information from the data.

Table 1. SCADA datasets description.

Start timestamp	End timestamp	Measured data	Filtered data
1/08/2010 00:00 AM	31/08/2010 23:50 PM	4400	2068

Criterion outlined in [15], such as missing values, values that are out of range, invalid values, mismatch timestamp, and negative power values are taken to minimize misleading data. Using this filtration approach, measured SCADA data sets were reduced by 2068 from 4400 and these pre-processed data points used to develop blade pitch curve based on GP and binning methods. The filtered and air density corrected blade pitch curve is shown in Figure 3.

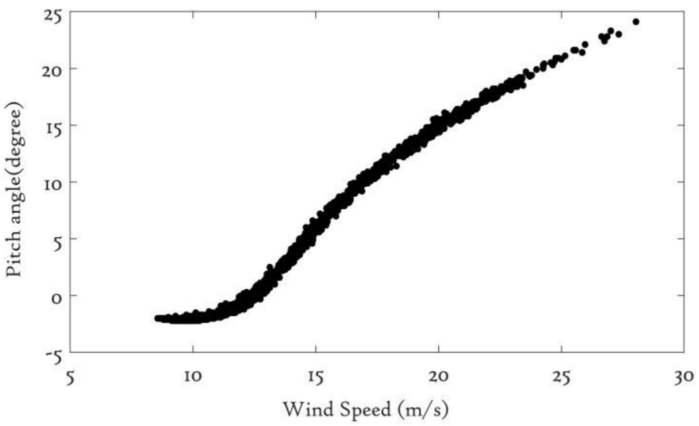

Figure 3: Pre-processed blade pitch curve

4. Methodologies

The two methods namely; binning and Gaussian Process used to build an effective blade pitch angle curves for the purpose of condition monitoring of a wind turbine are described as follows.

4.1. Gaussian Process based blade pitch curve

A Gaussian Process (GP) is a nonparametric stochastic approach widely used in classification and regression related problems. A GP describes a distribution over functions and is a collection of random variables, any finite number of which have a joint Gaussian distribution, [12]. A GP model is specified by a mean function and covariance function using the following equation,

$$f(x) \sim GP\big(m(x), k(x, x')\big) \tag{4}$$

where $m(x)$ is the mean function and $k(x, x')$ is the covariance function.

The covariance function describes the similarity between any set of data points relates to a multivariate Gaussian distribution as per the GP assumption. Covariance function selection determines the GP model effectiveness. Various covariance functions are available depending upon the need and application and are well described in [12]. The squared exponential covariance function (k_{SE}) with a fitted hyper-parameter is used in this study to predict the blade the pitch curve and is defined as:

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

In a Gaussian process, the incoming SCADA data can be considered noisy which affects the covariance function, k_{SE} , hence it is necessary to add a noise term to the covariance function in order to minimize the effect of noisy data, and hence equation (5) is modified to:

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

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 the distance between points. The squared exponential covariance function is technically a smooth sample function being infinitely differential. The modified squared exponential covariance function equation is used to estimate the blade pitch curve of a wind turbine and then this is compared with a traditional pitch curve in terms of uncertainty and accuracy, and is used to identify how quickly faults can be identified. The GP pitch curve with an estimated 95% confidence intervals (CIs) have been used in this paper, as shown in Figure 4. Furthermore, Figure 5 suggest that the estimated GP pitch curve follows the expected pattern.

