Energy Reports 11 (2024) 2577–2591

Contents lists available at ScienceDirect

Energy Reports

journal homepage: www.elsevier.com/locate/egyr

Research paper

Conditional monitoring and fault detection of wind turbines based on Kolmogorov–Smirnov non-parametric test

Olayinka S. Ohunakin [a,b,*], Emerald U. Henry [a,**], Olaniran J. Matthew [c], Victor U. Ezekiel [a], Damola S. Adelekan [a], Ayodele T. Oyeniran [d]

[a] The Energy and Environment Research Group (TEERG), Mechanical Engineering Department, Covenant University, Ogun State, Nigeria
[b] Faculty of Engineering & the Built Environment, University of Johannesburg, South Africa
[c] Institute of Ecology and Environmental Studies, Obafemi Awolowo University, Ile-Ife, Osun State, Nigeria
[d] Department of Mechanical Engineering, Obafemi Awolowo University, Ile-Ife, Osun State, Nigeria

ARTICLE INFO

Keywords:
Wind turbine
Condition monitoring
Power curve
Cochran test
Kolmogorov–Smirnov test
SCADA

ABSTRACT

This research presents a new method for conditional monitoring based on the wind turbine power curve. The Kolmogorov-Smirnov (K-S) distribution test is employed in the assessment of turbine data and the detection of abnormality (faults) in wind turbines. The process begins with anomaly detection and filtration of faulty SCADA data by a quantile-based filtration approach. Suitable data comprising wind speed, air density, ambient temperature, and pitch angle are utilized in the development of wind turbine power curve models that represents actualities within wind farms. The radial basis function (RBF), multi-layer Perceptron (MLP), and gradient boosting (GBR) methods utilized for model development are compared for predictive accuracy using Mariano-Preve test. The null hypothesis assumes equal predictive ability (EPA); if rejected, an algorithm compares the coefficients of correlation of the models and selects the closest to one (unity). The most accurate model is utilized for the creation of a bin-wise distribution from past data, and bin-wise confidence levels from the plot of wind speed and output power. Cochran's method was utilized to validate the minimum sample size that will possess a sampling distribution similar to that of the population, and a fault is detected if there is a reasonable difference between the sample distribution and population distribution. The K-S test, having a null hypothesis of equivalent distributions, signals a fault if the null hypothesis is rejected. Two wind turbine SCADA datasets associated with two fault events are used for the assessment of our method. The results indicate that our method effectively discovers abnormalities in power output relating to increased bearing temperature and reduced generator rpm, thereby aiding in the detection of faults long before they occur.

1. Introduction

Wind energy has acquired much attention in recent times because it is available anywhere and lacks pollution, aside from being one of the most potent renewable energy sources (Ohunakin et al., 2023). Wind energy development could address the environmental problems associated with fossil fuels, and lead to a more sustainable future (Matthew and Ohunakin, 2017). The Global Wind Energy Council (GWEC) recorded an increase in wind power installations by 93 GW for 2021 alone, thus making the total installed capacity 743 GW for both offshore and onshore (Global Wind Energy Council, 2021). However, due to the complexity of wind turbine assembly and the ever-changing harsh operating conditions on sites where they are deployed, they experience high failure rates due to failures associated with gearbox, generators, and blades (Akay et al., 2013; Kusiak and Verma, 2012; Ohunakin et al., 2011). Because of the size of these components, their failures are very expensive and usually lead to long downtime (Kusiak and Li, 2011). It is therefore necessary to monitor and detect faults before they transit to critical failures (Sun et al., 2016). Hence, condition-based monitoring of wind farms has been considered the most efficient solution for the effective maintenance of wind turbines (Sun et al., 2016; García Márquez et al., 2012; Castellani et al., 2017; Stetco et al., 2019; Artigao et al., 2018). A detailed review of condition-based monitoring as well as challenges and future progress is given in the work of Sun et al. (2022a).

* Corresponding author at: The Energy and Environment Research Group (TEERG), Mechanical Engineering Department, Covenant University, Ogun State, Nigeria.
** Corresponding author.
E-mail addresses: olayinka.ohunakin@covenantuniversity.edu.ng (O.S. Ohunakin), emerald.henry@stu.cu.edu.ng (E.U. Henry).

https://doi.org/10.1016/j.egyr.2024.01.081
Received 26 April 2023; Received in revised form 17 November 2023; Accepted 31 January 2024
Available online 20 February 2024

The available methods employed for wind turbine (WT) condition monitoring (CM) can be grouped into signal-based, physical model, and data-driven approaches. Signal-based methods require the installation of sensors for information capture, leading to a significant increase in O&M costs. The methods include acoustic emission (Soua et al., 2013; Feng et al., 2015), infrared (Gómez Muñoz et al., 2016), vibration analysis (Wei et al., 2015; de Novaes Pires Leite et al., 2021; Xu et al., 2020), etc. Physical model methods often require expert knowledge of various systems and their interactions, in order to develop an accurate physical model; it is also very difficult to obtain accurate physical models because of the system-to-system interactions (Dey et al., 2015). Therefore, the physical model approach focuses on developing component-specific physical models, such as physical model representations for gears and bearings (Odgaard and Stoustrup, 2015). Data-driven approaches develop methods that rely on the data collated by the turbine's supervisory control and data acquisition (SCADA) system (Stetco et al., 2019; Maldonado-Correa et al., 2020; Dao et al., 2018a; Gonzalez et al., 2019; Wang et al., 2014; Dao, 2022a; Tautz-Weinert and Watson, 2017; Dao, 2022b; Kusiak and Zhang, 2010; Yang et al., 2013; Schlechtingen et al., 2013; Meyer, 2021; Qu et al., 2020). This is a cost-effective approach because no additional cost is incurred for the installation of sensors or other devices (Stetco et al., 2019). Additionally, due to the diversity of parametric and nonparametric techniques available in the literature, they are considered as the most effective solution to WT conditional monitoring; the techniques utilize change in the data distribution (decrease or increase) of the parameters, to indicate and validate any impending fault. SCADA data contains a number of parameters that are captured at mostly 10-minute intervals, each of which has to be meticulously and continuously analyzed with the aim of detecting faults early. It is often a daunting endeavor for the most skillful analyst leading to poor assessment because of the mental strain caused by the volume of data available for analysis. Researchers have thus attempted to develop various approaches to aid in the classification and detection of abnormalities from turbine SCADA. Numerous research studies have employed machine learning (ML) methods for classification and regression operations aimed at distinguishing between faulty and proper readings in the context of wind turbines. These methods encompass a range of techniques, including k-nearest neighbor (Lin et al., 2013), support vector machines (Pei and Li, 2019), decision trees (Abdallah et al., 2018), and neural networks (Sun et al., 2016; Stetco et al., 2019; Morshedizadeh et al., 2017; Pliego Marugán et al., 2019; Kong et al., 2020). In addition, deep learning (DL), and representation learning methods have also found their place in the literature, such as long short-term memory (LSTM) models, multi-head attention (MSA) from transformer architectures (Wang et al., 2022a), health learning based on self-supervised learning (SSL) (Sun et al., 2022b), and generative adversarial networks (GANs) (Wang et al., 2022b). While these advanced approaches have demonstrated competitive performance, they also come with inherent complexities, thus demanding extended training times, requirement of large amounts of data, and substantial computational costs. Due to these drawbacks, researchers have started exploring more efficient solutions drawing from the fields of statistics and econometrics (Dao, 2022a, 2022b).

In recent times, artificial intelligence (AI), ML, and DL-based solutions have been increasingly employed for condition monitoring. However, these methods are not without their limitations. One of the challenges lies in the need for multiple parameters to confirm a fault and the fact that each turbine possesses a unique fault representation that is distinct from others within a wind farm. This complexity implies that wind farm analysts must manage a multitude of models alongside various SCADA parameters. In addition, the representations for normal functioning and faulty operations tend to evolve over time, primarily due to factors like turbine ageing, the recalibration of sensors, and the replacement of critical components. As a result, previously developed models become less effective. In such cases, it becomes essential to create new models tailored to specific turbines. Unfortunately, wind farm analysts often lack the necessary skill set for these tasks. Even if analysts possess these skills, there exists no standardized metric to determine when a model requires replacement or retraining. Hence, the performance of AI and ML-based approaches for condition monitoring gradually diminishes in efficiency over time, making them increasingly less practical for sustained use (Dao, 2022a). Another approach that has been attempted in the literature is the use of parametric models for CM (Dao, 2021; Dao et al., 2018b; Dao, 2018); some deep learning-based models have been argued to be parametric to some degree (Dao, 2022b). The problem with parametric approaches is that they generally assume that the data follows a normal distribution (Dao, 2021). Such tests like Fisher exact, student's t, and ANOVA, usually assume homogeneous variance leading to very accurate analysis of normally distributed data. The co-integration method (Dao et al., 2018a, 2018b; Dao, 2018), CUSUM-based approach (Dao, 2021), and the Chow test (Dao, 2022b) are examples of parametric implementations for CM. However, in cases where the data does not follow a normal distribution such as the case with most SCADA parameters analyzed during CM, parametric methods end up with misleading results. Most recently, researchers have attempted to develop nonparametric techniques in order to accommodate for both normal and non-normal parameter distributions that a given SCADA parameter may follow. Nonparametric approaches only require the data to follow a continuous distribution (Dao, 2022a).

This present study therefore utilizes a nonparametric method in an attempt to close the gaps associated with AI/ML/DL and parametric approaches. The wind turbine power curve alongside Kolmogorov-Smirnov's test (a nonparametric statistical approach) are utilized for the development of a new CM technique. Several algorithms and methods have been applied to developing wind turbine power curve that is a true representation of the actual conditions experienced in real life; such models are discrete models (Llombart et al., 2005), stochastic models (Gottschall and Peinke, 2007), parametric (Marčiukaitis et al., 2017; Kusiak and Verma, 2013; Villanueva and Feijóo, 2018), and nonparametric models (Karamichailidou et al., 2021; Pelletier et al., 2016; Manobel et al., 2018). These models are mainly focused on the problems of power prediction and forecasting with little consideration to monitoring and troubleshooting or predictive control and optimization. Attempts have been made to include performance monitoring in works that are focused on power prediction; but we found that in such works, there are no clear techniques to specifically address performance monitoring based on the developed power curve models. It is worth noting that performance monitoring is closely related to condition monitoring but not exactly identical; while performance monitoring indicates under-performance of a wind farm, it is not aimed at fault detection. In essence, an assessment of the predicted power and the actual power output of a certain turbine over a time period could suffice as performance monitoring because any reasonable difference between the actual power output and the predicted power output will indicate if the turbine or wind farm is performing acceptably or underperforming, but it will not provide enough information on the imminence of a fault. This is because various field conditions and turbine variables (e.g., turbine age) could be responsible for the noticeable underperformance. On the other hand, condition monitoring for fault detection requires a more specialized approach. To the best of our knowledge, no research within the literature provides a method for condition monitoring and fault detection based on the developed power curve. In addition, the technique developed in this paper employs a bin-wise approach for SCADA analysis by attempting to discretize the continuous power variable into wind speed intervals and identify the frequency distribution of each interval (bin), while simultaneously extracting significance levels (α). A few research in the literature have employed the concept of binning in the development of wind turbine power curves (WTPC). In Llombart et al (Llombart et al., 2005)., modifications were made to the IEC 61400–12 bins method that developed a single line power curve by the least squares method and binning. A comparison between the

binning approach and support vector regression for estimating the rotor speed-based power curve of a wind turbine is carried out in Pandit et. al (Pandit et al., 2020)., with the aim of comparing efficiencies between the two methods. There is no research in the literature to the best of the authors' knowledge, that utilizes power values that fall within wind speed intervals (bins) for the formation of frequency distributions, and the extraction of confidence levels (α) in any form and for any other application.

