Operational Variables for improving industrial wind turbine Yaw Misalignment early fault detection capabilities using data-driven techniques

Ravi Pandit, David Infield and Tim Dodwell

***Abstract*—Offshore wind turbines are complex pieces of engineering and are, generally, exposed to harsh environmental conditions that are making them to susceptible unexpected and potentially catastrophic damage. This results in significant down time, and high maintenance costs. Therefore, early detection of major failures is important to improve availability, boost power production and reduce maintenance costs.**

This paper proposes a SCADA data based Gaussian Process (GP) (a data-driven, machine learning approach) fault detection algorithm where additional model inputs, called operational variables (pitch angle and rotor speed) are used. Firstly, comparative studies of these operational variables are carried out to establish whether the parameter leads to improved early fault detection capability; it is then used to construct an improved GP fault detection algorithm. The developed model is then validated against existing methods in terms of capability to detect in advance (and by how much) signs of failure with a low false positive rate.

Failure due to yaw misalignment results in significant down time and a reduction in power production was found to be a useful case study to demonstrate the effectiveness of the proposed algorithms. Historical SCADA 10-minute data obtained from pitch-regulated turbines were used for models training and validation purposes. Results show that (i) the additional model inputs were able to improve the accuracy of GP power curve models with rotor speed responsible for a significant improvement in performance; (ii) the inclusion of rotor speed enhanced early failure detection without any false positives, in contrast to the other methods investigated.

***Index Terms*— Fault detection, condition monitoring, Gaussian Process, wind turbine.**

I. Introduction

ACCORDING to a World Wind Energy Association (WWEA) [1], wind power capacity worldwide reached 650.8 GW in 2019 out of which 59.7 GW was added in 2019 alone. Compared to onshore, offshore wind turbines (WTs) are subjected to harsher environmental and operational conditions and as more and more WTs are installed further out to sea, maintenance related activities becomes more challenging, resulting in a higher rate of catastrophic failures, significant down time and high operation and maintenance (O&M) costs [2]. Furthermore, as turbines get older, O&M cost is going to increase eventually affecting the profitability of offshore wind farms. The O&M cost further increases in case of unplanned maintenance caused by unexpected failures; this results in loss of revenue due to downtime and increases the overall Cost of Energy (CoE) [3]. Studies have shown that the spending on offshore WT O&M accounts for 25-30% of the life cycle cost of energy as compared with 10-15% for onshore wind. Part of the offshore O&M cost is accounted for by transport and logistic complexity [2]. For all these reasons, WT manufacturers and operators are continuously seeking cost-effective advanced technologies that improve WT reliability, availability to thereby minimise O&M costs.

Many state-of-art predictive maintenance as well as condition monitoring techniques [4] for various industries in past and recently started finding application in improving WTs performance and optimization related activities (e.g., early detection of failures) at reduced costs [5]. In WTs, commercial condition monitoring systems (CMS) such as, acoustic emission; oil debris analysis and vibration signal analysis are offline techniques and are costly as they require expensive sensors and extensive analysis, thereby making WT condition monitoring less cost-effective [6].By contrast, Supervisory Control and Data Acquisition (SCADA) data analysis based condition monitoring is a cost-effective approach with little or no additional cost to the wind farm operator [7, 8]. Because of rapid rise in WT installation, a huge amount of SCADA data has been collected by the wind energy industry. However, due to confidentiality and a lack of any data-sharing platform and engagement between research community and industry; access of these data is problematic [9]. Despite these challenges, the development of data-driven and big data computational

This paragraph of the first footnote will contain the date on which you submitted your paper for review. It will also contain support information, including sponsor and financial support acknowledgment. For example, "This work was supported in part by the U.S. Department of Commerce under Grant BS123456."

The next few paragraphs should contain the authors' current affiliations, including current address and e-mail. For example, F. A. Author is with the National Institute of Standards and Technology, Boulder, CO 80305 USA (e-mail: author@ boulder.nist.gov).

S. B. Author, Jr., was with Rice University, Houston, TX 77005 USA. He is now with the Department of Physics, Colorado State University, Fort Collins, CO 80523 USA (e-mail: author@lamar.colostate.edu).

T. C. Author is with the Electrical Engineering Department, University of Colorado, Boulder, CO 80309 USA, on leave from the National Research Institute for Metals, Tsukuba, Japan (e-mail: author@nrim.go.jp).

technologies support turbine condition monitoring based on SCADA data that as a result is getting more and more attention.

There have been several different approaches proposed for the WT condition monitoring based on SCADA data and these are broadly divided into parametric and nonparametric techniques. Nonparametric methods do not make strong assumptions while constructing the mapping function and therefore they are free to learn any functional form from the training data [10]. Because of this, nonparametric techniques have been found to be the most accurate in identifying key nonlinear relationships. Four methods in particular: Artificial neural network (ANNs); Gaussian Process (GP); support vector machine (SVM); and random forest (RF) are extensively used nonparametric techniques have been applied to WT performance. General reviews of these techniques in context to WTs can be found in [11-15]. In [16], a deep learning neural network was proposed to forecast wind power based on high-frequency SCADA data. Furthermore, the author of [17] used wavelets with a recursive least square (RLS) filter and a random forest model to develop a new integrated analytic framework for WT fault detection based on SCADA data.

