

Data-Driven Learning Framework for Associating Weather Conditions and Wind Turbine Failures

Maik Reder^{*,1}, Nurseda Y. Yürüşen^{*,1}, Julio J. Melero¹

** Shared first authorship - the authors contributed equally to this work.*

¹ CIRCE - Universidad de Zaragoza, C/ Mariano Esquillor 15, 50018, Zaragoza, Spain

Corresponding emails: mreder@fcirce.es; nyildirim@fcirce.es

Abstract

The need for cost effective operation and maintenance (O&M) strategies in wind farms has risen significantly with the growing wind energy sector. In order to decrease costs, current practice in wind farm O&M is switching from corrective and preventive strategies to rather predictive ones. Anticipating wind turbine (WT) failures requires sophisticated models to understand the complex WT component degradation processes and to facilitate maintenance decision making. Environmental conditions and their impact on WT reliability play a significant role in these processes and need to be investigated profoundly. This paper is presenting a framework to assess and correlate weather conditions and their effects on WT component failures. Two approaches, using (a) supervised and (b) unsupervised data mining techniques are applied to pre-process the weather and failure data. An apriori rule mining algorithm is employed subsequently, in order to obtain logical interconnections between the failure occurrences and the environmental data, for both approaches. The framework is tested using a large historical failure database of modern wind turbines. The results show the relation between environmental parameters such as relative humidity, ambient temperature, wind speed and the failures of five major WT components: gearbox, generator, frequency converter, pitch and yaw system. Additionally, the performance of each technique, associating weather conditions and WT component failures, is assessed.

Keywords: wind turbine, failure, weather, big data, association rule mining, k-means clustering, data mining, machine learning, operation & maintenance

1. Introduction

Throughout the past years the wind energy sector has been growing rapidly. New wind turbine (WT) technologies and higher installed capacity are resulting in increased cost and complexity of operation and maintenance (O&M) tasks. Current maintenance strategies are mainly focusing on corrective and preventive actions, yet, operators could benefit extremely from applying predictive maintenance approaches.

This issue applies to many industrial fields and the solutions proposed in previous research are mainly based on the system age. For this purpose, e.g. Martorell et al. [1] studied age-dependent reliability models such as the proportional age set back (PAS) and proportional age reduction (PAR) frameworks. Both of them take into account maintenance efficiency, environmental and operational conditions as generic imperfect maintenance functions. In their study, three different environmental setups are considered indicating good, normal and bad conditions, which were found to highly influence the age and reliability functions. In Martorell et al. [2] these maintenance models have been extended by including the systems failure causes. It was shown that failures can be related to certain root causes, which need to be taken into consideration. Carlos et al. [3] applied the PAS models in wind energy, with the aim of minimizing the O&M cost of wind farms while maximising their revenue considering the maintenance frequency as decision variable. However, they neglected operational and environmental conditions in the maintenance scheduling framework. A similar situation can be seen also for age reduction models. Ding et al. [4] studied an opportunistic maintenance approach for wind turbines by using an imperfect maintenance model, which depends on an age reduction ratio. However, they also were not including the environmental conditions. In general, environmental conditions are being neglected because of their complexity although having a very high impact on maintenance actions, as demonstrated by XiaoFei et al. [5]. They model the effect of system age and environmental conditions on the hazard rate function by introducing a stochastic process with two states, indicating normal and severe environment. They conclude that the system reliability degrades more rapidly when exposed to severe environmental conditions. Furthermore, the history of the environmental events is said to affect the hazard rate. Albeit, their study did not include the specific meteorological parameters themselves.

Thus, previous research shows that for developing effective maintenance decision models, a deep understanding of the component degradation processes including the highly variable environmental conditions is required. Complex en-

environmental and operational conditions must be defined and studied in detail to generate realistic ageing and reliability functions. Especially for systems that are interacting strongly with their surroundings. This paper shows how to complement these models by associating failure information with significant environmental variables by providing a tool that enables the interpretation of failures based on dynamic meteorological conditions. With this, the presented framework aims at supplementing existing maintenance and reliability models and enhancing the outcome of maintenance decision models. The framework will be applied to wind energy but can easily be adapted for other areas.

1.1. Background in Wind Energy

As stated in Kuik et al. [6], environmental conditions and their complicated combinations still have to be fully understood and methodologies have to be developed in order to analyse and correlate them to WT failures. Especially short term environmental variations, such as for example high-speed wind gusts, can cause a severe impact on components and need to be analysed extensively. This knowledge is essential for further use in WT reliability models and complex maintenance decision making tools.

Much research effort has been dedicated to identifying weather dependent failure rates and downtimes of WT components. Hahn et al. [7] were the first to carry out an extensive analysis on the effects of weather on WT reliability. They showed that with rising average daily wind speed the failure rates of certain components increased as well. Tavner et al. [8] analysed the effect of monthly averaged wind speed conditions on component failure rates, using the Danish reliability database WindStats. Assuming the monthly mean wind speed across Denmark to be representative for the considered wind farms, they concluded that there is an annual periodicity in failure occurrences due to seasonal variation in weather conditions. Faulstich et al. [9] carried out an analysis on the effects of wind speed on WT downtimes, taking into account energy- and time-based availability.