The Kolmogorov-Smirnov (K-S) test is a nonparametric test in statistics that measures the goodness of fit. It compares the cumulative distribution functions of two data samples, or one sample and a population in order to assess whether they were drawn from the same distribution. The chi-square test is an alternative; however, it is most sensitive at the center of the distribution and least sensitive at the edges. Furthermore, as the sample size decreases, the chi-square test becomes inapplicable whereas, K-S test retains its efficiency on small as well as large samples. It also does not suffer from reduced efficiency around the edges of a distribution, thus making it a better alternative for comparing data samples. The K-S test has been applied in a diversity of fields for comparing sample with sample, and sample with distribution. In the work of Zhang et al. (1993), K-S test was utilized for fast and robust sensing of spectrums in radio systems by computing the empirical cumulative distribution functions (ECDF) of some decision statistics and comparing it with the ECDF of the noise signal. The K-S test and convolutional neural networks were applied in (Guo and Fu, 2018) for fault diagnosis of turbines. However, this work is subject to the limitations of ML and DL approaches earlier discussed (such as enormous computation cost, long training time, and performance degradation over the long term). Additionally, the fault signal is not validated by other significant SCADA variables. The K-S test has also been applied for drift detection in machine learning (ML) (Dos Reis et al., 2016), explanation of unreliable ML survival models (Kovalev and Utkin,.), identification of the distribution of earthquake data to predict magnitude (Oktaviana and Irhamah, 2021), and for detecting changes in maps of gamma spectra in radioactivity (Reinhart et al., 2015).

1.1. Contribution and outline

This study is aimed at developing a new method for conditional monitoring of wind turbines by utilizing the wind turbine power curve and three well-developed test methods. Fault detection is validated by the K-S test; the method comprises a sequence of steps. First, SCADA data obtained from the operations of a wind turbine is processed for anomaly detection and removal. A quantile-based algorithm sets user-defined quantiles that differentiate between normal and faulty data. Afterward, useful SCADA parameters are utilized in developing power curve models that accurately represent actual field conditions within a wind farm. A superior predictive ability (SPA) of one of the compared models is asserted using the Mariano-Preve test of equal predictive ability (EPA), and by comparing their coefficients of determination. The most accurate model is utilized for the creation of bin-wise frequency distribution to serve as the ground truth data sample or population, and bin-wise confidence levels to serve as the decision factor for the K-S test. The work of Cochran (Cochran, 1977) developed a method of identifying the minimum sample size with the capacity to retain distribution information from the population. This minimum sample size represents the minimum number of acquired SCADA temporal instances needed to assert a fault while also signaling the time rate of fault detection. This method detects faults directly from the output power of a wind turbine in relation to an increase in bearing temperature and a reduction of generator speed. Two case studies of one year SCADA data from two onshore wind turbines [Supplementary data [A] and [B]], are used to validate the developed method. These SCADA data are associated with two faults or abnormal events; in one of the cases, we analyze the detected fault from the output power by indicating the simultaneous increase in bearing temperature and reduction in generator speed. Other than the K-S test method, the methods and tests utilized in this study have not been investigated in the literature for condition monitoring of wind turbines or any other aspect of wind energy. This research is motivated by its efficacy over the long-term when compared to methods based on ML algorithms as earlier discussed. It is aimed at improving the current state of condition-based maintenance measures that would in turn optimize energy generation by reducing turbine downtime and associated expenses incurred on corrective maintenance, thereby ultimately reducing the total cost of generating energy.

The rest of this paper is outlined as follows: Section 2 begins with a flowchart that graphically explains the methods utilized in this study. A detailed description of the methods and algorithms is also provided for the concepts used for the development of this CM technique. A step-by-step procedure is provided for the provision of a sequential flow and breakdown of the developed method. Section 3 presents results on the application of all adopted algorithms and tests, and a description of what these represent. Section 4 is the conclusion of the paper.

2. Methodology and algorithms

The flowchart for the detection of faults in condition monitoring of wind turbines from SCADA data, and the developed power curve is shown in Fig. 1.

2.1. Quantile filtering

Quantiles usually define a particular part of a data, in relation to other parts of the same data within a distribution. The simplest representation is a dividing plane that serves as a limiting condition to an assertion about the nature of the data. In Fig. 2, we are considering normally distributed data with no skewness (i.e., the LHS is an identical replica of the RHS) i.e., the distribution of q-quantile plots for all values $a \epsilon S$; the probability that x falls within quantile q is given by $P[X < x] \leq k/q$ (where x is a k-th q-quantile for a variable X), and the probability that x falls without the quantile q is given by $P[X < x] \geq 1 - k/q$ considering also that x is the k-th q-quantile for a variable X. The normal distribution is represented mathematically in Eq. (1).

$$P[X < x] = \frac{1}{\sqrt{2\pi}} \int_{\infty}^{x} e^{-\frac{t^2}{2}} \quad dt \tag{1}$$

The α-th quantile $\theta_\gamma(\propto), 0 < \propto < 1$ of a finite population vector $y = (y_1, \ldots, y_N)$ is defined as:

$$\theta_\gamma(\propto) = \inf\{t : F_\gamma(t) \geq \alpha\} \tag{2}$$

where $F_\gamma(t)$ is the distribution functionγ. In case$\dot{F}_\gamma(t)$, an estimator of $F_\gamma(t)$, is a monotonic non-decreasing function of t, the customary estimator of $\theta_\gamma(\propto)$ is obtained as:

$$\theta_\gamma(\propto) = \inf\{t : \dot{F}_\gamma(t) \geq \alpha\} \tag{3}$$

Let $\dot{F}_\gamma(t)$ be the customary estimator of $F_x(t)$. In case the population α-th quantile $\theta_x(\propto)$ of x is known, the ratio estimator of $\theta_\gamma(\propto)$ is given by:

$$\ddot{\Theta}_{r\gamma}(\propto) = \frac{\ddot{\Theta}_\gamma(\propto)}{\ddot{\Theta}_x(\propto)} \theta_x(\propto) \tag{4}$$

Similarly, a different estimator of $\theta_\gamma(\propto)$ is given by:

$$\ddot{\Theta}_{d\gamma}(\propto) = \ddot{\Theta}_\gamma(\propto) - R\{\ddot{\Theta}_x(\propto) - \theta_x(\propto)\} \tag{5}$$

where $R = \frac{\sum_{i \in S} \frac{y_i}{\pi_i}}{\sum_{i \in S} \frac{x_i}{\pi_i}}$ is a consistent estimator of the population ratio $R = Y/X$.

Both estimators $\ddot{\Theta}_{r\gamma}(\propto)$ and $\ddot{\Theta}_{d\gamma}(\propto)$ reduce to $\theta_\gamma(\propto)$ if $y_i \propto x_i \forall i \in U$. In this case, the variance become zero. The case is similar to a variety of distributions regardless of the nature, skewness, or shape.

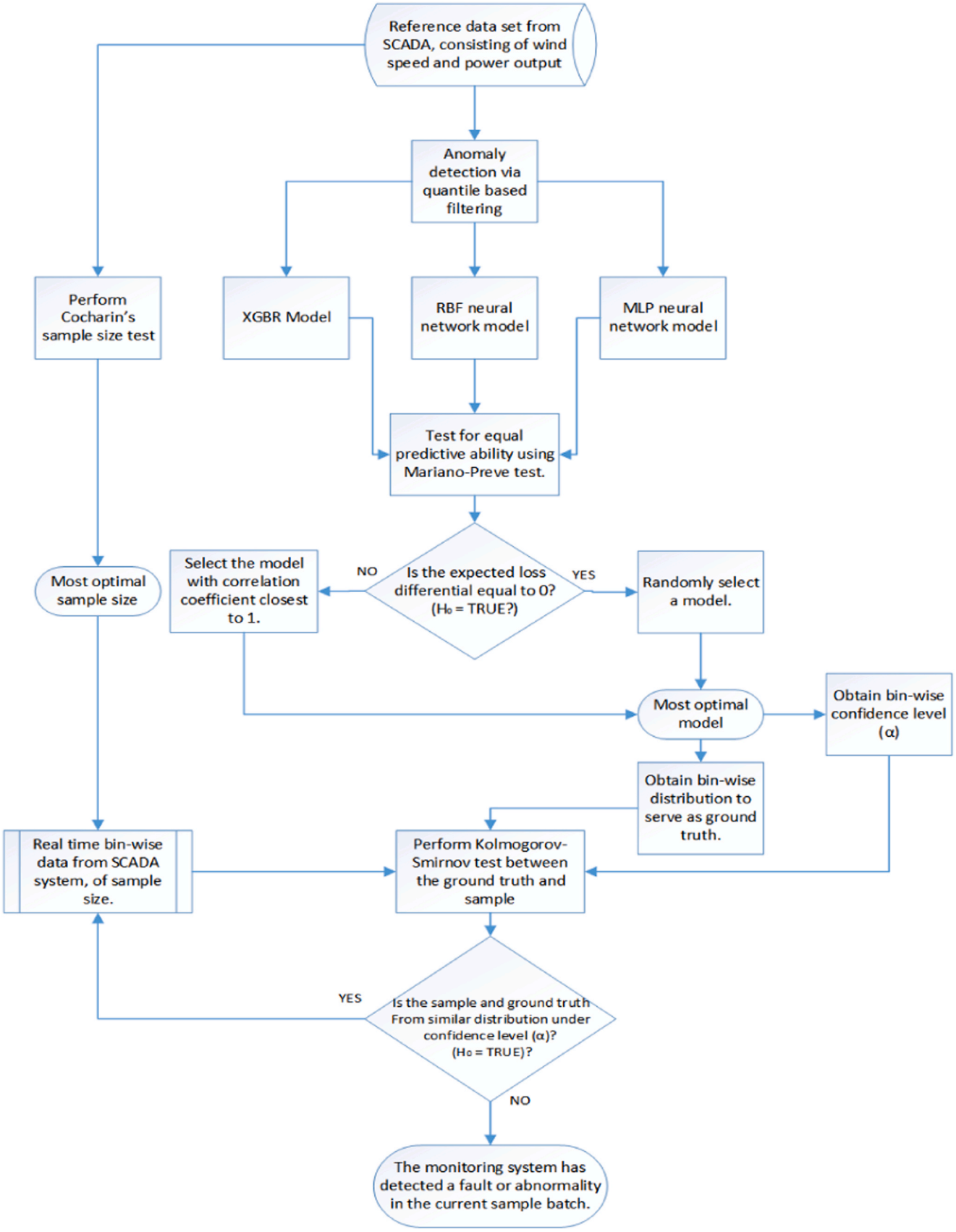

Fig. 1. Computation flowchart for the detection of fault in condition monitoring of wind turbines from SCADA data and the developed power curve.

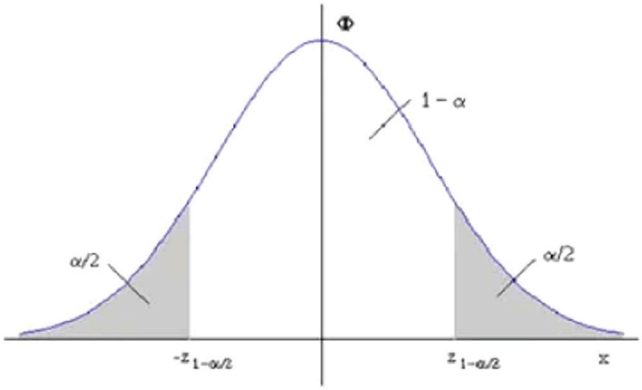

Fig. 2. A normally distributed case for quantile specification.

