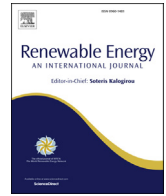




ELSEVIER

Contents lists available at ScienceDirect

Renewable Energy

journal homepage: [www.elsevier.com/locate/renene](http://www.elsevier.com/locate/renene)

# Data annotation and feature extraction in fault detection in a wind turbine hydraulic pitch system

Panagiotis Korkos<sup>a,\*</sup>, Matti Linjama<sup>b</sup>, Jaakko Kleemola<sup>c</sup>, Arto Lehtovaara<sup>a</sup>

<sup>a</sup> Tribology and Machine Elements, Materials Science and Environmental Engineering, Faculty of Engineering and Natural Sciences, Tampere University, P.O. Box 589, 33014, Tampere, Finland

<sup>b</sup> Automation Technology and Mechanical Engineering Unit, Faculty of Engineering and Natural Sciences, Tampere University, P.O. Box 589, 33014, Tampere, Finland

<sup>c</sup> Suomen Hyötytuuli Oy, P.O. Box 305, 28601, Pori, Finland

## ARTICLE INFO

### Article history:

Received 12 March 2021

Received in revised form

7 September 2021

Accepted 12 December 2021

Available online 17 December 2021

### Keywords:

Pitch system

Wind turbine

Fault detection

ANFIS

SCADA

## ABSTRACT

The performance of wind turbines can be improved by processing supervisory control and data acquisition (SCADA) data. SCADA data can be processed in a reasonable time to enhance decisions made about maintenance schedules. The pitch system is critical in improving wind turbine operation by analysing data of the most relevant SCADA features. This study gathers the most significant pitch faults, and by implementing the adaptive neuro fuzzy inference system (ANFIS) technique it demonstrates the fault detection potential of this technique. The proposed approach includes the detailed pre-processing of SCADA data, emphasising the labelling process, in which a modified power curve monitoring method is used. During the implementation of the ANFIS, different combinations of the selected parameters were tested for their effects on the performance of fault detection. This methodology was implemented at a windfarm, commissioned in 2004, in five 2.3 MW fixed-speed onshore wind turbines equipped with a traditional servo-valve controlled hydraulic pitch system. Overall, data on 10 years of the operation of each wind turbine were utilised, and a total of nine pitch events were considered. Individual measurement for each blade angle was available for detecting pitch faults. Results demonstrated above 86% achievement of F1-score for pitch fault detection.

© 2021 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Wind energy is one of the most promising renewable energy sources. Engineers have made many attempts to use it to produce both kinetic and electric energy. In recent decades, many wind turbines have been installed globally to eliminate the dependence of energy production on the coal industry and fossil fuels. According to WindEurope's annual report in 2019, the total capacity of installed wind power in Europe was 205 GW. In European Union countries, 15% of the electricity demand was met by wind energy [1]. Denmark achieved the highest score in the share of wind energy, as 48% of its electricity demand was met by wind energy, followed by Ireland, Portugal, Germany, and the UK, all of which had wind energy shares above 20%. According to the Global Wind Energy Council [2], in 2019, the global capacity of installed wind

turbines was 650 GW. In addition to onshore wind turbines, many countries have also developed windfarms in the sea. These offshore wind turbines demand very tight scheduling and accurate condition monitoring to compensate for the cost of offshore structures and operations.

A wind turbine is a complex mechanical system that comprises many subassemblies and components. It consists of a rotating shaft, generator, pitch system, yaw system, gearbox, hydraulic system, and electronic system. One of the most critical subsystems is the pitch system. The pitch system is responsible for the rotation of the blade, which controls the angle of attack of the wind to extract the maximum outcome of the provided wind energy. Pitch systems are either hydraulic or electric. The hydraulic pitch system, which is the type examined in the present study, includes several components, such as hydraulic cylinders, accumulator tanks, valves, pumps, pitch bearing, pitch pawl, and slip ring. According to a recent survey of the ReliaWind project [3], the faults and failures occurring in this system included 15.5% of total failures and 20% of total downtime. Similar conclusions have been drawn based on a survey of modern

\* Corresponding author.

E-mail address: [panagiotis.korkos@tuni.fi](mailto:panagiotis.korkos@tuni.fi) (P. Korkos).

multiMW offshore wind turbines [4], where the pitch and hydraulics subsystem showed the highest failure rate, accounting for ~13% of the overall failure rate. An interesting finding of this survey was that 17% of the overall pitch/hydraulic failures were caused by oil issues, 13.9% of which were valve issues, followed by accumulator, sludge issues, and pump repair and replacements, accounting for 10.7%, 6.4%, and 5.9%, respectively. Moreover, an earlier survey of Swedish wind power plants conducted from 2000 to 2004 indicated that almost 27% of the overall failures appeared at the pitch/hydraulic/hub subsystem [5]. Based on the findings of these previous surveys, the pitch system was considered the largest contributing factor.

Condition-based maintenance depends on the installation of condition monitoring systems in the preferred subsystems of a wind turbine. The development of various condition monitoring systems for every wind turbine component drastically increases the cost and complexity of condition-based maintenance. Nevertheless, every wind turbine is equipped with an integrated system of sensors and acquisition system called supervisory control and data acquisition (SCADA) system. Initially, they have been built to control electricity generation, by providing time-series signals in a low sampling frequency. Typically, the signals are recorded with 1 Hz sampling frequency, but they are available as statistical measures at 10-min intervals including average, standard deviation, maximum and minimum. SCADA system monitors the main components by recording a plethora of parameters including temperatures of bearings, lubricating oil and winding [6–8]. However each manufacturer has its own SCADA systems, leading to the conclusion that not all the wind turbines from different manufacturers record the same set of SCADA signals and their taxonomy differs a lot as well [9]. The usefulness of using SCADA data is affected by the facts that SCADA systems have been installed in almost every wind turbine. Thus it is cheap to use the signals instead of installing additional hardware and increasing the cost of maintenance [10]. However, disadvantages of using SCADA systems include the low sampling frequency which lowers the possibility to capture transient dynamic phenomena and the errors that may contain due to possible software updates. In addition, not all the SCADA systems follow the same taxonomy for signals and parameter names, making difficult the implementation of the developed approaches to every wind turbine, regardless of the manufacturer.

The monitoring of hydraulic pitch systems by using artificial intelligence techniques has been reported in relevant publications in the literature. Chen et al. [11,12] developed an online fault detection system for both hydraulic and electric pitch systems using an adaptive neuro fuzzy inference system (APK-ANFIS). They selected this technique after studying several options, such as fuzzy inference systems (FIS), k-means clustering, a self-organising map, artificial neural networks, naïve Bayes, Bayesian networks, support vector machines (SVM), and the adaptive neuro fuzzy inference system (ANFIS). In this study, five features (i.e., power output, wind speed, blade angle, rotor speed and motor torque for only the case of electric pitch system) were utilised, which showed that the methodology could predict faults 21 days prior to a catastrophic failure.

Schlechtingen et al. [13,14] proposed using the ANFIS to build normal behaviour models (NBM) by manipulating SCADA data to detect faults in various components of wind turbines. This approach can be implemented in various components of wind turbines, and the model was trained using nine months of operational data (29,513 10 min average values). In the second part of their published study [14], their results showed that their methodology detected hydraulic oil leakage, which is relevant for the present study. The only drawback of this technique is that the fuzzy expert application module yields only a single diagnosis at a time

for each defined component/subsystem. Moreover, in the case of two overlapping faults, the diagnosis is not adequate.

Leahy et al. [15] proposed a promising SVM method for detecting faults. However, his study pointed out that several improvements should be performed to tackle some problems, including the bias of the classifiers. In addition, in a more recent study [16], the same authors compared different SVM methods, including undersampling, oversampling, and ensemble methods, according to their ability to detect, diagnose, and predict a fault. Concerning the use of SVM, L. Hu et al. [17] aimed to enhance the feature set by adding features that were based on domain knowledge and the understanding of the physics of wind turbines and measurements. Kusiak and Verma [18] used genetic algorithms to detect blade pitch-related faults using SCADA data but did not specify the causes of faults; instead, only some effects of faults were reported, such as blade angle asymmetry and blade angle implausibility.

Pandit and Infield [19] investigated the use of the support vector regression (SVR) method to detect faults by implementing it on the pitch curve. Specifically, binned and SVR-based pitch curves were studied, and the results showed that the binned pitch curve was far too slow, in contrast to the SVR, which detected anomalies quickly. The same authors [20] proposed a Gaussian process (GP) algorithm to estimate operational curves based on key turbine critical variables, which could be used as a reference model to identify critical wind turbine failures and improve power performance. Three parameters (i.e., power curve, rotor speed curve, and blade pitch angle curve) were used to detect critical wind turbine failures.