Along with high wind speeds, literature also frequently addresses temperature and humidity as critical parameters. Especially humidity can cause corrosion and other highly dangerous degradation processes. Costa et al. [10] stated that correlating relative humidity and failures using classical analysis approaches, such as the analysis of variance (ANOVA) technique, is one of the most challenging tasks. Tavner et al. [11] cross-correlated component failures with average monthly maximum and mean wind speed, maximum and minimum air temperature and average daily mean relative humidity. Their study refers to three specific wind farms using yearly and monthly mean weather conditions taken from closely located meteo-

rological stations. However, neither the actual conditions at the time of failure occurrence nor short-time weather events, e.g. high-speed gusts, were considered. Also, only old wind turbine technologies with a rated power of 300 kW were analysed. They suggested that future studies should ensure to include modern WTs and short-time weather variations, along with details on maintenance strategies. Wilkinson et al. [12] showed the impact of environmental conditions on WT reliability using an availability database and WT downtimes. The environmental data were taken from the *Modern-Era Retrospective Analysis for Research and Applications* (MERRA) database and the results showed that there is a strong relationship between wind speed, temperature and high downtime. Wilson et al. [13] modelled the relationship between failures and weather conditions using Markov Chains, based on data for one wind farm over a relatively short period of time. Recently, the possibilities of using Supervisory Control and Data Acquisition (SCADA) data for these purposes have been discovered. SCADA systems are among the standard equipment of modern wind turbines and can provide a huge amount of information on many operational WT parameters, conditions of its components, as well as external weather parameters. Such extensive data require sophisticated analysis techniques and high computational effort. Thus, a number of research projects have been trying to apply data mining and machine learning techniques using WT SCADA data. For example Kusiak et al. [14] used apriori rule data mining to find frequent item sets in SCADA data in order to identify WT pitch system faults. Machine learning can reveal very interesting results on component degradation, however, they strongly depend on the availability of a considerable amount of input data.

Previous studies evaluating weather effects on component failures are mostly limited to a very low number of analysed turbines and/or out-dated WT technologies. The environmental data considered are mostly comprised of yearly or monthly average values obtained at close-by located weather stations, and short-term changes in these parameters are not taken into account. However, the turbines' SCADA systems provide a rich data source and can give a realistic representation of the conditions at the WTs. In an earlier study by Reder et al. [15], the wind speed effects on WT component reliability have been assessed using 10-minute mean SCADA data taken from failed WTs. It has been found, that short term wind changes need to be analysed further, in order to obtain a clear image of the conditions leading to component failures. Also, modern WTs and further environmental parameters should be taken into account. Hence, there is a significant need for extending previous studies, such as [12] and [11]. Certainly, failures are often caused by cumulative stress over a large period of time. How-

ever, especially short-time weather events including high-speed gusts and ambient temperature changes, can cause the final impact on a component and lead to its failure and are an important factor in weather dependent failure analysis.

Considering the fact that this will be based on a significantly higher amount of data than previous analyses, methodologies capable of handling big data and deriving meaningful results have to be developed. Thus, in this study an effective framework for processing and correlating the different inputs is introduced. This applied to a historical failure database of modern wind turbines located in Spain as well as environmental data taken from the affected wind turbines' SCADA system and weather stations close to the wind farm. Five components are considered: gearbox, generator, pitch system, yaw system and frequency converter. These were chosen according to the results of previous studies, e.g. Reder et al. [16], where the components with the highest share of the total WT failure rates and downtime were identified.

A summary of the relevant literature and the problem statement were presented in this section. The remainder of this paper is structured as follows. The input data for a case study, to which the framework will be applied to, are described in Section 2, comparing the available operational years, sample sizes and WT technology types to previous studies. The established framework including the data pre-processing and the methodology behind the supervised and unsupervised learning algorithms are described in Section 3. Here, also an example for the calculation of the evaluation metrics for interpreting the results is given. The results of applying the established framework to the case study are presented in Section 4. Conclusions are given in Section 5, where the key findings and the possible application areas are emphasized.

2. Case Study - Data Set Description

The framework was tested and verified using the data sets presented in this section. These are comprised of historical failure data, SCADA data, historical weather data and expert judgements.

2.1. Historical Failure Data:

As shown in Table 1, the historical failure database contains a significantly higher number of turbines compared to previous studies. It represents modern three bladed, pitch regulated WTs with rated power between 660 kW and 2000 kW, produced by the same WT manufacturer. All turbines are equipped with a doubly fed induction generator (DFIG). In total 146 failure events were considered,

consisting of around 30 failures for each of the five components. The failure data contains the exact time and duration of the failures (downtime) for each turbine. Failures are defined as WT stops due to component faults that require interventions such as repair or replacement, except inspections and cleaning activities.