2.2. Gradient boosting regressor (GBR)

The concept of boosting aims at combining multiple base regressors to form a sequential ensemble for the purpose of developing a committee with better performance than any single regressor.[1] Boosting is achieved by a step-wise training of a new learner, a weak learner, and a base learner model with respect to the error realized at that step. GBR utilizes the concept of boosting for the development of an ensemble model that is a collection of tree models arranged sequentially. In this arrangement, the succeeding model learns from the errors of the preceding model, the performance of the preceding weak learning model is said to be boosted by the succeeding learner model. This ensemble is usually achieved by decision tree algorithm (Rao et al., 2019). Considering a gradient boost regressor with N number of trees, Eq. (6) can thus be stated:

$$f_N(x_j) = \sum_{n}^{N} \beta_n \quad h_n(x_j) \tag{6}$$

where h_n represents the weak learner model that has performed poorly on its own, β_n will represent the contribution of the model tree to the performance of the weak model, and is identified as a scaling factor. The loss function employed by XGBR to minimize errors is the gradient descent loss function. It achieves this by updating initial estimation with newer ones thereby improving the performance of the final output tremendously.

2.3. Multi-layer perceptron (MLP) neural networks

MLP networks are a type of feed forward neural networks that consist of three layers i.e., input layer, hidden layer, and output layer. They have been widely used for regression and classification tasks. However, their efficacy is experienced in their ability to perform accurate regression analysis. A perceptron acquires a total of n features as input $x = x_1, x_1, \ldots, x_n$, each of which has a weight associated with it. All features inputted into the network must be numeric in nature; all non-numeric features must be first converted into numbers before being inputted into the network. The input features are passed on to an input function u, this function computes the weighted sum of the input features.

$$u(\boldsymbol{x}) = \sum_{i=1}^{n} w_i x_i \tag{7}$$

The result $u(\mathbf{x})$ is passed onto an activation function f, this function assists in producing the output of the perceptron. The activation function utilized in this step is a RELU.

$$y(x) = MAX(\mathbf{0}, x) \tag{8}$$

$$y(x) = \begin{cases} 0 for & x < 0 \\ x \quad for & x \geq 0 \end{cases} \tag{9}$$

Learning in MLPs consist of adjusting the weights in order to reduce the error in predicting the training data. Learning is a back propagation task achieved by a back propagation algorithm (optimizer), that attempts to minimize the loss in predicting the ground truth.

2.4. Radial basis function (RBF)

The RBF architecture was first proposed by Broomhead and Lowe in their work entitled: Radial Basis Functions, Multivariate Functional Interpolation and Adaptive Networks in 1988 (Broomhead and Lowe, 1988). RBFs consist of three layers by design: the input layer, the hidden layer, and the output layer. Fig. 3 represents the typical structure of a RBF network with a single output node for single value tasks (e.g., regression etc.). The input node distributes the k input variables to the m nodes of the hidden layer. In the hidden layer, each node has a center with the same dimensions as the number of input variables. The hidden layer applies a non-linear transformation to the input space, transforming it into a higher-dimensional space. The activity $\mu_l(\mathbf{x}(\mathrm{f}))$of the lth node is the Euclidean normal of the difference between the f-th input vector and the node center, and is given in Eq. (10) as:

$$\mu_l(\mathbf{x}(\mathrm{f})) = ||\boldsymbol{x}(f) - \dot{\mathbf{x}}_l|| = \sqrt{\sum_{i=1}^{k} (\boldsymbol{x}(f) - \dot{\mathbf{x}}_{l,i})^2}, f = 1, \ldots, f \tag{10}$$

where f is the total number of available data, $\boldsymbol{x}^T(f) = [x_1(f), x_2(f), \ldots, x_k(f)]$ is the input vector, and $\boldsymbol{x}_l^T = [\dot{\mathbf{x}}_{Il,} \quad \dot{\mathbf{x}}_{2l,\ldots,}\dot{\mathbf{x}}_{k\,l}]$ is the centre of the lth node.

The activation function for each node is a radially symmetric function. In this work, we employ the sigmoid function given in Eq. (11):

$$g(\mu) = \frac{1}{1 + e^{-\mu}} \tag{11}$$

The hidden node response is denoted by $\mathbf{z}(\mathrm{f})$ (Eq. (12)):

$$\mathbf{z}(\mathrm{f}) = [g(\mu_1(\boldsymbol{x}(f))), g(\mu_2(\boldsymbol{x}(f))), \ldots, g(\mu_m(\boldsymbol{x}(f)))] \tag{12}$$

The output of a RBF network contains y unit, where y is the singular possible output value. The numerical output y(f) is produced by a linear combination of the hidden nodes' response (Eq. (13)):

$$y(f) = \mathbf{z}(\mathrm{f}).\mathbf{w}_n = \sum_{i=1}^{m} w_{l,n} g(\mu_1(\boldsymbol{x}(f))) \tag{13}$$

where $\mathbf{w}_n = [w_{1,n}, w_{2,n}, \ldots, w_{m,n}]^T$ is a vector containing the synaptic weights corresponding to the output n.

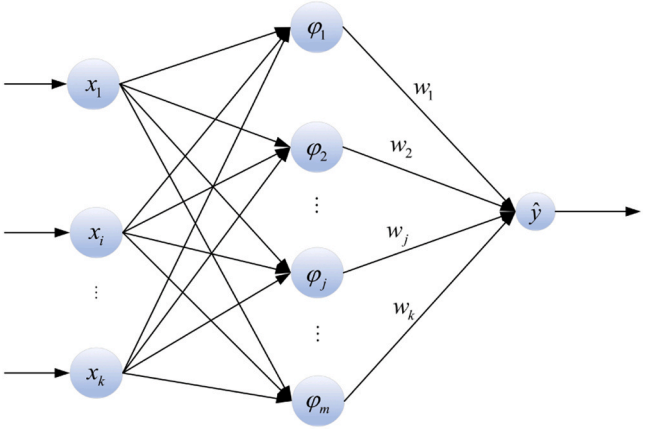

Fig. 3. Radial Basis Function Network.

[1] https://jerryfriedman.su.domains/ftp/stobst.pdf

The synaptic weights are commonly determined using linear regression of the hidden layer outputs to the real measured output after the RBF centers and non-linearities in the hidden layer have been fixed. In most cases, linear least squares in matrix form can be used to solve the regression problem.

$$\mathbf{W} = (\boldsymbol{Z}^T.\boldsymbol{Z})^{-1}.\boldsymbol{Z}^T.\boldsymbol{Y} \tag{14}$$

where $\boldsymbol{Z} = [z(1), z(2), \ldots, z(F)]^T$ is a matrix containing the hidden layer responses for all input vectors. $W = [w_1, w_2, \ldots, w_n]$ is a matrix containing all the synaptic weights for the output layer and converges to a scalar containing the target vector. The target vector y(f) carries the information of the value predicted by the f-th input vector.

2.5. Mariano-preve test

An explicit test of the null hypothesis for the purpose of validating equal predictive ability (EPA) of two competing forecasting models was introduced in the field of econometrics by Diebold and Mariano in their work: 'Comparing Predictive Accuracy, 1995′. This test method does not require any symmetric or quadratic relationship for the loss function and can be applied when the error distribution is non-Gaussian, has a non-zero mean, and serially and contemporaneously correlated (Diebold, 2008). However, this proposed asymptotic and exact finite sample test only compared two competing forecasts; in cases where competing forecasts were more than two, inferior methods had to be employed. The concept was further expanded for three or more competing forecast in the work of Mariano and Preve (Mariano and Preve, 2012). This test is model free in the sense that it assumes that the only information available to the analyst is a time series of forecast and actual values of the prediction. The task of such analyst is to ascertain if all the models perform equally in terms of a specific loss function, which could be squared error or absolute error. Let Eq. (15) represent forecast errors of k competing models:

$$\{f_{it}\} = \{\hat{Y}_{it} - y_t\}, i = 1, 2, 3, \ldots k \tag{15}$$

and if $g : R \rightarrow R$ represents the utilized loss function. The null hypothesis states that all the models have equal predictive ability under the specified loss function defined in Eq. (16) as:

$$Eg(f_{1t}) = Eg(f_{2t}) = \ldots = Eg(f_{kt}) \tag{16}$$

Consider the loss differential series $\{d_{jt}\}$ as expressed in Eq. (17), the null hypothesis requires that the expectation of the loss differential $Ed_t = 0$.

$$d_{jt} = g(f_{it}) - g\left(f_{i+1,t}\right), i = 1, 2, 3, \ldots, k \tag{17}$$

The test statistics d_s is based on the vector of observed sample means. It is represented in Eq. (18) as:

$$d_s = \frac{1}{s}\sum_{s=1}^{s} d_t \tag{18}$$

where s is the sample size.

2.6. Cochran's test

Finite population is usually being represented by a sample that will possess characteristics approximate to that of the population. In this study, our concern is a sampling distribution asymptotic to that of the population. The work of Cochran (Cochran, 1977) developed a formula to aid the calculation of the minimum sample size with the ability to imitate the population. Cochran's method assumes that the population is normally distributed and attempts to verify this after computing the minimum required sample size. Let n denote the minimum sample size, Cochran (Cochran, 1977) proposes Eq. (19) as:

$$n = \frac{n_o}{1 + \frac{n_o}{N}} \tag{19}$$

where N denotes the size of a finite population and n_o can be represented by Eq. (20) as:

$$n_o = \frac{Z^2 P(1-P)}{e^2} \tag{20}$$

Here Z denotes the z-score at confidence interval e, and P represents the portion of the population assumed to generally represent the population characteristics. In this work, P is taken to be 50% of the entire population.

2.7. Estimation of significance level

There exists a threshold value usually relating to the degree of significance, that validates the rejection of a hypothesis in most statistical tests. For the case of Kolmogorov-Smirnov's test used in this study, we attempted at defining a threshold value to serve as sufficient proof for the rejection of a certain assertion. In this approach, confidence levels are calculated separately for each wind speed bin based on the modeled plot of wind speed and power output for each turbine; this value is usually found to be approximately equal for similar brand of turbines. The process involves binning the two-dimensional power curve developed by the WTPC model, on wind speed basis and obtaining the geometric median for each bin. Afterwards, the Euclidean distance between the median point and all other data points within the bin are computed. The percentage variation between the largest distance value and the Euclidean distance from each median point to the reference x-plane will be utilized in calculating the confidence level. Fig. 4 depicts a modeled plot of wind speed and power output after binning. Let $\dot{X}$ represent the median point obtained after calculating the geometric median of data points within a specified bin (for instance, 7.5 m/s to 10 m/s) and let the data points within the bin be represented by $X(P) = X_1, X_2, X_3, X_4, \ldots, X_i$, the vector $Y(P)$ contains the distance of all data points within a bin to their bin-wise median point.

$$Y_i = EUCLIDEAN(|X_i - \dot{X}|) \tag{21}$$

Here Y_i denotes the distance between a data point and the median value $\dot{X}$. The distance between $\dot{X}$and the reference x-plane is denoted by B. Therefore, the confidence level (CL) is expressed in Eq. (22) as:

$$CL = \left[1 - (100 \times \frac{A_i}{B_i})\right] \tag{22}$$

A_i denotes the data point with the largest distance from the median point within a specific bin, while B_i represents the distance between the median point and the reference x-plane for the specific bin.