Guo and Infield [21] developed a multivariable power curve model with a modified Cholesky decomposition GP, which detected faults via the power curve and identified them by processing raw signals. In this method, the inputs were wind speed, wind direction, pitch angle, yaw error, rotor speed, and tip speed ratio; the only output was power. This model was compared with the binning method and the sixth-order polynomial regression method. Their performances were evaluated based on the mean absolute percentage error (MAPE). For that reason, a sequential probability ratio test (SPRT) with two groups of hypotheses was introduced to analyse and detect abnormal changes. Skrimpas et al. [22] attempted to diagnose pitch faults by processing vibration signals from the main bearing using k-means clustering. In particular, pitch issues were detected by analysing the effects of vibration on the main bearing accelerometers and applying environmental noise and speech recognition techniques. Wu [23] used an asymmetric support vector machine (ASVM) to diagnose the fault of cylinder internal leakage. The developed ASVM model was adopted to reduce the possibility of missed not-fault prediction. The results showed that fewer support vectors and a lower order kernel could be chosen to derive the model of the fault map.

The review of the relevant literature showed that 10-min SCADA data have been frequently used in previous studies. The total available dataset, from which the training dataset was extracted, was between nine months and two years. However, because adequate data were lacking to complete a fault identification study in the pitch system of wind turbines, some regions were sparsely represented. The parameters varied between two (a simple model), four (more complex pitch system analysis), and 33 in the implementation of techniques that did not suffer from the “curse of dimensionality”, such as SVM and ANFIS-based normal behaviour modelling. Regarding the techniques used, ANFIS, with the incorporation of a priori knowledge in the case of sparse data and SVM, were the most promising methods for fault detection in the pitch system.

However, previous studies [11–23] had provided no or limited detail about the faulty cases, leading to sparse representation in

specific areas of SCADA representations such as in the power curve. In other words, no adequate information about the component of the pitch system, in which the fault occurred, was available. Thus, it is not clear if they had included all the possible faulty cases in their studies, related to the pitch system of a wind turbine. Also other scientists have limited their work in doing anomaly detection, taking into account only the healthy operation of a wind turbine [13,14]. Consequently, this approach does not provide the possibility for researchers to identify fault type. On the contrary, using faulty data from each faulty case for fault detection in the pitch system is a first stage towards fault identification which will be investigated as a future topic by the authors. Regarding data annotation, most of the researchers who have performed fault detection, have not presented a detailed approach of labelling the data. When a dataset consists of periods before and after maintenance, data points are mixed in the power curve, thus, in this study, the detailed approach of labelling the data is elaborated based on the maintenance log.

The objective of the present study is to develop a method for fault detection in a hydraulic pitch system. The available dataset contains 10 years of 10-min SCADA data on five wind turbines. From this dataset, nine pitch events of different types were selected, and semiautomatic data labelling was conducted using a modified version of a power curve monitoring method. Fault detection was accomplished using the ANFIS technique, and the effects of selected features and their combination were considered in the evaluation of the model. To determine the potential of the method for fault detection, separate measurements of the angle of each blade were considered instead of averaging them.

## 2. Description of data

### 2.1. Available data

In this study, the data were collected from the SCADA system of a windfarm in north-western Finland. The windfarm consisted of five fixed-speed 2.3 MW wind turbines with a hydraulic pitch system. Each blade had an independent pitch system, and each blade angle was measured separately. The reference power curve, provided by the manufacturer, is presented in Fig. 1(a), which shows the performance capabilities of the wind turbine. The power curve represents the power output against wind speed, and three regions were distinguished based on the wind speed values. These three values defined the operation of the wind turbine, and they are elaborated as follows. At wind speeds below the cut-in speed, the wind turbine does not work. Typical values for this parameter are 3 or 4 m/s. In the area between the cut-in speed and the rated speed, where the maximum power output was first observed, the wind turbine started working, and the power output ideally followed a polynomial curve that was proportional to the cubic power of the wind speed. The value of rated speed in the studied system was 17 m/s. In the last region, which was defined by the rated speed and the cut-out speed (equal to 25 m/s), the wind turbine operated at a stable power output that was equal to the nominal power output. At wind speeds above the cut-out speed, the wind turbine stopped operating.

The stored data covered a period of almost 10 years from 1 July 2007 to 1 April 2017. The available data contained 10-min values of various operational and non-operational parameters. These data were stored in the SCADA system as average, standard deviation, maximum, and minimum values, which were stored every 10 min. The signals were collected at a higher frequency, but the SCADA system includes only these four statistical quantities. An important characteristic of these data was that the values were stored in the SCADA system at the precision of one decimal digit, which

conformed to the quasi-industry standard. SCADA data were available in csv format and retrieved from an SQL database. In addition to the SCADA data, the maintenance log was available, and this information provided guidance in determining faulty and non-faulty status as well as the alarm log, in which only hub and hydraulic-related pitch alarms were considered in this study. In the study period, 16,399 hub-related and 1,007 hydraulic-related alarms were recorded.

Regarding feature selection, although a plethora of parameters could be recorded in the SCADA system, only features related to the pitch system were considered in this study. Other features, which had a direct relation to other subsystems such as the yaw system, drivetrain were ignored. The selection was performed by a team of wind turbine operators as a first step to reduce the dimension of the available features and focus on the monitoring of pitch system. Therefore, the list of features was selected by taking into account domain knowledge. Among them were environmental parameters, such as wind speed, ambient temperature, and wind direction, as well as operational data, such as power output, pitch blade angles, rotor speed, hub and hydraulic temperatures, and hub and hydraulic pressures. Typical curves of blade angles against wind speed and rotor speed against wind speed are shown in Fig. 1 (b) and (c), respectively. These features, as well as the power curve, were selected for use in the fault detection process. Some additional parameters indicated the status of the generator, such as whether the generator speed was either 1,000 or 1,500 rpm and the status of breaks to inform the operator about possible stops. Moreover, other parameters indicated, for example, the number of times valves were opened and closed during a period of 10 min and the condition of the lubrication system.

### 2.2. Pre-processing of data

Before the data analysis was conducted, the data underwent a pre-processing procedure. This study focuses only on the average values and the standard deviations of the available data. Even though both maximum and minimum values in a 10-min interval are available for all the parameters, they are not used in the current study. The reason is that maxima and minima within a 10-min period may occur a single time instant that the value of a parameter had increased rapidly. This behaviour is not indicative of wind turbine operation, as it does not necessarily mean that a fault occurred. Summarizing, different combinations of average and standard deviation values of the selected features are incorporated in the model to make a decision about the best feature set for a pitch fault detection task. Regarding the data, the values were missing in some timestamps, which may have been because of errors in the sensors or the recording system. It would have been better to replace these NaN values by approximate values based on the parameter's prior and posterior values, but because a large amount of data was available, this omission was not problematic. Furthermore, a validity check [24] was performed, in which wind speeds above 25 m/s and blade angles outside the range of  $[-90, +2]$  degrees were filtered out. Some interesting cases were present at the power curve, which required investigation. Because the studied wind turbines were equipped with a fixed-speed generator, in cases of wind gusts, the pitch angle controller was very slow in adjusting the blade angle. As a result, the power output was larger than normal. Although it deviated from the rest of the data, it was linked to the normal operation. In addition, a significant attribute of the power curve was that if points that exceeded the nominal power were eliminated, then the power curve was not formed. The reason was that there were only a couple of points in the rated power region, as shown in Fig. 2. It should be noted that at points where the power was above the nominal, the generator

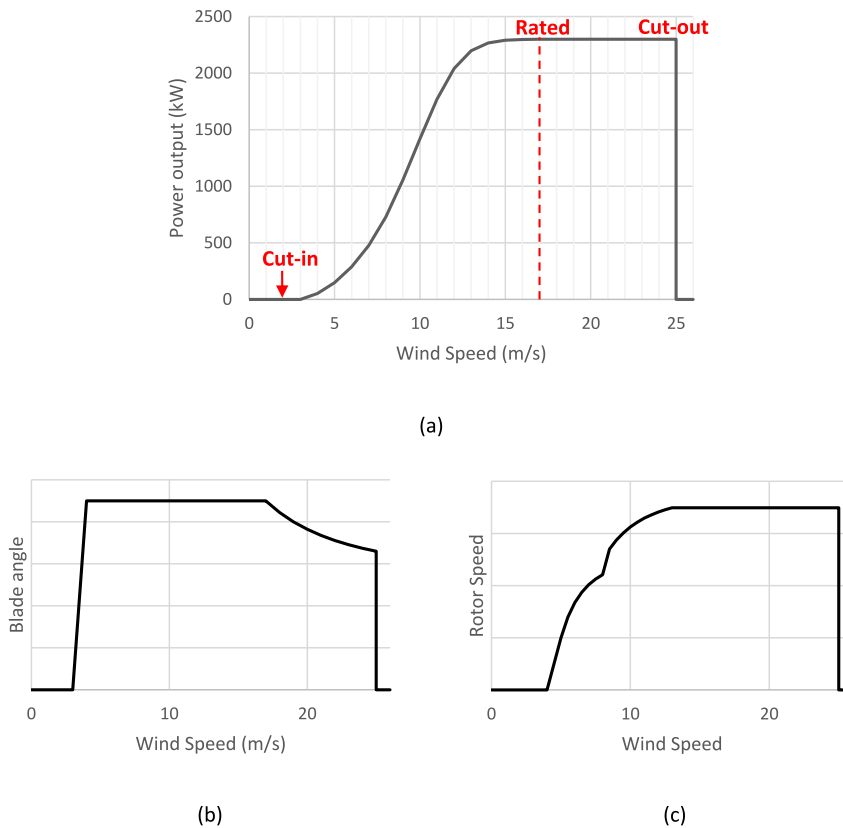


Fig. 1. (a) Power curves of studied wind turbines; (b) typical curve of blade angle as a function of wind speed; (c) typical curve of rotor speed as a function of wind speed.

