



Article Wind Fleet Generator Fault Detection via SCADA Alarms and Autoencoders

Mattia Beretta ^{1,2,*}, Juan José Cárdenas ², Cosmin Koch ² and Jordi Cusidó ^{2,3,*}

- ¹ Unitat Transversal de Gestió de l'Àmbit de Camins UTGAC, Universitat Politécnica de Catalunya, 08034 Barcelona, Spain
- ² SMARTIVE S.L., 08204 Sabadell, Spain; juan.cardenas@geniussportsmedia.com (J.J.C.); cosmin.koch@smartwires.com (C.K.)
- ³ Enginyeria de Projectes i de la Construcció EPC, Universitat Politécnica de Catalunya, 08028 Barcelona, Spain
- * Correspondence: mattia.beretta@smartive.eu (M.B.); jordi.cusido@upc.edu (J.C.)

Received: 27 October 2020; Accepted: 1 December 2020; Published: 3 December 2020



Featured Application: Novel approach to wind fleet generator fault detection using Supervisory Control and Data Acquisition (SCADA) data and alarm logs.

Abstract: A hybrid health monitoring system for wind turbine generators is introduced. The novelty of this research consists in approaching a 115-wind turbine fleet by using the fusion of multiple sources of information. Analog SCADA data is analyzed through an autoencoder which allows to identify anomalous patterns within the input variables. Alarm logs are processed and merged to the anomaly detection output, creating a reliable health estimator of generator conditions. The proposed methodology has been tested on a fleet of 115 wind turbines from four different manufacturers located in various locations around Europe. The solution has been compared with other existing data modeling techniques offering impressive results on the fleet. An accuracy of 82% and a Kappa of 56% were obtained. The detailed methodology is presented using one of the available windfarms, composed of 13 onshore wind turbines rated 2 MW power. The rigorous evaluation of the results, the utilization of real data and the heterogeneity of the dataset prove the validity of the system and its applicability in an online operating scenario.

Keywords: alarms; anomaly detection; autoencoder; fault detection; SCADA data; generator; predictive maintenance; wind turbines; renewable energy

1. Introduction

Wind energy is one of the main enablers of the ongoing renewable energy revolution. It was reported by WindEurope that in 2016, wind energy production overtook coal as the second largest form of power capacity in Europe, right behind natural gas. The strong increasing trend suggests that it is just a matter of time for wind energy to take the lead [1].

Many challenges are yet to be solved to increase wind energy profitability, and operation and maintenance (O&M) in particular has to be improved. It was reported that unexpected breakdowns typically cause 10–15% of production losses, with extreme peaks of 30% [2]. These losses cripple the profit of energy companies, thus it is not surprising to find optimization of O&M through big data, cloud solutions and innovative technologies as one of the top priorities of the industry [3].

Historically, maintenance has been performed via a reactive approach, based on preventive inspections and corrective interventions once failures were acknowledged. New approaches providing predictive maintenance solutions have emerged both in the academia and the industrial scene.

Turbines are commonly equipped with a Supervisory Control and Data Acquisition (SCADA) system, which was initially installed to monitor and operate the system, but lately has been utilized to

assess and predict the health status of the turbines as well. SCADA data is recorded by a network of sensors located in the main components of the turbine, the typical sampling frequency is 10 min, making it relatively cheap to collect, transmit and store in a database. All wind fleet operators collect data on their centralized control. SCADA data is collected and stored on Structured Query Language (SQL) databases from SCADA providers or OsiSoft PI system.

Early fault detection can be achieved, as shown by Schlettingen and Santos, by building a model that captures normal operation of the system and by comparing the difference between predicted and measured values of a key variable, to detect anomalies [4]. This approach does not fully take advantage of the high dimensionality of the SCADA dataset and focuses only on the behavior of a single key variable, while component failures are typically complex and can manifest themselves in different failure modes.

The literature is rich with examples based on power curve modeling of wind turbines [5–8]. This approach is based on tracking the relation between wind speed and output power, the function that describes the relation between these two variables can be inferred from operational data and compared to the one provided by the manufacturer, and significant deviations from the theoretical power curve can be hints of problems in the turbine. Different algorithms, as well as the introduction of context variables, have been studied in order to get a reliable picture of the turbine behavior. The main drawback of this approach is its incapacity to determine which component is causing underperformance since the turbine is studied as a whole.

Solutions based on control monitoring systems (CMS) are available and have been studied in the literature [9–11]. These analyses typically use vibration, sound and acceleration measurements to detect anomalies in the behavior of bearings, gearboxes and other mechanical components. The frequency of the data used for these analyses is much higher than the typical SCADA data, thus bringing more information for the detection of failures. That being said, most turbines are not provided with vibration sensors, the installation of these instruments disrupts the operation of the turbine and can cost a windfarm owner thousands of euros per turbine. The authors of Reference [12] presented a thorough analysis of the available monitoring techniques for wind turbine; regarding the CMS, they highlighted as main challenges: financial cost, difficulty of interpretation of the results and not-trivial integration with all the existent monitoring systems, as well as its scalability.

For these reasons, solutions based on the usage of SCADA data can be particularly interesting for owners of old turbines, since no installation of additional sensors or interruption of their operations is needed. Value can be created from the large quantity of unutilized SCADA data stored in their databases.