2.8. Kolmogorov–Smirnov (K-S) Test

One of the ways of constructing limits for a set of probability distribution function and for taking the amount of statistical data into account is by using the Kolmogorov-Smirnov (K-S) test for the empirical cumulative distribution function (CDF) which is constructed for m observation and denoted by $F_m(x)$. Consider the function $F(x)$ that represents a true probability distribution function of the observation, which in this case represents temporal instances of SCADA data and under an assumption that the PDF is unknown. If the observation set consists of m number of instances, a critical value of a test statistics $d_{m,1-\gamma}$ can be calculated such that a width band $\pm d_{m,1-\gamma}$ as relating to $F_m(x)$ will entirely contain $F(x)$ under a significance of$(1-\gamma)$ interpreted as a confidence statement that signifies belief in a statistical framework. In such cases, a measure of the test statistics $D_m = \max|F_m(x) - F(x)|$, known as the K-S test statistic is relevant for asserting the test under $\Pr\{D_m \geq d_{m,1-\gamma}\} = \gamma$. Several ways for computing $d_{m,1-\gamma}$, and for

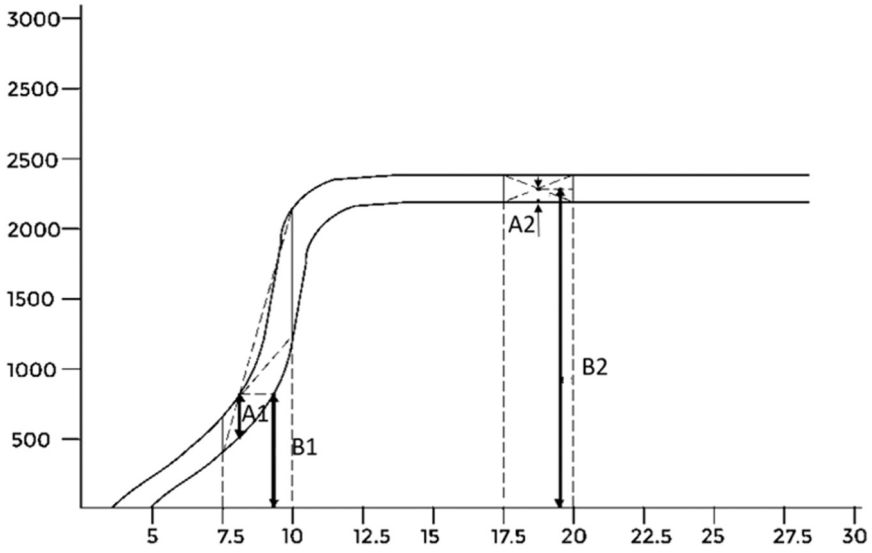

Fig. 4. Significance level estimation from the modeled power curve.

numerous values of m and γ, are detailed in the work of Baselice et al. (2019). A good approximation of the test statistics for $m > 10$ is shown by the two Eqs. (23) and (24), according to Zhang et al. (1993).

$$d_{m,1-\gamma} \approx \frac{(1-\gamma)}{\sqrt{m}} \tag{23}$$

and

$$d_{m,1-\gamma} \approx (1-\gamma)(\sqrt{m} + 0.12 + 0.11\sqrt{m})^{-1} \tag{24}$$

In both cases the limits are the cumulative distribution functions (CDF) which are lower $F_m^l(x)$ and upper $F_m^u(x)$ bounds, and are members of a known distribution function $F(x)$:

$$F_m^l(x) \le F(x) \le F_m^u(x) \tag{25}$$

where,

$$F_m^l(x) = \max\left(F_m(x) - d_{m,1-\gamma}, 0\right), \tag{26}$$

$$F_m^u(x) = \min\left(F_m(x) + d_{m,1-\gamma}, 1\right) \tag{27}$$

It is important to note that the K-S boundaries depend on the training examples m. It is seen from the inequality that the left of the upper boundary is $d_{m,1-\gamma}$ and the right of the lower boundary is $1 - d_{m,1-\gamma}$; these boundaries are located between boundary point of the sample space far apart.

2.9. Procedure

For a conceptual understanding of the technique proposed in this research, we defined the method by a sequence of six steps:

Step 1:	From a SCADA dataset of previous operations of a wind turbine containing necessary process parameters, use quantile-based filtering technique to detect and remove anomalous data. The algorithm is stated below: • Divide DataFrame into sub-frames by iterating over the power variable in steps of 50. • Define the probability distribution. • Apply quantiles to the probability distribution for each sub-frame in order to detect and remove outliers. • Merge all data frames. • END
Step 2:	Use the filtered SCADA dataset for training and validation of three choice model types. • Multiple variables are inputted for model development other than wind speed only, with the aim of capturing actual field conditions. The variables utilized are: wind speed, density, blade, pitch angle, and temperature. • For this study, the models considered are architectures of RBF, MLP, and GBR. These models are considered because of their high estimation performance as recorded in the literature.
Step 3:	Compare the loss distribution of the competing models for EPA. • The Mariano-Preve's test for EPA is utilized in this step. The null hypothesis states that all competing models have EPA, the alternate hypothesis states that the expected loss differential between the models is not equal to zero under a 0.05 significance level. • If the null hypothesis is rejected, an algorithm compares the correlation coefficient of the models and select the model with correlation coefficient closest to one as exhibiting SPA.
Step 4:	The superior model is used in generating power values corresponding to various wind speed, to serve as a modeled population from which inference about a normally functioning turbine can be drawn. • The population is separated into bins of wind speed intervals to serve as distributions representing normal operating condition (NOC) of the wind turbine. • The plot of wind speed and power output is used in generating bin specific significance levels to aid in K-S test decision making.
Step 5:	The most suitable sample size is calculated using Cochran's formula. This sample size represents the number of SCADA temporal instances required for the K-S test of similar distributions.
Step 6:	The K-S test ascertains if a sample is drawn from a certain distribution or not, this is used to differentiate normal and abnormal operation.

See Supplementary material [C] for the link to the code that contains data for both modeled and unmodeled SCADA parameters alongside implementations of all the test methods used in this work.

3. Results and discussion

3.1. Anomaly detection

Fig. 5a and b represent data obtained from wind turbine SCADA systems. It can be seen from the plot that data obtained directly from SCADA systems contains numerous erroneous readings of various types including: Type I errors i.e., errors generated when no power output is recorded at times when wind speed is significantly greater than the cut-in wind speed; Type II errors i.e., errors generated when the output power is constrained at higher values of wind speed, and Type III errors which are errors typically generated by unsteady readings and are usually close to the designed value. Therefore, there is need for an anomaly detection and filtering approach before being utilized for modelling. This study makes use of a quantile-based approach which is

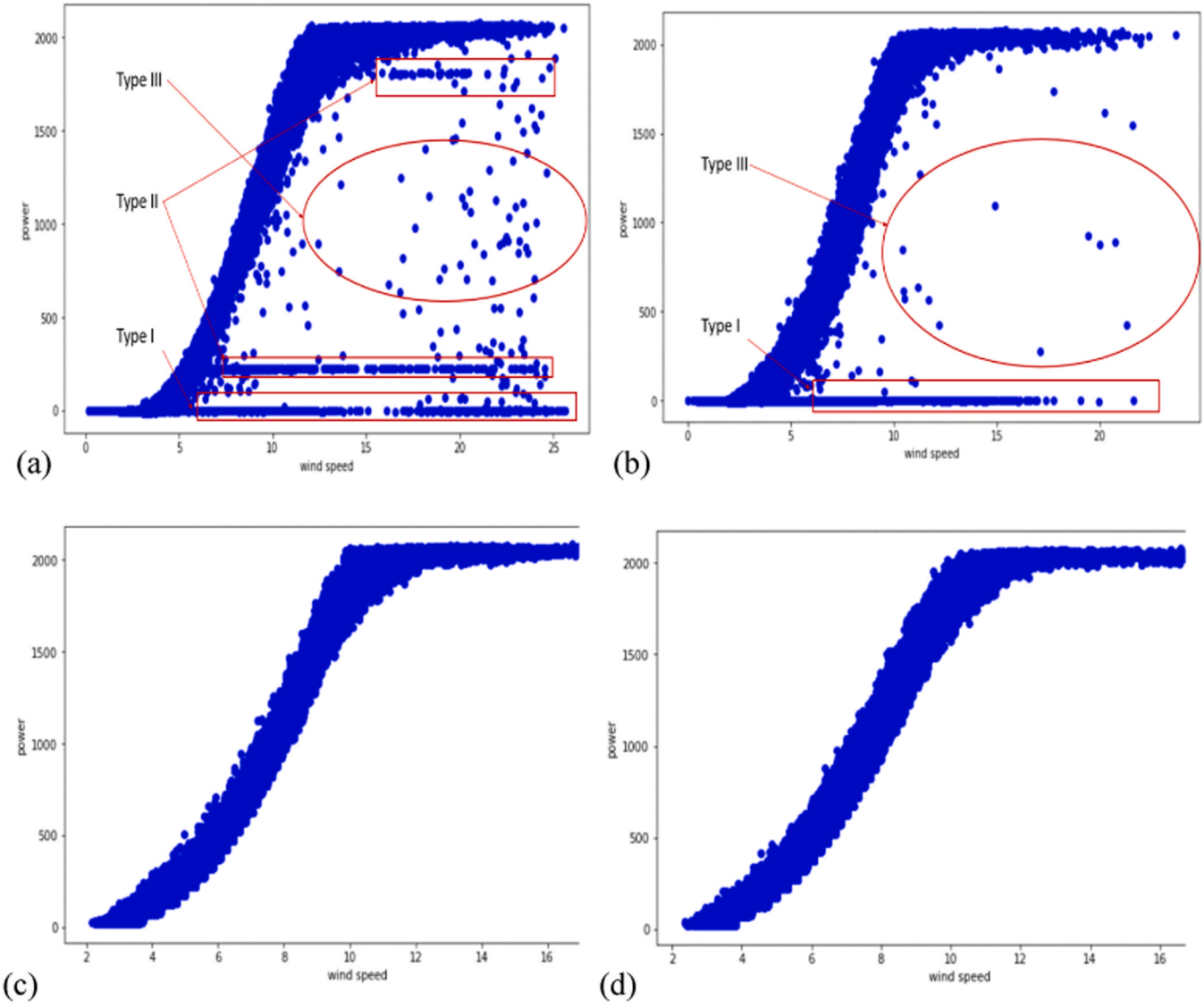

Fig. 5. SCADA population showing (a) and (b) unfiltered SCADA data, and (c) and (d) filtered representations.

established upon a hypothesis about the probability distribution of the SCADA population. It was discovered that faulty data appeared with less frequency compared to normal ones; if represented by a distribution, we could set quantiles to aid in the separation of normal data from abnormal data. Using the quantile-based filtration technique, proper filtration results was achieved as shown in Fig. 5c and d.

The visualization of a filtered plot is usually not enough evidence that a filtration technique is performing optimally or that it can be compared to alternatives developed and used in the literature. One method for validating the efficiency of a filtration technique is to compare its elimination rate with that of its numerous alternatives found in the literature. Table 1 shows a comparison of the elimination rates of the filtration techniques adopted in this work, with existing techniques found in the literature. The results highlighted can be found in Morrison et al., (Morrison et al., 2022). They include quantile-based filtering (QF) adopted in this work, isolation forest (iForest), Gaussian Mixture Modelling (GMM), and local outlier factor (LOF).

Table 1
The elimination rates of various filtration techniques are compared. This is performed with the aim of providing sufficient evidence about the performance of *quantile-based* filtration technique used in this work. The results show that the elimination rate of the QF based method is comparative to those utilized in the literature. This further confirms the efficiency of the utilized filtration technique.

Method	Elimination rate (%)
QF	16
iForest	27.47
GMM	15
LOF	12

3.2. Developed power curves

Two test metrics are used in evaluating superior predictive ability (SPA) among the utilized models. The metrics are mean absolute error (MAE) and coefficient of determination (R^2) expressed in Eqs. (28) and (29), respectively.

$$MAE = median(|\hat{Y}_n - Y_n|) \tag{28}$$

$$R^2 = 1 - \frac{\sum_{n=1}^{N}(\hat{Y}_n - Y_n)^2}{\sum_{n=1}^{N}(\hat{Y}_n - Y_n)^2} \tag{29}$$

Table 2 shows the average performance of all the models utilized in this study. The MLP network gave an average MAE and R^2 score of 20.07 and 0.995, respectively; RBF network resulted in MAE and R^2 values of 21.00 and 0.9949, respectively, while the GBR method gave 19.53 as MAE value, and 0.9953 as R^2 value. In a case where only the MAE and R^2 scores are analyzed, no consideration is made about the distribution of model losses. This analysis may be sufficient when there is considerable difference between the observed MAE and R^2 values. However, a problem arises when inference about the strength or weakness of the compared model is made, and when there is no significant difference in their observed MAE and R^2 values. In such cases, the forecast error distribution should be considered. This study considers the MAE values to assert that the choice models are competitive, while SPA is validated by Mariano-Preve's test of the null hypothesis. Fig. 6 indicates the comparison between values of actual power and estimated values. The 1:1 line indicates perfect correlation between actual power and model prediction. As corroborated by the R^2 values in Table 2, it can be seen that there exists a high correlation between actual power and model prediction.