WT manufacturers and operators extensively use the power curve to quantify turbine performance for range of applications, for example, condition monitoring, performance optimization, forecasting and improving asset life. A brief review of methods applied to SCADA data for power curve modeling can be found in [11], [18] and recently work in [14] [19]. In general, researchers have exploited any significant deviation from a reference power curve to infer operational anomalies or component failure for condition monitoring purposes. In recent years, data-driven wind turbine power curves (WTPCs) have become vital for many applications such as condition monitoring, forecasting, see for example [20],[21]. Examples of data-driven techniques for monitoring WTs are also presented in [8] and [22-25].

Recently, as an effective nonparametric, data-driven approach, GPs have been applied in a wide range of application, both in regression [26] and classification [27]. GPs provide intrinsic uncertainty estimates and can learn the noise and smoothness parameters from training data [28]. Despite these significant advantages, GP have not received much attention for WTs condition monitoring or performance monitoring activities. Recent applications of GPs models for various WTs issues can be found in [11], [14] and [29-32].

Power curves are generally provided by the turbine manufacturer for commercial purposes and most of research used only the mean wind speed at hub height and the air density as relevant input parameters for WTs condition monitoring purposes [11,14] and ignored the impact of operational variables (rotor speed & blade pitch angle) on power output. In recent years, air density, turbulence intensity and wind direction have been used to improve the power curve modeling accuracy [33-35]. For example, the latest edition of the IEC test standard, [36], though including turbulence and wind shear, disregards the importance of operational variables. All these studies are also neglected the studies on operational variables, more importantly in context to WTs condition monitoring activities. The aim of this research is to fill this gap by studying the impacts of these operational parameters and based on that proposed GP fault detection algorithm. Thereafter, the developed fault detection algorithm is then compared with existing methods in order to identify the impact of inclusion of these operational variables on improving early fault detection capability.

II. Wind Turbine Power Curve modelling

Power curve is a key tool with which to assess any underperformance issues associated with wind turbine operation. For example, severe blade erosion causes loss in power production and careful monitoring of changes to the power curve can a provide simple and cost-effective approach to condition monitoring. The power curve describes the nonlinear relationship between hub height wind speed and the power produced by a WT. A typical power curve can be divided into three regions separated by specific wind speed values, namely: i) cut in speed (the minimum wind speed at which turbine delivers useful power output); ii) rated speed (at which the maximum power of the turbine is obtained) and iii) cut out speed (at which power generation is stopped for engineering design and safety constraint reasons. Even though a power curve gives useful information about turbine performance that can be used in energy yield estimation, it exclude technical details such as such as local terrain, wind direction, turbine wakes and other factors, [36]. The power output of a WT has a roughly cubic relationship with the wind speed, which is underpinned by following equation:

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

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 tip speed ratio (TSR) λ is a dimensionless variable that depends on rotor speed and wind speed while C_p is a function of blade pitch angle and TSR. A plot of rotor speed against wind speed is called the rotor speed curve and is monotonically increasing with respect to wind speed. At the optimal rotor speed, a turbine extracts maximum power whereas pitch angle is used to limit the generator power at rated power output by reducing the angle of the blades [19]. Both these variables affect the operational behavior of a WT, and hence are called operational variables.

III. SCADA data preparation

SCADA datasets are used in this study come from an operational variable pitch regulated turbine manufactured by Siemens and rated at 2.5 MW. They record 10-minute mean, max and standard deviation values of more than 100 variables such as timestamp, wind speed, rotor speed, blade pitch angle power output, ambient temperature, atmospheric pressure and so on, and a sample of these data is shown in Table 1. Due to computation and storage issues, records comprise in the main10 min averaged data.

TABLE I
SAMPLE SCADA DATA OF A WIND TURBINE

TimeStamp	Wind speed (Avg.) m/sec	Power (Avg.) kW	Ambient temp (Avg.) ℃	Atmospheric pressure (Avg.) mbar	Rotor speed (Avg.) m/sec	Blade pitch angle (Avg.) ℃
12/ 03/2009 10:00:00	5.05	270.93	7.44	986.35	9.57	-0.99
12/ 03/2009 10:10:00	5.07	230.45	7.85	986.45	8.75	-0.99
12/ 03/2009 10:20:00	6.09	150.72	7.90	986.47	7.98	-0.99
12/ 03/2009 10:30:00	6.10	255.20	8.40	986.55	9.30	-0.99
12/ 03/2009 10:40:00	6.15	240.15	8.80	986.58	8.35	-0.99

TABLE II
WIND TURBINES SCADA DATA DESCRIPTIONS

Start timestamp	End timestamp	Measured data	Filtered data
11/3/2009 14:30 PM	30/03/2009 15:20 PM	4725	3274

The raw data obtained from SCADA systems incorporate a number of different kinds of errors due to sensor malfunction and communication failures that, if not excluded, can affect model accuracy and condition monitoring effectiveness. Thus, the first step is to reprocess these data prior to use in condition monitoring. At first, samples with missing values or negative power values are filtered out. Data points where maximum wind speed has reached more than 25 m/s are also filtered out because, beyond this wind speed, the turbine is stopped. Besides, data sampling during frequent start-up or stop in the low-wind-speed period may have a different variation. Overall, criterion such as timestamp mismatch, negative power values, out of range values and turbine curtailment is used to filter out such misleading data similar to the one described in [14, 15]. Table II summarises a SCADA data file beginning with time stamp ''11/3/2009 14:30 PM'' and ending at time stamp ''30/03/2009 15:20 PM'' that records 4725 measured values which were reduced to 3274 data points after pre-processing using criterion as stated above.