speed was 1,500 rpm and the brakes were released. Furthermore, at these points, the blade angles and rotor speed were within normal ranges. Hence, these points were included in the study. Regarding the blade angles, a few values that exceeded the aforementioned range were recorded, but because their statuses corresponded to the parked position of the blades with attached brakes and very low power output, they were eliminated. In addition, concerning the wind speed measurement, each wind turbine was equipped with two anemometers; thus, measurements of wind speed by the primary and secondary anemometers were available. In some cases, the primary anemometer may have failed, so the secondary

anemometer would have been used to measure the wind speed.

The data were stored in the SCADA system in different tables based on the meaning of the parameters. These tables did not include exactly the same list of timestamps; therefore, unlike the rest of the values, the values of some parameters at specific timestamps were not stored. For that reason, the timestamps, which did not exist in some of them, were filtered out to ensure that the same timestamps and sizes of all parameters were recorded.

After the above steps were completed, the values of the data were normalised using the max-min normalisation equation (Eq. (1)). The normalised average values are shown in Fig. 3. It is

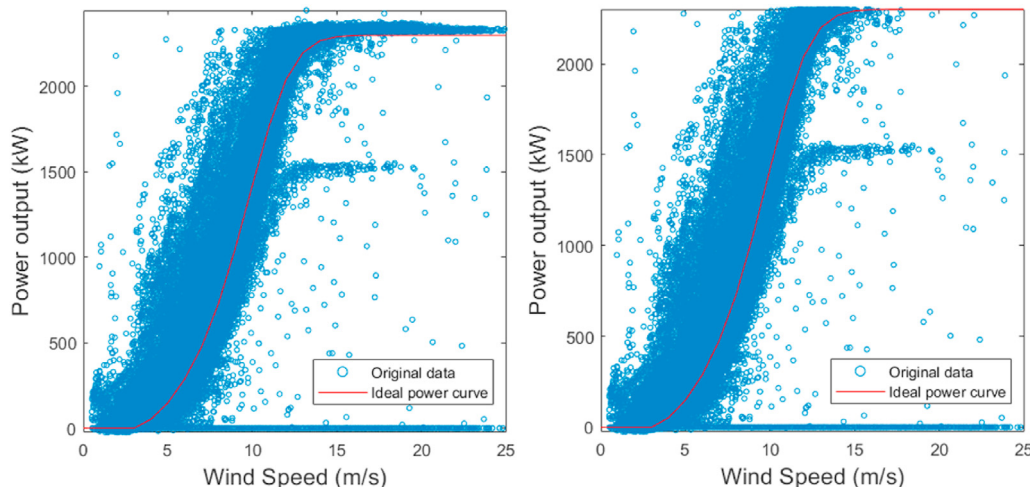


Fig. 2. Original data of power curve (left) and original data limited to nominal power output (right) accompanied by the reference curve (red).

important to point out that each parameter was normalised based on the maximum value across all the wind turbines and not individually, which ensured that the data were not wind turbine-dependent or environment-dependent. Different conditions prevail at each location where a wind turbine is placed. In addition, the recorded parameters were in different ranges compared with the others. Hence, in the feature-scaling process, the range of all parameters was set between zero and one (0,1). Consequently, all the algorithms were run more quickly. The results of the max-min normalisation of the power curve are shown in Fig. 3 (a), which includes the power curve of each wind turbine. As shown in Fig. 3, the dataset of each wind turbine contained different faults, as there were different clusters of abnormal points to the left and right of the imaginary power curve. Fig. 3 (b) shows the entire normalised dataset, which includes the dataset of all the wind turbines.

$$x_{new}^i = \frac{x^i - x_{min}}{x_{max} - x_{min}} \tag{1}$$

### 3. Data labelling

After the pre-processing was completed, every data point was associated with a label that indicated the presence or absence of a fault. In the case of a binary classification, these labels were “0” or “1”, indicating the absence and presence of faults, respectively. The remaining question was how to determine these labels in performing the data annotation.

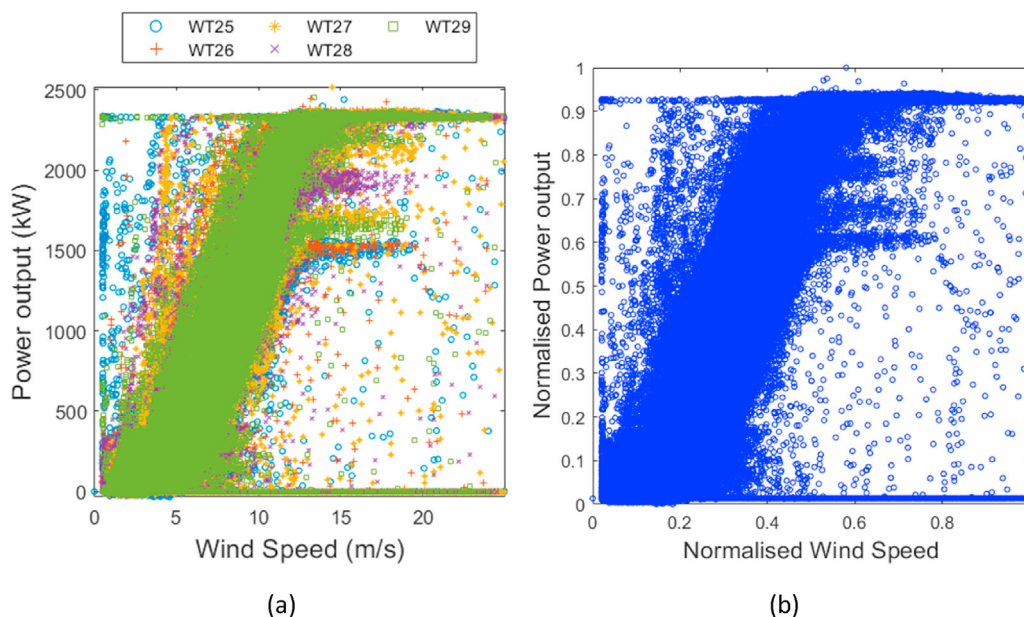
For the purpose of data annotation, the maintenance log and the alarm log were checked. The maintenance log included the maintenance of all the wind turbines, but the start and end times of each task were not recorded. Nevertheless, some maintenance actions may not have solved the problem because the technicians may not have taken the right decision, or they may have had poor information about the condition of the wind turbine. Moreover, their decisions are based on generated alarms. However, because their trigger rate was very high, it was very difficult to investigate the causes of the alarms. Therefore, in these cases, abnormal points after maintenance were observed at the power curve. The events

that were recorded in the maintenance log were gathered. The most frequent events, regarding the pitch system, are shown in Table 1.