The rapid growth of the Deep Learning field led many researchers to apply neural networks to solve data challenges. Autoencoders in particular appear to be a good fit for anomaly detection. Autoencoders have been applied in multiple practical applications, such as anomaly detection of seasonal Key Performance Indicators (KPIs) in web application [13], cyber-security monitoring [14] and monitoring of gas turbine conditions [15].

In the wind energy sector, Jiang et al. stacked multiple autoencoders to extract new representations of vibration data in the event of gearbox failures [16]. Successively, they also utilized denoising autoencoders, enriched with temporal information to assess turbine conditions in a laboratory and online scenario [17]. Finally, autoencoders have been successfully used for ice-detection on turbines' blades by Liu et al. [18].

Alarms and events records have been used to determine the remaining useful life of wind turbines [19]. In Reference [20], the time-sequence of the alarms is analyzed to detect relations between the different alarms, determining the causal relationship between the different events and helping to determine the root-cause of failures.

This research aims to explore the capabilities of autoencoders and SCADA alarms as a hybrid fault detection system for wind turbines' generators. While in the literature examples of predictive strategies based only on SCADA data or alarms are present, no holistic approach using both sources of

information is present. This paper reports a methodology that takes advantage of both SCADA and alarm logs in the same algorithm.

As a benchmark, other typical anomaly detection algorithms are implemented, and their results are compared with the autoencoder's results. Additionally, the overall methodology is compared to a normality model, one of the most common predictive maintenance approaches available in the literature. Given the practical nature of the project, SCADA and alarm logs of existing windfarms are used. Results are validated using maintenance logs and verifying the concordance between the predictions and the available information.

A key aspect of this investigation is the thorough analysis of real data from a heterogeneous sample of data. The dataset includes four different turbine brands, from seven different windfarms, located in different nations and climates (Spain, United Kingdom and Poland). Moreover, the size of the sample is remarkable, as more than a hundred turbines are studied. These factors are rare in the relevant literature, as most of the time, a single turbine or windfarm is analyzed. All these considerations support the applicability of the approach in real-life scenarios and its ability to generalize results to heterogeneous conditions.

2. Materials and Methods

2.1. Data Description

The source of information used for this research are the SCADA and alarm datasets as inputs to the model, and the maintenance task logs as ground-truth material to evaluate the effectiveness of the methodology. Two years of operation data for more than 100 turbines rated 2 MW and different manufacturers was available. Data has been received directly from the windfarm operator in the form of comma-separated values (csv) and text archives and uploaded in a SQL database.

The dataset was split into a training and test set, maintaining a train/test split ratio of 70–30%. The last 9 months of data have been used as the test dataset, and the remaining data was used for training the algorithms. The utilized data is a real-life dataset of various windfarms operating under common conditions, it is not the results of a simulation. As a consequence, the data required thorough cleaning and pre-processing to get rid of inconsistencies due to sensors' errors and communication malfunctions.

2.1.1. SCADA Dataset

The SCADA dataset contains more than 300 variables as the main systems of the turbine are all monitored (pitch, main shaft bearing, gearbox, generator, etc.). Sampling frequency is 10 min and quantities such as the arithmetic mean, minimum, maximum and standard deviation are computed with the data acquired for this period. The format of the SCADA dataset, as well as the name of the variables and position of the sensors, may vary according to the manufacturer of the turbine. An example of the dataset used in this research is provided in Table 1.

Table 1. Sample of the Supervisory Control and Data Acquisition (SCADA) dataset. Average (avg.) and standard deviation (std.) values are reported.

Timestamp	Wind Speed (avg.) (m/s)	Power (avg.) (kW)	Power (std.) (kW)	Generator Stator Temperature (avg.) (°C)		
2018-10-01 00:10:00	4.945	282.8	28.524	64.653		
2018-10-01 00:20:00	5.361	331.433	20.253	64.322		
2018-10-01 00:30:00	5.01	289.525	47.297	61.16		

2.1.2. Alarm Dataset

Alarms are typically triggered whenever an operating parameter, most typically a temperature, exceeds its normal operation range. Table 2 is an example of the information contained in the alarm

dataset. The alarm description field contains standardized text data, generated by the control system of the turbines.

		-	
Start time	End time	Turbine	Alarm Description
2017-05-24 10:39:29	2017-05-24 10:40:27	WT05	Gen brushWear Warn
2018-07-05 11:47:29	2018-07-05 11:59:57	WT03	GenRot RpmMonitor Stop
2018-04-13 08:00:58	2018-04-13 12:37:04	WT02	Gear OilFilt Warn (75% clogged)

Table 2. Alarm dataset sample.

2.1.3. Work Orders Dataset

All the maintenance tasks that have been carried on in the windfarm, including inspections, regular checks as well as extraordinary interventions, are registered in the work order logs. An example of the available work order logs is provided in Table 3. This information has been used for labeling turbines' SCADA data. Records preceding critical interventions to the turbines have been removed from the training dataset. Work orders have also been used for the prediction evaluation. The information of the work orders is not provided in any form to the predicting algorithm, it is uniquely utilized to process data, assigning labels, and finally, evaluate the predictions, thus being the ground truth for the algorithm.