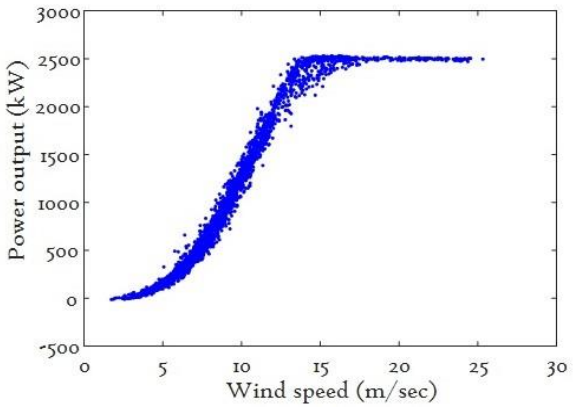

Fig. 1. Filtered and air density corrected power curve

The next pre-processing step is to undertake air density correction to adjust the data to reflect the fact that according to equation 1, wind turbine power output at a given wind speed depends on air density. For a variable pitch regulated wind turbine and as per IEC standard 61400-12-1 [36] the following equations are used for calculating the air density correction:

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

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

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. Fig 1 shows the air density corrected and pre-processed power curve and will be used in next section to construct the reference power curve model based on a GP algorithm.

IV. GAUSSIAN PROCESS METHODOLOGY

A GP is a data-driven, probabilistic technique that includes Gaussian- distributions over the function; it has strengths in uncertainty quantification and function approximation. GP models are flexible in that they not predefine the relationship between input and output variables to a specific form. The theoretical description of GP models are well covered in [28]. In this study, a GP for wind turbine power curve modelling is outlined as follows. GP regression is defined in terms of a mean function, $m(x) := E[f(x)]$ and covariance functions, $K(x,x') := E\big[\big(f(x) - m(x)\big)\big(f(x') - m(x')\big)\big]$ for a given values x, x' and if $f(x)$ is a GP distributed function, then the relationship between these two functions can be expressed as:

$$f(x) \sim GP\big(m(x), K(x,x')\big) \quad (4)$$

The mean function $m(x)$ often constructed be zero for notational simplicity, however, its value can be arbitrarily selected. Covariance function $K(x,x')$ quantifies the joint variability of the random variables, used to measure distance or similarity between given data points (x,x'). There are number of different covariance functions available; these are well described in [28]. However, the squared exponential covariance function was found to be effective as suggested by [37], and will be be used in this study. The squared exponential function is mathematically expressed as:

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

To compensate for the effect of measurement noise, a noise term added into the squared exponential and thus, equation (5) modified to be:

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

Where σ_f^2 and l are the hyper-parameters in which σ_f^2 signifies

the signal variance while l (length scale) describes how quickly the covariance decreases with distance between points. The model uncertainty is quantified by σ_n, the standard deviation of the noise fluctuations, and δ is the Kronecker delta. Let us consider that $A = \{(x_i, y_i), i = 1, \dots \dots, N\}$ of n observations is the training dataset. x is the input vector of dimension D , and y is the scalar output. The $A \times n$ matrix defines the input datasets. Target output is y, therefore, $A = (X, y)$. Theoretically, the relationship between input and target values for a GP can be expressed as:

$$y_i = f(x_i) + \epsilon_i \quad (7)$$

Equation (7) used to define the underlying function of the data modeled where x are values from the training datasets and ϵ is Gaussian white noise of variance σ_n^2 so that, $\epsilon = N(0, \sigma_n^2)$. And the prior to y becomes:

$$E|\mathrm{y}| = \mathrm{E}|f + \epsilon| = 0 \quad (8)$$

$$cov\ |\mathrm{y}| = K|X, X| + \sigma_n^2 I \quad (9)$$

The prior distribution contains vital information about uncertain parameters and it can be uninformative or informative and since GP regression is based on Bayesian analysis. The prior distribution along with the probability distribution of new incoming data is used to generate the posterior distribution. Thereafter, the estimated posterior distribution will be used for future inference and any decisions involving the uncertain parameters [28,31]. To predict the output f, for a given new input x^*, the distribution can be defined as follows :

$$\begin{pmatrix} y \\ f^* \end{pmatrix} \sim N\left(0, \begin{bmatrix} K(X,X) + \sigma_n^2 I & k(X, x^*) \\ k(x^*, X) & k(x^*, x^*) \end{bmatrix}\right) \quad (10)$$

Where,

$k(X, x^*) = k(x^*, X)^T = [k(x_1, x^*), k(x_2, x^*), \dots . k(x_n, x^*)]$, which is for the sake of simplicity, denoted by k^*. Then, from the joint Gaussian distribution, the estimation of target values is given by:

$$\overline{f*} = k_*^T (K + \sigma_n^2 I)^{-1} y \quad (11)$$

$$Var[f*] = \mathrm{k}(x^*, x^*) - k_*^T k^* (K + \sigma_n^2 I)^{-1} \quad (12)$$

The obtained posterior variance ($Var[f*]$) is inversely proportional to the distance between test and training data points while estimation of the mean ($\overline{f*}$) is a linear combination of the output y in which linear weights are defined $k_*^T (K + \sigma_n^2 I)^{-1}$. Equation (11) and (12) estimate the mean and variance of the model for a given data points.