3.3. Test for superior predictive ability

The multivariate Diebold and Mariano test can be used in comparing more than two competing forecasts for EPA. The original test properties inherited by the multivariate version includes no quadratic or symmetric requirement for the loss function, the errors realized during prediction should have a non-zero mean and can be non-Gaussian. The test statements are as detailed:

Null hypothesis:	H_0 will denote null hypothesis. The null hypothesis states that all the models are performing equally or that all forecasting models have equal predictive ability. In statistics terms, the expectation for a *k-th* loss differential series $\boldsymbol{E} \quad d_t$ will equal zero to confirm EPA.
Alternate hypothesis:	H_1 denotes the alternate hypothesis. When the expectation of the *K-th* loss differential series is not equal to zero, the alternate hypothesis is confirmed. This means the forecasting models do not have EPA.
If Null:	If the null hypothesis is confirmed, an algorithm randomly selects a model.
If Alternate:	If the null hypothesis is rejected, an algorithm compares the R^2 of all competing forecasts and select the model with an R^2 value closest to one.

In this study, we utilized a significance level of 95% to serve as sufficient deviation that indicates a rejection of the null hypothesis. The q value is obtained experimentally. In the work of Mariano *et al* (Mariano and Preve, 2012)., q values from 1 to 4 were utilized, the most optimum observed value was 3, which we employed in this research.

The test statistic is denoted by $S_{C.}$ The values of the test statistic and p-value are calculated and utilized to confirm the acceptance or rejection of a hypothesis. A rejection of the null hypothesis is validated under significance level α whenever $S_C > \chi^2_{k,1-\alpha}$ where $\chi^2_{k,1-\alpha}$ is the (1-α) quantile of the chi-squared distribution, represented by (p-values). Table 3 presents the results from Mariano-Preve's test. For Data 1 the test statistic is 0.28449, which is significantly lower than the obtained p-value of 0.8672. In this case, we accept the null hypothesis, leading to random selection of a forecasting model from the list. Data 2 presents a different result; the test statistic of 3.0217 is greater than the p-value of 0.2207, validating the rejection of the null hypothesis. In this case, the model with R^2 value closest to one is selected as having SPA.

Table 2
Comparison of the models based on MAE and R^2. This signifies that the compared models are close to each other in terms of accuracy, with very little calculated differences. In this case, considerations have to be made regarding the distribution of model loss instead of visual comparison.

Model	No of Nodes	MAE	R^2
MLP-NN	1	20.07	0.9950
RBF-NN	1	21.00	0.9949
XGBoost	-	19.53	0.9953

3.4. Estimation of Sample Size

A sample composed of data instances from real time operation of a wind turbine is needed for the validation of a fault by K-S test. A larger sample size can also be used for K-S assessment; however, it will take a long time for such sample to be generated from real time instances to aid the K-S test. At this point, the concern is acquiring the smallest sample that can inherit population characteristics in order to reduce the time taken for asserting faults. Cochran's formula (see Section 2.6) for obtaining the smallest sample size that will retain all the information of its population was utilized in this study. To utilize this method, certain parameters must be obtained; a data percentage that will have characteristics very close to that of the population and denoted by P, is selected to be 0.5 as it is obvious that a sample size half that of the population will inherit all the population characteristics to a large extent. The sample characteristics n_0 is estimated from confidence and error values, taken to both be 0.1. The test is carried out for the two turbine SCADA dataset [Supplementary data [A] and [B]]. We found that the minimum sample size with the ability to inherit population characteristics with error ε and under confidence level α, is 65. This is shown in Table 4.

3.5. Estimation of confidence level

There exists a level of significance above which the K-S test should sufficiently reject the null hypothesis. Because the magnitude and frequency of the data varies between bins, the required significance level should be bin specific. This study presents a method for obtaining bin-wise confidence levels based on the modeled plot of wind speed and power output as shown in Fig. 7. A detailed explanation of this process is provided in Section (2.7). Bins contain wind speed intervals of 2.5 m/s. The farthest data point from the middle a (Watt) is used alongside the distance from the middle point to the reference x-plane b (Watt) for calculating the level of significance for each bin. It can be seen from Tables 4 and 5 that the significance value α_s reduces with an increase in bin wind speed range, corresponding to increased power output. However, from the rated wind speed to the cut-out wind speed, the significance level remains the same. One caveat to this method is that the first (Table 5) and second (Table 6) bins cannot be used in asserting a fault. This is because their wind and power values are small and relatively close to zero. Hence, they do not possess deviations reasonable enough to confirm a claim, as can be seen from their recorded α_s values which are found to be relatively high. It is also worth noting that significance levels are similar, regardless of the turbine being investigated.

3.6. Fault detection

In this research, it was hypothesized that abnormality (faults) could be detected by comparing the distribution of modeled power estimates within a specified bin with a sample of real time SCADA temporal instances of minimum sample size and wind speed values within the specified interval (bin). One such test method applied to this type of problem is the K-S goodness of fit. This test compares the cumulative distribution functions (CDF) of the modeled power output with the CDF of a sample of real time SCADA data within a specified bin, by calculating the deviation in CDF and comparing it to a critical value defined by significance levels.

Null hypothesis	H_0 will denote null hypothesis. The null hypothesis states that the population and sample were drawn from similar distributions or that the test statistic D is a value less than K-S

(*continued on next page*)

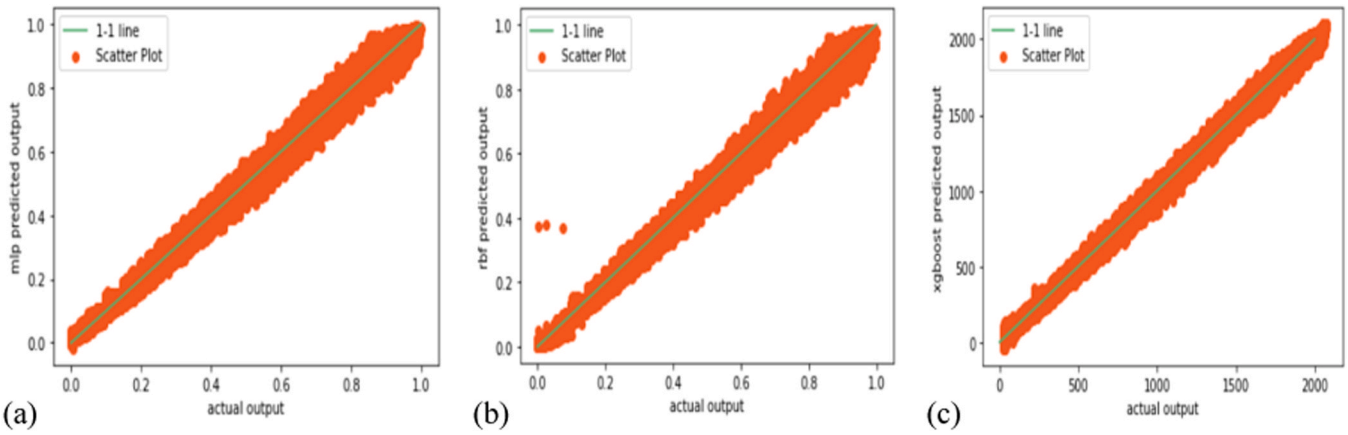

Fig. 6. Actual vs Predicted power output for (a) MLP based model, (b) RBF based model, and (c) gradient boosting based model. The y = x (1:1) identity line represents perfect correlation.

Table 3
The Mariano-Preve's table of values. It indicates the test statistics (Sc), q-value, significance level (α), P-value and decision.

Dataset	S_C	q	Significance (α)	P-value	Decision
Data 1	0.28449	3	0.05	0.8672	accept H_0
Data 2	3.0217	3	0.05	0.2207	Reject H_0

Table 4
Cochran's test for obtaining the minimum sample size with the capacity to possess all the information of the population.

Dataset	Population	P	α	Z	ε	n_0	Sample size
Data 1	56, 560	0.5	0.1	1.65	0.1	68.2	65
Data 2	56, 500	0.5	0.1	1.65	0.1	67.4	64

(*continued*)

	critical value D_α, defined by significance level α_s. In this case, the test asserts that the turbine is performing normally.
Alternate hypothesis	H_1 will denote the alternate hypothesis. Whenever the test statistics D reports a value that is greater than the K-S critical value D_α, the null hypothesis is rejected in favor of the alternate. This signals a fault or abnormality.

For K-S test, one can assert a claim by comparing the test statistic D and K-S critical value D_α or by comparing the p-value and significance level α_s. In either case, the result will be the same. Fault is confirmed whenever $(D \geq D_\alpha)$ or $(\alpha_s \geq p - value)$. The application of K-S test for validating normal and faulty operations is detailed in Table 7. For each data sample, three tests were performed; two of these tests were normal operation without any associated faulty feedback, whereas one is a sample taken few days before SCADA system recorded faults that led to downtime. In cases where normal samples were analyzed, the K-S test displayed results that validated an acceptance of the null hypothesis, while in cases where the test sample was taken few days before the emergence of a fault leading to downtime, the results obtained from K-S validation indicated the rejection of the null hypothesis in favor of the alternate.

Table 5
Bin-wise confidence level estimates for Data 1.

	Dataset 1			
S/n	Bin (m/s)	*a* (Watt)	*b* (Watt)	α_s
1	2.5-5.0	10	10	100
2	5.0-7.5	150	500	0.3
3	7.5-10.0	150	1050	0.14
4	10.0-12.5	150	1750	0.085
5	12.5-15.0	120	2000	0.06
6	15.0-17.5	120	2000	0.06
7	17.5-20.0	120	2000	0.06
8	20.0-22.5	120	2000	0.06
9	22.5-25.0	120	2000	0.06

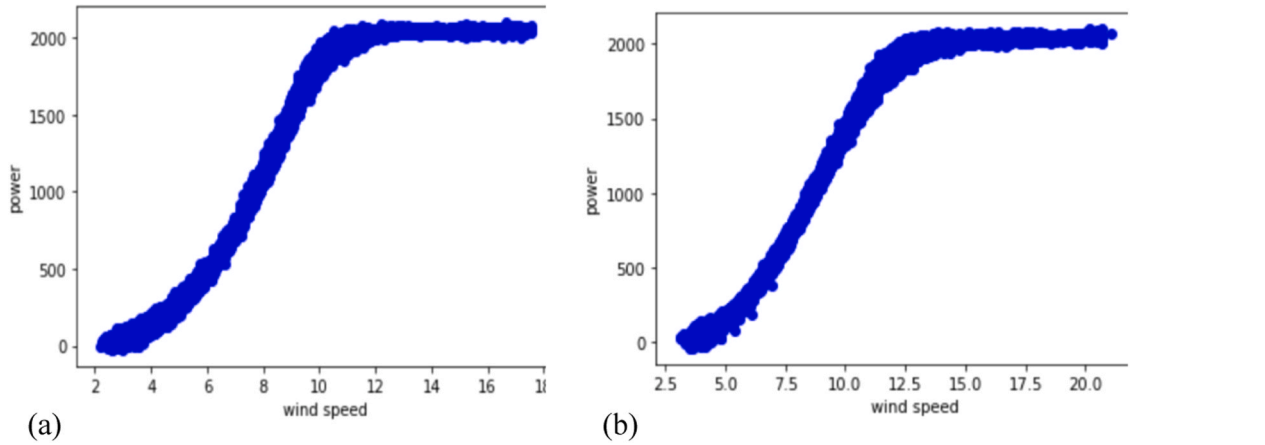

Fig. 7. Modeled power curve of most optimal model for (a) Data 1, and (b) Data 2, utilized for the development of bin-wise confidence level and bin-wise ground truth for validating the K-S test.