The selection of periods was accomplished by observing the power curve and the other three critical characteristic features (CCF). Following [11], three periods were distinguished based on the maintenance period: “Generating Fault” prior to the maintenance task; “Maintenance” during the maintenance task; and “After Maintenance” after the maintenance task. Because the recorded date of maintenance was not precise, the maintenance period was selected by taking into account both the date of the recorded maintenance task in the maintenance log and the most relative alarms that were generated in that period. The selection was then based on the deviation from the ideal power curve and the other two critical characteristic features. Typical CCF curves are shown in Fig. 1. However, it was expected that the real data points would not match the ideal curves because of the dynamic nature of wind energy and the dynamic state of the wind turbine [25]. Fig. 4 shows an example of selecting the “After Maintenance” period, where the data points, which are distanced from the imaginary ideal power curve, are assumed to be abnormal. Therefore, the usage of upper and lower bounds was necessary to correctly label the data points [25]. Moreover, it is worth noting that the labelling process would have been easier if data from the beginning of operation of the new wind turbines were available. The reason is that these data would represent a normal operation, as it was assumed that no fault

**Table 1**  
Event list.

Fault	Frequency
Hydraulic hoses and oils replacement	22
Hub oil leakage + Hyd. Oil replacement + Bl. valve 6 replacement	1
Block replacement at blade B (No3)	9
Block leakage in blade B(No1)	1
Replacement of A- blade valve 102 (No3)	3
Replacement of A, B, C- blade valve 116 (No3)	6
Nitrogen accumulator (No 4) replacement of Blade A (No5)	5
Blade tracking error during stop/operation of Blade A (No1)	34
Replacement of hyd. cylinder (No2)	26



**Fig. 3.** (a) Non-normalised and (b) normalised power curve.

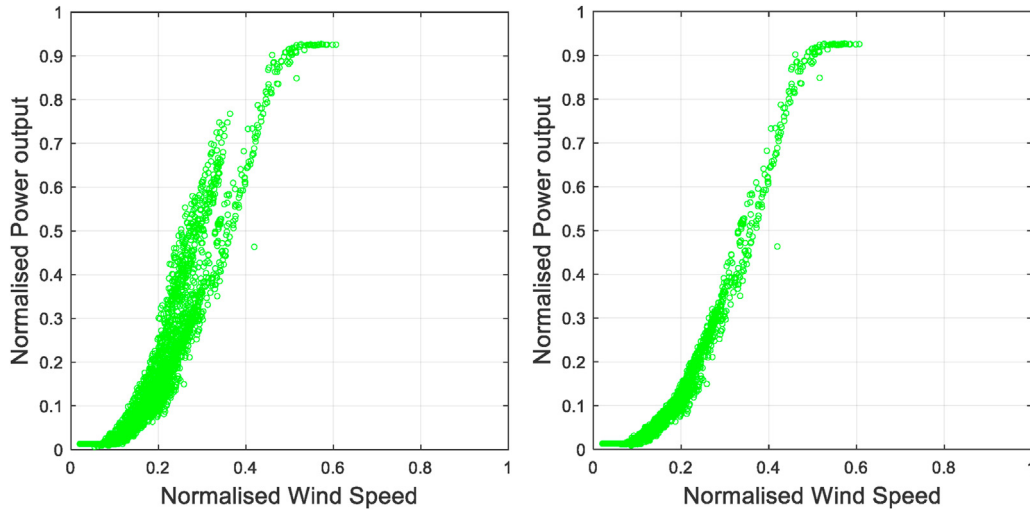


Fig. 4. Wrong (left) and correct (right) “After Maintenance” period.

occurred when a new turbine began operating. However, in this study, the study period was after the commencement of operations, and it was unclear whether the first data corresponded to normal data.

As shown in Table 1, because one event was selected from each event type, our dataset included nine incidents. The power curves of the selected events are shown in Fig. 5, which includes the “Generating Fault” and “After Maintenance” data points. It should be noted that the use of the entire dataset was not recommended

because it contained faults in other subsystems, which would have skewed our data. Hence, periods referring only to pitch subsystem-related faults were chosen. The challenge in setting up the training dataset was the annotation of the “Generating Fault” data points, which were in the normal region. If the data points in this period were among the points in the “After Maintenance” period, they were assigned as normal points. To tackle this challenge, a filter was implemented by using a modified version of Park et al.’s [26] power curve monitoring method to construct two boundary curves that

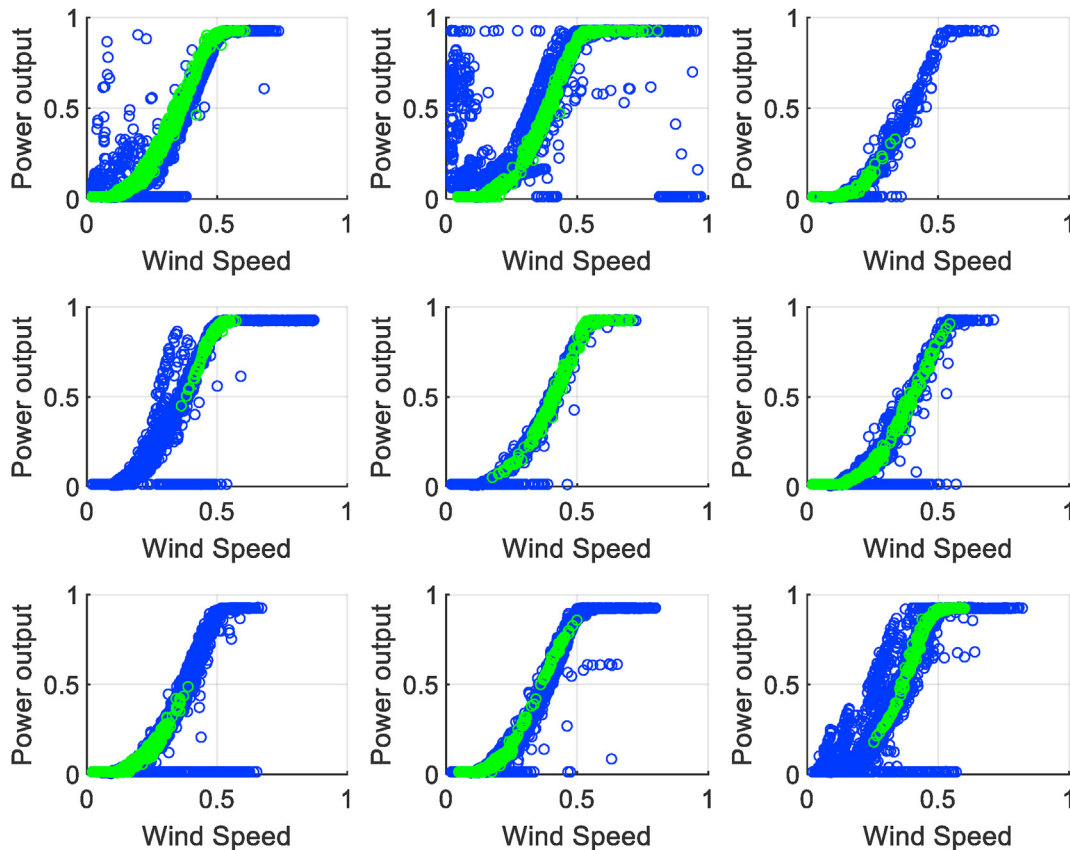


Fig. 5. “Generating Fault” (blue) and “After Maintenance” (green) data points in each fault incident.

included the data on the “Generating Fault” period, which were in the normal region.

This power curve monitoring method was applied to perform an optimised estimation of power output with respect to wind speed, thus resembling the ideal power curve. Specifically, the input data were sorted according to a variable speed bin, the value of which decreased after each iteration. Because the input data were normalised, the width of the wind speed bin was equal to 0.1 divided by the number of iterations of the overall algorithm loop to be adapted to smaller values between 0 and 1. The average power and standard deviation were calculated per each bin, and the average values were interpolated by applying a cubic spline. In the fourth and fifth stages of this method [26], the parameter, which was used to move the estimated power curve left or right, was  $\Delta v = 0.004$ , and a similar one,  $\Delta P = 0.0025$ , was then selected to move it up or down. A modification at the 4th stage was performed which directly affects the 5th stage. Therefore, at wind speeds larger than the normalised rated speed, if the value of upper boundary curve is non positive, the estimation of power for this specific wind speed will replace it. This change was performed to solve the problem of non positive power after a certain point, as was expected. Finally, regarding the parameters that determined whether the optimal positions of limits were attained during their movement,  $\beta_{\text{shift}} = 1\%$  and  $y_{\text{offset}} = 0.05\%$  were selected. The flow chart in Fig. 6 shows the steps in this filter.

Fig. 7(a) presents the “Generating Fault” and “After Maintenance” data points, including the estimated power curve. To assign the “Generating Fault” data points, which were among the “After Maintenance” data points, as normal points, the estimated power curve was moved right and left equally by 0.02 and up and down equally by 0.0375.

#### 4. Fault detection and data analysis

The ANFIS [27] has a distinct effect on interpretability, especially in systems where the operation is based on historical data. Insights into underlying physical phenomena are indispensable; this capability is provided by ANFIS. This hybrid model consists of an artificial neural network (ANN) component that enables training a model based on historical data, as well as a fuzzy logic component that connects linguistic statements to data, thus resembling human logic in terms of constructing IF-THEN rules (e.g., see Rule 1, Rule 2 below). Such rules are defined by an expert on the field, where fuzzy logic is applied, in order to convert uncertain statements, made by human, to mathematical expressions.