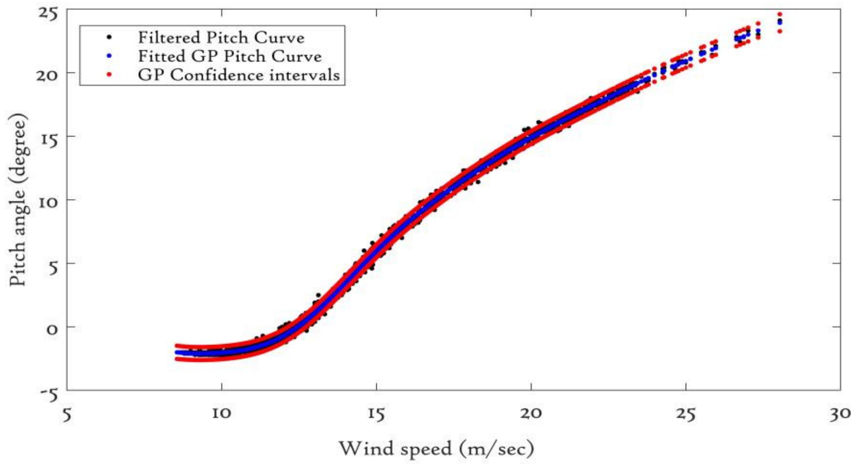

Figure 4: Estimated GP pitch curve with confidence intervals

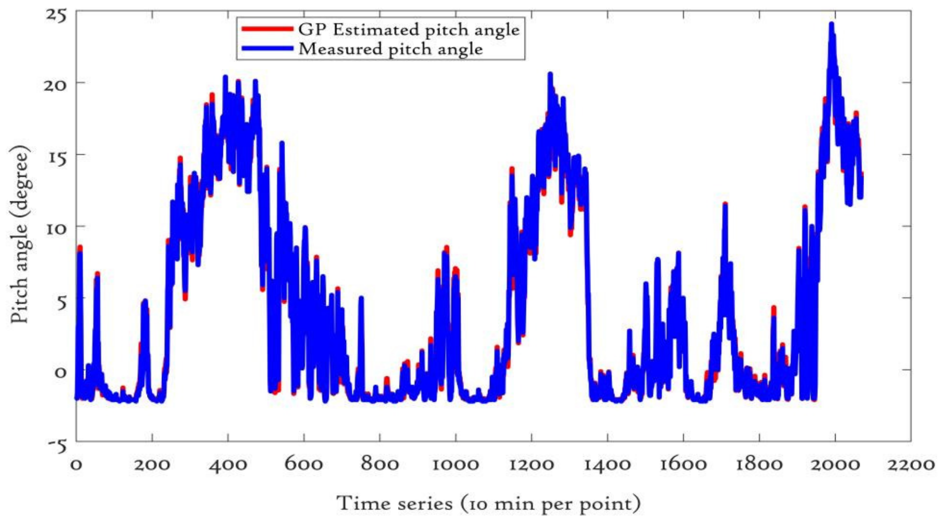

Figure 5: Comparative analysis of estimated and measured pitch angle in time series

Due to the nonparametric nature of the GP model, the residual analysis is essential to find out whether GP distribution is Gaussian or not. Residual is the difference between measured and estimated value. The frequency distribution of the residuals is shown in figure 6 together with a fitted Gaussian distribution. As expected, the distribution of residuals is closely following Gaussian.

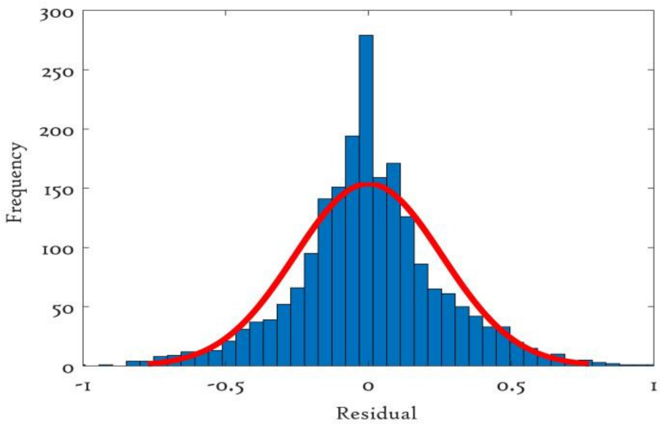

Figure 6: GP residuals histogram fitting

4.2. Binned based blade pitch curve

A Gaussian The IEC 61400-12 [14] recommended data reduction technique called the 'binning' which is generally used to calculate the power curve of a wind turbine and its associated uncertainty. In this paper, the 'method of bins' technique is applied to calculate the blade pitch curve. In the 'method of bins' 10-min average SCADA data is grouped into wind speed intervals of 0.5 m/s in order to get an average output pitch angle for each bin by using the following equations,

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

$$B_i = \frac{1}{N_i}\sum_{j=1}^{Ni} B_{n,i,j}$$

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.
$B_i =$ normalised and averaged pitch angle in bin i.
$B_{n,i,j} =$ normalised pitch angle of data set j in bin i.
$N_i =$ number of 10 min average data sets in bin i.

It should be noted that the wind speed is the most significant source of uncertainty and including more data points give more certainty to the average value in the pitch curve. Type B uncertainties would be difficult to treat in a consistent manner without greater knowledge of the instrumentation used. Therefore, in this paper, we used the statistical spread evident in the binned data. The pitch curve is calculated together with error bars and shown in Figure 7. The two standard deviations (i.e., 95% confidence intervals) of measured pitch angle values are used to obtain the error bars which is used to measure the uncertainty associated with each bin of the pitch curve. However, the 'binning' is not the most effective technique since its accuracy compromised by choosing bin width of 0.5 m/sec because within each bin the measured power or pitch will depend strongly and non-linearly on wind speed and a wide bin would result in systematic bias. Also, there is a need in practice to get sufficient data points in each bin to be of statistical significance.