Table 3. Work order sample.

Start Time	End Time	Turbine	Component	Work Description
2017-02-18 07:52:00	2017-02-19 13:30:00	WT06	Generator bearing	Generator bearings replacement
2017-06-30 10:46:00	2019-06-30 14:03:00	W108	Blade	Scheduled inspection
2017-08-27 08:50:00	2018-09-03 15:28:00	WT07	Gearbox	Gearbox replacement due to fractured gear tooth

2.2. Autoencoder Anomaly Detection

Anomaly detection via autoencoder is performed providing the network a training dataset composed of normal data, that can be represented as $\{x(1), x(2), ..., x(m)\}$. Autoencoders can be divided into two parts: an encoder and a decoder.

The encoder's goal is to reduce the dimension of the data, mapping data into lower dimensional spaces, reducing the number of neurons in each successive layer, until the bottleneck is reached. The number of layers and neurons in the network is determined by a tradeoff between the compression of the input information and the ability to reconstruct the input sufficiently well. Neurons are activated by an activation function such as the one presented by the following equation [21]:

$$a_i^{(l)} = f\!\left(\sum_{j=1}^n W_{ij}^{(l-1)} a_j^{(l-1)} + b_i^{(1)}\right) \tag{1}$$

where *W* and *b* are the weight and bias of the model, and the indexes *i* and *j* denote the unit and the layer, respectively. Non-linear activation functions are typically utilized to allow the network to represent non-linear characteristics of the data. In this research, the rectified linear unit (ReLU) function has been used, and is defined as follows:

$$f(x) = \max(0, x) \tag{2}$$

The decoder's function is to reconstruct the encoded data at the best of its possibilities. The entire structure, encoder and decoder, is in fact optimized, minimizing the following cost function, presented in Reference [21]:

$$J(W,b) = \frac{1}{m} \sum_{i=1}^{m} \left(\frac{1}{2} \|x(i) - \hat{x}(i)\|^2\right) + \frac{\lambda}{2} \sum_{l=1}^{n_l-1} \sum_{i=1}^{s_l} \sum_{j=1}^{s_l+1} \left(W_{ji}^{(l)}\right)^2$$
(3)

in which n_l is the number of layers, s_l is the number of units in layer L_l and λ is the regularization parameters that keep a balance between the memorization and generalization capabilities of the network. As the equation shows, a larger and more complex network would be penalized by the factor λ . The first part of the equation defines the difference between the input and output vectors, and thus a priority of the network will be to minimize this difference.

As explained in Reference [22], anomaly detection using autoencoders can be seen as a semi-supervised learning problem. The autoencoder is trained with normal data and learns its representation in a reduced dimensional space. The reconstruction error is utilized as a metric to determine abnormal data. Data that does not fit the representation learned in the training phase results in higher reconstruction error and can be marked as anomalous.

2.3. Methodology

Fusion of multiple sources of information, namely SCADA data anomaly detection and alarm registers, is the core of this research. First, the initial processing of the SCADA data is presented, then the processing of the alarms and the final step of merging the autoencoder and alarms' predictions in unique indicators are discussed separately.

2.3.1. SCADA Data Processing

Of the entire dataset, a subset of six variables is used to model the generator: active and reactive power, temperature of nacelle and generator stator, as well as wind and generator speed. While the dataset was composed of more than 300 variables, just a small selection was kept. Processing all the variables would result in very large computation time and likely lead to overfitting of the data, interpretability of the predictions would also be not trivial since the number of inputs would be very large. The selection of the variables has been done choosing measurements related to the system under evaluation (generator speed, generator stator temperature) as well as context signals that determine the operating status of the turbine (active and reactive power, nacelle temperature and wind speed).

The dataset is split into a training and a test set, the first 70% of the data was used for training and the remaining 30% for test. Data shuffling has been avoided, since the dataset is composed of timeseries and random selection of data could result in information leakage.

Analysis of the maintenance and alarm logs allows to filter out abnormal operating conditions from the training set, as well as remove outliers caused by sensor malfunctions, thus creating a training set composed only by normal operation records. No imputation of missing data was performed. To filter data, pre-processing algorithms [23] are applied. In practice, a range of acceptable values for the input variable of the model is defined and all the data not conforming with this range has been filtered, considered as communication errors.

A crucial part of pre-processing is normalization of data, the training set is used to determine the minimum and maximum value for each input variable, and these values are then stored to be used later on in the test set.

2.3.2. Autoencoder Architecture Selection and Training Process

To determine the optimal architecture (number of layers and neurons, activation function, etc.) of the autoencoder, a grid search approach is used, multiple configurations are tested and the one obtaining the lowest reconstruction error is chosen. Training time and complexity of the network have been considered. A process of trial and error of different configurations is necessary to determine the best structure for the available data; thus, a different dataset could result in a different network structure. The best network layout is a fully connected network composed of six layers, having respectively 7–12–4–12–7 neurons activated by the rectified linear unit (ReLU) function and mean squared error was used to measure the distance between the input and output, and the optimization algorithm is "adam".

Having found the best network layout, its predictions on the training data are created to obtain the distribution of the reconstruction error, which is the difference between the original and the processed

data. The assumption that the reconstruction error does not contain systematic errors is verified, analyzing its distribution that resembles a normal distribution. Using this information, it is possible to determine a critical value to identify anomalies. Three standard deviations from the central value are utilized.