Concerning the ANFIS architecture, the first-order Takagi–Sugeno fuzzy system was presented as assuming two inputs  $x$  and  $y$  and one output  $z$ . In this study, these two inputs were features on the CCF list (e.g., power output and wind speed). Then two fuzzy if–then rules were constructed as follows:

**Rule 1:** If  $x$  is  $A_1$  and  $y$  is  $B_1$ , then  $f_1 = p_1x + q_1y + r_1$

**Rule 2:** If  $x$  is  $A_2$  and  $y$  is  $B_2$ , then  $f_2 = p_2x + q_2y + r_2$

where  $A_i$  and  $B_i$  for  $i = 1, 2$  are linguistic labels of input membership function (e.g., “low”, “medium”, and “high”);  $\{p_i, q_i, r_i\}$  for  $i = 1, 2$  are the consequent parameters presented at the fourth layer, as shown below. The consequent function is a first-order polynomial because it was assumed that a first-order Takagi–Sugeno model would be built. These parameters were determined during the training of the model.

The ANFIS architecture was presented, which clearly indicated the combination of ANN and the FIS system by using layers, nodes,

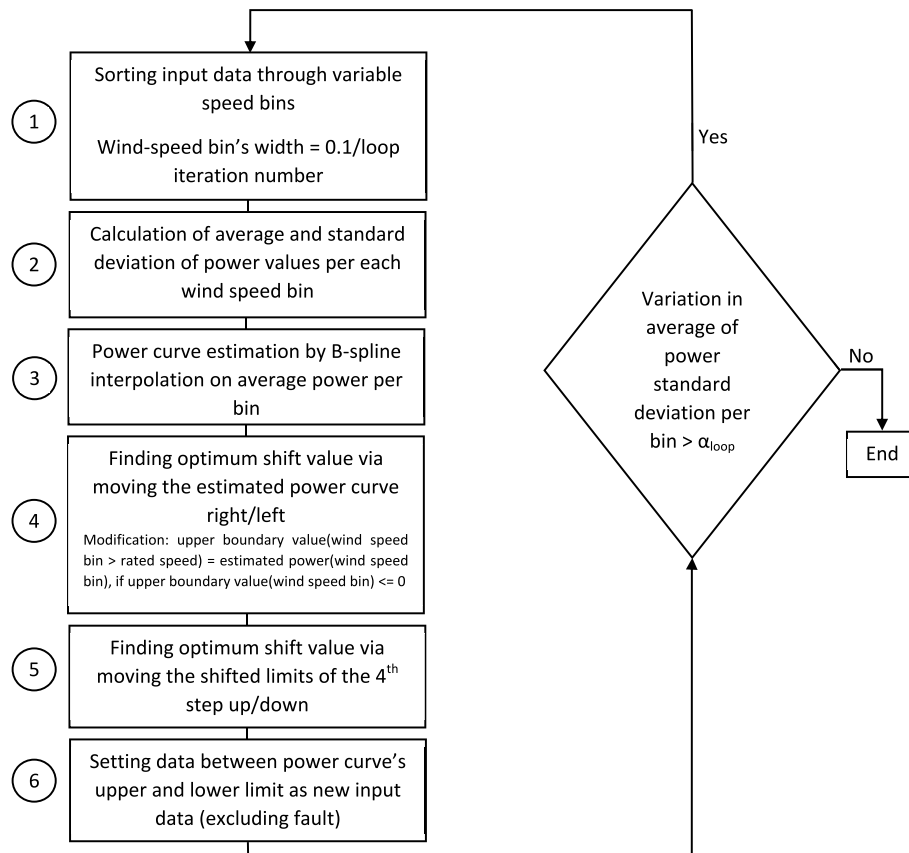


Fig. 6. Modified algorithm for automatic power curve limit computation.

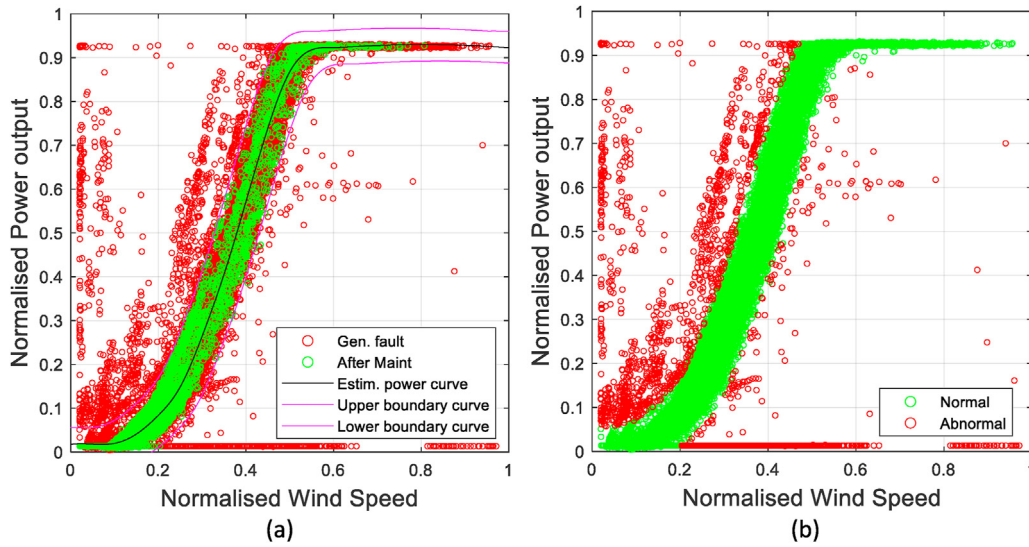


Fig. 7. (a) Dataset including “Generating fault” and “After Maintenance” data points; (b) dataset after the labelling process.

and antecedents, as well as the consequent part of the rule base, as shown in Fig. 8. This architecture is called type-3 ANFIS, which was derived by type-3 fuzzy reasoning.

As shown in Fig. 8, there were two inputs:  $x$  and  $y$ , one output, and five layers. The first and fourth layers were adaptive. The description of each layer is presented below.

**Layer 1:** This layer corresponded to the fuzzification process, wherein the crisp input values were transformed to fuzzy values. This task was accomplished by computing the value of the appropriate membership function, which described the linguistic label of the input or crisp variable. The value of this membership function implied the degree to which the input was compatible with the assumed linguistic label or not. For example, if the input was indeed low as it was assumed at the “if” condition of the rule base, the value of the membership function describing the ‘low’ label would be greater than the value of membership function describing the ‘high’ label. Thus, the outputs of this layer  $O_A^i$  and  $O_B^i$  for  $i = 1, 2$  were equal to the membership function  $\mu(x)$ , which was often a generalised bell-shaped membership function (Eq. (2)).

$$\mu(x) = \frac{1}{1 + \left| \frac{x - c_i}{a_i} \right|^{2b_i}} \quad (2)$$

where  $\{a_i, b_i, c_i\}$  are premise parameters, and their values were adjusted to the training data. In other words, the shape of these functions was altered based on the premise parameters of reaching a sufficient degree of satisfaction with the training dataset. It was highly significant that the premise parameters had a physical

meaning in the background:  $c$  represents the centre of the corresponding membership function;  $a$  is the half width of the curve; and  $b$  is a parameter that determines, in conjunction with the value of  $a$ , the slopes at the crossover points.

**Layer 2:** The second layer was fixed and not adaptive, as a common algebraic operation was implemented. The result of this layer was the product of the inputs’ fuzzy values, which were used as the firing strength of each rule to connect them to each other. In Fig. 8, the product is symbolised by the Greek capital letter  $\Pi$  inside the circle. The outcome  $w_i$  of this layer was as follows:

$$w_i = \mu_{A_i}(x) \times \mu_{B_i}(y), \text{ for } i = 1, 2 \quad (3)$$

**Layer 3:** This layer was responsible for defining the effect of each rule’s firing strength on the fuzzy set of the output. Its product was simply the normalisation of each rule’s firing strength, and it was calculated by Eq. (4). The name of this layer’s output was normalised firing strength, which was represented by the capital letter  $N$ .

$$\bar{w}_i = \frac{w_i}{\sum_{i=1}^2 w_i} = \frac{w_i}{w_1 + w_2}, \text{ for } i = 1, 2 \quad (4)$$

**Layer 4:** The aim of the so-called defuzzification layer was to obtain the crisp output of each rule. The output of this layer for every rule was the product of the normalised firing strength, computed at the third layer, and the consequent part of the rule, which was defined in the “THEN” section of each rule. Specifically, according to the aforementioned rules, the consequent function was a first-order polynomial function. The output value of this layer was calculated as follows.

$$\bar{w}_i f_i = w_i(p_i x + q_i y + r_i), \text{ for } i = 1, 2 \quad (5)$$

The consequent parameters  $\{p_i, q_i, r_i\}$ , as previously mentioned, were adaptive, which means that their values varied according to the training dataset.