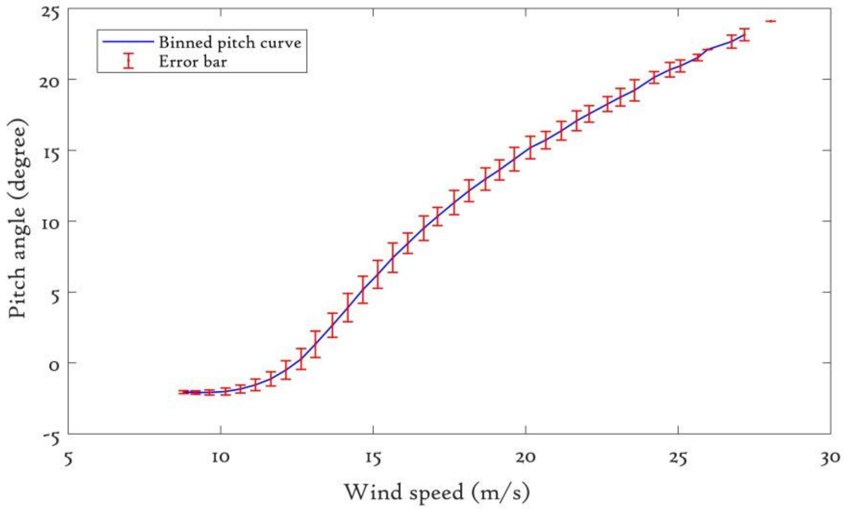

Figure 7: Binned blade pitch curve with error bars

5. Comparative studies

Conventional IEC Standard binned pitch curves described above are compared with a GP pitch curve in Figure 8; this shows that GP pitch curve model closely follows the binned pitch curve. This observation confirms that a GP is competent in estimating the wind turbine pitch curve. However, the accuracy of a GP model depends upon the quantity and quality of the data, as well as the appropriate method used. A low number of pitch angle-wind speed pairs may not give a smooth pitch curve while a high number is also not desirable because of the mathematical challenge (due to cubic inversion issue) of calculating a Gaussian Process for a large number of data points, [12].

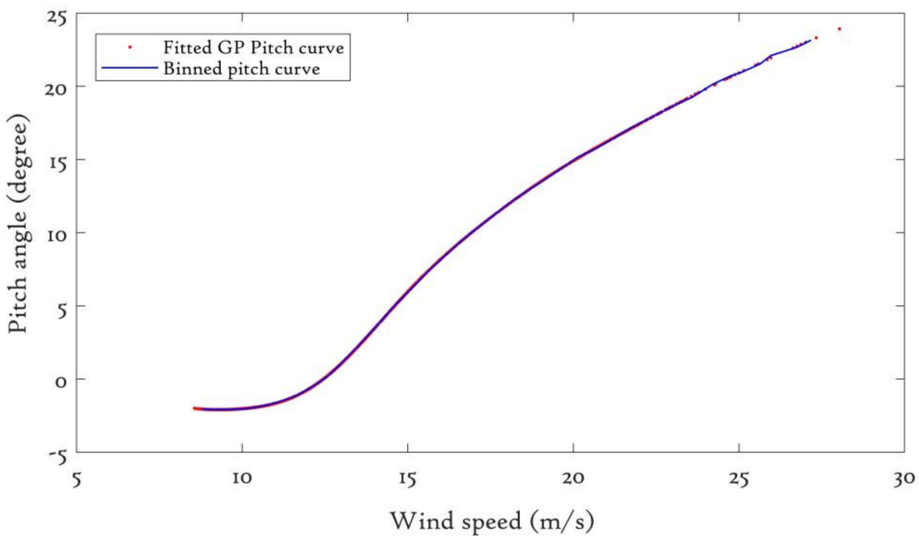

Figure 8: Comparison between binned and GP based pitch curve

A GP model comes with intrinsic confidence intervals (CIs), and this is key to wind turbine anomaly detection. These GP confidence intervals give significant information on the uncertainty surrounding an estimation, but they are also model-based estimates, [16]. Data points that lie outside of the confidence intervals can be considered anomalous, signifying a potential malfunction of the wind turbine. The GP estimated pitch curve compared with the binned pitch curve (with 95% confidence interval of error bars) in terms of uncertainty and is presented in Figure 9. This comparative analysis suggests that GP CIs are smaller than those from the binned pitch curve in almost all cases. Due to this smaller CIs, ability to reject the unhealthy or faulty data is better in GP as compared to the binned power curve, see Figure 9.