Table 6
Bin-wise confidence level estimate for Data 2.

	Dataset 2			
S/n	Bin (m/s)	*a* (Watt)	*b* (Watt)	α_s
1	2.5-5.0	5	5	100
2	5.0-7.5	100	500	0.200
3	7.5-10.0	135	1000	0.121
4	10.0-12.5	140	1550	0.090
5	12.5-15.0	125	1950	0.064
6	15.0-17.5	125	1950	0.064
7	17.5-20.0	125	1950	0.064
8	20.0-22.5	125	1950	0.064
9	22.5-25.0	125	1950	0.064

The power distribution as specified within the wind speed intervals is that which is analyzed by the K-S test. Out of the seven parameters investigated in the work of Dao (Dao, 2022a), two other process parameters including gearbox bearing temperature, and generator speed that provide optimal fault detection were identified, apart from output power. With the aim of providing further validation for faults detected using K-S analysis of the output power, an assessment of both gearbox bearing temperature and generator speed is provided. For this investigation, SCADA data corresponding to four process parameters, and captured few days before a fault is detected were utilized. The data point captured is of minimum sample size as required for K-S evaluation.

Fig. 8 shows that the wind speed at the captured time interval is stochastic in nature, falling between (13.0 m/s and 24 m/s). We also observed that after the 24th instance, the wind speed exhibits higher frequency around 20 m/s. It can be observed that as the wind speed in Fig. 8 frequently exhibits high values around 20 m/s from the 24th instance, the output power in Fig. 9 is seen to decrease slightly. It is also observed that the output power attempts to correct itself back to the rated power range. This change in the behavior of power within specified bins causes a significant shift in the frequency distribution that leads to large distance between the observed CDFs as calculated by K-S test statistics. The anomaly in power distribution from Fig. 9 can be related to an increase in gearbox bearing temperature as shown in Fig. 10. It is observed that as the output power decreased slightly, there was a simultaneous increase in gearbox bearing temperature, from an interval between 64 °C and 74 °C to an interval between 71 °C and 80 °C. This increase is rather minimal and may not be detected by a wind farm operator. Furthermore, it was identified that the K-S test could be performed for gearbox bearing temperature and will yield similar results as obtained from the K-S test performed on the output power.

The slight decrease in output power and increase in gearbox bearing temperature as shown in Figs. 9 and 10 is accompanied by a decrease in generator speed (see Fig. 11). The generator speed generally exhibits a downward trend, the rate of which increases after the 24th instance. The analysis entails 65 instances corresponding to the minimum sample size that will possess the ability to retain population information as proposed by Cochran (Cochran, 1977). The generator speed is observed to decrease from 1806 revolution/minute (rpm) at the 2nd instance to about 1798 rpm at the 20th instance. The rate of this decrease is observed to improve around the 20th instance, from about 1804 rpm at the 23rd instance to 1785 rpm at the 60th instance. This shows that the results obtained from an analysis of generator speed further corroborate

Table 7
The results detailing the outcomes of K-S test. It can be seen that the K-S test accurately detects *Normal* and *Faulty* SCADA samples by either accepting the null hypothesis when a normal sample is tested, or by rejecting the null hypothesis when the sample is faulty.

Dataset	Bin (m/s)	Nature of Sample	Sample Time	α_s	D	D_α	P-value	Decision
Data 1	10-12.5	Normal	05/06/2020	0.090	0.0876	0.115	0.6788	Accept H_0
Data 1	12.5-15	Normal	01/07/2020	0.064	0.0459	0.1455	0.9986	Accept H_0
Data 1	10-12.5	Faulty	03/11/2020	0.090	0.178	0.115	0.0248	**Reject** H_0
Data 2	10-12.5	Normal	15/03/2020	0.090	0.0831	0.115	0.7394	Accept H_0
Data 2	12.5-15	Normal	19/09/2020	0.064	0.1235	0.1455	0.2360	Accept H_0
Data 2	12.5-15	Faulty	01/06/2020	0.064	0.1787	0.1455	0.0248	**Reject** H_0

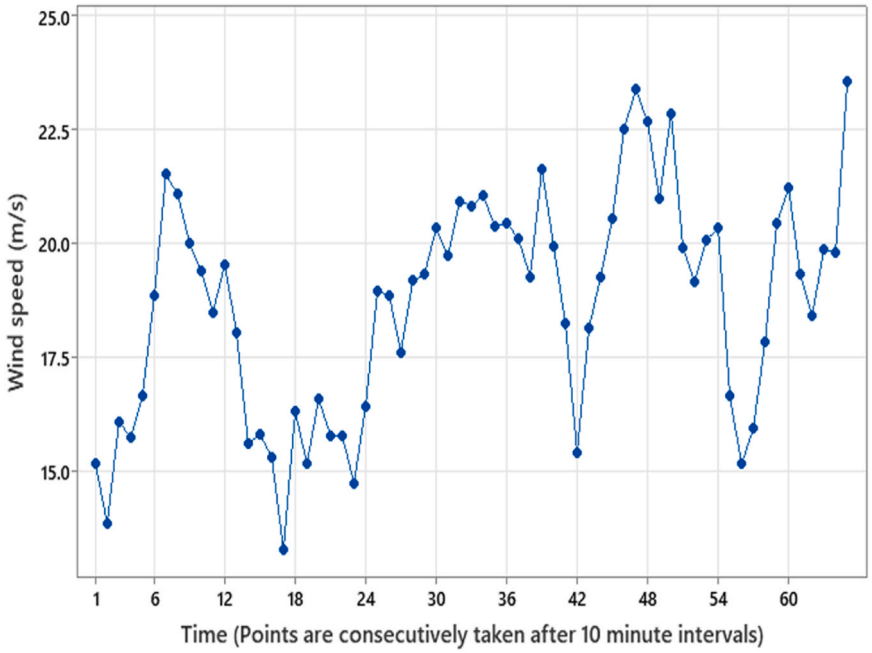

Fig. 8. Temporal plot of wind speed, of minimum sample size and collected few days before fault was realized in Data 2.

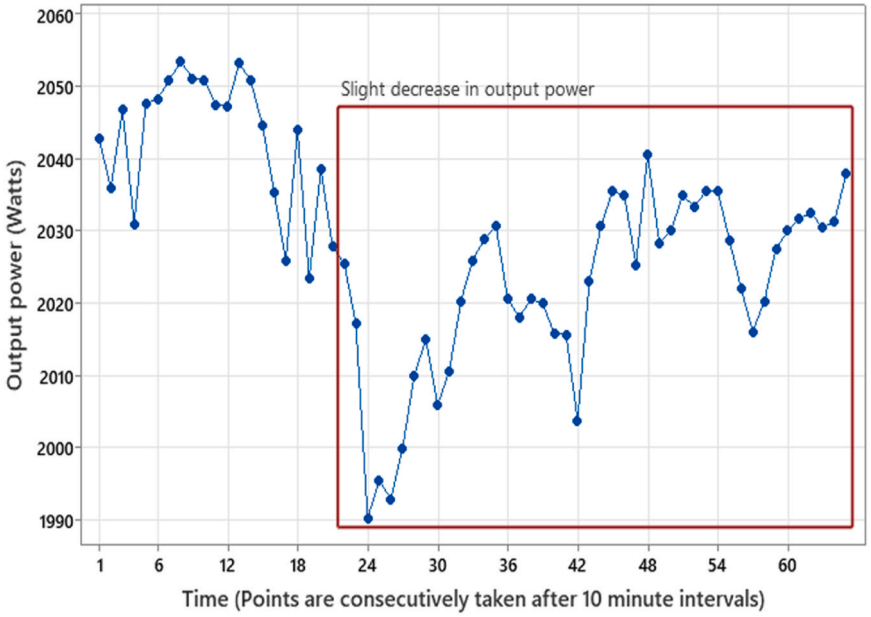

Fig. 9. Temporal plot of power output of minimum sample size collated few days before a fault was detected in Data 2. It corresponds to the wind speed plot.

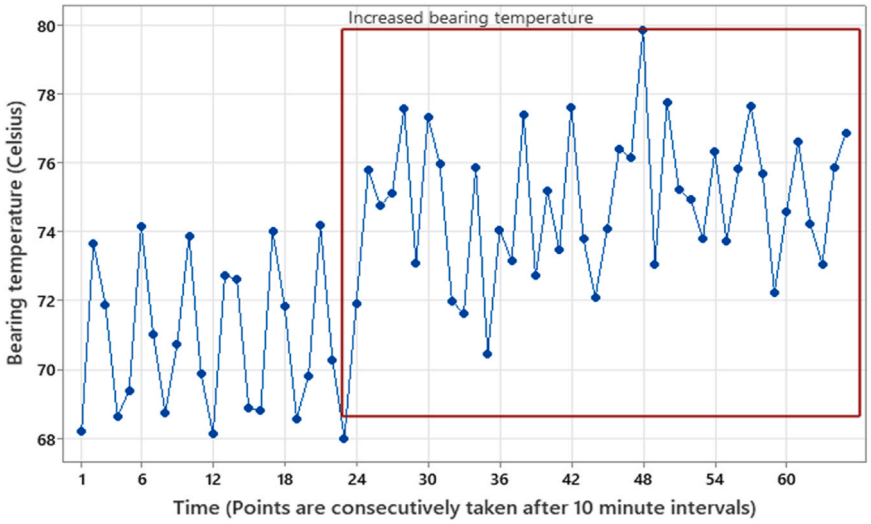

Fig. 10. Temporal plot of bearing temperature. Data points are of minimum sample size and collated a few days before fault was detected in Data 2.

the result obtained from performing the K-S test on the output power. Generator speed will also yield favorable results if utilized as a variable for the K-S test.

3.7. Discussion

Discussions of benefits and limitations of this work are drawn based on the results from each of the sections as given:

The quantile-based filtration technique (Sections (2.1) and (3.1)), detects and eliminates faulty data by defining quantiles on a distribution of SCADA data, on the basis of a process parameter. We utilized output power in this research. It is important to note that the most optimal quantile must be obtained experimentally and is user defined. In other words, the efficiency of this technique depends on several iterations performed with different quantile values alongside an assessment of the filtered plot of wind speed and power output, and the elimination rate. If a good quantile value is realized, this method could yield better performance than other techniques employed for data filtration in the literature.

This study employs RBF, MLP and GBR for developing power curves that are compared for the extraction of bin-wise information required for the K-S test. The MLP and RBF are neural network based methods while GBR is based on a decision tree algorithm. While it is possible that one of the competing models was disadvantaged due to the selection of sub-optimal hyper-parameters, we attempted using the most optimal hyper-parameter. The superior forecasting model will possess better bin information than any inferior alternative. In turn, this will reduce the tendency of realizing a Type 1 (incorrect rejection of the null hypothesis) or Type 2 (failure to reject the null hypothesis when true) errors. In cases where forecasting options do not have EPA, it was found that by analyzing the extent to which the predicted power linearly correlates to the observed values for each of the forecasting options, SPA could be confirmed.