The optimal values of the hyperparameters are going to be identified through maximising GP model accuracy and hence, in this paper, hyperparameters are tuned using Bayesian optimization techniques where optimization attempts to minimize the cross-validation loss or error for GP regression by varying the parameters. To do this, the 'fitrgp' function of the MATLAB with the 'automatic hyperparameters optimization' option has been used [38]. Furthermore, the initial value of σ_n is calculated by, $\frac{std(y)}{\sqrt{2}}$ where y is the response variable, realized in MATLAB. To calculate log-likelihood and gradient, the QR factorization technique is adopted as it yields better accuracy as compared to V-method-based technique [28, 38]. GP models estimate confidence intervals (CIs) (that reflect the uncertainty of the model) for the predicted function that are useful in uncertainty quantification. Using equation (12), these are calculated as follows [28],

$$CI_n = \mu_n \pm 2\sigma_n \quad (9)$$

It should be noted that GP uncertainty uses probabilistic descriptions of the model input that can be used to derive probability distributions of model outputs and system performance indices. GPs are multivariate models where the covariance matrix, K, gives the variance of each variable along the leading diagonal, and the off-diagonal elements measure the correlations between the different variables using following relationships:

$$K = \begin{bmatrix} k_{11} & \cdots & k_{1n} \\ \vdots & \ddots & \vdots \\ k_{n1} & \cdots & k_n \end{bmatrix} \quad \text{where } k_{ij} = k(x_i, x_j)$$

where K is of size $n \times n$, where n is the number of input data points considered, and it must be symmetric and positive semidefinite i.e. $K_{ij} = K_{ji}$. $n = number\ of\ inputs\ for\ GP\ regression\ mode$ used to incorporate numbers of input variables where x is the wind speed along with operational parameters, rotor speed and blade pitch angle in our case. In [39, 40], rotor speed was found to be a fundamental covariate for improving data-driven model accuracy whose target is WT power. Therefore, rotor speed along with wind speed are used as input variables to train and validate the GP power curve model using the data outlined in section 3 and the methodologies described above.

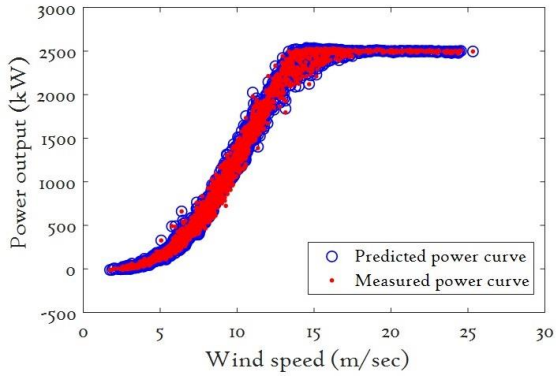

Fig. 2. GP power curve incorporating rotor speed

Fig. 2 depicts the estimated and measure power curve and suggests strongly that GP model accuracy is improved by incorporating rotor speed. In addition, blade pitch angle has also been incorporated together with rotor speed together in the model. The inclusion of blade pitch angle makes insignificant improvement does in GP model accuracy as well as uncertainty which further supports the conclusion of [39,40]. In addition, Fig 3 and calculated performance error metrics (shown in Table III) suggest that the inclusion of rotor into GP model makes significant improvement in accuracy as well as uncertainty as compare to others. Thus, hinting, rotor speed will be used in developing GP fault detection algorithm.

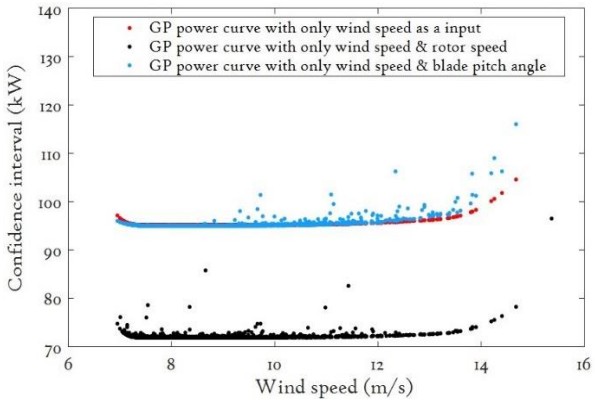

Fig. 3. Impact of operational variables on GP model uncertainty

TABLE III
PERFORMANCE ERROR METRICS

Models	RMSE	MAE	R^2
GP without rotor speed	45.56	36.52	00.9810
GP with rotor speed	32.36	27.12	00.9979

Computational difficulties in dealing with extensive data sets (the cubic inversion issue), and restrictive modelling assumptions for complex data sets are considered to be two main disadvantages of GPs. Many methods have been proposed to address these issues [41, 42], but these methods require high processing power and computational cost. For robust and effective GP modelling, a balance between the number of data points and computational cost must be struck.

V. YAW ERROR MISALIGNMENT – A CASE STUDY

Yaw misalignment is due to the difference between nacelle direction and wind direction, termed as yaw error. Early detection of yaw misalignment improves power generation, minimises damaging stress loads and fatigue on the wind turbine rotor and drive train and thus increases performance and profitability. An example of yaw failure provides an excellent test case since increased yaw error diminishes wind turbine performance. The yaw failure is exhibited in Fig 4, where it can be seen that the nacelle is stuck in a fixed position (roughly 200 degrees) for an extended period of time, with no yaw activity, whereas the wind direction changed in a normal manner during this period.