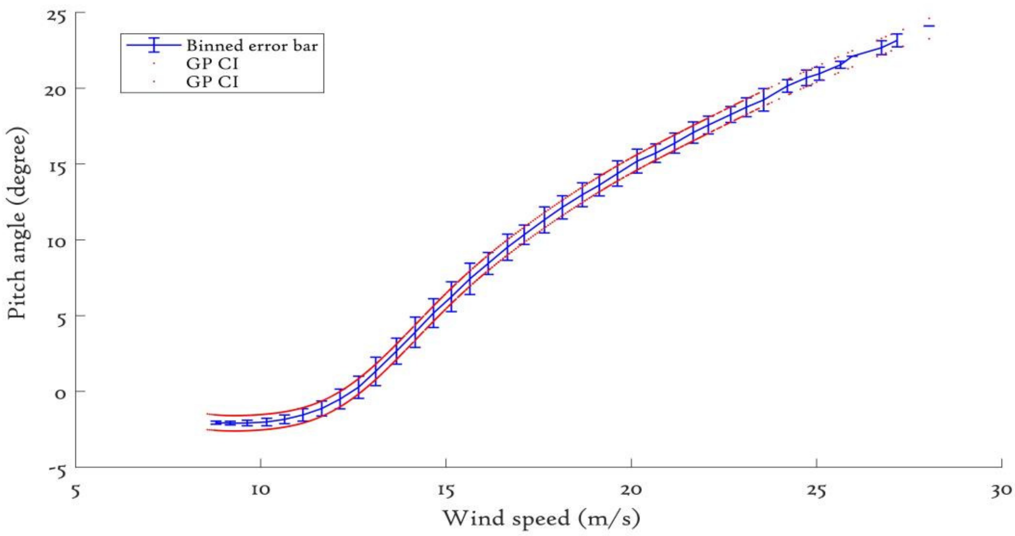

Figure 9: Comparative studies of pitch curve uncertainty based on binning and GP

6. Conclusion and Discussion

In this paper, a novel approach using a Gaussian Process to estimate blade pitch curve has been presented and assessed in comparison with a binning approach. The comparison suggests that the GP model is an effective approach, and can produce accurate results based on limited data. For a binned pitch curve, the data variation is highest between a cut in and cut out wind speed range of the pitch curve. The GP confidence limits are useful for uncertainty analysis and are crucial for robust anomaly detection. By comparing a binned pitch curve with a GP pitch curve, it has been found that the uncertainty is smaller as compared to the binned pitch curve in the cut in to cut out wind speed range.
Future work constructs a robust anomaly detection algorithm based on a Gaussian Process.

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] Global Wind Energy Council. Global Wind Report Annual Market Update. 2016. (Acessed on 1st June 2018)
[2] K. Fischer, F. Besnard, L. Bertling. Reliability-centered maintenance for wind turbines based on statistical analysis and practical experience. IEEE Trans Energy Convers, **27** (1) (2012).
[3] M. Lydia, S.S. Kumar, A.I. Selvakumar, G.E. Prem Kumar. A comprehensive review on wind turbine power curve modeling techniques. Renew. Sustain. Energy Rev., **30** (2014), pp. 452-460.
[4] J. Gottschall, J. Peinke. How to improve the estimation of power curves for wind turbines. Environ. Res. Lett., **3** (2008), Article 015005.
[5] Singh, P. (2013). Analytical Techniques of SCADA Data to Assess Operational Wind Turbine Performance. Glasgow: University of Strathclyde.
[6] 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.

[7] 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.
[8] A. Marvuglia and A. Messineo. Monitoring of wind farms power curves using machine learning techniques. Appl. Energy, vol. **98**, no. 0, pp. 574–583, 2012.
[9] R. K. Pandit and D. Infield. SCADA-based wind turbine anomaly detection using Gaussian process models for wind turbine condition monitoring purposes. IET Renewable Power Generation, vol. **12**, no. 11, pp. 1249-1255, 20 8 2018. doi: 10.1049/iet-rpg.2018.0156.
[10] R. K. Pandit and D. Infield. Gaussian Process Operational Curves for Wind Turbine Condition Monitoring. Energies, vol 11,no 7, pp. 1631, 2018. doi: 10.3390/en11071631.
[11] A. Kusiak, H. Zheng, and Z. Song. Models for monitoring wind farm power. Renew. Energy, vol. **34**, no. 3, pp. 583–590, 2009.
[12] C.E. Rasmussen, C.K.I. Williams. Gaussian processes for machine learning. the MIT Press (2006). ISBN 026218253X.
[13] T. Burton, D. Sharpe, N. Jenkind, E. Bossanyi. Wind Energy Handbook. Wiley (2001)
[14] Wind Turbines—Part 12-1: Power Performance Measurements of Electricity Producing Wind Turbines, British Standard, IEC 61400-12-1, 2006.
[15] 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.
[16] R. K. Pandit and D. Infield. Performance Assessment of a Wind Turbine Using SCADA based Gaussian Process Model. The International Journal of Prognostics and Health Management (IJPHM), vol 9, no 2, pp. 6, 2018. doi: https://www.phmsociety.org/node/2492.