2.3.3. Alarms' Processing

Alarms' data is processed by selecting, from all the alarms available in the dataset, the ones that are more relevant for the generator assembly, such as high temperature, overspeed and overload of the generator or its auxiliaries, such as cooling fans. The alarm description field of the dataset was analyzed by keywords, terms such as: "high-temperature", "error", "warning", "over speed", "overload", etc., were searched. In this step, expert knowledge played an important role in excluding from the initial selection those alarms that do not represent truly critical conditions and not simple communication errors.

Once the list of alarms has been defined, it is possible to count how many times any of the selected alarms has occurred during the period under evaluation. In this research, the authors decided not to assign a different weight to the various alarms and simply counted the occurrences. More refined strategies involving rankings of the alarms, as well as detection of patterns or study of the time separating two consecutive alarms, could be implemented in future studies. According to this indicator, turbines having a higher number of alarms should be prioritized for maintenance.

2.3.4. Indicators' Merging Process

The health predictions are made for the entire period of time comprised in the test set and information is aggregated to construct a generator health indicator. Anomalies are summarized to a weekly resolution, by comparing the number of anomalies detected in each turbine with respect to the windfarm. The distribution of anomalies within the windfarm is calculated and turbines lying at a distance superior to two standard deviations from the central value are considered anomalous. This is done because particular external conditions lead the entire windfarm to behave anomalously while not undergoing a real fault in the generator system.

The generator's health indicator is a vector defined in a two-dimensional space. The components of the vector are the processed output of the autoencoder and the counter of key alarms per turbine during the period of the analysis, the module of the vector is calculated as the Euclidean Sum of the two components. A threshold is defined to determine and prioritize the turbines that require maintenance. Alarms' data is used directly in the model, hybridizing and complementing the results of the numerical analysis performed with the autoencoder. The generated status vector considers anomalies in the numerical data and information from the alarm system.

Figure 1 summarizes all the steps of the proposed methodology showing data reception, its storage and preprocessing and the predicting algorithm.



Figure 1. Alarm and autoencoder hybrid predictive system pipeline.

3. Results

The methodology has been proven on a fleet of more than 100 wind turbines, from four different manufacturers, located in very different geographical locations ranging from hot climates, such as south of Spain, to colder ones such as Poland and the United Kingdom. While adjustments were required due to the different variables and characteristics of the turbines, the overall methodology was not modified.

3.1. KPIs Definition

A brief explanation of the indicators utilized for the presentation of the results is provided in this subsection.

In order to assess the prediction power of the predictive models, we have used the confusion matrix (CM) as a basic unit of evaluation. The CM consists of four labels given to each prediction according to its veracity. In summary, these labels are true positives (TP, a failure occurs when a failure was predicted), false positives (FP, no failure when a failure was predicted), true negatives (TN, no failure when no failure was predicted) and false negative (FN, failure when no failure was predicted). Using the count of these basic evaluation units, the main KPIs are calculated.

The main KPIs used in this project are sensitivity, specificity, accuracy, Kappa, precision and F1 score. Sensitivity, Recall is the ratio of predicted events over the total of events:

Sensitivity,
$$Recall = \frac{TP}{TP + FN}$$
 (4)

Specificity is the ratio of well-predicted negative events over the total of negative events:

$$Specificity = \frac{TN}{TN + FP}$$
(5)

Accuracy is the ratio of the total well-predicted observations over the total number of observations:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$
(6)

Cohen's Kappa is defined as follows:

$$K = \frac{p_0 - p_e}{1 - p_e}$$
(7)

where p_0 is the relative observed agreement among raters, which is analogous to accuracy, and p_e is the hypothetical probability of chance agreement, using the observed data to calculate the probabilities of each observer randomly seeing each category. For categories, k, number of items, N, and n_{ki} , the numbers of times the rater i predicted category k, p_e can be calculated as follows:

$$p_e = \frac{1}{N^2} \sum_k n_{k1} n_{k2} \tag{8}$$

A low value of *K* means that there is no agreement among the raters other than what would be expected by chance. A *K* value close to one is an indication of good performance of the classifier.

Precision is the ratio of predicted events over the total of positive predictions:

$$Precision = \frac{TP}{TP + FP} \tag{9}$$

F1 score is defined as the harmonic mean of precision and recall and it is typically used to measure the accuracy of a test:

$$F1 = 2\frac{precision * recall}{precision + recall} = \frac{TP}{TP + \frac{1}{2}(FN + FP)}$$
(10)

3.2. Autoencoder and Alarms Results

As one of the main goal of this research is to demonstrate the advantages of merging different sources of information, the results of the autoencoder and an alarm-based predictive system are presented and compared to the numbers obtained using a unique predictor made by the fusion of the two individual methods.

Table 4 presents the results obtained using the autoencoder as a unique predictor of the generator status. It can be seen that various failures are anticipated, but the rate of FPs is quite high, as well as the FNs.

Table 4. Results obtained using the autoencoder. WF stands for Windfarm, TP True Positive, FN False Negative, FP False Positive and TN True Negative.