**Layer 5:** The final layer was the outcome of this process, which was calculated as follows.

$$\text{Total output} = f = \sum_i \bar{w}_i f_i = \frac{w_1 f_1 + w_2 f_2}{w_1 + w_2} \quad (6)$$

After all layers were calculated, an adapted type-3 fuzzy inference system was built. In the training process, during the

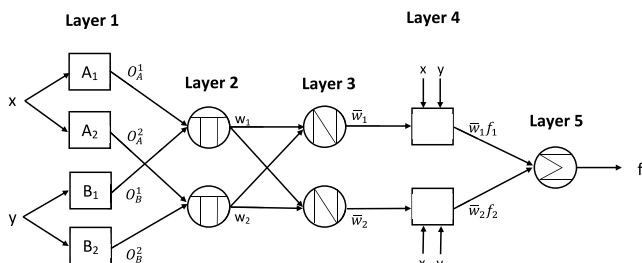


Fig. 8. Type-3 ANFIS architecture.



elaboration of the ANFIS architecture, the values of the premise parameters at the first layer and consequent parameters at the fourth layer were updated. Therefore, the learning algorithm was presented to clarify how the values of these parameters were optimised so that the ANFIS output was adjusted as well as possible to the training data. Based on the architecture presented above, the ANFIS output  $f$  was a linear combination of the consequent parameters if the values of the premise parameters were assumed to be fixed. Thus, the ANFIS output was expressed as follows:

$$\begin{aligned}
 f &= \bar{w}_1 f_1 + \bar{w}_2 f_2 = \bar{w}_1 (p_1 x + q_1 y + r_1) + \bar{w}_2 (p_2 x + q_2 y + r_2) \\
 &= (\bar{w}_1 x) p_1 + (\bar{w}_1 y) q_1 + (\bar{w}_1) r_1 + (\bar{w}_2 x) p_2 + (\bar{w}_2 y) q_2 \\
 &\quad + (\bar{w}_2) r_2
 \end{aligned}
 \tag{7}$$

Equation (7) presents output  $f$  as a linear expression with respect to the consequent parameters ( $\{p_i, q_i, r_i\}$  for  $i = 1, 2$ ). This expression demonstrated that the premise parameters were nonlinear and the consequent parameters were linear. Consequently, a hybrid learning algorithm was applied using the least squares method and the gradient descent method. Specifically, a two-pass learning algorithm was implemented, in which, during the forward pass, the least squares method was applied at the fourth layer to optimise the values of the consequent parameters under the assumption that the premise parameters were fixed at the first layer. The backward pass was activated immediately after setting up the optimal values of the consequent parameters. In this specific pass, the consequent parameters were assumed to be fixed, and the optimal values of the premise parameters at the first layer were determined using the gradient descent method to match the input as perfectly as possible. The entire procedure of this hybrid two-pass learning algorithm was recursive until the overall squared error between the desired and the actual output was below a limit value or the learning algorithm had exceeded the maximum iteration value set by the user.

## 5. Results and discussion

### 5.1. Training

This study used the ANFIS architecture presented by Chen et al. [12,25]. The input of the ANFIS model was five CCFs separately in each training accompanied by the label of each data point: power output vs. wind speed; blade angle a vs. wind speed; blade angle b vs. wind speed; blade angle c vs wind speed; and rotor speed vs. wind speed. Each feature was symbolised by  $F_i$ , where  $i$  refers to the couple of CCFs and the output is represented as  $O_i$  (see Eq. (8)). As a result, five models were trained, and five ANFIS coefficients were computed. The final result is the aggregation of these five coefficients, as shown in Eq. (9).

$$P_i = [F_{i,1}, F_{i,2}, O_i]^T, \quad i \in [1, 2, 3, 4, 5]
 \tag{8}$$

$$Result = \frac{\sum_{i=1}^5 k_i \cdot c_{ANFIS}^i}{\sum_{i=1}^5 k_i}
 \tag{9}$$

where  $k_i$  is the corresponding weight. It is worth noting that in this study, the result (see Eq. (9)) was calculated as the average of the ANFIS coefficients since all  $k_i$  were assigned a unary value ( $k_i = 1$ ).

The training dataset consisted of data that were selected from the maintenance log of a wind farm. Specifically, nine pitch events were selected to test the methodology. Regarding the hybrid learning algorithm, to reach relative convergence, the minimum

value error, which should be attained, was set at 0.01, and the maximum iterations value was set at 150. Regarding the structure of the model, an optimisation test was performed by evaluating the root mean square error (RMSE) of different numbers of membership functions (MF) in each feature over the maximum allowed number of epochs. The RMSE curves of each CCF are shown in Fig. 9, and the final optimal structure, which was selected, is listed in Table 2. For instance, the optimal structure of wind speed vs. power output was  $5 \times 4$ , which yielded a total of 20 rules in the 2D input space.

To apply the ANFIS technique and evaluate its performance, the dataset containing the nine pitch events was randomly shuffled and split into two parts. Then 80% of this dataset was used in training, and the remaining 20% was used in testing [28]. Evaluation metrics accuracy, precision, recall, and F1-score were computed. The formulae used to calculate these metrics are as follows:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}
 \tag{10}$$

$$Precision = \frac{TP}{TP + FP}
 \tag{11}$$

$$Recall = \frac{TP}{TP + FN}
 \tag{12}$$

$$F1 = \frac{2 \cdot TP}{2 \cdot TP + FP + FN}
 \tag{13}$$

where TP are True Positive, indicating that the faulty points (label “1”) were diagnosed correctly, TN are True Negative, showing that the faulty points were diagnosed incorrectly as normal. The same was applied to FP (False Positive) and FN (False Negative), whose actual label was “0”.

### 5.2. Testing and evaluation metrics

After the dataset was trained on the ANFIS model, its performance was evaluated on the test dataset using a threshold of 0.5. In each case, different combinations of average and standard deviations of the CCFs were used to test the performance of the ANFIS. The first case was the average values of the CCFs. In the second case, the average and standard deviation values were calculated against the average wind speed in 10 constructed models. In the third case, the average values vs. the average values and standard deviation values vs. standard deviation values of CCFs. 4th case contains three CCFs, namely power output vs. wind speed, rotor speed vs. wind speed, as well as the average of the three blade angles (i.e., A, B, and C) vs. wind speed. Fig. 10 shows the ANFIS performance in the test dataset using a threshold equal to 0.5 in case 1. Fig. 11 presents the values of the results (see Eq. (9)) of the bins and the threshold, which is shown as a red line. These results indicated that when the result exceeded the red line, an abnormal point was detected. Fig. 12 shows the confusion matrix in case 1, and the values of the performance metrics in each case are summarised in Table 3. The confusion matrix is a practical means of summarizing in absolute numbers the TP, FP, TN, and FN. It is worth noting that a data scientist should decide which of the available performance metrics are the most suitable for evaluating performance.

Accuracy is not a good candidate for model evaluation because it is strongly affected by the large number of normal points compared to the lesser number of abnormal points. This situation leads to unbalanced dataset, building biased models and making biased conclusions. Accuracy will be dominated by the True Negatives

