https://doi.org/10.5194/wes-2022-111
Preprint. Discussion started: 1 February 2023

Open Access WIND ENERGY SCIENCE DISCUSSIONS

Anomaly-Based Fault Detection in Wind Turbine Main Bearings

Lorena Campoverde-Vilela[1], María del Cisne Feijóo[1], Yolanda Vidal[2,3], José Sampietro[4], and Christian Tutivén[1]

[1] ESPOL Polytechnic University, Escuela Superior Politécnica del Litoral, Faculty of Mechanical Engineering and Production Science (FIMCP), Mechatronic Engineering, Campus Gustavo Galindo Km. 30.5 Vía Perimetral, P.O. Box 09-01-5863, Guayaquil, Ecuador.
[2]Control, Data, and Artificial Intelligence (CoDAlab), Department of Mathematics, Escola d'Enginyeria de Barcelona Est (EEBE), Universitat Politècnica de Catalunya (UPC), Campus Diagonal-Besós (CDB), Eduard Maristany, 16, 08019 Barcelona, Spain.
[3] Institute of Mathematics (IMTech), Universitat Politècnica de Catalunya (UPC), Pau Gargallo 14, 08028 Barcelona, Spain.
[4]Universidad Ecotec, Km. 13.5 Samborondón, Samborondón, EC092302 Ecuador.

Correspondence: Yolanda Vidal (yolanda.vidal@upc.edu)

Abstract. Renewable energy is a clean and inexhaustible source of energy, so every year interest in the study and the search for improvements in production increases. Wind energy is one of the most used and therefore the need for predictive maintenance management to guarantee the reliableness and operability of each of the wind turbines has become a great study opportunity. In this work, a fault detection system is developed by applying an anomaly detector based on principal component analysis (PCA), in order to state early warnings of possible faults in the main bearing. For the development of the model, SCADA data from a wind park in operation are utilized. The results obtained allow detection of failures even months before the fatal breakdown occurs. This model requires (to be constructed) only the use of healthy SCADA data, without the need to obtain the fault history or install additional equipment or sensors that require greater investment. In conclusion, this proposed strategy provides a tool for the planning and execution of predictive maintenance within wind parks.

1 Introduction

High consumption of fossil fuels increases environmental pollution and increases the effects of climate change. For this reason, it is important to deepen the study and development of renewable energies to achieve a green energy economy (Baloch et al., 2022; May, 2017). Some authors, such as Yii and Geetha (2017) and Kang et al. (2020), focus their research on the fact that technological innovation reduces environmental

https://doi.org/10.5194/wes-2022-111
Preprint. Discussion started: 1 February 2023

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

degradation and reduces the effects of pollution. It can be seen from the documents of the International Energy Agency of 2019, that the amount of renewable energy as a supply of the OECD (Organization for Economic Cooperation and Development) reached 10.8 % of its primary energy. Similarly, renewable energy is expected to provide one third of total energy generation by 2035. This growth motivates the study and implementation of new technologies and algorithms for their development (Wang, 2022).

One of the topics that has caught the interest of many researchers is the development of new technologies that allow for improved maintenance management in wind turbines (WTs), since a good maintenance strategy is the basis for supporting production with reliability, safety, environmental preservation, and adequate costs. In addition, it guarantees continuous operation and avoids economic losses due to emergency stops or equipment damage (Velmurugan and Dhingra, 2015).

According to May et al. (2015), operating and maintenance costs regularly represent 30 % of the total cost of industrial plants, so poor management or maintenance generates additional costs that can significantly affect their operation and economy. Currently, there are different maintenance strategies, such as corrective, preventive, predictive, and proactive (Bahar et al., 2021; Velmurugan and Dhingra, 2015). In general, wind parks use preventive and corrective maintenance; however, considering industrial environments, where stops can cause large losses (Hu and Chen, 2020), predictive maintenance is becoming a method of growing interest, especially with its new approach geared toward the use of industry 4.0 standards. This new technology has allowed integration between physical and digital systems, facilitating the collection of a large amount of data (Borgi et al., 2017).

Hand in hand with this technology, artificial intelligence (AI) is used, which is based on a set of machines that, through defined algorithms, develop tasks that would normally require the intelligence of a person (Jakhar and Kaur, 2020). In addition, AI techniques such as machine learning (ML) are applied, which is defined as computational procedures that with the help of input values allow a machine to simulate human intelligence by adapting and learning from its environment without being explicitly programmed to do so (El Naga and Murphy, 2015). Within this branch there are different learning models, including normality models that are trained only with normal data without failures, i.e. healthy data, which represents a relevant advantage over a supervised model, which is trained with a history of previously labeled data between healthy or damaged and thus learns to predict the output value (Kotsiantis et al., 2006).

Processing and analysis of the data allows for a better understanding of the behavior and dynamics of the system. Predictive maintenance allows you to take advantage of this information to intervene only when the machine really needs it, providing advantages such as reduced maintenance costs, emergency stops,

https://doi.org/10.5194/wes-2022-111
Preprint. Discussion started: 1 February 2023

inventory, and an increase in the useful life of spare parts and production efficiency, among others. (Peres et al., 2018; Sezer et al., 2018).

Table 1 shows scientific investigations based on the detection of failures classified according to the principal component or system of the WT, such as the gearbox, the pitch system, the yaw system, among others. Similarly, Table 2 shows the research based on fault detection but classified according to the monitoring method used.

Table 1. Papers based on the detection of faults classified according to the main component or system of the WT.

1. Gearbox	(Jiang et al., 2019; Zhang et al., 2012; Teng et al., 2016; Feng et al., 2014) (Dameshghi et al., 2019)
2. Electrical and electronics	(Kim et al., 2011; Borchersen and Kinnaert, 2016; Astolfi et al., 2019)
3. Blades/pitch angle	(Rezamand et al., 2020; Guo et al., 2021; Oliveira et al., 2020; Ou et al., 2017) (Dervilis et al., 2014)
4. Tower	(Leon-Medina et al., 2021; Nguyen et al., 2015, 2018; Motlagh et al., 2021)
5. Bearing	(Natili et al., 2021; Encalada-Dávila et al., 2021; Hu et al., 2018; Fuentes et al., 2020) (Zhang et al., 2022)
6. Yaw system	(Chen et al., 2020)

Table 2. Papers of investigations based on fault detection but classified according to the monitoring method used

1. Temperature (oil, generator, bearing)	(Fu et al., 2019; Li et al., 2021; Zeng et al., 2020) (Guo et al., 2018)
2. SCADA data	(Xiang et al., 2021; Chacón and Márquez, 2021; Wang et al., 2019) (Encalada-Dávila et al., 2022; Tutivén et al., 2022) (Dao, P. B., 2022; Dameshghi et al., 2019)
3. Vibration signals	(Son et al., 2018; Ren et al., 2019; Joshuva et al., 2020) (Xiuli et al., 2018)
4. Electrical signals	(Dahiya, R., 2018; Zhang et al., 2019)
5. Acoustic emission	(Karabacak and Özmen, 2022; Yao et al., 2021)
6. Strain measurements	(Hubbard et al., 2021)
7. Other non-destructive testing	(Wang et al., 2022)

https://doi.org/10.5194/wes-2022-111
Preprint. Discussion started: 1 February 2023

For WTs, one of the predominant systems is the supervisory control and data acquisition (SCADA) and condition monitoring (CM) system (Artiago et al., 2018; Leite et al., 2018). CM systems are born from the need to guarantee a more controlled and safe operation. Most of these systems available on the market for WTs are mainly based on vibration measurements or in some works, such as Wang et al. (2019); Hammed et al. (2009), external variables and tower oscillations are also analyzed. This system allows moving from preventive maintenance to predictive maintenance, which, based on the study of behavior, allows a deviation to be detected, warning of a possible failure in the machine. However, its use requires the installation of additional sensors, which allow the acquisition and analysis of new variables, which causes an increase in the costs of operation, maintenance, and complexity. Regarding SCADA data, it has approximately 200 stored variables, which are acquired with a sampling frequency of 1 Hz, but, in general, to optimize space and bandwidth, the variables are only recorded every 10 minutes, storing the mean, standard deviation, minimum, and maximum values of each variable. This information, added to the work order records, results in the generation of complete operation histories, faults, and alarms for each WT.

Recently, there has been lot of interest in the analysis of SCADA data because although the system was not developed for condition monitoring, the extraction of relevant information through data analysis and processing can become a great alternative that allows one to ensure reliability, maintaining the costs and complexity of the machine, in this case WT. However, it is necessary to consider relevant points when using this information, such as the fact that there is great variability in the data due to changes in environmental conditions such as wind speed and direction, ambient temperature, turbulence, etc. Also, as explained above, in most cases the data is stored only in 10-minute intervals, so there is a loss of information. At the same time, work order history and entry are not always standardized and are prone to human error. Some researchers have already started to use SCADA data for fault detection in WTs using different tools and methodologies. As, for example, in Dao et al. (2018), methodologies based on SCADA data cointegration analysis are presented for condition monitoring and fault diagnosis, using data obtained from a single WT with a nominal power of 2 MW under environmental and operating conditions. In Xiang et al. (2021) an early warning model for fault detection is proposed, using a convolutional neural network (CNN) and an attention mechanism (AM) based on a long-short-term memory (LSTM) neural network, where SCADA data are used as input variables and build the CNN and LSTM architecture, and AM is applied to strengthen the impact of important information. Furthermore, with respect to studies on bearing failures in WT, in Zhang and Wang (2014), a bearing failure detection method based on existing SCADA data is proposed using an artificial neural network (ANN). Similarly, in Encalada-Dávila et al. (2022), a methodology is shown that detects main bearing failures using a closed recurrent unit (GRU) neural network. In the present work, a methodology

https://doi.org/10.5194/wes-2022-111
Preprint. Discussion started: 1 February 2023

Open Access WIND ENERGY SCIENCE DISCUSSIONS

is developed that allows predicting main bearing failures through a principal component analysis (PCA) anomaly detector, with the help of only SCADA data and with the aim of achieving continuous operation of the equipment for as long as possible.