	Brand	Location	ТР	FN	FP	TN	Accuracy	Kappa	Sensitivity	Specificity	Precision	F1
WF1	Vestas	Spain	3	1	4	3	55%	15%	75%	43%	43%	55%
WF2	Vestas	Spain	1	3	2	3	44%	-15%	25%	60%	33%	29%
WF3	Siemens	Poland	1	0	6	11	67%	17%	100%	65%	14%	25%
WF4	Siemens	Poland	3	3	1	8	73%	41%	50%	89%	75%	60%
WF5	Senvion	Poland	4	1	4	4	62%	27%	80%	50%	50%	62%
WF6	Senvion	Poland	3	1	6	12	68%	28%	75%	67%	33%	46%
WF7	Nordex	UK	2	4	7	13	58%	-1%	33%	65%	22%	27%
TOTAL			17	13	30	54	62%	18%	57%	64%	36%	44%

Table 5 shows the results obtained using an alarm-based predictor. The results are not so different from the autoencoder's ones, a slightly higher Kappa is achieved by this method, and one more TP was found, while the FPs rate is almost equal. It is clear that neither of the two techniques, on its own, would be sufficiently reliable in a real-life scenario.

Table 5. Results obtained using an alarm-based predictor. WF stands for Windfarm, TP True Positive, FN False Negative, FP False Positive and TN True Negative.

	Brand	Location	ТР	FN	FP	TN	Accuracy	Kappa	Sensitivity	Specificity	Precision	F1
WF1	Vestas	Spain	2	2	1	6	73%	38%	50%	86%	67%	57%
WF2	Vestas	Spain	2	2	1	4	67%	31%	50%	80%	67%	57%
WF3	Siemens	Poland	1	0	17	0	6%	0%	100%	0%	6%	11%
WF4	Siemens	Poland	2	4	2	7	60%	12%	33%	78%	50%	40%
WF5	Senvion	Poland	4	1	0	8	92%	83%	80%	100%	100%	89%
WF6	Senvion	Poland	4	0	6	12	73%	42%	100%	67%	40%	57%
WF7	Nordex	UK	3	3	4	16	73%	28%	50%	80%	43%	46%
TOTAL			18	12	31	53	62%	19%	60%	63%	37%	46%

3.3. Overall Results

As the results of the individual predictors are not sufficiently good, the authors present a hybrid technique that merges the two systems in a more complete predictor, as detailed in Section 2.3. Table 6 presents a summary of the results. The turbines that were obtaining higher values for the health KPIs were reported. Examining the reported turbines and the maintenance log, the results table was done. During the test period, problems such as broken generators, consumed generator brushes or generators bearing damages were encountered.

It can be seen that most of the reported turbines were found to have some problems; moreover, the results across the various windfarms are consistent. The accuracy never gets lower than 70% and the overall Kappa is 56%. The advantages of using a hybrid predictor are clear when its results are compared to the ones of the autoencoder and alarm predictors. The number of TPs increased substantially, and remarkably, the number of FPs was halved. The two components of the composed predictors are complementary, allowing for more accurate and reliable predictions.

	Brand	Location	ТР	FN	FP	TN	Accuracy	Kappa	Sensitivity	Specificity	Precision	F1
WF1	Vestas	Spain	4	0	0	7	100%	100%	100%	100%	100%	100%
WF2	Vestas	Spain	3	1	1	4	78%	55%	75%	80%	75%	75%
WF3	Siemens	Poland	1	0	3	14	83%	34%	100%	82%	25%	40%
WF4	Siemens	Poland	4	2	0	9	87%	71%	67%	100%	100%	80%
WF5	Senvion	Poland	5	0	2	6	85%	70%	100%	75%	71%	83%
WF6	Senvion	Poland	3	1	4	14	77%	41%	75%	78%	43%	55%
WF7	Nordex	UK	3	3	4	16	73%	28%	50%	80%	43%	46%
TOTAL			23	7	14	70	82%	56%	77%	83%	62%	69%

Table 6. Key Performance Indicators (KPIs) results summary for all the available windfarms (WF), sorted by turbine manufacturer (Brand) and location. TP stands for True Positive, FN False Negative, FP False Positive and TN True Negative.

The Receiving Operator Curve (ROC) is calculated to represent the predictive power of the proposed methodology and its response to adjustments in the cutoff value to apply to the health status vector. In Figure 2, the ROC curves of the different windfarms are presented. The cutoff values are adjusted for each wind farm to obtain optimal results.



Figure 2. Receiving Operator Curve (ROC) of the various windfarms, the points and the corresponding values represent possible cutoff values to make a decision.

Figure 3 represents the dataset as a whole, without distinction between the different windfarms and simulating the effect of a unique cutoff value. The two dashed line defines the values of the false positive rate and true positive rate that can be obtained by selecting the optimal cutoff value for each windfarm. It can be seen that fixing a unique threshold value yields good results while being a simpler decision strategy, but in applications where the reliability of the prediction is the key objective, the additional complexity provides better outputs.



Figure 3. ROC curve of the entire dataset, using a unique threshold value for all the windfarms. The point defined by the intersection of the dashed lines is the result reported in Table 6.