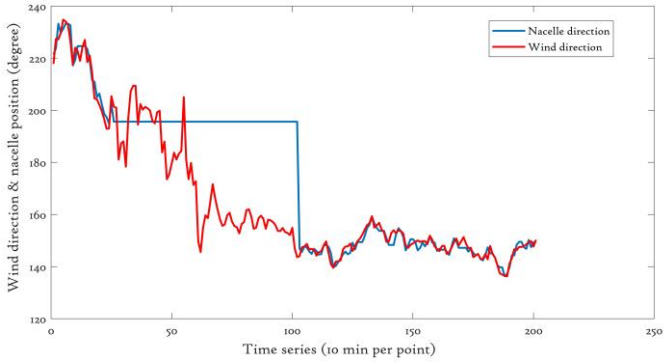

Fig. 4. Time series of wind direction and nacelle position

VI. GP FAULT DETECTION ALGORITHM INCORPORATING ROTOR SPEED

As stated above, power curves are considered to be a key indicator and can be used to identify specific component anomalies if interpreted with care. The GP algorithm incorporating rotor speed is applied to automated detection of yaw misalignment. A reference power curve model based on GP modelling of a healthy turbine was constructed using SCADA data based on the mythologies outlined in section IV. Fig 5 and fig 6 shows the estimated GP power curve with and without the inclusion of rotor speed and suggest that the incorporating rotor speed into GP model narrowed down the CIs significantly and thus should be advantageous in identifying anomalous performance quickly. To validate this, fault detection makes use of the 99.5% confidence level (i.e. a significance level of 0.05) and was used to calculate sequential anomalous data point values at each time for the reference GP power curve of Fig 6.

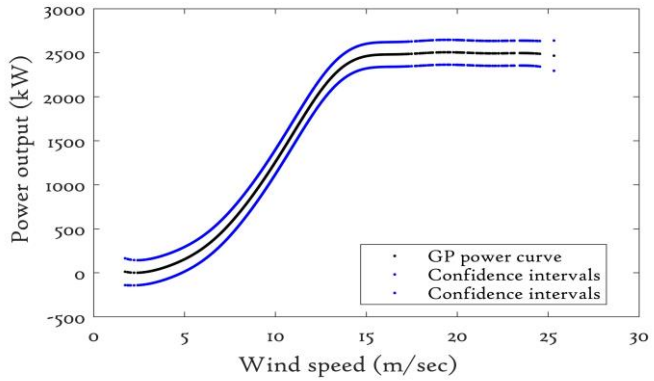

Fig. 5. GP power curve without rotor speed

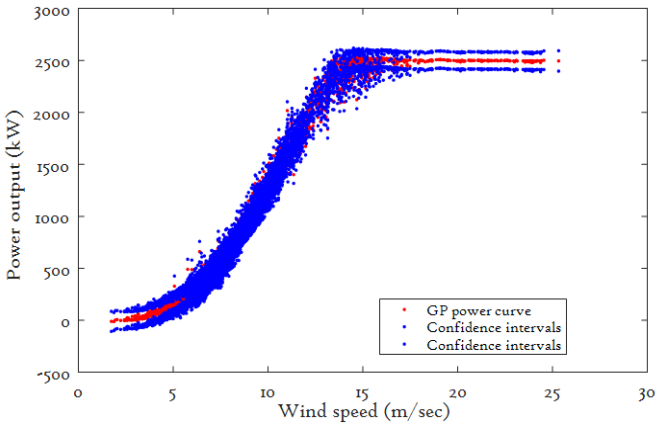

Fig. 6. GP power curve incorporating rotor speed

It should be noted that both Fig 5 and 6 shows the modified CIs in order to compensate for the impact of noise as described in equation 6. Modified CIs are then used to assess incoming sequential data points based on a point-by-point probabilistic calculation to identify yaw failure data points. The hypothesis testing p-value or probability value was kept at a threshold of 0.003 to filter individual faulty incoming points. False alarms affect both reliability and O&M costs of WTs, therefore

attention is paid to ensure that the algorithm generates no false alarms. This is done by adjusting the probability threshold until no false alarms occur. The value of 0.003 yields the most accurate results with no false positives, and is therefore, used in this algorithm.

Fig 7 shows a time series plot of the absolute yaw error; it indicates that the yaw error exceeds 20 degrees consistently for timestamps 50 to 100 due to faulty yaw control or drive. This is confirmed by fig 4 where it is clear that the nacelle is stuck and does not follow the wind direction. The alarms generated by the GP fault detection algorithm are also plotted in Fig 7 and it is found that an alarm is first raised at 21:40 on 14/04/2009, just after 40 minutes after the start of yaw misalignments at 21:00 on 14/04/2009 without any false positive, confirming that the proposed GP fault detection algorithm robust.

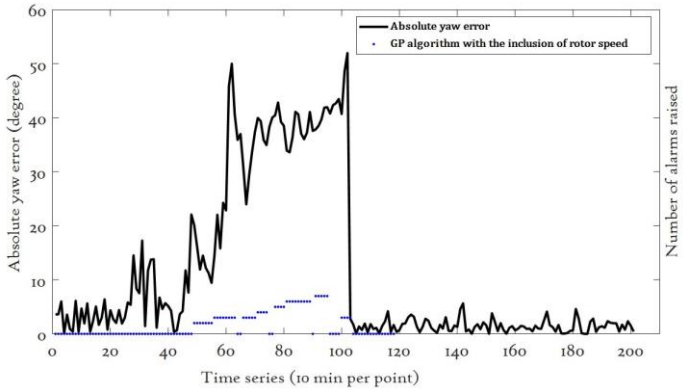

Fig. 7. Absolute yaw error detection using online power curve model

VII. COMPARATIVE STUDIES

Author of [14], proposed an initial GP algorithm for yaw error detection in which power curve relations were used to some effect; this thus provided a good benchmark against which to validate the effectiveness of the proposed improved model. They considered only wind speed and air density for constructing GP fault detection algorithm and used a Fisher test with a threshold of 0.008 to filter the individual p-values. This former model is plotted together with proposed model for the yaw error time-series and is shown in Fig 8. By comparing them, it has been found that whilst the GP incorporating rotor speed fault detection algorithm took only 40 minutes to detect the yaw failure, this simpler model took 1.5 hrs to detect the same yaw failure, as summarized in Table IV.