The organization of this article is as follows. The general description of the wind park is described in Section 2. Section 3 presents the acquisition and analysis of the SCADA data used. In Section 4, the preprocessing of the data (data cleaning and variable selection) is presented. Then, in Section 5, the anomaly detection algorithm is shown. In Section 6 the indicators are analyzed, and the threshold value is defined. The results are presented and discussed in Section 7 and finally, in Section 8, the conclusions are detailed.

2 Wind park description

In this study, data are collected from a wind park composed of 18 WTs located in Poland. All WTs have the same characteristics, which are detailed in Table 3.

Table 3. The main technological features of the type of WT in the wind park considered.

Number of blades	3
Nominal power	2300 kW
Voltage	690 V
Rotor diameter	101 m
Wind class	IEC IIb
Swept area	8000 m^2
Rotor speed	6-16 rpm
Tower height	80 m
Cut-in wind speed	3-4 m/s
Rated wind speed	12-13 m/s
Cut-out wind speed	25 m/s
Gearbox type	3-stage planetary/helical
Gearbox ratio	1:91
Power regulation	pitch regulation with variable speed

As can be seen, all the WTs on the park under study have a three-bladed rotor with pitch regulation with variable speed that allows them to operate closer to the limit of aerodynamic effectiveness for a greater fraction of time. Figure 1 shows the power curve for these WTs.

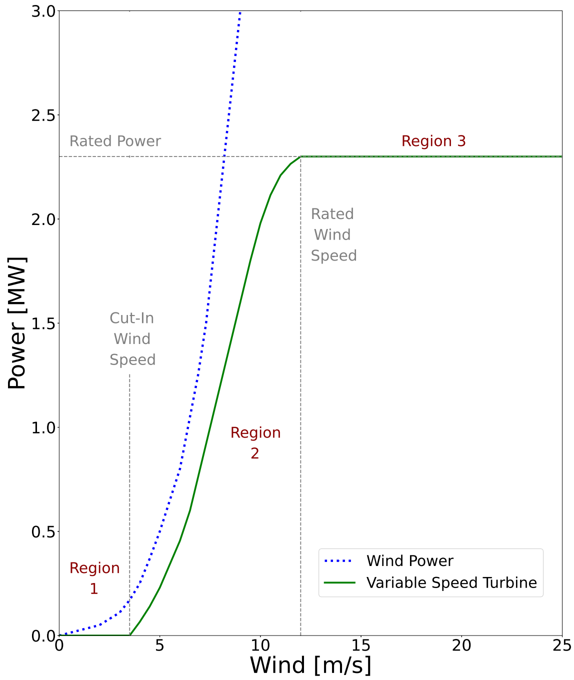

Figure 1. Wind turbines calculated power curve.

100 The estimated power curve data are valid for an air temperature of 15 degrees Celsius, an atmospheric pressure of 1,013 hPa, an air density of 1,225 kg/m^3, a clean rotor blade, and an undisturbed horizontal air flow. Figure 2 shows the principal components of these WTs.

In addition, each turbine is equipped with a SCADA system, thus plant operators can get an overview of the entire wind park. The main features of the SCADA system are the following: use of the standard communication protocol OPC UA, standardized data structure according to IEC61400-25, alarm and data history, online/offline trends. These SCADA-collected data are used in this work and are further explained in Section 3.

3 Data description, acquisition, and analysis

Wind Park SCADA data were recorded every 10 minutes from 1 January 2014 to 12 December 2019, obtaining the average, standard deviation, highest value, and lowest value of the data sampled from each variable.

https://doi.org/10.5194/wes-2022-111
Preprint. Discussion started: 1 February 2023

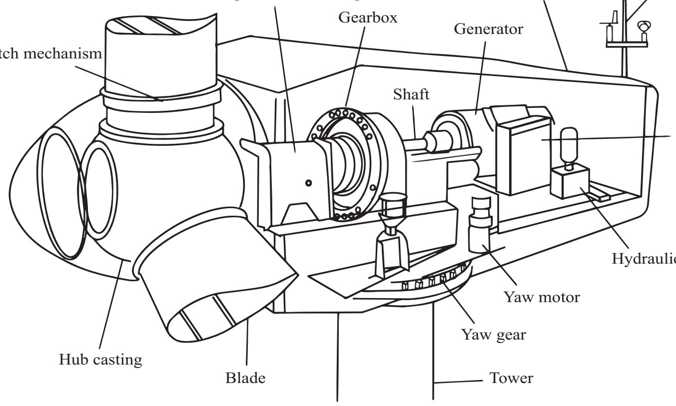

Figure 2. Main components of WTs. (Jiang et al., 2017).

These variables can be divided into different groups: energy, hydraulic, climatic, control, and heat emanating from the elements (see Encalada-Dávila et al. (2021)).

For the analysis and development of this work, it was decided to use only the mean of the external SCADA measurements (which can be broadly grouped into the so-called environmental variables; see Table 4) and the mean of the variable MainBTmp (low speed shaft temperature), related to our failure of interest, with the objective that detection alerts are not affected by other types of failure other than the one under study. It is important to know that there is additional information (work orders) such as the maintenance and repair work carried out on each of the turbines, the dates on which they were carried out, and different references that help determine which elements were replaced or repaired. This information allows for the validation of the proposed methodology.

3.1 Data division

To avoid biases in the normality model to be developed, before analyzing the data and performing the data preprocessing, it is important to perform the data division.

The data division procedure is a crucial step in any deep learning model, as it has a very relevant effect on the whole process. To separate the data for this study, the training data is selected such that:

- The model is trained using only healthy data.

https://doi.org/10.5194/wes-2022-111
Preprint. Discussion started: 1 February 2023

Table 4. External selected measurements.

Variable	Description
AmbieTmp	Environmental temperature
A1ExtTmp	Environmental temperature of a different detector
SeWindSp	Associated to wind speed
AcWindSp	Associated to wind speed
PrWindSp	Associated to wind speed
PriAnemo	Associated with anemometer readings
SecAnemo	Associated with anemometer readings

- The model is trained using data from all operating regions of the WT (see Figure 1). There is no selection of a specific region or conditions of operation, or specific year season, thus the model is robust to all these variations.

130 As a result, a normality model is constructed that is tolerant to the various environmental and working conditions of WTs.

Based on the work orders, for this analysis, the selection of WT number 5 of the wind park (from now on denoted as "WT5") is made, since it is known that in 2018 it suffered a failure of the main bearing, which is the failure of interest in this work.

135 For data division, it is taken into account that the main bearing failure occurs on 11 June 2018, and the model must be trained for at least one year to ensure that it is seasonally robust. Therefore, according to the conditions detailed above, the data used to train the model are selected from 1 January 2014 to 11 December 2017, while the test data from 11 December 2017 to 12 December 2019, as shown in Figure 3.

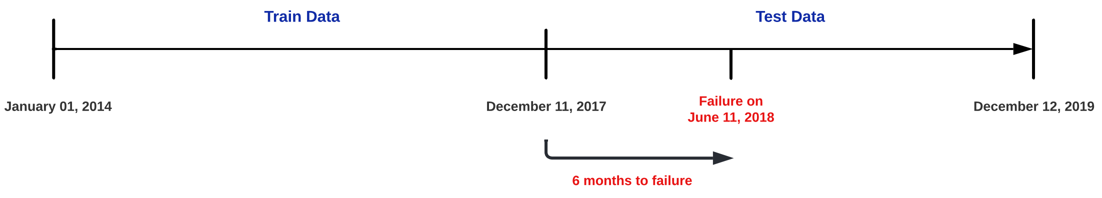

Figure 3. Data splitting for model training and testing.

https://doi.org/10.5194/wes-2022-111
Preprint. Discussion started: 1 February 2023

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

Before performing the pre-processing of the data, the training data is explored to visualize the behavior of the different variables, as detailed in Section 3.2.

3.2 Exploratory data analysis

Exploratory data analysis (EDA) is a process of analyzing and understanding the input information, its structure, selecting relevant variables, cleaning outliers, etc. (Vanawat et al., 2021). That is, it seeks to explore the data and discover important information through statistical graphs or some other visualization method, ensuring the quality and efficiency of the subsequent analysis (Shin, 2020).

It should be noted that, in general, failure to perform an EDA on the study data set results in the development of inaccurate models due to data bias, which are missing or inconsistent values. Therefore, in this study, the EDA is based on understanding the behavior of each of the variables relevant to the study, taking into account that only training data are used. Figure 4 details the behavior of the eight selected variables, and the following deductions are obtained:

– The plot of the MainBTmp variable shows seasonal variation as well as A1ExtTmp and AmbieTmp, indicating that the low-speed shaft temperature is altered by the environment temperature.

– The graphs of the SeWindSp and SecAnemo variables show a very similar behavior. The same is observed for the variables AcWindSp and PrWindSp.

– In general, all variables show an increase in variations in wind speed at the same stages, so the measurements could be affected by wind currents.

– For the PriAnemo variable, it can be observed in its behavior that it is a staggered response, where from 2014 to the beginning of 2018 it remained at the value of 1.2 while the remainder at almost 0. This variable can be interpreted as an on- and off state, so it is decided to eliminate this variable, as it does not provide relevant information.

To correct for missing values, blanks, or outliers, pre-processing techniques must be used. For this reason, data pre-processing is proposed, which allows solving the problems mentioned above by converting them into quality data.