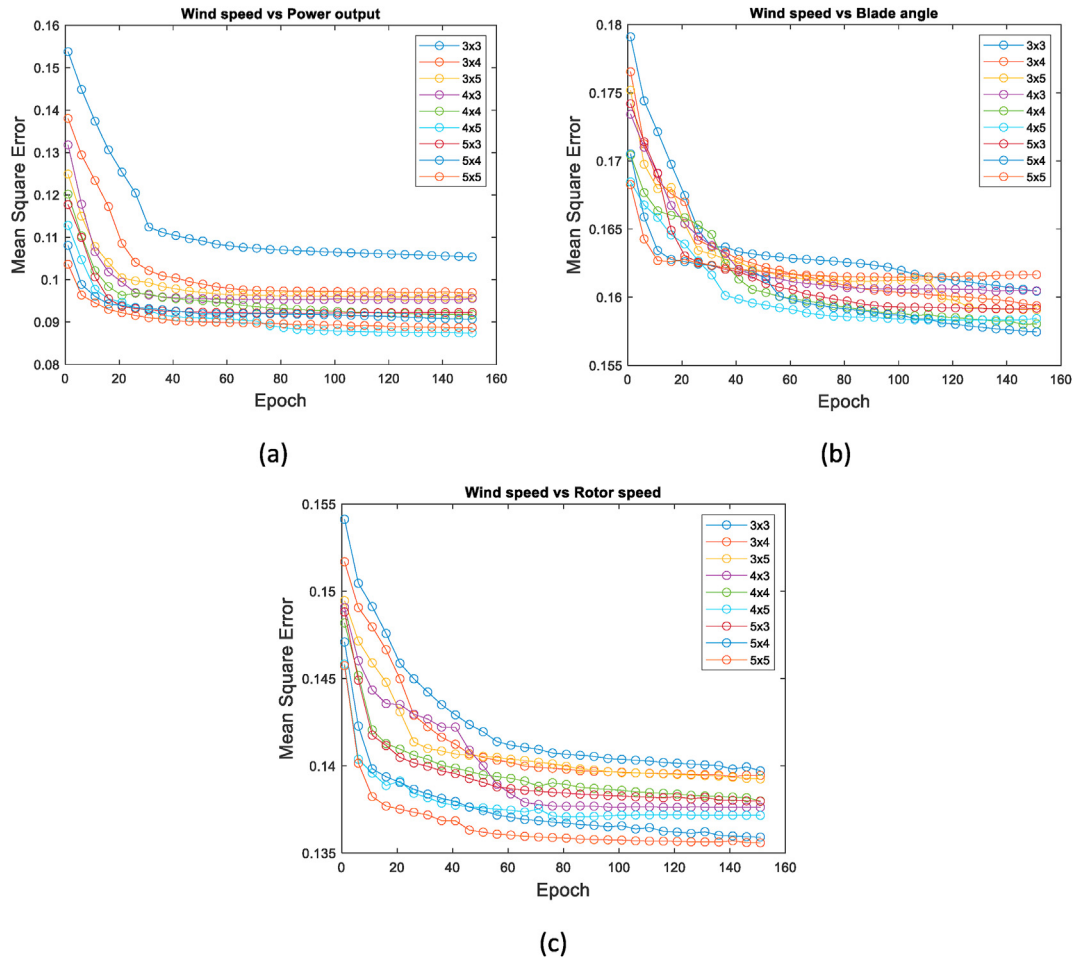


Fig. 9. RMSE curves of different ANFIS structures of (a) power output vs. wind speed, (b) blade angle vs. wind speed, and (c) rotor speed vs. wind speed.

Table 2  
ANFIS optimal structure of each CCF.

ANFIS model	Optimal structure
Wind Speed vs. Power output	4x5
Wind Speed vs. Blade angle	5x4
Wind Speed vs. Rotor speed	5x5

(points which were normal and predicted as normal) because it depends on True Negatives and True Positives (points which were abnormal and predicted as abnormal). This means that the information about detecting a fault or not will be lost. Even if a faulty point is not detected correctly, the accuracy will have a very high value because it is more possible that the normal points will be detected successfully which is not the scope of fault detection. This conclusion indicates that accuracy is a poor means of evaluating the performance of the model of such a system [28]. Other possible evaluation metrics were precision and recall, which were the fraction of correct detections reported by the model and the fraction of true events that were detected, respectively. Because both precision and recall are important, another performance metric, the F1-score, was applied, which combined the effects of these two scores. The goal in this study is to get as high F1-score as possible. High F1-score means that most of the points in the test dataset will be predicted correctly. This metric was selected to compare the performance of the ANFIS model in each case. As shown in Table 3, when the average values of the three blade angles (case 4) were

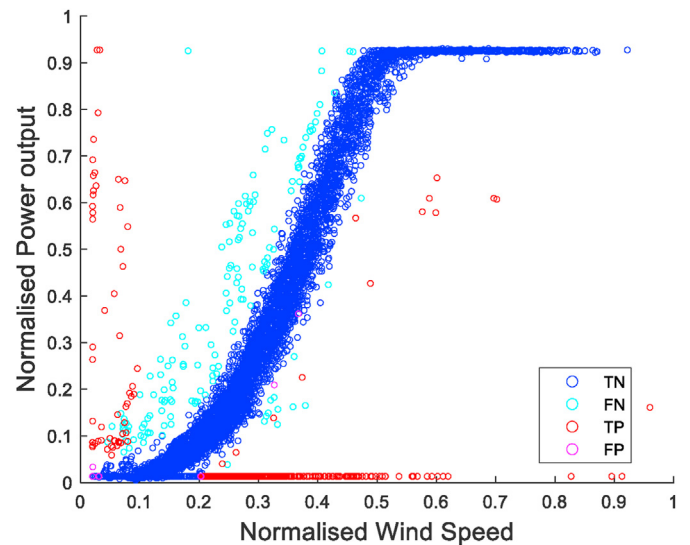


Fig. 10. ANFIS performance in the test dataset using threshold = 0.5 in case 1.

averaged, the F1-score was higher than the ones in the other cases. This result implied that in fault detection tasks but not diagnosis (i.e., the identification of specific faults), having separate values of the angle of each blade did not benefit the system.

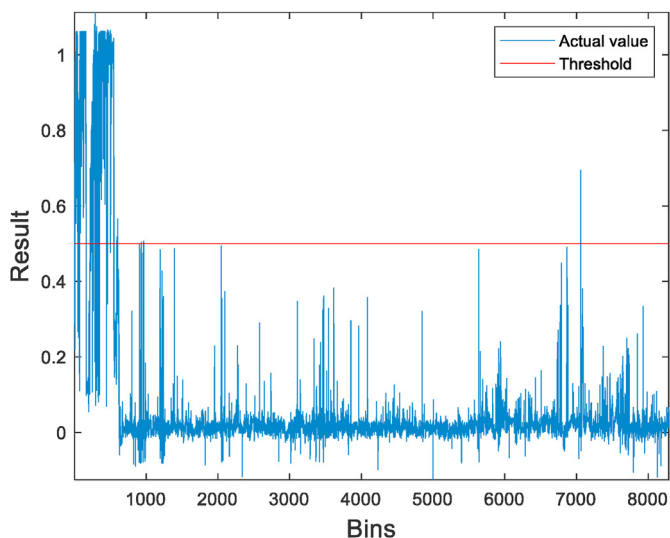


Fig. 11. Aggregation of ANFIS coefficients (blue) at the time of test dataset and threshold mapping = 0.5 (red) in case 1.

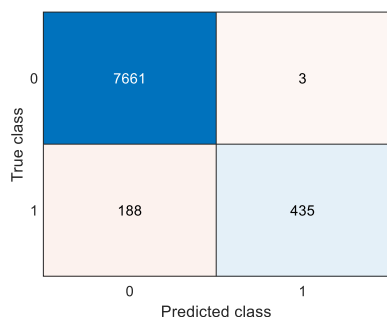


Fig. 12. Confusion matrix in case 1.

Table 3 Performance metrics in each case.

No.	Accuracy	Precision	Recall	F1-score
Case 1	0.9770	0.9932	0.6982	0.8200
Case 2	0.9777	0.9953	0.7005	0.8223
Case 3	0.9714	0.9890	0.6058	0.7513
Case 4	0.9823	0.9959	0.7687	0.8677

The attained result of the current approach cannot be compared directly with those of previous approaches due to the usage of different datasets and pre-processing strategies. In addition, data may differ from studies to studies as wind turbines are experiencing different environmental conditions, regarding their location, thus making hard to compare their results with each other. However, some indicative results may be presented to provide a benchmark for the comparison. Leahy et al. [15] managed to achieve 0.65 F1-score, without showing more details about the faults. Additionally, 0.9 F1-score was attained by Hu et al. [17] when enhancing the previous feature set. Regarding a similar approach [11], Chen et al. attained 0.5 F1-score for fixed-speed wind turbines using some pitch faults, providing no information about them. Therefore, the attained F1-score of almost 0.87 of the current approach demonstrates its potential for the purpose of fault detection. The proposed approach refers to a traditional servo-valve controlled hydraulic pitch system, equipped in fixed-speed wind turbines. This means that if it is applied in a different type

of wind turbine, i.e., variable speed, or in hydraulic pitch systems, comprised of hydraulic motors instead of hydraulic cylinders may affect the results. Nevertheless, apart from the technology of the hydraulic pitch system and wind turbine as a total, the current study presents satisfying results for pitch fault detection due to the extensive dataset, being rich of different pitch faults.

### 6. Conclusion

The aim of this research was to detect pitch events in a hydraulic pitch system of wind turbines utilising 10 years of 10-min SCADA data derived from five fixed-speed wind turbines. The entire dataset was pre-processed, excluding points beyond specified ranges and including points on the power curve associated with normal operation but deviated from the nominal power because of blade controller issues. The pre-processing procedure ended with the normalisation of the features using max-min normalisation. From this dataset, nine representative pitch events were selected, and the periods before and after the maintenance events were determined. The challenge of labelling the data was addressed by implementing a modified version of a power curve monitoring method. The power curve was estimated using the dataset of the nine pitch events, and then the boundary curves were set to include the data points belonging to the periods before maintenance, which were among the data points included in the periods after maintenance. Points within the boundary curves were assigned as normal points, and the rest were assigned as abnormal. Each pitch event, which was used in training the model represented a different type of pitch fault, such as valve fault, hydraulic cylinder fault, and so on. Because of the diversity of these faults, the model was more robust to pitch faults. After the data annotation, the ANFIS model was built using 80% of the dataset for training and 20% for testing. The results were assessed using the F1-score.