3.4. Normality Model Comparison

An additional validation of the results is presented. A normality model using the same input data is trained and utilized to make health predictions of the generators. Details on how to build a normality model are available in Reference [4]. The value of the generator stator temperature is predicted by a ridge regression model and the prediction error is used as a metric for the generator status. Details on the algorithm can be found in Reference [24], the decision of using this algorithm is dictated by its capacity to deal with multicollinearity in the inputs. The results of the normality model are presented in Table 7.

	Brand	Location	ТР	FN	FP	TN	Accuracy	Kappa	Sensitivity	Specificity	Precision	F1
WF1	Vestas	Spain	4	0	2	5	82%	65%	100%	71%	67%	80%
WF2	Vestas	Spain	1	3	0	5	67%	27%	25%	100%	100%	40%
WF3	Siemens	Poland	1	0	12	5	33%	4%	100%	29%	8%	14%
WF4	Siemens	Poland	4	2	6	3	47%	0%	67%	33%	40%	50%
WF5	Senvion	Poland	0	5	2	6	46%	-28%	0%	75%	0%	NA
WF6	Senvion	Poland	2	2	5	13	68%	17%	50%	72%	29%	36%
WF7	Nordex	UK	5	1	6	14	73%	41%	83%	70%	45%	59%
TOTAL			17	13	33	51	60%	14%	57%	61%	34%	43%

Table 7. Normality model results. WF stands for Windfarm, TP True Positive, FN False Negative, FP False Positive and TN True Negative.

One can see that while the normality model yields reasonable results, it scores lower overall values for the tracked indicator when compared to the presented methodology. In particular, it should be noted that the number of FPs is more than double the proposed solution and the total number of TPs is lower. The only case in which the normality model performed better is WF7, where two additional TPs are found.

The presented results were obtained using a large sample of real data. The sample is extremely heterogeneous since it represents four different turbine brands, and the windfarms are located in different geographical locations (Poland, Spain and United Kingdom), characterized by very different climates and wind conditions. Such results are rare in the literature, as many algorithms have been tested either in laboratories or in a reduced sample of turbines.

In Section 4, the detailed analysis of windfarm 5 is proposed. This one was chosen since it has a high prevalence of failures of the generator and two predictions were classified as FN, so it is useful to analyze them in detail to determine the reason why the alarms were raised.

4. Discussion

The last 9 months of data available were used as a test set. The performance of the autoencoder as an anomaly detector was compared to other algorithms that have been widely used for anomaly detection tasks. Isolation forest and one-class support vector machine were tested. Details on these algorithms can be found in References [25,26].

The same post-processing methodology was applied to all algorithms. Results are presented in Figure 4. Three risk-areas were identified based on the generator's health indicator value distribution. Table 8 provides the information to assess the accuracy of the predictions, and major component replacements that took place during the testing phase are reported.



Figure 4. Comparison of the results obtained by the three implemented algorithms hybridized with alarms information. The higher the distance from the origin, the worse the conditions of the generator. Three areas are identified according to the health status: healthy (green), warning (yellow) and danger (red). The shape determines the presence and type of fault occurred.

Turbine	Maintenance Description	Component
WT13	Bearing High Speed Shaft replacement	Gearbox-Generator
WT11	Generator brushes replaced	Generator
WT11	Generator bearing Non-Drive End replaced	Generator
WT10	Generator bearing Non-Drive Endreplaced	Generator
WT08	Generator bearing Non-Drive Endreplaced	Generator
WT07	Generator brushes replaced	Generator

Table 8. Principal maintenance intervention occurred during the testing phase.

All three algorithms, when merged with alarm information, are able to satisfactorily isolate faulty turbines from the rest. Autoencoder is selected as the algorithm of choice to analyze SCADA data, since it is able to better diagnose faulty turbines even in the absence of alarms data, as in the case of turbine WT13. Moreover, the autoencoder better identifies the high-speed shaft-bearing fault, where isolation forest could not separate it sufficiently and one class Support Vector Machine (SVM) positioned it on the frontier between the warning and safe areas, the ability to identify various failure modes holds large relevance in the selection of the algorithm. Analyzing the results of the autoencoder, it can be noticed that most of the turbines in the critical (red) and dangerous (yellow) areas required replacement of the bearings or brushes of the generator. None of the windmills located in the safe (green) area required maintenance.

A detailed study of the data of WT09 and WT12 was done due to their high anomaly count and absence of maintenance intervention. The input variable distributions of all the signals and some other key variables of the generator have been reviewed thoroughly to understand the reason why the autoencoder has found these turbines to be anomalous. The most relevant relationships related with generator failure are presented here and discussed.

In Figure 5, the distribution of the probability density of the temperature difference across the two sides of the generator bearing of turbine WT09 are represented, compared with the mean value of the

windfarm and the characteristic curve of this temperature difference with respect to nominal power. It can be observed that the behavior of turbine WT09 is widely different from the rest of the windfarm. These considerations lead us to categorizing this prediction as early fault alert of the generator bearing conditions, rather than a false alarm.



Figure 5. On the left, the characteristic curve that relates the active power and the temperature difference on the two sides of the generator bearing. On the right, the distribution of the probability density function of the temperature difference across the generator bearing. In red, the data belonging to turbine WT09, in black, the mean of the windfarm.