4 Data pre-processing

When working with SCADA data, there is the possibility that incomplete, inconsistent, unprocessed, or error-prone values may be present, so pre-processing is of great importance to ensure data quality through

https://doi.org/10.5194/wes-2022-111
Preprint. Discussion started: 1 February 2023

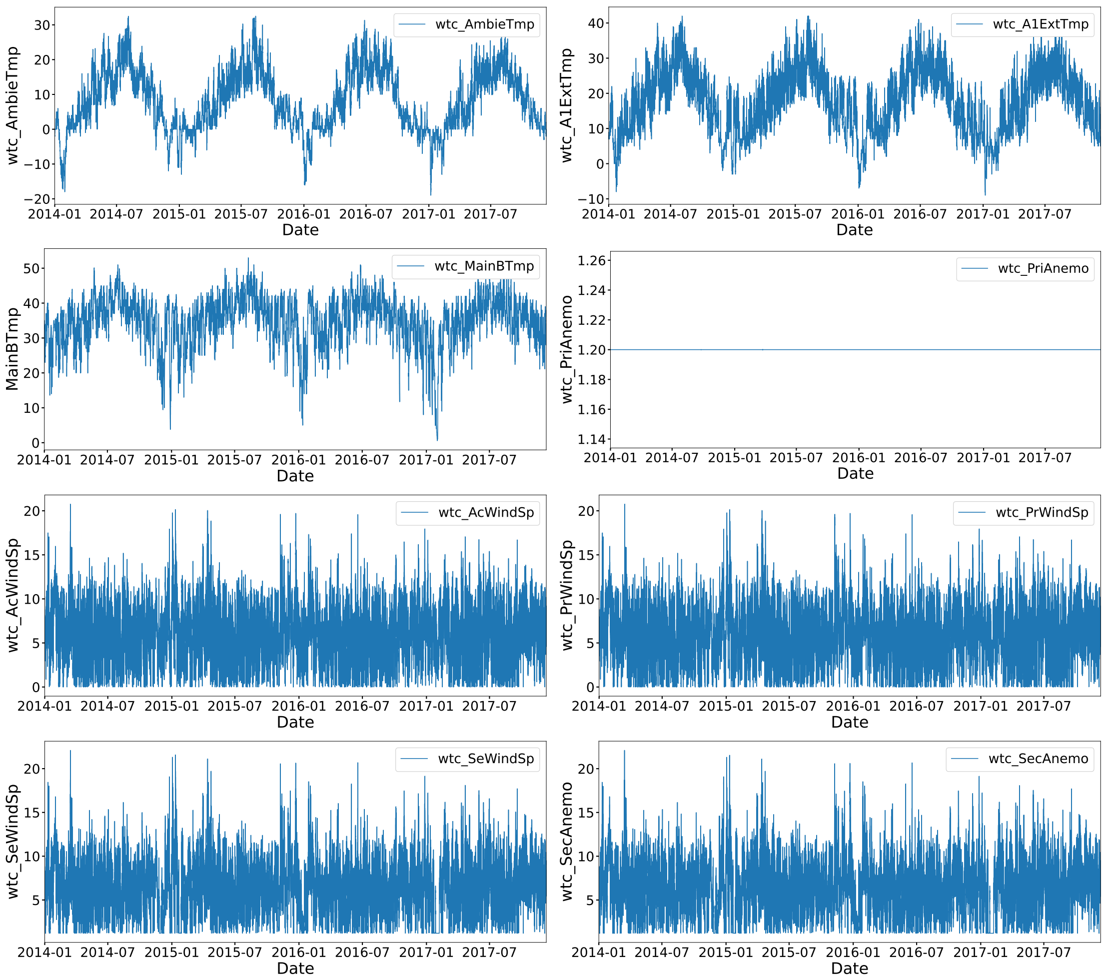

Figure 4. Exploratory data analysis for the selected variables.

https://doi.org/10.5194/wes-2022-111
Preprint. Discussion started: 1 February 2023

data cleaning and preparation (García and Luengo, 2015). In this study, data cleaning and variable selection preprocessing techniques are used and explained in detail in the next subsections.

4.1 Data cleaning

Data cleaning is based on the determination and correction of missing or atypical values, where data are treated to give coherence to the data set (Brownlee, 2020). Several forms of outlier identification are available. For the development of this work, the method related to the realistic value range filter is used (see Encalada-Dávila et al. (2021)). Table 5 shows the realistic possible values for each of the selected variables. All detected outliers are converted to missing values and subsequently filled with realistic values, preventing data waste and without introducing errors into the model.

Table 5. Range of possible values for each of the variables under analysis.

Variable	Range
AmbieTmp	[-19 43]
A1ExtTmp	[-19 43]
MainBTmp	[0 100]
SeWindSp	[0 23]
AcWindSp	[0 23]
PrWindSp	[0 23]
PriAnemo	[0 24]
SecAnemo	[0 24]

The method described in Encalada-Dávila et al. (2021) is used for the imputation. To complete the missing data for two samples, the piecewise cubic Hermite polynomial interpolation method is used, which allows new values to be included, ensuring that the function will behave monotonically and that the first derivative will be continuous. Also, since there may be missing values at the boundaries, the closest data that follows or precedes the missing values is used.

An example of the imputation strategy applied to complete the absent values in the low-speed shaft temperature is detailed in Figure 5.

https://doi.org/10.5194/wes-2022-111
Preprint. Discussion started: 1 February 2023

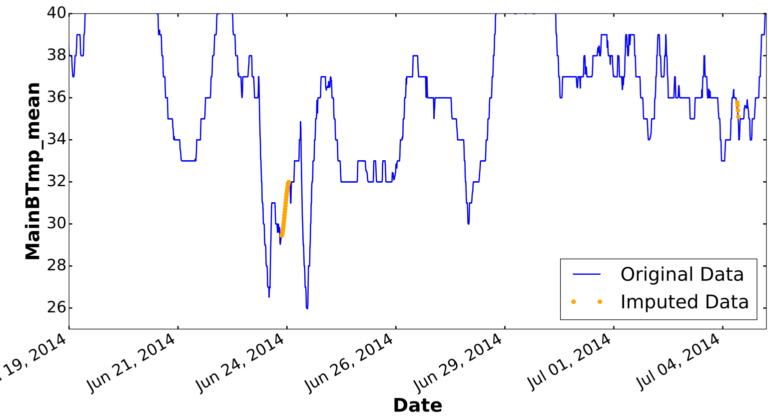

Figure 5. Imputation strategy used to fill missing values in the low-speed shaft temperature.

4.2 Variable selection

Recall that in Section 3 a preselection of variables is carried out based on the use of exogenous variables and the variable most related to the failure under study, but when working with large data sets (in this case several years), a new process of selecting variables is needed to reduce the number of features, eliminating non-informative or redundant variables that may slow down the development and training of the model or also can introduce errors (Kim et al., 2011). Therefore, this procedure is carried out with the objective of improving the prediction efficiency of the model, providing faster and more reliable predictions.

One way to determine which variables are useful when training the model is through a correlation analysis, which allows one to know the relation between the variables. In this work, Pearson's correlation coefficient is used, which is based on a measure of the linear relation between variables; this value is in the range of -1 (negative correlation) to $+1$ (positive correlation). Obviously, a correlation equal to zero means there is no relation between the variables.

According to Sabilla et al. (2019) depending on the coefficient value, it can be determined whether the correlation between two variables is strong or weak, considering a strong relationship for values from 0.50 to 0.69, a very strong relationship from 0.70 to 0.89 and almost perfect from 0.90 to 1. For the selection of the variables that are to be analyzed and studied, in this work it has been decided to select variables with a correlation greater than $+0.8$ and less than -0.8, selecting one of them and eliminating the other, given that, as observed when analyzing the data, there are variables that offer the same information as others since they

https://doi.org/10.5194/wes-2022-111
Preprint. Discussion started: 1 February 2023

WIND ENERGY SCIENCE DISCUSSIONS
Open Access

are highly correlated. Therefore, the selected variables are MainBTmp, SecAnemo and AmbieTmp, already described in Table 4. Figure 6 shows the correlation analysis performed between all variables.

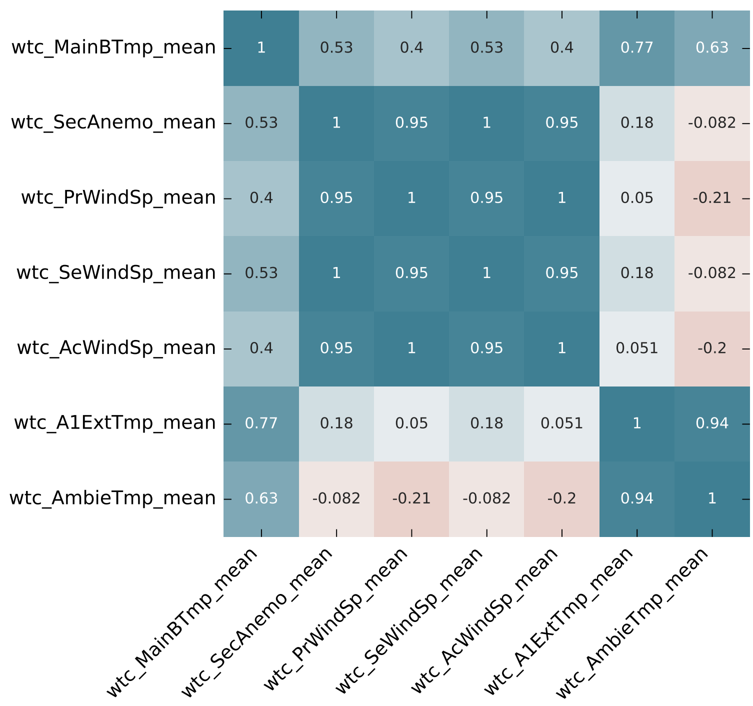

Figure 6. Correlation heat map of preselected variables.