The performance of fault detection based on the F1-score was evaluated using statistical quantities of the features and a combination of them. Generally, among the parameters stored in the SCADA system, only six were used. These parameters formed five CCFs: power output vs. wind speed; blade angles A, B and C vs. wind speed; and rotor speed vs. wind speed. The case containing the average values of all the aforementioned parameters, in which the three blade angles were aggregated into one, was demonstrated to have the best performance among the other cases, where F1 = 86.77%, followed by the case that consisted of the same set of parameters without the aggregation of the blade angles in addition to their standard deviation against average wind speed, where F1 = 82.23%. These results demonstrated that a pitch fault could be successfully detected.

### CRediT authorship contribution statement

**Panagiotis Korkos:** Conceptualization, Methodology, Software, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Matti Linjama:** Conceptualization, Writing – review & editing, Supervision. **Jaakko Kleemola:** Conceptualization, Investigation, Resources, Writing – review & editing, Supervision. **Arto Lehtovaara:** Conceptualization, Writing – review & editing, Supervision, Project administration.

### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

This research was funded by the Doctoral School of Industry Innovations (DSII) of Tampere University and Suomen Hyötytuuli Oy.

## References

- [1] WindEurope, Wind energy in Europe in 2019 - trends and statistics, brussels, in: [https://gwec.net/global-wind-report-2019/](https://proceedings.wind-europe.org/biplatform/rails/active_storage/disk/eyJmFpbHMlOnsibWVzc2FnZSI6IkBaDdDRG9jYTJWNVNTSWhObTg0Y0RaNGMyc3dPV0ZwYm14dmVHbDNaElZ2WWpaa1Y5Smxid1k2QmtWVU9oQmthWE53YjNOcGRHbHZia2tpZFdsdWJHbHVvaVHNnWm1sc1pXNWWhiV1U5SWxcKjltUkZkWEp, 2020. (Accessed 5 November 2020).</a></li>
<li>[2] Global Wind Energy Council, Global Wind Report 2019, Glob. Wind Energy Council, 2020. <a href=). (Accessed 5 November 2020).
- [3] M. Wilkinson, B. Hendriks, F. Spinato, K. Harman, E. Gomez, H. Bulacio, J. Roca, P. Tavner, Y. Feng, H. Long, Methodology and results of the Reliawind reliability field study, *Eur. Wind Energy Conf. Exhib. EWEC*. 3 (2010) (2010) 1984–2004.
- [4] J. Carroll, A. McDonald, D. McMillan, Failure rate, repair time and unscheduled O & M cost analysis of offshore wind turbines, *Wind Energy* 19 (2016) 1107–1119, <https://doi.org/10.1002/we.1887>.
- [5] J. Ribrant, L.M. Bertling, Survey of failures in wind power systems with focus on Swedish wind power plants during 1997–2005, *IEEE Trans. Energy Convers.* 22 (2007) 167–173, <https://doi.org/10.1109/PES.2007.386112>.
- [6] A. Zaher, S.D.J. McArthur, D.G. Infield, Y. Patel, Online wind turbine fault detection through automated SCADA data analysis, *Wind Energy* 12 (2009) 574–593, <https://doi.org/10.1002/we.319>.
- [7] B. Chen, D. Zappala, C.J. Crabtree, P.J. Tavner, Survey of Commercially Available SCADA Data Analysis Tools for Wind Turbine Health Monitoring, 2014.
- [8] J. Tautz-Weinert, S.J. Watson, Using SCADA data for wind turbine condition monitoring - a review, *IET Renew. Power Gener.* 11 (2017) 382–394, <https://doi.org/10.1049/iet-rpg.2016.0248>.
- [9] A. Stetco, F. Dinmohammadi, X. Zhao, V. Robu, D. Flynn, M. Barnes, J. Keane, G. Nenadic, Machine learning methods for wind turbine condition monitoring: a review, *Renew. Energy* 133 (2019) 620–635, <https://doi.org/10.1016/j.renene.2018.10.047>.
- [10] W. Yang, R. Court, J. Jiang, Wind turbine condition monitoring by the approach of SCADA data analysis, *Renew. Energy* 53 (2013) 365–376, <https://doi.org/10.1016/j.renene.2012.11.030>.
- [11] B. Chen, P.C. Matthews, P.J. Tavner, Automated on-line fault prognosis for wind turbine pitch systems using supervisory control and data acquisition, *IET Renew. Power Gener.* 9 (2015) 503–513, <https://doi.org/10.1049/iet-rpg.2014.0181>.
- [12] B. Chen, P.C. Matthews, P.J. Tavner, Wind turbine pitch faults prognosis using a-priori knowledge-based ANFIS, *Expert Syst. Appl.* 40 (2013) 6863–6876, <https://doi.org/10.1016/j.eswa.2013.06.018>.
- [13] M. Schlechtingen, I.F. Santos, S. Achiche, Wind turbine condition monitoring based on SCADA data using normal behavior models. Part 1: system description, *Appl. Soft Comput. J.* 13 (2013) 259–270, <https://doi.org/10.1016/j.asoc.2012.08.033>.
- [14] M. Schlechtingen, I.F. Santos, Wind turbine condition monitoring based on SCADA data using normal behavior models. Part 2: application examples, *Appl. Soft Comput. J.* 14 (2014) 447–460, <https://doi.org/10.1016/j.asoc.2013.09.016>.
- [15] K. Leahy, R.L. Hu, I.C. Konstantakopoulos, C.J. Spanos, A.M. Agogino, Diagnosing wind turbine faults using machine learning techniques applied to operational data, *IEEE Int. Conf. Progn. Heal. Manag. ICPHM* (2016) 1–8, <https://doi.org/10.1109/ICPHM.2016.7542860>, 2016.
- [16] K. Leahy, R.L. Hu, I.C. Konstantakopoulos, C.J. Spanos, A.M. Agogino, D.T.J. O'Sullivan, Diagnosing and predicting wind turbine faults from SCADA data using support vector machines, *Int. J. Prognostics Health Manag.* 9 (2018) 1–11.
- [17] R.L. Hu, K. Leahy, I.C. Konstantakopoulos, D.M. Auslander, C.J. Spanos, A.M. Agogino, Using domain knowledge features for wind turbine diagnostics, *Proc. 2016 15th IEEE Int. Conf. Mach. Learn. Appl. ICMLA* (2016) 300–305, <https://doi.org/10.1109/ICMLA.2016.172>, 2017.
- [18] A. Kusiak, A. Verma, A data-driven approach for monitoring blade pitch faults in wind turbines, *IEEE Trans. Sustain. Energy* 2 (2011) 87–96, <https://doi.org/10.1109/TSTE.2010.2066585>.
- [19] R.K. Pandit, D. Infield, Comparative assessments of binned and support vector regression-based blade pitch curve of a wind turbine for the purpose of condition monitoring, *Int. J. Energy Environ. Eng.* 10 (2019) 181–188, <https://doi.org/10.1007/s40095-018-0287-3>.
- [20] R. Pandit, D. Infield, Gaussian process operational curves for wind turbine condition monitoring, *Energies* 11 (2018), <https://doi.org/10.3390/en11071631>.
- [21] P. Guo, D. Infield, Wind turbine power curve modeling and monitoring with Gaussian process and SPRT, *IEEE Trans. Sustain. Energy* 11 (2020) 107–115, <https://doi.org/10.1109/TSTE.2018.2884699>.
- [22] G.A. Skrimpas, K.S. Marhadi, R. Gomez, C.W. Sweeney, B.B. Jensen, N. Mijatovic, J. Holboell, Detection of pitch failures in wind turbines using environmental noise recognition techniques, *Proc. Annu. Conf. Progn. Heal. Manag. Soc. PHM*. (2015) 280–287.
- [23] X. Wu, R. Su, C. Lu, X. Rui, Internal Leakage Detection for Wind Turbine Hydraulic Pitching System with Computationally Efficient Adaptive Asymmetric SVM, 2015 34th Chinese Control Conf. C.C.C., 2015, pp. 6126–6130, <https://doi.org/10.1109/ChiCC.2015.7260599>.
- [24] M. Schlechtingen, I. Ferreira Santos, Comparative analysis of neural network and regression based condition monitoring approaches for wind turbine fault detection, *Mech. Syst. Signal Process.* 25 (2011) 1849–1875, <https://doi.org/10.1016/j.ymssp.2010.12.007>.
- [25] C. Bindi, Automated On-Line Fault Prognosis for Wind Turbine Monitoring Using SCADA Data, Durham University, 2014. <http://theses.dur.ac.uk/10772/>.
- [26] J.Y. Park, J.K. Lee, K.Y. Oh, J.S. Lee, Development of a novel power curve monitoring method for wind turbines and its field tests, *IEEE Trans. Energy Convers.* 29 (2014) 119–128, <https://doi.org/10.1109/TEC.2013.2294893>.
- [27] J.R. Jang, Anfis : adaptive-network-based fuzzy inference system, *IEEE Trans. Syst. Man. Cybern.* 23 (1993).
- [28] I. Goodfellow, Y. Bengio, A. Courville, *Deep Learning*, MIT Press, 2016.