Furthermore, inclusion of rotor speed not only increase the early detection capability but also records no false positives in contrast to that from the former model highlighted by the circle in Fig 8.

TABLE IV
ALARM RECORD AND DETECTION BY GP MODELS

Model	Alarm detected	Time is taken to identify the fault
Probabilistic assessment using GP	22:30 on 14/4/2009	1.5 hours
Probabilistic assessment using GP with the inclusion of rotor speed	21:40 on 14/4/2009	40 minutes

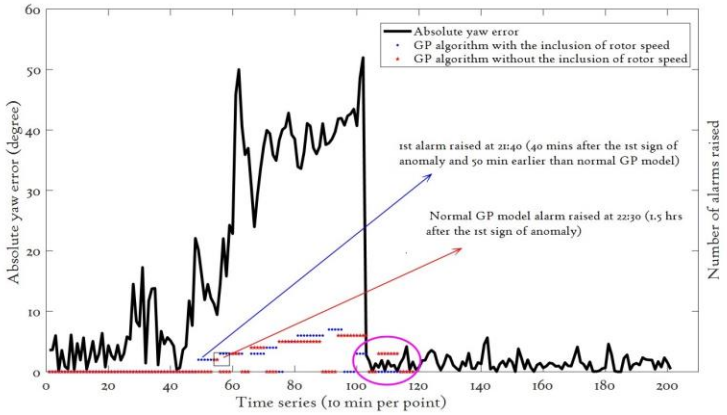

Fig. 8. Impact of rotor speed on GP fault detection algorithm

VIII. CONCLUSIONS

Wind turbine power curves have been used extensively for energy assessments, warranty formulations and forecasting purposes, and have recently started to find applications in condition monitoring related activities based on SCADA data. This study uses operational parameters (rotor speed and blade pitch angle) to improve GP model based fault detection accuracy and thereby help to reduce O&M costs. The SCADA data collected from an operational WT was used to validate the effectiveness of the proposed approach. Results show that the significant improvement in GP model accuracy as well as reduced uncertainty is achieved through the inclusion of rotor speed as shown in Fig 2-3 and Table III. Based on this outcome, a GP fault detection algorithm based on SCADA data incorporating rotor speed is proposed for early detection of yaw failures. This is then compared with existing and effective method to validate the improved effectiveness of the proposed model. The comparative analyse suggest that a GP fault detection algorithm incorporating rotor speed significantly enhances the early fault detection capability of the GP model (which took only 40 minutes to detect the first sign of yaw error) while other earlier approach without rotor speed took 1.5 hrs to detect the same yaw failure as shown in Table IV. In addition, importantly the improved GP based fault detection algorithm incorporating rotor speed generated no false positives. In summary proposed technique not only improves early failure detection preventing catastrophic damage but also provides a significant time window for the turbine operator to carry out repair work thereby reducing downtime and also reducing O&M costs. Future work will apply the approach to a range of different wind turbine faults and test with other data-driven algorithms. Future work will also look, compared other machine learning algorithms (such as SVM, Random forest), and compare them against proposed techniques to validate its effectiveness.

ACKNOWLEDGMENT

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] World Wind Energy Association report on "Global Wind Installation," https://library.wwindea.org/global-statistics/.

[2] Yang W, Tavner P J, Wilkinson M R, "Condition monitoring and fault diagnosis of a wind turbine synchronous generator drive train," IET Renewable Power Generation, vol 3, DOI: 10.1049/iet-rpg:20080006, no 1,pp. 1–11, March 2009.

[3] Article on 'US wind O&M costs estimated at $48,000/MW; Falling costs create new industrial uses: IEA'. https://analysis.newenergyupdate.com/wind-energy-update/us-wind-om-costs-estimated-48000mw-falling-costs-create-new-industrial-uses-iea.

[4] H. M. Hashemian and W. C. Bean, "State-of-the-Art Predictive Maintenance Techniques," in IEEE Transactions on Instrumentation and Measurement, vol. 60, no. 10, pp. 3480-3492, Oct. 2011, doi: 10.1109/TIM.2009.2036347.

[5] B. Manobel, F. Sehnke, J.A. Lazzús, I. Salfate, M. Felder, S. Montecinos, "Wind turbine power curve modeling based on Gaussian Processes and Artificial Neural Networks," Renewable Energy, vol 125, DOI: 10.1016/j.renene.2018.02.081, pp. 1015-1020, 2018.

[6] S. Sheng, "Monitoring of wind turbine gearbox condition through oil and wear debris analysis: A full-scale testing perspective," *Tribol. T.*, vol. 59, no. 1, pp. 149-162, 2016.

[7] J. Tautz-Weinert and S. J. Watson, "Using SCADA data for wind turbine condition monitoring – a review," in IET Renewable Power Generation, vol. 11, doi: 10.1049/iet-rpg.2016.0248, no. 4, pp. 382-394, 2017.

[8] M. Rezamand, M. Kordestani, R. Carriveau, D. S. . -K. Ting, M. E. Orchard and M. Saif, "Critical Wind Turbine Components Prognostics: A Comprehensive Review," in IEEE Transactions on Instrumentation and Measurement, vol. 69, no. 12, pp. 9306-9328, Dec. 2020, doi: 10.1109/TIM.2020.3030165.