It is important to consider that to avoid errors due to seasonality, the temperature of the low speed shaft (MainBTmp) is subtracted from the environmental temperature (AmbieTmp). As you can see in Section 5, the model is trained with this new processed variable and the SecAnemo variable.

5 Principal Component Analysis (PCA)

Principal component analysis (PCA) is a statistical technique widely used for the study of data in almost all scientific areas (Kurita, 2019). This method is based on determining the most significant base of a given data set to decrease its dimensionality for large data sets, simplifying the complexity of multidimensional spaces while keeping as much information as possible (Rodrigo, 2020). Its objective is to be able to differentiate which features are the most relevant, which are redundant and which are just noise (Tibaduiza et al., 2011).

https://doi.org/10.5194/wes-2022-111
Preprint. Discussion started: 1 February 2023

Suppose that there is a sample of n observations with dependent variables that are interrelated and have noise (Jolliffe and Cadima et al., 2016), where each has p variables or dimensions. PCA is expected to obtain hidden and meaningful data from these observations and filter out the noise. To achieve this purpose, PCA determines a set of new orthogonal variables called principal components, z, which are the result of linear combinations of the initial variables (Kurita, 2019).

In short, with PCA it is possible to describe an observation, for which p values were previously needed with only z values, where $z < p$, as shown in Figure 7.

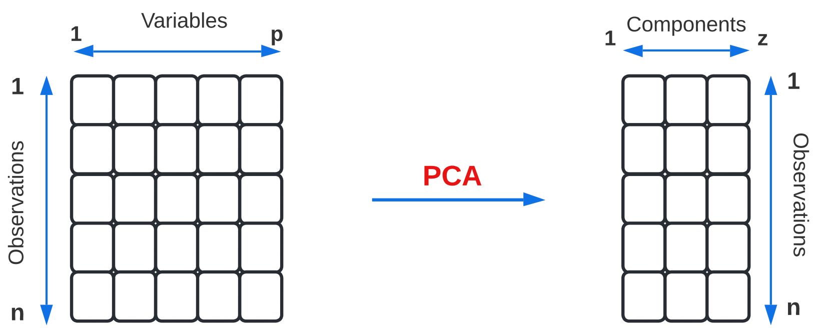

Figure 7. Dimensionality reduction applying PCA.

Suppose that there is a data set $X = [x_1, x_2, .., x_N]$, where each column is an observation n with p variables or dimensions. The mean vector of the sample $\bar{x}$ and the sample covariance matrix Σ are defined in Equations (1) and (2), respectively, where the matrix $\widetilde{X} = [x_1 - \widetilde{x}, .., x_N - \widetilde{x}]$.

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{1}$$

$$\Sigma = \frac{1}{N}\sum_{i=1}^{N} (x_i - \bar{x})(x_i - \bar{x})^T = \frac{1}{N}\widetilde{X}\widetilde{X}^T \tag{2}$$

A principal component analysis stands to reason if the variables have strong correlations, as this is an indication of the presence of redundant information, and therefore few components could explain much of the total variance. The selection of the number of components is determined based on % of the variability that is considered appropriate. The first component collects the greatest possible variability of the original

https://doi.org/10.5194/wes-2022-111
Preprint. Discussion started: 1 February 2023

data, the second collects the maximum possible variability (in an orthogonal direction to the first one) not considered by the first, and the following components repeat it until an adequate variability is reached.

In (Kurita, 2019), the PCA technique is explained in greater depth, describing the mathematical formulas used to calculate each principal component and its projection in a new vector space.

One of the most important applications of PCA is its use in anomaly detection, where the normalized reconstruction error (RE) is used as a scoring level to determine whether a sample is healthy or anomalous. For this study, this technique is used, as described in Section 5.1.

5.1 Anomaly Detection with PCA

Anomaly detection with PCA is an unsupervised statistical strategy to identify anomalous samples present in a set of data elements, based on RE results (Rodrigo, 2020). In particular, the goal is to decompose the source data set into its principal components and then reconstruct the original data using the necessary principal components.

For this work, the preselected variables are projected using the PCA strategy, obtaining three principal components.

Equation (3) describes the first principal component, where $a_1^T = (a_{11}, .., a_{N1})^T$ is a set of weights for linear combinations.

$$y_{1i} = a_1^T (x_i - \bar{x}), (i = 1, .., N) \tag{3}$$

The vector of this component has a tendency in the direction of the observations with greater variance. The projection of each observation in that direction represents the value of the first component of that observation.

For the calculation of the following main components, the same principle is used iteratively until the original number of variables. Considering that the vector of each next principal component will represent the next direction in which the data have more variance and that it should not be correlated with the previous component, which implies that the direction of their vectors is perpendicular/orthogonal to each other (Rodrigo, 2020).

Each new input is processed by the anomaly detector, where it first calculates the projection of the eigenvector and then the RE (Microsoft, 2020). The anomaly detection with PCA indicates that the anomalous elements will be the reconstructed data that differ the most from their original counterparts. See Figure 8.

https://doi.org/10.5194/wes-2022-111
Preprint. Discussion started: 1 February 2023

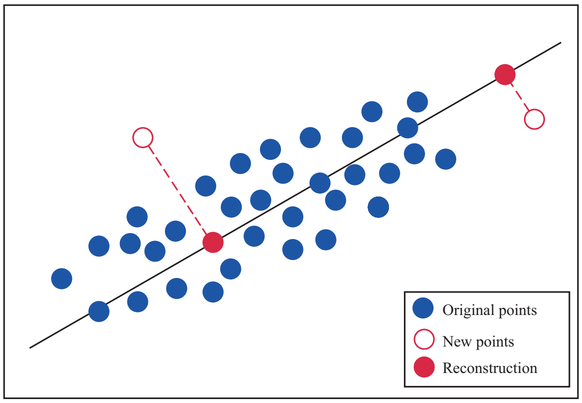

Figure 8. Anomaly detection employing PCA RE.

255 Let x_n, where $n = (1, 2, .., N)$ and $x_n \; \epsilon \; \mathbb{R}^d$ be the set of data observations and $\widetilde{x}$ the reconstruction produced based on an implicit variable of smaller dimension. Then, the RE is defined by Eq. (4).

$$RE\,(x) = \frac{1}{N}\sum_{n=1}^{N} \parallel x_n - \bar{x}_n \parallel^2 \tag{4}$$

In this study, the PCA method is used with the objective of identifying observations that differ from a large part of the values, these being the healthy data used for the training process. When using this approach, 260 prefailure data (anomalies) are expected to have qualitatively distinct attributes from normal data. Therefore, the PCA anomaly detection model is trained only on data from one class (healthy samples), and subsequently predictions are made to find anomalies in the test data set.

Once the PCA anomaly detection algorithm has been created, all phases discussed above are represented in a diagram, as shown in Figure 9. Unlike other approaches, such as the neural autoencoder, the main utility 265 of using PCA as an anomaly detector is its simplicity. Figure 9 shows the data cleaning, variable selection, and training stages of the PCA anomaly detection model.

6 Fault detection indicator (FDI)

For FDI, a threshold is typically established. When the model's discovery of anomalies (outliers) exceeds the predetermined threshold, an alert is produced. However, given that the study only used available 10-minute 270 SCADA data, there may be an excessive number of false alarms, leading to alarm fatigue for WT operators. The FDI that is suggested to bypass this problem is computed in three phases, which are explained in this

https://doi.org/10.5194/wes-2022-111
Preprint. Discussion started: 1 February 2023

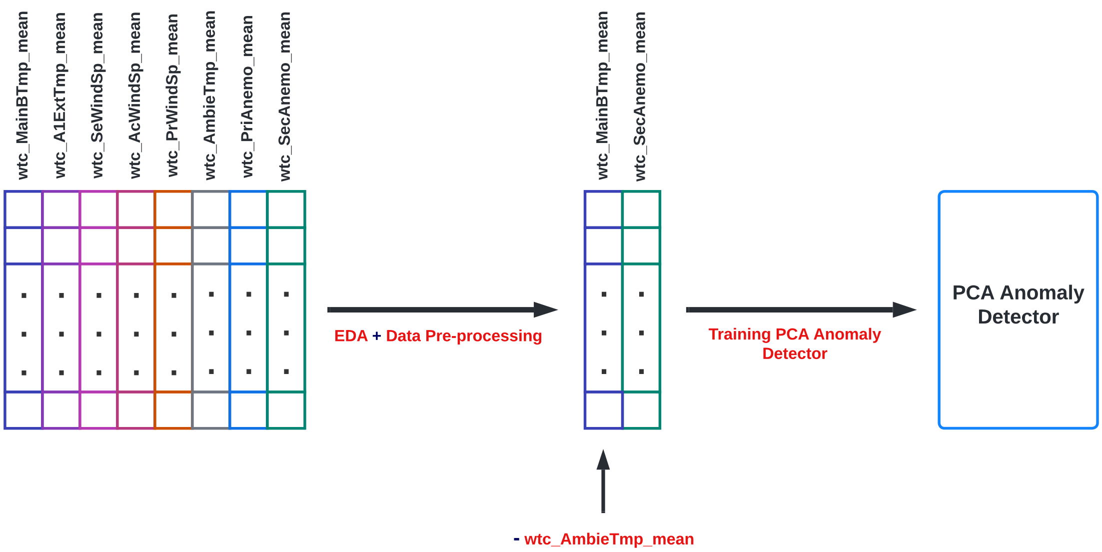

Figure 9. First phase of the stated methodology.

section. Weekly outliers are aggregated (counted) using the PCA anomaly detector approach, after which the exponential weighted moving average (EWMA) filter (Ratner, 2017) is applied to the weekly outlier count, and finally a threshold is defined.