Cochran's method for sample size estimation aids the acquisition of a minimum sample size with sampling distribution similar to that of the population under a specified significance level. A significance level (α)

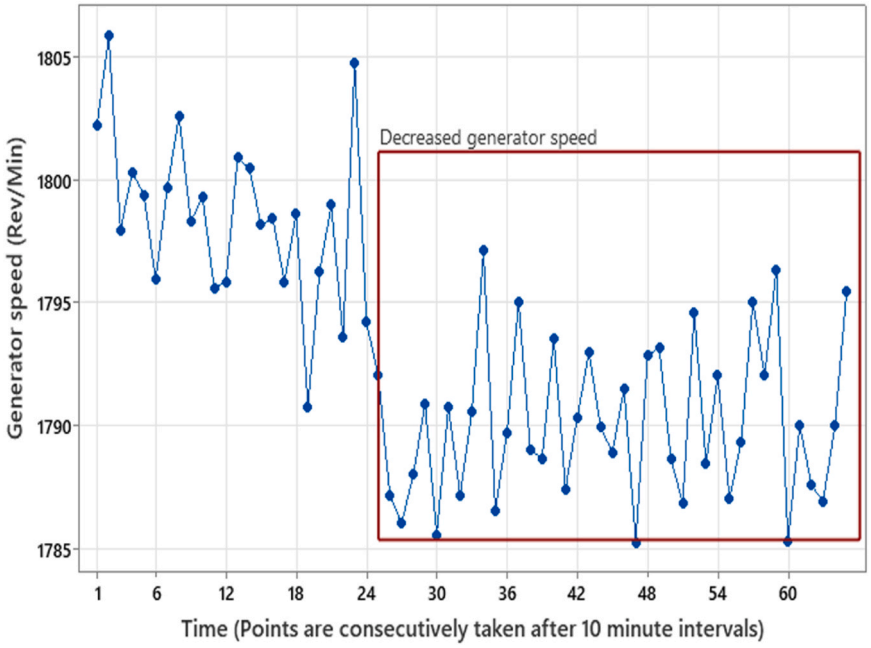

Fig. 11. Temporal plot of generator speed containing instances equivalent to the minimum sample size and taken a few days before a fault was detected.

of 0.05 indicates that 1 out of 20 samples will not possess adequate information from the population. If this aberrant sample is used for the K-S test, it will result in a false positive or Type 1 error. An analysis of other related process parameters, as discovered in the work of Dao (Dao, 2022a), can be used to identify such samples. If the fault obtained from the K-S test is not validated by the other process parameters, the sample should be regarded as a false positive.

A bin comprises parameters corresponding to a wind speed of 2.5 m/s intervals. The first bin contains parameters corresponding to wind speed values between 2.5 m/s and 5.0 m/s, and the last bin is defined by wind speed values between 22.5 m/s and 25.0 m/s. All bins situated after the rated wind speed possess the same K-S significance value (α_s) while the value of α_s decreases rapidly for bins before the rated wind speed. In cases where the wind speed is highly dynamic and intersects several bins located after the rated wind speed, the analysis could be carried out by merging them together because their α_s values are the same. On the other hand, the first and second bin represented with intervals of (2.5 m/s - 5.0 m/s) and (5.0−7.5 m/s) respectively, should not be used for K-S analysis as the results obtained using them will be erroneous due to the nature of their α_s.

The K-S goodness of fit test proves to be an efficient method for comparing the CDF of samples. The K-S evaluation is performed on the output power in this work; this is because α_s was generated based on the wind turbine power curve. We observed that K-S evaluation can be performed on gearbox bearing temperature as well as generator rpm, achieving optimal results. However, it is not clear whether α_s obtained from analyzing the power curve could be applied for such analysis.

4. Conclusion

This study presents a new method for operational state monitoring and fault detection of wind turbines based on the modeled power curve and Kolmogorov-Smirnov's nonparametric goodness of fit test. ML and DL have been applied in recent times for conditional monitoring and fault detection of wind turbines. However, they are known to possess some limitations including long training time, enormous computational cost, and large data requirement. In addition, ML and DL methods decrease in performance over time owing to factors such as turbine ageing, change of critical components, and sensor recalibration. These reasons present the need for more reliable and efficient solutions to the problem of CM and automatic fault detection of wind turbines. The method investigated in this study attempts the validation of faults by a comparison of distributions. Filtered SCADA datasets [Supplementary data [A] and [B]] are used for the development of a WTPC model of optimal performance. Three modelling architectures namely: radial basis function, multi-layer perceptron, and gradient boosting are compared in order to identify the most optimal alternative. A comparison of the models is performed using Mariano-Preve and linear correlation test. Distribution information from each wind speed interval (bin), alongside confidence levels, is extracted from the most optimal model alternative to be applied for the K-S test of similarity between CDFs. The minimum sample size is calculated using Cochran's formula and is used to specify the number of SCADA data temporal instances required to assert a fault or abnormality, aiding in the reduction of time for fault detection. The null hypothesis assumes that both samples are drawn from similar distributions or have similar CDFs; if the null hypothesis is rejected, a fault claim is asserted. This claim is assessed in relation to other process parameters known to provide useful information on the imminence of a fault before it is confirmed; these are gearbox-bearing temperature and generator speed.

Two SCADA data associated with two fault events were utilized as case studies for confirming the technique proposed in this research. One major advantage of this method, apart from its fast computation, is its invariance to performance degradation over time. This sets it apart from ML, DL, and other parametric approaches, including neural networks, decision trees, support vector machines, SSL, MSA, GANs, cointegration methods, CUSUM-based approach, and the Chow test approach. The monitoring and fault detection process is also very simple. Furthermore, other than in wind energy, this method can find application in a plethora of sectors in engineering provided there is sufficient historical data for the formation of a frequency distribution, and a system for acquiring real time data to be used for comparison.

The certainty of a fault is confirmed after analyzing other process parameters known to provide acceptable fault signals. A system can be developed where K-S test is performed on the three significant process parameters simultaneously, to completely eliminate Type 1 and Type 2 errors. Currently, there exists no method for obtaining the significance levels needed for decision making for gearbox bearing temperature and

generator speed data. Hence, future works should focus on developing a method for obtaining confidence levels that will form decision boundaries for the acceptance or rejection of the hypothesis.

CRediT authorship contribution statement

Olayinka S. Ohunakin: Conceptualization of the work, Supervision, Model development, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Emerald U. Henry**: Data curation, Model development, Model input preparation, Writing – original draft. **Olaniran J. Matthew**: Validation, Writing – review & editing. **Victor U. Ezekiel**: Model input preparation, Writing – review & editing. **Damola S. Adelekan**: Data curation, Writing – review & editing. **Ayodele T. Oyeniran**: Writing – review & editing.

Declaration of Competing Interest

Authors declare that there are no known competing interests.

Data availability

I have include a link to the data in the manuscript.

Acknowledgements

The authors wish to acknowledge Covenant University, Ota, Lagos State, Nigeria in the actualization of this research work.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.egyr.2024.01.081.

References

Abdallah, I., Dertimanis, V., Mylonas, H., Tatsis, K., Chatzi, E., Dervili, N., Worden, K., Maguire, E., 2018. Fault diagnosis of wind turbine structures using decision tree learning algorithms with big data. Saf. Reliab. Safe Soc. Chang. World Proc. 28th Int. Eur. Saf. Reliab. Conf. ESREL 2018 (iii), 3053–3062. https://doi.org/10.1201/9781351174664-382.

Akay, B., Ragni, D., Ferreira, C.S., Van Bussel, G.J.W., 2013. Investigation of the root flow in a Horizontal Axis. Wind Energy 1–20. https://doi.org/10.1002/we.

Artigao, E., Martín-Martínez, S., Honrubia-Escribano, A., Gómez-Lázaro, E., 2018. Wind turbine reliability: a comprehensive review towards effective condition monitoring development. Appl. Energy 228, 1569–1583. https://doi.org/10.1016/j.apenergy.2018.07.037.

Baselice, F., Ferraioli, G., Pascazio, V., Sorriso, A., 2019. Denoising of MR images using Kolmogorov-Smirnov distance in a Non Local framework. Magn. Reson. Imaging 57, 176–193. https://doi.org/10.1016/j.mri.2018.11.022.

Broomhead, D.S., Lowe, D., 1988. Radial basis functions, multi-variable functional interpolation and adaptive nctworks. Emerg. Technol. Situ Process 139 (4148), 221.

Castellani, F., Astolfi, D., Sdringola, P., Proietti, S., Terzi, L., 2017. Analyzing wind turbine directional behavior: SCADA data mining techniques for efficiency and power assessment. Appl. Energy 185, 1076–1086. https://doi.org/10.1016/j.apenergy.2015.12.049.

Cochran W.G. Samling Techniques, 1977: 448.

Dao, P.B., 2018. Analysis of gearbox and generator temperature data. Diagnostyka 19 (1), 63–71. https://doi.org/10.29354/diag/81298.

Dao, P.B., 2021. A CUSUM-based approach for condition monitoring and fault diagnosis of wind turbines. Energies 14 (11). https://doi.org/10.3390/en14113236.

Dao, P.B., 2022a. On Wilcoxon rank sum test for condition monitoring and fault detection of wind turbines. Appl. Energy 318, 119209. https://doi.org/10.1016/j.apenergy.2022.119209.

Dao, P.B., 2022b. Condition monitoring and fault diagnosis of wind turbines based on structural break detection in SCADA data. Renew. Energy 185, 641–654. https://doi.org/10.1016/j.renene.2021.12.051.

Dao, P.B., Staszewski, W.J., Barszcz, T., Uhl, T., 2018a. Condition monitoring and fault detection in wind turbines based on cointegration analysis of SCADA data. Renew. Energy 116, 107–122. https://doi.org/10.1016/j.renene.2017.06.089.

Dao, P.B., Staszewski, W.J., Barszcz, T., Uhl, T., 2018b. Condition monitoring and fault detection in wind turbines based on cointegration analysis of SCADA data. Renew. Energy 116, 107–122. https://doi.org/10.1016/j.renene.2017.06.089.

de Novaes Pires Leite, Gustavo, Tenorio, Guilherme, da Cunha, Maciel, dos Santos Junior, Jose Guilhermino, Araújo, Alex Maurício, Andre, Pedro, Rosas, Carvalho, Stosic, Tatijana, Stosic, Borko, Rosso, Osvaldo Anibal, 2021. Alternative fault detection and diagnostic using information theory quantifiers based on vibration time-waveforms from condition monitoring systems: application to operational wind turbines. Renew. Energy 164, 1183–1194. https://doi.org/10.1016/j.renene.2020.10.129.

Dey, S., Pisu, P., Ayalew, B., 2015. A comparative study of three fault diagnosis schemes for wind turbines. IEEE Trans. Control Syst. Technol. 23 (5), 1853–1868. https://doi.org/10.1109/TCST.2015.2389713.

Diebold F.X. Comparing Predictive accuracy. 2008; 20: 134–144.

Dos Reis, D., Flach, P., Matwin, S., Batista, G., 2016. Fast unsupervised online drift detection using incremental kolmogorov-smirnov test. Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min. (1), 1545–1554. https://doi.org/10.1145/2939672.2939836.

Feng, Z., Chen, X., Liang, M., 2015. Iterative generalized synchrosqueezing transform for fault diagnosis of wind turbine planetary gearbox under nonstationary conditions. Mech. Syst. Signal Process. 52–53, 360–375. https://doi.org/10.1016/j.ymssp.2014.07.009.

García Márquez, F.P., Tobias, A.M., Pinar Pérez, J.M., Papaelias, M., 2012. Condition monitoring of wind turbines: Techniques and methods. Renew. Energy 46, 169–178. https://doi.org/10.1016/j.renene.2012.03.003.

Global Wind Energy Council. Global Wind Report: Annual Market Update; 2021. ⟨https://gwec.net/global-wind-report-2021/⟩.

Gómez Muñoz, C.Q., García Márquez, F.P., Sánchez Tomás, J.M., 2016. Ice detection using thermal infrared radiometry on wind turbine blades. Meas. J. Int. Meas. Confed. 93, 157–163. https://doi.org/10.1016/j.measurement.2016.06.064.

Gonzalez, E., Stephen, B., Infield, D., Melero, J.J., 2019. Using high-frequency SCADA data for wind turbine performance monitoring: a sensitivity study. Renew. Energy 131, 841–853. https://doi.org/10.1016/j.renene.2018.07.068.