[9] A. Kusiak , "Share data on wind energy," Nature, vol. 529, no. 7584, pp. 19–21, 2016.

[10] Stuart Russell and Peter Norvig, "Artificial Intelligence: A Modern Approach ," 3rd edition. Pearson Publisher, 2009. ISBN-13: 978-0136042594.

[11] R. Pandit, D. Infield, A. Kolios, "Comparison of advanced nonparametric models for wind turbine power curves," in IET Renew. Power Generation., vol 13,DOI: 10.1049/iet-rpg.2018.5728, no 9, pp. 1503-1510, 2019.

[12] Yang and Z. Zhang, "Wind Turbine Gearbox Failure Detection Based on SCADA Data: A Deep Learning-Based Approach," in IEEE Transactions on Instrumentation and Measurement, vol. 70, pp. 1-11, 2021, Art no. 3507911, doi: 10.1109/TIM.2020.3045800.

[13] N. Huang, Q. Chen, G. Cai, D. Xu, L. Zhang and W. Zhao, "Fault Diagnosis of Bearing in Wind Turbine Gearbox Under Actual Operating Conditions Driven by Limited Data With Noise Labels," in IEEE Transactions on Instrumentation and Measurement, vol. 70, pp. 1-10, 2021, Art no. 3502510, doi: 10.1109/TIM.2020.3025396.

[14] R. K. Pandit and D. Infield, "SCADA-based wind turbine anomaly detection using Gaussian process models for wind turbine condition monitoring purposes," in IET Renewable Power Generation, vol. 12, DOI: 10.1049/iet-rpg.2018.0156, no. 11, pp. 1249-1255, 2018.

[15] Z. Wang, L. Wang and C. Huang, "A Fast Abnormal Data Cleaning Algorithm for Performance Evaluation of Wind Turbine," in IEEE Transactions on Instrumentation and Measurement, doi: 10.1109/TIM.2020.3044719.

[16] Zi Lin, Xiaolei Liu, "Wind power forecasting of an offshore wind turbine based on high-frequency SCADA data and deep learning neural network," in Energy, vol. 201, DOI: doi:10.1016/j.energy.2020.117693, 117693, ISSN 0360-5442..

[17] Shiyao Qin, Mengzhou Zhang, Xiaojing Ma & Mei Li., "A new integrated analytics approach for wind turbine fault detection using wavelet, RLS filter and random forest," Energy Sources, Part A: Recovery, Utilization, and Environmental Effects, DOI: 10.1080/15567036.2019.1677815.

[18] 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, vol 30, DOI: 10.1016/j.rser.2013.10.030, pp. 452-460,2014.

[19] A. Kusiak, W. Li., "The prediction and diagnosis of wind turbine faults," Renewable Energy, vol 36, DOI: 10.1016/j.renene.2010.05.014, pp. 16-23, 2011.

[20] Z. Lin, X. Liu, M. Collu., "Wind power prediction based on high-frequency SCADA data along with isolation forest and deep learning neural networks," Int J Electr Power Energy Syst, vol 118, DOI: 10.1016/j.ijepes.2020.105835, p no. 105835.

[21] Y. Zhao, L. Ye, W. Wang, H. Sun, Y. Ju and Y. Tang, "Data-Driven Correction Approach to Refine Power Curve of Wind Farm Under Wind Curtailment," in *IEEE Transactions on Sustainable Energy*, vol. 9, no. 1, doi: 10.1109/TSTE.2017.2717021, pp. 95-105, Jan. 2018.

[22] M. Catelani, L. Ciani, D. Galar and G. Patrizi, "Optimizing Maintenance Policies for a Yaw System Using Reliability-Centered Maintenance and Data-Driven Condition Monitoring," in IEEE Transactions on Instrumentation and Measurement, vol. 69, no. 9, pp. 6241-6249, Sept. 2020, doi: 10.1109/TIM.2020.2968160.

[23] X. Yu, B. Tang and K. Zhang, "Fault Diagnosis of Wind Turbine Gearbox Using a Novel Method of Fast Deep Graph Convolutional Networks," in IEEE Transactions on Instrumentation and Measurement, vol. 70, pp. 1-14, 2021, Art no. 6502714, doi: 10.1109/TIM.2020.3048799.

[24] L. Lu, Y. He, Y. Ruan and W. Yuan, "Wind Turbine Planetary Gearbox Condition Monitoring Method Based on Wireless Sensor and Deep Learning Approach," in IEEE Transactions on Instrumentation and Measurement, vol. 70, pp. 1-16, 2021, Art no. 3503016, doi: 10.1109/TIM.2020.3028402.

[25] A. Kusiak and A. Verma, "Monitoring wind farms with performance curves," IEEE Trans. Sustain. Energy, vol. 4, DOI: 10.1109/TSTE.2012.2212470, no. 1, pp. 192–199, Jan 2013.

[26] Arthur, C.K., Temeng, V.A. & Ziggah, Y.Y., "Novel approach to predicting blast-induced ground vibration using Gaussian process regression," in Engineering with Computers, vol 36, DOI: 10.1007/s00366-018-0686-3, pp. 29–42, Jan 2020.

[27] J. Li, G. Lu, B. Zhang, J. You and D. Zhang, "Shared Linear Encoder-Based Multikernel Gaussian Process Latent Variable Model for Visual Classification," in IEEE Transactions on Cybernetics, DOI: 10.1109/TCYB.2019.2915789, Early access.

[28] Rasmussen C, Williams C. Gaussian Processes for Machine Learning. Publisher: MIT Press. 2005. ISBN 026218253X.