6.1 Weekly gathering

SCADA samples for the current work are collected every 10 minutes, totaling 1008 samples per week. The anomaly detector with PCA algorithm's weekly grouping comprises counting the number of samples out of a total of 1008 that are labeled as anomalies. Formally, the number of samples identified as anomalies for the given i-th week is counted and indicated as C_i

6.2 Exponential weighted moving average (EWMA)

A well-known technique for smoothing historical data is the moving average (MA) approach (Hunter, 1986). The MA has several expansions, each with its own unique features, but their fundamental goal is still the same. In this work, a common MA extension known as EWMA is applied, which combines the calculation of the weight factor for the weighted moving average (WMA) and exponential moving average (EMA). A WMA assigns each observation the same weight, but an EWMA responds more strongly to recent sample

https://doi.org/10.5194/wes-2022-111
Preprint. Discussion started: 1 February 2023

changes. A parameter, α, must be defined to determine the importance of each sample in the EWMA calculation. This technique follows the starting time series more accurately, the higher the value of alpha. Weekly grouping time series, C_i, is applied to the EWMA formula used in this study as

$$\mathrm{EWMA}_t = \alpha C_t + (1-\alpha)\mathrm{EWMA}_{t-1}, \tag{5}$$

where EWMA_0 is the mean of previous data and α is the user-determined weight. One way to specify the parameter an is in terms of spans, generally known as an n-day EWMA,

$$\alpha = \frac{2}{s+1}. \tag{6}$$

In this work, $s = 4$, indicating that 4-week groups are taken into account (around a month). Actually, the conclusions of McKinnon et al. (2020) have an impact on this choice. They evaluated three different moving windows, daily, weekly, and monthly, in their study of the impact of time history on WT failures using SCADA data. The weekly moving window performs best in terms of discovering failures compared to the others.

The FDI reports alarms when the total number of anomalies exceeds a specified threshold. The EWMA filter and weekly counting are then used. The EWMA over the mean (μ) and standard deviation (σ) of the training set is calculated. Recalling that values within three standard deviations of the mean make up roughly 99.7% of a data set with an approximately normally distributed distribution, the so-called three-sigma rule conveys a common heuristic that it is desirable to regard 99.7% of probabilities as being close to certain. As a result, the criterion used in this study specifies that results greater than three standard deviations from the mean should be regarded as anomalies. Consequently, the threshold is stated as

$$\mathrm{threshold} = \mu + 3\sigma. \tag{7}$$

The flow chart shown in Figure 10 describes the latter phases of the approach (from the model of the PCA anomaly detector already trained). The graphic shows how the inferences of X_{train} are grouped by week and then filtered to determine a threshold. Then the same procedure is used with the X_{test} inferences, but this time the goal is to utilize the previously established threshold (from X_{train}) to trigger an alert when the output goes over it.

https://doi.org/10.5194/wes-2022-111
Preprint. Discussion started: 1 February 2023

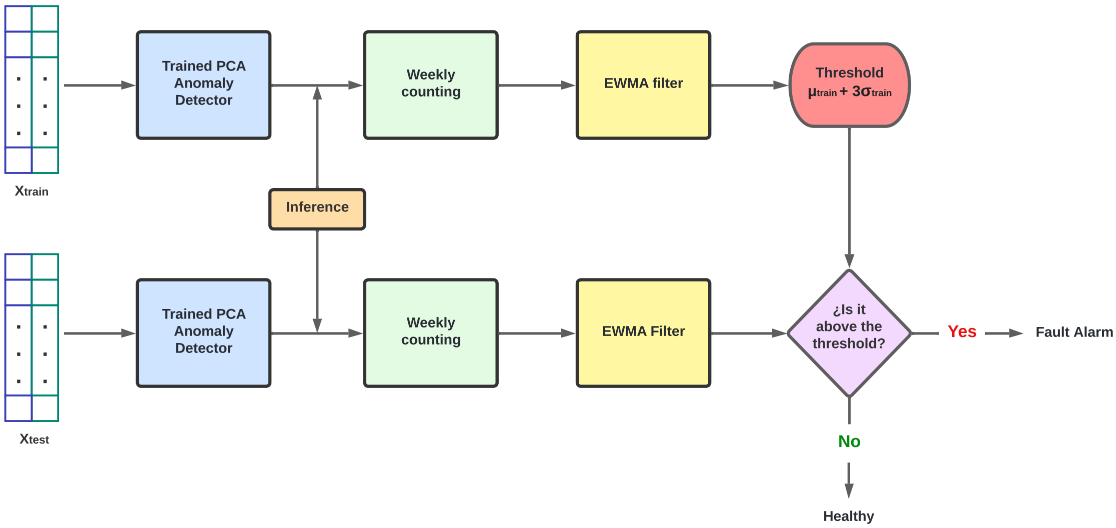

Figure 10. Second phase of the stated methodology.

7 Results

As a result of the proposed study, a predictive maintenance plan is developed to predict failures in the main bearing of the wind park. The model only needs healthy data to be trained and, as a result, allows for detecting anomalies before the fatal failure occurs.

Figures 11,12 and 13 show the figures of the weekly indicator, using the test data, for each of the 18 WTs, to which the detection threshold defined for the evaluation of the model is applied.

When analyzing the figure of each turbine, it can be seen that WT5 and WT16 have data above the established threshold, indicating the presence of alarms and predicts a failure during the indicated time.

In the case of the WT5 turbine, based on the information provided by the wind park, a work order was generated on 11 June 2018 for immediate maintenance of the WT5 turbine, corresponding to a main bearing failure (failure of interest in this study). Using the model, it can be seen in Figure 11 that 3 alerts were generated. The first alert generated is given on 9 April 2018, the next on 16 April 2018 and the last on the day of failure (10 June 2018). This allows inferring that the system is capable of alerting about this type of failure months in advance, which allows corrective measures to be taken in time, to avoid emergency shutdowns, additional costs, and loss of efficiency in energy production.

https://doi.org/10.5194/wes-2022-111
Preprint. Discussion started: 1 February 2023

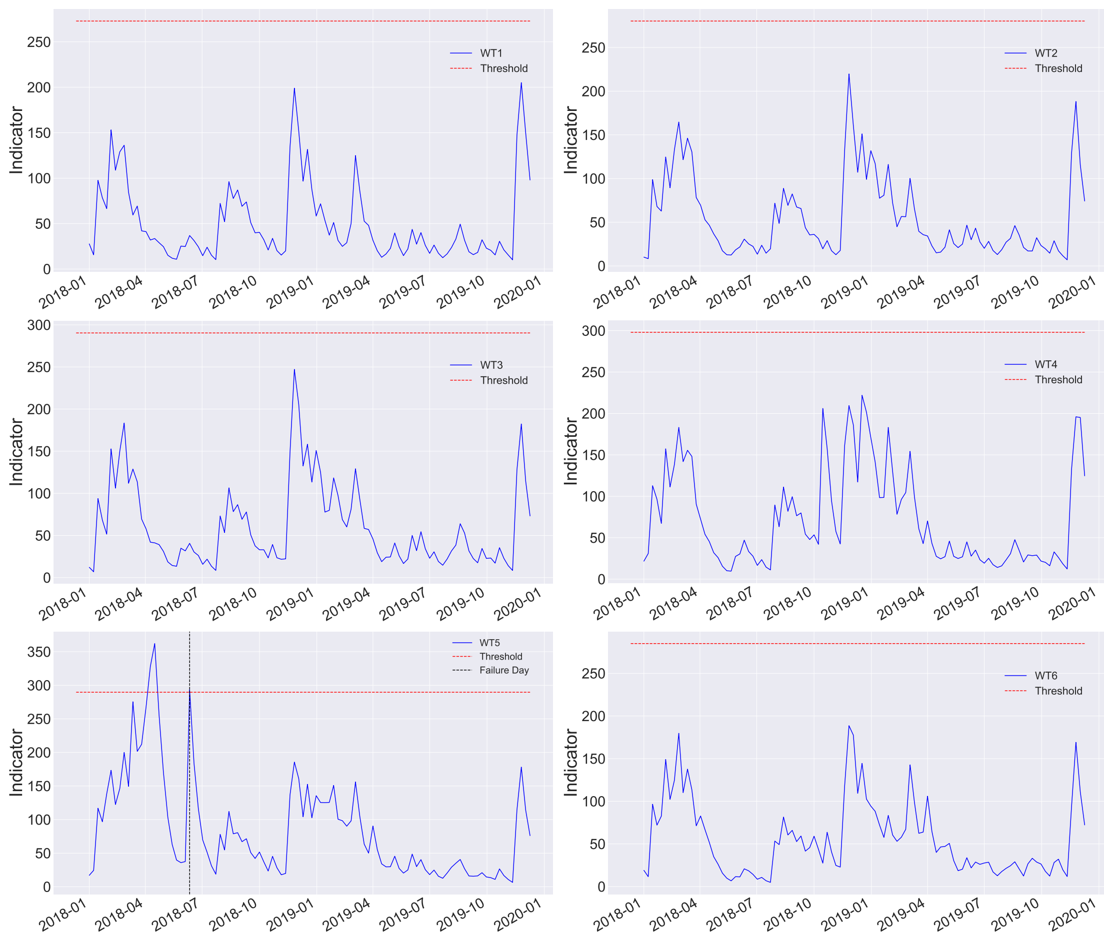

Figure 11. Weekly indicator (WT1-WT6).