Gottschall, J., Peinke, J., 2007. Stochastic modelling of a wind turbine's power output with special respect to turbulent dynamics. J. Phys. Conf. Ser. 75 (1) https://doi.org/10.1088/1742-6596/75/1/012045.

Guo P., Fu J. Condition Monitoring and Fault Diagnosis of Wind Turbines Gearbox Bearing Temperature Based on Kolmogorov-Smirnov Test and Convolutional Neural Network Model. 2018, doi: ⟨10.3390/en11092248⟩.

Karamichailidou, D., Kaloutsa, V., Alexandridis, A., 2021. Wind turbine power curve modeling using radial basis function neural networks and tabu search. Renew. Energy 163, 2137–2152. https://doi.org/10.1016/j.renene.2020.10.020.

Kong, Z., Tang, B., Deng, L., Liu, W., Han, Y., 2020. Condition monitoring of wind turbines based on spatio-temporal fusion of SCADA data by convolutional neural networks and gated recurrent units. Renew. Energy 146, 760–768. https://doi.org/10.1016/j.renene.2019.07.033.

Kovalev M.S., Utkin L.V. A robust algorithm for explaining unreliable machine learning survival models using the Kolmogorov–Smirnov bounds. Neural Networks; 132: 1–18, https://doi: ⟨10.1016/j.neunet.2020.08.007⟩.

Kusiak, A., Zhang, Z., 2010. Analysis of wind turbine vibrations based on SCADA data. J. Sol. Energy Eng. Trans. ASME 132 (3), 0310081–03100812. https://doi.org/10.1115/1.4001461.

Kusiak, A., Li, W., 2011. The prediction and diagnosis of wind turbine faults. Renew. Energy 36 (1), 16–23. https://doi.org/10.1016/j.renene.2010.05.014.

Kusiak, A., Verma, A., 2012. A data-mining approach to monitoring wind turbines. IEEE Trans. Sustain. Energy 3 (1), 150–157. https://doi.org/10.1109/TSTE.2011.2163177.

Kusiak, A., Verma, A., 2013. Monitoring wind farms with performance curves. IEEE Trans. Sustain. Energy 4 (1), 192–199. https://doi.org/10.1109/TSTE.2012.2212470.

Lin, D.F., Chen, P.H., Williams, M., 2013. Measurement and analysis of current signals for gearbox fault recognition of wind turbine. Meas. Sci. Rev. 13 (2), 89–93. https://doi.org/10.2478/msr-2013-0010.

Llombart, A., Watson, S.J., Llombart, D., Fandos, J.M., 2005. Power curve characterization I: improving the bin method. Renew. Energy Power Qual. J. 1 (3), 367–371. https://doi.org/10.24084/repqj03.304.

Maldonado-Correa, J., Martín-Martínez, S., Artigao, E., Gómez-Lázaro, E., 2020. Using SCADA data for wind turbine condition monitoring: a systematic literature review. Energies 13 (12). https://doi.org/10.3390/en13123132.

Manobel, B., Sehnke, F., Lazzús, J.A., Salfate, I., Felder, M., Montecinos, S., 2018. Wind turbine power curve modeling based on Gaussian Processes and Artificial Neural Networks. Renew. Energy 125, 1015–1020. https://doi.org/10.1016/j.renene.2018.02.081.

Marčiukaitis, M., Žutautaitė, I., Martišauskas, L., Jokšas, G., Gecevičius, B., Sfetsos, A., 2017. Non-linear regression model for wind turbine power curve. Renew. Energy 113, 732–741. https://doi.org/10.1016/j.renene.2017.06.039.

Mariano, R.S., Preve, D., 2012. Statistical tests for multiple forecast comparison. J. Econom. 169, 123–130. https://doi.org/10.1016/j.jeconom.2012.01.014.

Matthew, O.J., Ohunakin, O.S., 2017. Simulating the effects of climate change and afforestation on wind power potential in Nigeria. Sustain. Energy Technol. Assess. 22, 41–54. https://doi.org/10.1016/j.seta.2017.05.009.

Meyer, A., 2021. Multi-target normal behaviour models for wind farm condition monitoring. Appl. Energy 300, 117342. https://doi.org/10.1016/j.apenergy.2021.117342.

Morrison, R., Liu, X., Lin, Z., 2022. Anomaly detection in wind turbine SCADA data for power curve cleaning. Renew. Energy 184, 473–486. https://doi.org/10.1016/j.renene.2021.11.118.

Morshedizadeh, M., Kordestani, M., Carriveau, R., Ting, D.S.K., Saif, M., 2017. Improved power curve monitoring of wind turbines. Wind Eng. 41 (4), 260–271. https://doi.org/10.1177/0309524×17709730.

Odgaard, P.F., Stoustrup, J., 2015. A benchmark evaluation of fault tolerant wind turbine control concepts. IEEE Trans. Control Syst. Technol. 23 (3), 1221–1228. https://doi.org/10.1109/TCST.2014.2361291.

Ohunakin, O.S., Adaramola, M.S., Oyewola, O.M., 2011. Wind energy evaluation for electricity generation using WECS in seven selected locations in Nigeria. Appl. Energy 88, 3197–3206. https://doi.org/10.1016/j.apenergy.2011.03.022.

Ohunakin, O.S., Matthew, O.J., Adaramola, M.S., Atiba, O.E., Adelekan, D.S., Aluko, O. O., Henry, E.U., Ezekiel, V.U., 2023. Techno-economic assessment of offshore wind energy potential at selected sites in the Gulf of Guinea. Energy Convers. Manag. 288, 117110 https://doi.org/10.1016/j.enconman.2023.117110.

Oktaviana, P.P., Irhamah, 2021. Kolmogorov-Smirnov goodness-of-Fit test for identifying distribution of the number of earthquakes and the losses due to earthquakes in Indonesia. J. Phys. Conf. Ser. 1821 (1) https://doi.org/10.1088/1742-6596/1821/1/012045.

Pandit, R.K., Infield, D., Kolios, A., 2020. Gaussian process power curve models incorporating wind turbine operational variables. Energy Rep. 6, 1658–1669. https://doi.org/10.1016/j.egyr.2020.06.018.

Pei, S., Li, Y., 2019. Wind turbine power curve modeling with a hybrid machine learning technique. Appl. Sci. 9 (22) https://doi.org/10.3390/APP9224930.

Pelletier, F., Masson, C., Tahan, A., 2016. Wind turbine power curve modelling using artificial neural network. Renew. Energy 89, 207–214. https://doi.org/10.1016/j.renene.2015.11.065.

Pliego Marugán, A., Peco Chacón, A.M., García Márquez, F.P., 2019. Reliability analysis of detecting false alarms that employ neural networks: a real case study on wind turbines. Reliab. Eng. Syst. Saf. 191, 106574 https://doi.org/10.1016/j.ress.2019.106574.

Qu, F., Liu, J., Zhu, H., Zhou, B., 2020. Wind turbine fault detection based on expanded linguistic terms and rules using non-singleton fuzzy logic. Appl. Energy 262, 114469. https://doi.org/10.1016/j.apenergy.2019.114469.

Rao, H., Shi, X., Rodrigue, A.K., Feng, J., Xia, Y., Elhoseny, M., Yuan, X., Gu, L., 2019. Feature selection based on artificial bee colony and gradient boosting decision tree. Appl. Soft Comput. J. 74, 634–642. https://doi.org/10.1016/j.asoc.2018.10.036.

Reinhart, A., Ventura, V., Athey, A., 2015. Detecting changes in maps of gamma spectra with Kolmogorov-Smirnov tests. Nucl. Instrum. Methods Phys. Res. Sect. A Accel. Spectrometers Detect. Assoc. Equip. 802, 31–37. https://doi.org/10.1016/j.nima.2015.09.002.

Schlechtingen, M., Santos, I.F., Achiche, S., 2013. Wind turbine condition monitoring based on SCADA data using normal behavior models. Part 1: system description. Appl. Soft Comput. J. 13 (1), 259–270. https://doi.org/10.1016/j.asoc.2012.08.033.

Soua, S., Van Lieshout, P., Perera, A., Gan, T.H., Bridge, B., 2013. Determination of the combined vibrational and acoustic emission signature of a wind turbine gearbox and generator shaft in service as a pre-requisite for effective condition monitoring. Renew. Energy 51, 175–181. https://doi.org/10.1016/j.renene.2012.07.004.

Stetco, A., Dinmohammadi, F., Zhao, X., Robu, V., Flynn, D., Barnes, M., Keane, J., Nenadic, G., 2019. Machine learning methods for wind turbine condition monitoring: a review. Renew. Energy 133, 620–635. https://doi.org/10.1016/j.renene.2018.10.047.

Sun, P., Li, J., Wang, C., Lei, X., 2016. A generalized model for wind turbine anomaly identification based on SCADA data. Appl. Energy 168, 550–567. https://doi.org/10.1016/j.apenergy.2016.01.133.

Sun, S., Wang, T., Chu, F., 2022a. In-situ condition monitoring of wind turbine blades: a critical and systematic review of techniques, challenges, and futures. Renew. Sustain. Energy Rev., 112326 https://doi.org/10.1016/j.rser.2022.112326.

Sun, S., Wang, T., Yang, H., Chu F, F., 2022b. Condition monitoring of wind turbine blades based on self-supervised health representation learning: a conducive technique to effective and reliable utilization of wind energy. Appl. Energy 313, 118882. https://doi.org/10.1016/j.apenergy.2022.118882.

Tautz-Weinert, J., Watson, S.J., 2017. Using SCADA data for wind turbine condition monitoring - a review. IET Renew. Power Gener. 11 (4), 382–394. https://doi.org/10.1049/iet-rpg.2016.0248.

Villanueva, D., Feijóo, A., 2018. Comparison of logistic functions for modeling wind turbine power curves. Electr. Power Syst. Res. 155, 281–288. https://doi.org/10.1016/j.epsr.2017.10.028.

Wang, A., Qian, Z., Pei, Y., Jing, B., 2022b. A de-ambiguous condition monitoring scheme for wind turbines using least squares generative adversarial networks. Renew. Energy 185, 267–279. https://doi.org/10.1016/j.renene.2021.12.049.

Wang, A., Pei, Y., Qian, Z., Zareipour, H., Jing, B., An, J., 2022a. A two-stage anomaly decomposition scheme based on multi-variable correlation extraction for wind turbine fault detection and identification. Appl. Energy 321, 119373. https://doi.org/10.1016/j.apenergy.2022.119373.

Wang, K.S., Sharma, V.S., Zhang, Z.Y., 2014. SCADA data based condition monitoring of wind turbines. Adv. Manuf. 2 (1), 61–69. https://doi.org/10.1007/s40436-014-0067-0.

Wei, L., Wenxiang, Z., Chuncheng, L., Guohu, G., et al., 2015. Thermal degradation mechanism of poly(hexamethylene carbonate. Polym. Degrad. Stab. 112, 70–77. https://doi.org/10.1016/j.polymdegradstab.2014.12.013.

Xu, M., Feng, G., He, Q., Gu, F., Ball, A., 2020. Vibration characteristics of rolling element bearings with different radial clearances for condition monitoring of wind turbine. Appl. Sci. 10 (14) https://doi.org/10.3390/app10144731.

Yang, W., Court, R., Jiang, J., 2013. Wind turbine condition monitoring by the approach of SCADA data analysis. Renew. Energy 53, 365–376. https://doi.org/10.1016/j.renene.2012.11.030.

Zhang, G., Wang, X., Liang, Y.C., Liu, J., 1993. Fast and robust spectrum sensing via Kolmogorov-Smirnov test. IEEE Work. Local Metrop. Area Netw. 58 (12), 3410–3416. https://doi.org/10.1109/TCOMM.2010.11.090209.