[29] X. Jiang, B. Dong, L. Xie, and L. Sweeney, "Adaptive Gaussian process for short-term wind speed forecasting," in 19th European Conference on Artificial Intelligence, Proceedings, ser. Frontiers in Artificial Intelligence and Applications, vol. 215, pp. 661–666, August, 2010,.

[30] E. Papatheou, N. Dervilis, A. E. Maguire, I. Antoniadou and K. Worden, "A performance monitoring approach for the novel Lillgrund offshore wind farm," *IEEE Trans. Ind. Electron.*, vol. 62, no. 10, pp. 6636-6644, Oct. 2015.

[31] P. Guo and D. Infield, "Wind Turbine Power Curve Modeling and Monitoring With Gaussian Process and SPRT," in IEEE Transactions on Sustainable Energy, vol. 11, no. 1, pp. 107-115, Jan. 2020, doi: 10.1109/TSTE.2018.2884699.

[32] R. K. Pandit and D. Infield, "Performance Assessment of a Wind Turbine Using SCADA based Gaussian Process Model," in The International Journal of Prognostics and Health Management (IJPHM), vol 9, doi: https://www.phmsociety.org/node/2492, no 2, pp. 6, 2018.

[33] R.K. Pandit, D. Infield, J. Carroll, "Incorporating air density into a Gaussian process wind turbine power curve model for improving fitting accuracy," in Wind Energy, DOI:10.1002/we.2285, pp.1–14, 2018.

[34] L.M. Bardal, L.R. Sætran, "Influence of turbulence intensity on wind turbine power curves," in Energy Procedia, vol 137, DOI: 10.1016/J.EGYPRO.2017.10.384, pp. 553-558, 2017.

[35] M Jafarian, A Soroudi, M. Ehsan, "The effects of environmental parameters on wind turbine power pdf curve," CCECE (2008), pp. 001193-001198.

[36] IEC Standard 61400–12-1, 2017.Wind turbines—part 12–1: power performance measurements of electricity producing wind turbines (IEC 61400-12-1:2017).

[37] R. K. Pandit and D. Infield, "Comparative analysis of Gaussian process power curve models based on different stationary covariance functions for the purpose of improving model accuracy," in Renewable Energy, vol. 140, DOI: 10.1016/j.renene.2019.03.047,pp. 190–202, 2019.

[38] Gaussian Process Regression Models, Matlab toolbox.

[39] Astolfi, D., Castellani, F., Fravolini, M. L., Cascianelli, S., & Terzi, " Precision Computation of Wind Turbine Power Upgrades: An Aerodynamic and Control Optimization Test Case," in Journal of Energy Resources Technology, vol 14, DOI: 10.1115/1.4042450, no 5, May 2019.

[40] Ravi Kumar Pandit, David Infield, Athanasios Kolios, "Gaussian process power curve models incorporating wind turbine operational variables," in Energy Reports, vol 6, DOI:10.1016/j.egyr.2020.06.018, pp. 1658-1669, ISSN 2352-4847, June 2020.

[41] Hartikainen J, Särkkä S, "Kalman filtering and smoothing solutions to temporal Gaussian process regression models," In: Proceedings of IEEE International Workshop on Machine Learning for Signal Processing, DOI:10.1109/MLSP.2010.5589113, pp. 379-384, 2010.

[42] Sarkka S, Solin A, Hartikainen," Spatiotemporal Learning via Infinite-Dimensional Bayesian Filtering and Smoothing: A Look at Gaussian Process Regression Through Kalman Filtering," in IEEE Signal Processing Magazine, vol 30, DOI: 10.1109/MSP.2013.2246292, no 4, pp. 51–61, 2013.

Dr Ravi Pandit is a Research Fellow in the Institute of Data Science and Artificial Intelligence (IDSAI) of the Computer Science Department, University of Exeter. Dr Pandit received PhD from the University of Strathclyde in 2019. Dr Pandit Bachelor and Master degree is from Jadavpur University (2009) and Indian Institute of Technology, Kharagpur (2011) respectively.

From 2011 to 2016, he worked as assistant professor at Jadavpur University and Vellore Institute of Technology. His areas of Interest Are data-driven application on offshore wind including condition monitoring, predictive maintenance, forecasting & Prediction and SCADA data statistical analysis. Dr. Pandit have more than 4 years of direct research experiences in these areas.

Professor David Infield received a B.A. degree in mathematics and physics from the University of Lancaster, Lancaster, U.K. and the PhD degree in applied mathematics from the University of Kent, Canterbury, U.K.

He worked for the Rutherford Appleton Laboratory in Oxfordshire, U.K., from 1982 to 1993 researching into wind electricity systems. From 1993 to 2007, he was with Loughborough University, Leicestershire, U.K., where he established CREST, the Centre for Renewable Energy Systems Technology. He is now Research Professor of Renewable Energy Technologies with the Institute for Energy at the University of Strathclyde.

Professor Tim Dodwell holds a chair in Computational Mechanics at the University of Exeter, and holds a prestigious 5-year Turing AI Fellowship at the Alan Turing Institute. He obtained a PhD in Applied Mathematics at Bath (2012), since then has sequence of fellowship as a faculty member at Bath and then Exeter. Now at Exeter he leads an interdisciplinary Data Centric Engineering Group of 17 researchers, which do leading research at the dynamic interface between applied mathematics, statistics and high-performance scientific computing; spanning all aspects of research from fundamental theory in data science and AI to applied industrial focused projects.