On the other hand, analyzing the case of the WT16 turbine, the model generated the alert on December 17, 2018, and on June 26, 2019, the wind park registered a work order in this WT, due to the presence of a fault in the gearbox. Although this is not the failure of the study, the detection of this type of failure may be caused by the high relationship between the main components where failures were recorded (low speed
330 shaft and gearbox) and the use of variables or characteristics of the model that are mostly exogenous.

From the results, it is concluded that this predictive maintenance study is effective in detecting major bearing failures and other related system damage through early warnings, generated months in advance.

https://doi.org/10.5194/wes-2022-111
Preprint. Discussion started: 1 February 2023

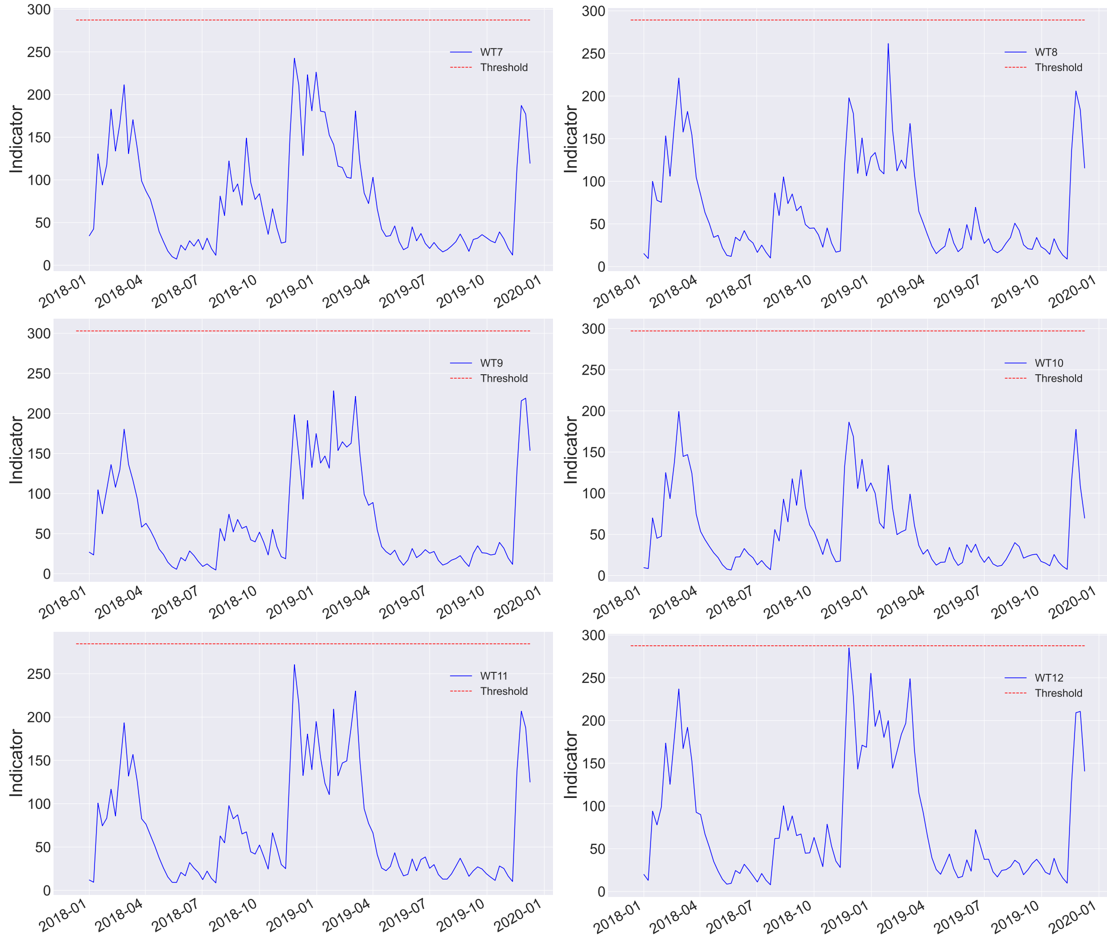

Figure 12. Weekly indicator (WT7-WT12).

8 Conclusions

In this work, an anomaly detection system is developed and evaluated by principal component analysis
335 (PCA), in order to obtain early warnings of possible failures in the main bearing of the WTs. The stated methodology only requires healthy data at the time of development of the model, and the achievements obtained show a functional and efficient system in the generation of alerts for the detection of failures. Among the main advantages of this system are:

– It is not necessary to implement or purchase new sensors or additional systems, since only the data
340 acquired by the SCADA system is used.

https://doi.org/10.5194/wes-2022-111
Preprint. Discussion started: 1 February 2023

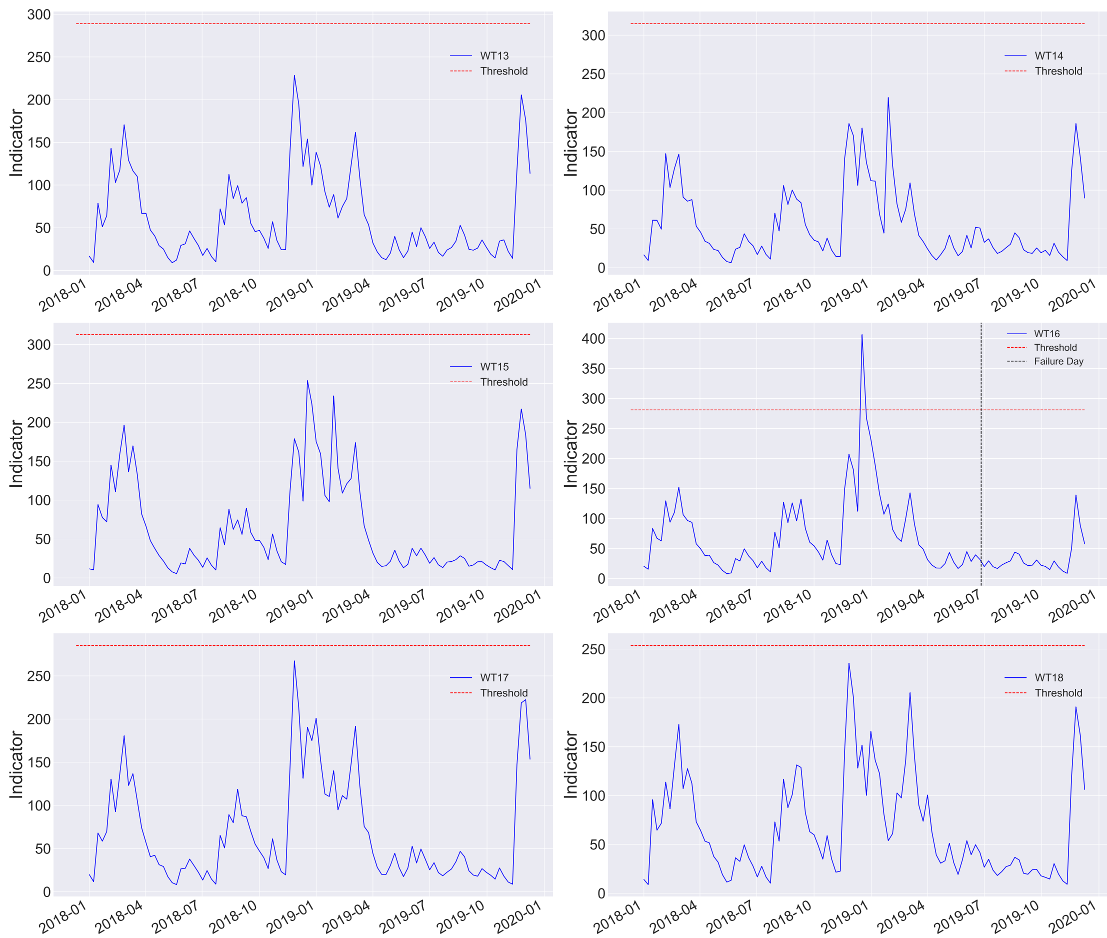

Figure 13. Weekly indicator (WT13-WT18).

– It is not necessary to have a turbine failure history, as this strategy is developed with a semi-supervised model that is developed with healthy information only.

– For the training and evaluation of the model, data from a real wind park in production are used.

Finally, it can be concluded that the stated methodology works accurately and efficiently since, when tested over a year in a complete wind park, it raised alerts months before the fatal breakdown occurs without false alarms. In addition, it allows for the detection of failures of not only the main bearing but also nearby components, thus offering the industry the benefit of predictive maintenance for early planning, guaranteeing long-term safe and continuous operation, reducing downtime, emergency stops, and additional costs.

https://doi.org/10.5194/wes-2022-111
Preprint. Discussion started: 1 February 2023

Author contributions. All authors have contributed equally.

Competing interests. No competing interests are present.

Acknowledgements. This work is partially funded by the Spanish Agencia Estatal de Investigación (AEI) - Ministerio de Economía, Industria y Competitividad (MINECO), and the Fondo Europeo de Desarrollo Regional (FEDER) through the research project PID2021-122132OB-C21. In addition, the authors express their gratitude and appreciation to the Smartive company (http://smartive.eu/), as this work would not have been possible without their support and ceded wind park data.

https://doi.org/10.5194/wes-2022-111
Preprint. Discussion started: 1 February 2023

References

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

Astolfi, D., Castellani, F., and Natili, F.: Wind turbine generator slip ring damage detection through temperature data analysis, Diagnostyka, 20(3), 3–9, 10.29354/diag/109968, 2019.

Bahar, K.P., Yıldız, G.B., and Soylu, B.: Predictive Maintenance System Integrated with Periodic Maintenance: Machine Learning and Classical Approaches, EasyChair, 5806, 2021.