Figure 6 shows that turbine WT12 is characterized by an anomalous distribution of the generator stator temperature, in fact the standard deviation of its recorded values is larger than the value of the windfarm, meaning that the generator of this turbine is subjected to less stable operating conditions. This case can also be considered anomalous and worthy of a technical review of the generator.



Figure 6. Relation between active power and standard deviation of the generator stator temperature (**left**) and the distribution of the probability density function of generator stator standard deviation (**right**). In red, values of turbine WT12, and in black are the mean values for the entire windfarm.

Merging the information of alarms with anomalies provides a more comprehensive health status of the generator. Looking at the plots, it can be seen that alarms are able to isolate most of the faulty turbines, that being said, there are also cases in which a low number, or no alarms are raised, but nonetheless, the turbine was found to be faulty. WT08 problems are detected mainly by the alarm counter, whereas WT13 is purely diagnosed by the anomaly count, the rest of the faults are found by a mix of the two information sources. Ultimately, merging the information from alarms and SCADA data proved a rewarding strategy able to better separate turbines according to their health status, making use of available and easily accessible data.

5. Conclusions

A hybrid fault detection system based on SCADA alarm logs and an anomaly detection autoencoder were presented and validated on a fleet of more than 100 wind turbines, from four different manufacturers, located in different parts of Europe. Real operating data has been used and most of the raised alarms corresponded to problems related to the generator that required the substitution of the component or some parts of it (bearings, brushes).

A detailed explanation of the most critical windfarm was presented to show how the methodology can be applied in practice and the kind of analyses that were carried out to corroborate the results.

It has been observed that the alarm counter is a valid tool to distinguish faulty turbines from healthy ones. That being said, the alarm counter alone cannot anticipate all failures. The fusion of anomalies and alarms information complements the individual approaches, providing a more reliable system.

All five failures that occurred during the test phase were correctly detected. Of the two "false positive" predictions that were obtained, detailed analyses suggested that they are likely early fault detections, rather than errors. Ultimately, this methodology provides windfarm operators a reliable tool to assess the health of generators and improve operation and maintenance of the turbines.

The results of the autoencoder as an anomaly detector were compared with other common algorithms in the literature, such as isolation forest and one-class support vector machine. The results showed that while the other two algorithms provide acceptable results, autoencoders are more confident in their predictions in cases where alarm information cannot help so much with separating faulty from healthy turbines. Autoencoders, having more tunable parameters and allowing for more elaborated structures, are capable to better interpret non-linear data, such as that of a turbines generator. Additionally, the overall methodology was tested against a normality model, and the results clearly showed that the proposed solution ranks better for all the tracked statistics.

This research contributes to present a novel methodology that makes use of data analysis techniques for anomaly detection and consolidates the results, merging the anomaly predictions with information from the alarm system. The large size of the datasets and its diversity contribute to prove the approach as a general solution that can work well in real-life conditions and is not only applicable to a niche of turbines.

Different network architecture, including temporal information and denoising autoencoders, should be explored in future research to boost the accuracy of the system. Interpretability of results is a key aspect that requires further improvements to ensure acceptance of this methodology in the market.

Author Contributions: Conceptualization, M.B. and J.J.C.; methodology, M.B., J.J.C. and C.K.; software, M.B., J.J.C. and C.K.; validation, M.B., J.J.C., J.C. and C.K.; formal analysis, M.B., J.J.C. and J.C.; investigation, M.B.; resources, J.C.; data curation, M.B. and J.C.; writing—original draft preparation, M.B.; writing—review and editing, J.C.; visualization, M.B.; supervision, J.J.C. and J.C.; project administration, J.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: The authors want to acknowledge the support of the other colleagues at SMARTIVE, Beatriz Jaramillo and Jose Sepulveda, for their support and technical advice in the design of this experiment and the analysis of the results. Additionally, the authors would like to acknowledge the work of the anonymous reviewers, who provided useful notes and critiques to improve the overall quality of the article.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Wind Energy in Europe in 2018—Trends and Statistics. 2019. Available online: https://windeurope.org/ wp-382content/uploads/files/about-wind/statistics/WindEurope-Annual-Statistics-2017.pdf (accessed on 12 March 2020).
- 2. Faulstich, S.; Hahn, B.; Tavner, P.J. Wind turbine downtime and its importance for offshore deployment. *Wind. Energy* **2010**, *14*, 327–337. [CrossRef]