Baloch, Z.A., Tan, Q., Kamran, H.W., Nawaz, M.A., Albashar, G., and Hameed, J.: A multi-perspective assessment approach of renewable energy production: policy perspective analysis, Environ. Dev. Sustain., 24, 2164-2192, https://doi.org/10.1007/s10668-021-01524-8, 2022.

Borchersen, A. B. and Kinnaert, M.: Model-based fault detection for generator cooling system in wind turbines using SCADA data, Wind Energy, 19(4), 593-606, https://doi.org/10.1002/we.1852, 2016.

Borgi, T., Hidri, A., Neef, B., and Naceur, M.S.: Data analytics for predictive maintenance of industrial robots, International Conference on Advanced Systems and Electric Technologies (IC_ASET), Hammamet - Tunisia, 412-417, 14-17 January 2017, 10.1109/ASET.2017.7983729, 2017.

Brownlee, J.: Data preparation for machine learning: data cleaning, feature selection, and data transforms in Python, Machine Learning Mastery, 2020.

Chen, B., Xie, L., Li, Y., Gao, B.: Acoustical damage detection of wind turbine yaw system using Bayesian network, Renew. Energ., 160, 1364-1372, https://doi.org/10.1016/j.renene.2020.07.062, 2020.

Chacón, A., Márquez, F.: SCADA data analytics for fault detection and diagnosis of wind turbines, 7th International Conference on Control, Instrumentation and Automation (ICCIA), pp. 1-6, IEEE, 2021.

Dahiya, R.: Condition monitoring of wind turbine for rotor fault detection under non stationary conditions, Ain Shams Eng. J., 9(4), 2441-2452, https://doi.org/10.1016/j.asej.2017.04.002, 2018.

Dameshghi, A., and Refan, M. H.: Wind turbine gearbox condition monitoring and fault diagnosis based on multi-sensor information fusion of SCADA and DSER-PSO-WRVM method, Int. J. Simul. Model., 39(1), 48-72, https://doi.org/10.1080/02286203.2018.1476008, 2019.

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

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

https://doi.org/10.5194/wes-2022-111
Preprint. Discussion started: 1 February 2023

Dervilis, N., Choi, M. , Taylor, S.G., Barthorpe, R.J., Park, G., Farrar, C.R., and Worden, K.: On damage diagnosis for a wind turbine blade using pattern recognition, J. Sound Vib., 333 (6), 1833-1850, https://doi.org/10.1016/j.jsv.2013.11.015, 2014.

Encalada-Dávila, Á., Puruncajas, B., Tutivén, C., and Vidal, Y.: Wind Turbine Main Bearing Fault Prognosis Based Solely on SCADA Data, Sensors, 21, 2228, https://doi.org/10.3390/s21062228, 2021.

Encalada-Dávila, A., Moyón, L., Tutivén, C., Puruncajas, B., and Vidal, Y.: Early Fault Detection in the Main Bearing of Wind Turbines Based on Gated Recurrent Unit (GRU) Neural Networks and SCADA Data, IEEE/ASME Transactions on Mechatronics, doi: 10.1109/TMECH.2022.3185675, 2022.

El Naqa, I. and Murphy, M.J.:What Is Machine Learning?, Springer,https://doi.org/10.1007/978-3-319-18305-3_1, 2015.

Feng, Y., Qiu, Y., Crabtree, C. J., Long, H., and Tavner, P. J.: Use of SCADA and CMS signals for failure detection and diagnosis of a wind turbine gearbox, In European Wind Energy Association Conference, EWEC 2011, 14-17 March 2014, Brussels, Belgium, 2014.

Fu, J., Chu, J., Guo, P., and Chen, Z.: Condition monitoring of wind turbine gearbox bearing based on deep learning model, IEEE, 7, 57078-57087, 10.1109/ACCESS.2019.2912621, 2019.

Fuentes, R., Dwyer-Joyce, R. S., Marshall, M. B., Wheals, J., and Cross, E. J.: Detection of sub-surface damage in wind turbine bearings using acoustic emissions and probabilistic modelling. Renew. Energ., 147, 776-797, https://doi.org/10.1016/j.renene.2019.08.019, 2020.

García, S. and Luengo, F.H.: Data Preprocessing in Data Mining, Springer, Intelligent Systems Reference Library, Springer Cham, 320, https://doi.org/10.1007/978-3-319-10247-4, first ed., 2015.

Guo, J., Liu, C., Cao, J., and Jiang, D.: Damage identification of wind turbine blades with deep convolutional neural networks, Renew. Energ., 174, 122-133, https://doi.org/10.1016/j.renene.2021.04.040, 2021.

Guo, P., Fu, J., and Yang, X.: Condition monitoring and fault diagnosis of wind turbines gearbox bearing temperature based on kolmogorov-smirnov test and convolutional neural network model, Energies, 11(9), 2248, https://doi.org/10.3390/en11092248, 2018.

Hameed, Z., Hong, Y.S., Cho, Y.M., Ahn, S.H., and Song, C.K.: Condition monitoring and fault detection of wind turbines and related algorithms: A review.Renew. Sust. Energ. Rev., 13(1), 1-39, https://doi.org/10.1016/j.rser.2007.05.008, 2009.

Hubbard, P. G., Xu, J., Zhang, S., Dejong, M., Luo, L., Soga, K., and Minto, C.: Dynamic structural health monitoring of a model wind turbine tower using distributed acoustic sensing (DAS), J. Civ. Struct. Health Monit., 11(3), 833-849, https://doi.org/10.1007/s13349-021-00483-y, 2021.

Hu, J. and Chen, P.: Predictive maintenance of systems subject to hard failure based on proportional hazards model, Reliab. Eng. Syst. Saf., 196, 106707, https://doi.org/10.1016/j.ress.2019.106707, 2020.

Hu, Y., Li, H., Shi, P., Chai, Z., Wang, K., Xie, X., and Chen, Z.: A prediction method for the real-time remaining useful life of wind turbine bearings based on the Wiener process, Renew. Energ., 127, 452-460, https://doi.org/10.1016/j.renene.2018.04.033, 2018.

Hunter, J.S.: The exponentially weighted moving average, J. Qual. Technol., 18(4), 203-210, https://doi.org/10.1080/00224065.1986.11979014, 1986.

425 Jakhar, D. and Kaur, I.: Artificial intelligence, machine learning and deep learning: definitions and differences, Clinical and Experimental Dermatology, 45(1), 131–132, https://doi.org/10.1111/ced.14029, 2020.

Jiang, Z., Hu, W., Dong, W., Gao, Z., and Ren, Z.: Structural reliability analysis of wind turbines: A review, Energies, 10, 2099, https://doi.org/10.3390/en10122099, 2017.

Jiang, G., He, H., Yan, J., and Xie, P.: Multiscale Convolutional Neural Networks for Fault Diagnosis of Wind Turbine Gearbox, IEEE, 66 (4), 3196-3207, 10.1109/TIE.2018.2844805, 2019.

Jolliffe, I.T. and Cadima, J.: Principal component analysis: a review and recent developments, Philos. Trans. Royal Soc. A., 374, 20150202, https://doi.org/10.1098/rsta.2015.0202, 2016.

Joshuva, A., and Sugumaran, V. : A lazy learning approach for condition monitoring of wind turbine blade using vibration signals and histogram features, Measurement, 152, 107295, 984, https://doi.org/10.1016/j.measurement.2019.107295, 2020.

Kang, J.N., Wei, Y.M., Liu, L.C., Han, R., Yu, B.Y., and Wang, J.W.: Energy systems for climate change mitigation: A systematic review, Appl. Energy, 263, 114602, https://doi.org/10.1016/j.apenergy.2020.114602, 2020.

arabacak, Y., Özmen, N.: Common spatial pattern-based feature extraction and worm gear fault detection through vibration and acoustic measurements, Measurement, 187, 110366, https://doi.org/10.1016/j.measurement.2021.110366, 2022.

Kim, K., Parthasarathy, G., Uluyol, O., Foslien, W., Sheng, S., and Fleming, P.: Use of SCADA data for failure detection in wind turbines, ASME 2011 5th International Conference on Energy Sustainability, Washington DC - USA, 7–10 August 2011, https://doi.org/10.1115/ES2011-54243, 2011.

Kramti, S., Ali, J., Bechhoefer, E., Takrouni, K., Darghouthi A., Sayadi, M.: Toward an online strategy for mechanical failures diagnostics inside the wind turbine generators based on spectral analysis, Wind. Eng., 45(4), 782-792, 10.1177/0309524X211028759, 2021.

Kotsiantis, S.B., Zaharakis, I.D. and Pintelas, P.E.: Machine learning: a review of classification and combining techniques, Artif Intell, Rev 26, 159–190, https://doi.org/10.1007/s10462-007-9052-3, 2006.

Kurita, T.:Principal component analysis (PCA), Computer Vision: A Reference Guide, Springer, 1–4, https://doi.org/10.1007/978-3-030-03243-2_649-1, 2019.

Leite, G.d.N.P., Araújo, A.M., and Rosas, P.A.C.: Prognostic techniques applied to maintenance of wind turbines: a concise and specific review, Renew. Sust. Energ. Rev., 81, 1917–1925, https://doi.org/10.1016/j.rser.2017.06.002, 2018.

Li, Y., Jiang, W., Zhang, G., and Shu, L.: Wind turbine fault diagnosis based on transfer learning and convolutional autoencoder with small-scale data, Renew. Energ., 171, 103-115, https://doi.org/10.1016/j.renene.2021.01.143, 2021.

https://doi.org/10.5194/wes-2022-111
Preprint. Discussion started: 1 February 2023

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

Leon-Medina, J. X., Anaya, M., Parés, N., Tibaduiza, D. A., and Pozo, F.: Structural damage classification in a Jacket-type wind-turbine foundation using principal component analysis and extreme gradient boosting, Sensors, 21(8), 2748, https://doi.org/10.3390/s21082748, 2021.

May, N.: The impact of wind power support schemes on technology choices, Energy Econ., 65, 343-354, https://doi.org/10.1016/j.eneco.2017.05.017, 2017.

May, A., McMillan, D., and Thöns, S.: Economic analysis of condition monitoring systems for offshore wind turbine sub-systems, IET Renew. Power Gener., 9, 900–907, https://doi.org/10.104Carroll20199/iet-rpg.2015.0019, 2015.

McKinnon, C., Turnbull, A., Koukoura, S., Carroll, J. and McDonald, A.: Effect of time history on normal behaviour modelling using SCADA data to predict wind turbine failures, Energies, 13(18), 4745, https://doi.org/10.3390/en13184745, 2020.

Microsoft: https://docs.microsoft.com/en-us/azure/machine-learning/component-reference/pca-based-anomaly-detection, last access: 07 March 2022, 2021.

Motlagh, M. M., Bahar, A., and Bahar, O.: Damage detection in a 3D wind turbine tower by using extensive multilevel 2D wavelet decomposition and heat map, including soil-structure interaction, Elsevier, 31, 842-861, https://doi.org/10.1016/j.istruc.2021.01.018, 2021.

Natili, F., Daga, A. P., Castellani, F., and Garibaldi, L.: Multi-Scale Wind Turbine Bearings Supervision Techniques Using Industrial SCADA and Vibration Data, Appl. Sci., 11(15), 6785, https://doi.org/10.3390/app11156785, 2021.

Nguyen, T., Huynh, T., and Kim, J.: Numerical evaluation for vibration-based damage detection in wind turbine tower structure, Wind Struct, 21(6), 657-675, http://dx.doi.org/10.12989/was.2015.21.6.657, 2015.

Nguyen, C., Huynh, T., and Kim, J.: Vibration-based damage detection in wind turbine towers using artificial neural networks, Struct. Monit. Maint., 5(4), 507, https:// doi.org/10.12989/smm.2018.5.4.507, 2018.

Oliveira, M., Simas, E., Albuquerque, M., Santos, Y., da Silva, I., and Farias, C.: Ultrasound-based identification of damage in wind turbine blades using novelty detection, Ultrasonics, 108, 106166, https://doi.org/10.1016/j.ultras.2020.106166, 2020.

Ou, Y., Chatzi, E. N., Dertimanis, V. K., and Spiridonakos, M. D.: Vibration-based experimental damage detection of a small-scale wind turbine blade, Struct Control Health Monit, 16(1), 79-96,10.1177/1475921716663876, 2017.

Peres, R., Rocha, A., Leitao, P., and Barata, J.: IDARTS – Towards intelligent data analysis and real-time supervision for industry 4.0, Comput. Ind., 101, 138-146, https://doi.org/10.1016/j.compind.2018.07.004, 2018.

Ratner, B.: Statistical and machine-learning data mining: Techniques for better predictive modeling and analysis of big data, CRC Press, Third Edition, Taylor & Francis Group, USA, 2017.

Rodrigo, J.A.: Análisis de componentes principles PCA con Python, https://www.cienciadedatos.net/documentos/py19-pca-python.html, last access: 03 March 2022,2020.

Ren, H., Liu, W., Shan, M., and Wang, X. : A new wind turbine health condition monitoring method based on VMD-MPE and feature-based transfer learning, Measurement, 148, 106906, https://doi.org/10.1016/j.measurement.2019.106906, 2019.

https://doi.org/10.5194/wes-2022-111
Preprint. Discussion started: 1 February 2023

Rezamand, M., Kordestani, M., Carriveau, R., Ting, D., and Saif, M.: A New Hybrid Fault Detection Method for Wind Turbine Blades Using Recursive PCA and Wavelet-Based PDF, IEEE Sensors Journal, 20 (4), 2023-2033, 10.1109/JSEN.2019.2948997, 2020.

Sabilla, S. I., Sarno, R., and Triyana, K.: Optimizing threshold using pearson correlation for selecting features of electronic nose signals, Int. J. Intell. Eng. Syst, 12(6), 81-90, 2019.

Sezer, E., Romero, D., Guedea, F., Macchi, M., and Emmanouilidis, C.: An Industry 4.0-Enabled Low Cost Predictive Maintenance Approach for SMEs, IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC), Stuttgart - Germany, 1-8, 17-20 June 2018, doi: 10.1109/ICE.2018.8436307, 2018.

Shin, T.: An Extensive Step by Step Guide to Exploratory Data Analysis, https://towardsdatascience.com/an-extensive-guide-to-exploratory-data-analysis-ddd99a03199e, last access: 10 January 2022, 2020.

Son, J., Kang, D., Boo, D., and Ko, K. : An experimental study on the fault diagnosis of wind turbines through a condition monitoring system, J. Mech. Sci. Technol., 32(12), 5573-5582, https://doi.org/10.1007/s12206-018-1103-y, 2018.

Teng, W., Ding, X., Zhang, X., Liu, Y., and Ma, Z.: Multi-fault detection and failure analysis of wind turbine gearbox using complex wavelet transform, Renew. Energ., 93, 591-598, https://doi.org/10.1016/j.renene.2016.03.025, 2016.

Tibaduiza, D., Mujica, L., and Rodellar, J.: Comparison of several methods for damage localization using indices and contributions based on PCA, J. Phys.: Conf. Ser., 305(1), 012013, 2011.

Tutivén, C., Vidal, Y., Insuasty, A., Campoverde-Vilela, L., and Achicanoy, W.: Early fault diagnosis strategy for WT main bearings based on SCADA data and one-class SVM, Energies, 15, 12, 4381 2022.

Vanawat, N.: How To Perform Exploratory Data Analysis -A Guide for Beginners, https://www.analyticsvidhya.com/blog/2021/08/how-to-perform-exploratory-data-analysis-a-guide-for-beginners/, last access:10 January 2022, 2021.

Velmurugan, R.S. and Dhingra, T.: Maintenance strategy selection and its impact in maintenance function: A conceptual framework, Int. J. Oper. Prod. Manag., 35, 1622-1661, https://doi.org/10.1108/IJOPM-01-2014-0028,2015.

Wang, Q., Dong, Z., Li, R., and Wang, L.: Renewable energy and economic growth: new insight from country risks, Energy, 238 (Part C), 122018, https://doi.org/10.1016/j.energy.2021.122018, 2022.

Wang, X., Zhang, L., and Heath, W. P. : Wind turbine blades fault detection using system identification-based transmissibility analysis, Insight-Non-Destructive Testing and Condition Monitoring, 64(3), 164-169, https://doi.org/10.1784/insi.2022.64.3.164, 2022.

Wang, T., Han, Q., Chu, F., and Feng, Z.: Vibration based condition monitoring and fault diagnosis of wind turbine planetary gearbox: A review, Mech. Syst. Signal Process., 126, 662–685, https://doi.org/10.1016/j.ymssp.2019.02.051, 2019.

Wang, H., Wang, H., Jiang, G., Li, J., and Wang, Y. : Early fault detection of wind turbines based on operational condition clustering and optimized deep belief network modeling, Energies, 12(6), 984, https://doi.org/10.3390/en12060984, 2019.

Xiang, L., Wang, P., Yang, X., Hu, A., and Su, H.: Fault detection of wind turbine based on SCADA data analysis using CNN and LSTM with attention mechanism, Measurement, 175, 109094, https://doi.org/10.1016/j.measurement.2021.109094, 2021.

Xiang, L., Wang, P., Yang, X., Hu, A., and Su, H. : Fault detection of wind turbine based on SCADA data analysis using CNN and LSTM with attention mechanism, Measurement, 175, 109094, https://doi.org/10.1016/j.measurement.2021.109094, 2021.

Xiuli, L., Xueying, Z., and Liyong, W. : Fault diagnosis method of wind turbine gearbox based on deep belief network and vibration signal, IEEE, SICE, 1699-1704, https://doi.org/10.1016/j.measurement.2019.106906, 2018.

Yii, K.J., Geetha, C.: The nexus between technology innovation and CO2 emissions in Malaysia: Evidence from granger causality test, Energy Procedia, 105, 3118–3124, https://doi.org/10.1016/j.egypro.2017.03.654, 2017.

Yao, J., Liu, C., Song, K., Feng, C., Jiang, D.: Fault diagnosis of planetary gearbox based on acoustic signals, Applied Acoustics, 181, 108151, https://doi.org/10.1016/j.apacoust.2021.108151, 2021.

Zeng, X. J., Yang, M., and Bo, Y. F.: Gearbox oil temperature anomaly detection for wind turbine based on sparse Bayesian probability estimation, Int. J. Electr. Power Energy Syst., 123, 106233, https://doi.org/10.1016/j.ijepes.2020.106233, 2020.

Zhang, Z., Verma,A., and Kusiak, A.: Fault analysis and condition monitoring of the wind turbine gearbox, IEEE, 27(2), 526-535, 10.1109/TEC.2012.2189887, 2012.

Zhang, Z.Y. and Wang, K.S.: Wind turbine fault detection based on SCADA data analysis using ANN. Adv. Manuf. 2, 70–78, https://doi.org/10.1007/s40436-014-0061-6, 2014.

Zhang, P., and Lu, D. : A survey of condition monitoring and fault diagnosis toward integrated O&M for wind turbines, Energies, 12(14), 2801, https://doi.org/10.3390/en12142801, 2019.

Zhang, B., Zhang, F., and Luo, H.: Virtual shaft-based synchronous analysis for bearing damage detection and its application in wind turbines, Wind Energy, 25(7), 1252-1269, https://doi.org/10.1002/we.2727, 2022.