- 3. Making Transition Work. 2016. Available online: https://windeurope.org/wp-content/uploads/files/about-386wind/reports/WindEurope-Making-transition-work.pdf (accessed on 12 March 2020).
- Schlechtingen, M.; Santos, I.F. Comparative analysis of neural network and regression based condition monitoring approaches for wind turbine fault detection. *Mech. Syst. Signal Process.* 2011, 25, 1849–1875. [CrossRef]
- 5. Gonzalez, E.; Stephen, B.; Infield, D.; Melero, J.J. On the use of high-frequency SCADA data for improved wind turbine performance monitoring. *J. Phys. Conf. Ser.* **2017**, *926*, 392. [CrossRef]
- 6. Ouyang, T.; Kusiak, A.; He, Y. Modeling wind-turbine power curve: A data partitioning and mining approach. *Renew. Energy* **2017**, *102*, 1–8. [CrossRef]
- 7. Shokrzadeh, S.; Member, S.; Jozani, M.J.; Bibeau, E. Wind Turbine Power Curve Modeling Using Advanced Parametric and Nonparametric Methods. *IEEE Trans. Sustain. Energy* **2014**, *5*, 1262–1269. [CrossRef]
- 8. Pandit, R.K.; Infield, D.; Kolios, A. Gaussian process power curve models incorporating wind turbine operational variables. *Energy Rep.* **2020**, *6*, 1658–1669. [CrossRef]
- 9. Teng, W.; Ding, X.; Zhang, X.; Liu, Y.; Ma, Z. Multi-fault detection and failure analysis of wind turbine gearbox using complex wavelet transform. *Renew. Energy* **2016**, *93*, 591–598. [CrossRef]
- He, G.; Ding, K.; Li, W.; Jiao, X. A novel order tracking method for wind turbine planetary gearbox 405 vibration analysis based on discrete spectrum correction technique. *Renew. Energy* 2016, *87*, 364–375. [CrossRef]
- 11. Mollasalehi, E.; Wood, D.; Sun, Q. Indicative Fault Diagnosis of Wind Turbine Generator Bearings 408 Using Tower Sound and Vibration. *Energies* **2017**, *10*, 1853. [CrossRef]
- 12. Dias, H.; de Azevedo, M.; Maurício, A.; Bouchonneau, N. A review of wind turbine bearing 410 condition monitoring: State of the art and challenges. *Renew. Sustain. Energy Rev.* **2016**, *56*, 368–379. [CrossRef]
- Xu, H.; Feng, Y.; Chen, J.; Wang, Z.; Qiao, H.; Chen, W.; Zhao, N.; Li, Z.; Bu, J.; Li, Z.; et al. Unsupervised Anomaly Detection via Variational Auto-Encoder for Seasonal KPIs in Web Applications. In Proceedings of the WWW 2018: The 2018 Web Conference, Lyon, France, 23–27 April 2018; pp. 187–196.
- 14. Yousefi-azar, M.; Varadharajan, V.; Hamey, L.; Tupakula, U. Autoencoder-based Feature Learning 415 for Cyber Security Applications. *Int. Jt. Conf. Neural Netw.* **2017**, 3854–3861. [CrossRef]
- 15. Yan, W.; Yu, L. On Accurate and Reliable Anomaly Detection for Gas Turbine Combustors: A Deep Learning Approach. In Proceedings of the Annual Conference of the Prognostics and Health Management Society, San Diego, CA, USA, 19–24 October 2015.
- Jiang, G.; He, H.; Xie, P.; Tang, Y. Stacked Multilevel-Denoising Autoencoders: A New Representation Learning Approach for Wind Turbine Gearbox Fault Diagnosis. *IEEE Trans. Instrum. Meas.* 2017, 66, 2391–2402. [CrossRef]
- 17. Jiang, G.; Xie, P.; He, H.; Yan, J. Wind Turbine Fault Detection Using Denoising Autoencoder with Temporal Information. *IEEE/ASME Trans. Mechatron.* **2017**, *23*, 89–100. [CrossRef]
- 18. Wang, Q. Intelligent wind turbine blade icing detection using supervisory control and data acquisition 427 data and ensemble deep learning. *Energy Sci. Eng.* **2019**, *7*, 2633–2645. [CrossRef]
- Luis, M.L.; Rodríguez-l, M.A. Development of indicators for the detection of equipment malfunctions and degradation estimation based on digital signals (alarms and events) from operation SCADA. *Renew. Energy* 2016, *99*, 224–236. [CrossRef]
- Gonzalez, E.; Reder, M.; Melero, J.J. SCADA alarms processing for wind turbine component failure detection SCADA alarms processing for wind turbine component failure detection. *J. Phys. Conf. Ser.* 2016, 753, 072019. [CrossRef]
- 21. Sakurada, M.; Yairi, T. Anomaly Detection Using Autoencoders with Nonlinear Dimensionality Reduction. In Proceedings of the MLSDA 2014 2nd Workshop on Machine Learning for Sensory Data Analysis, Gold Coast, Australia, 2 December 2014; pp. 4–11. [CrossRef]
- 22. An, J. Variational Autoencoder based Anomaly Detection using Reconstruction Probability. *Spec. Lect. IE* **2015**, *2*, 1–18.
- 23. Marti-Puig, P.; Blanco-M, A.; Cárdenas, J.J.; Cusidó, J. Effects of the pre-processing algorithms in fault diagnosis of wind turbines. *Environ. Model. Softw.* **2018**, *110*, 119–128. [CrossRef]
- 24. Hoerl, A.E.; Kennard, R.W. Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics* **1970**, *12*, 55. [CrossRef]

- 25. Liu, F.T.; Ting, K.M.; Zhou, Z.-H. Isolation Forest. In Proceedings of the 2008 Eighth IEEE International Conference on Data Mining, Pisa, Italy, 15–19 December 2008; pp. 413–422.
- Li, K.; Huang, H.; Tian, S.; Xu, W. Improving one-class SVM for anomaly detection. In Proceedings of the Second International Conference on Machine Learning and Cybernetics, Xi'an, China, 2–5 November 2003. [CrossRef]

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).