Table 1: Historical Failure Data in this study compared to previous studies with similar objectives

	Wilson et al. [13]	Tavner et al. [11]	This study
Number of WTs	140	32	448
Total WT Years	381.7	130	972
WT Technology	Variable speed, pitch regulated 2.3 MW	Enercon E32/33 (300 kW)	47 WTs with 0.66 MW, 289 WTs with 0.85 MW, 112 WTs with 2 MW

In failure analyses and reliability modelling the databases are often considering different turbine technologies separately, as their failure behaviour can differ substantially. Additionally, even WTs of the same technology behave differently under the influence of distinct weather conditions. The latter effect has not been investigated in previous studies, yet.

As the framework presented in this study needs a sufficiently large database in order to provide meaningful results, it was decided the data set was not going to be divided into different turbine technology groups. The division was made according to the failed component class. When comparing previous studies e.g. [8], [15], [13] [17] and [18], it can be seen that even though these studies are analysing different turbine technologies, the respective components behave similarly under the influence of certain weather conditions. At this point it is stressed, that this study examines different WT technologies made by the same manufacturer. Nonetheless, the difference in failure behaviour of different component brands could be investigated in further studies easily by using the herein presented framework.

2.2. SCADA Data:

The wind speed, the turbine’s power production and the downtimes were taken directly from the failed WTs’ SCADA system. The values were compared to the ones collected at close-by turbines, in order to ensure a properly functioning system. They showed very good correlation.

2.3. Environmental Conditions - NCEP/NCAR Reanalysis data set:

The NCEP/NCAR reanalysis data set [19, 20] is a continuously updated weather database using observations and numerical weather predictions. It is developed by

the National Centers for Environmental Prediction (NCEP) and the National Center for Atmospheric Research (NCAR). The data can be obtained free of charge from the National Oceanic and Atmospheric Administration Earth System Research Laboratory of U.S. This served as input for the humidity and temperature at the given wind farm locations and times of failure occurrences. The downloaded reanalysis data were compared to real data taken at three different sites and the results were more than acceptable for the herein used humidity and temperature ranges (mean error under 3 %).

2.4. Expert Judgement:

The opinion of two big European wind farm operators was consulted in order to define the maintenance strategy and the classification of failure severity according to its caused WT downtime. The availability of maintenance personnel has considerable effect on the component repair time.

2.5. Analysed Parameters for the Case Study

The exact times of failure occurrences, the wind speed, relative humidity and ambient temperature were extracted from the data set, as explained below. Additionally, the turbine performance is evaluated by considering its power production and maintenance schedules. The resulting values, labels and clusters for each of the parameters are discussed in Section 4.

- *Wind Speed (WS)*: The 10-minute mean values for a period of 80 minutes ahead of failure ($WS_{80}, WS_{70}, \dots, WS_{10}$), and one wind speed measurement at the time of the failure occurrence (WS_{atF}) were taken from SCADA data.
- *Relative Humidity (RH)*: The hourly values of 10 hours ahead of failure were included into the analysis and are indicated by RH_{atF}, RH_1, RH_2 , etc.
- *Ambient Temperature (T)*: This includes the monthly mean T_m , maximum T_{max} and minimum temperatures T_{min} for the 30 days ahead of failure, as well as the temperature at the exact time of failure T_{atF} .
- *Power Production (P)*: The power production at failure P_{atF} (last 10-minute mean value before failure occurrence) and the power production before failure P_{bF} (the previous 10-minute measurement) are also taken from the turbines' SCADA data. These values will be presented in relation to the measured monthly mean power curve. In other words, as the manufacturers power curves are not site and season specific, a measured monthly mean

representation of power versus wind speed is taken as reference to display the turbine's relative performance. Thus, the observed SCADA value P_o will be divided by the mean monthly value P_m at the same wind speed at the specific site within the month of failure occurrence. This can be seen as a measure of efficiency of the turbines' power production: $P_e = P_o / P_m$ with values between 0 and 1. Nonetheless, values slightly below 0 or above 1 can occur, as the reference power curve is only representing a mean value. Hence, values of $P_o/P_m \leq 0$ are set to 0 and values $P_o/P_m \geq 1$ are used as 1. Another approach to achieve site specific power curves could be accomplished by calibrating the manufacturer power curves to different control systems or site conditions. This was not done in the present study, as the information required for this process was not available, but could be subject to further studies. Furthermore, it has to be mentioned that power curves can change or degrade over the life-time of the wind turbine and assuming a constant power curve over the operational life of a WT is not always valid. However, as in the present study the average observation period is three years, the monthly mean power curve is assumed to be constant.

- *Downtime / Severity*: The downtime per failure is specified in the historical failure data and cross-checked with the SCADA data. It indicates the severity of each failure in terms of WT (un-)availability.
- *Maintenance Strategy*: indicates the availability of the maintenance personnel, which directly affects the repair time. It was classified after consulting expert judgement on O&M strategies; and the typical working hours of close-by located personnel, where the day shift stands for 8:00 to 18:00 and the night shift covers the remaining hours.

3. Methodology

A framework has been developed, capable of handling big environmental and historical failure data in order to quantify the impact of meteorological conditions on WT component failures. As visualised in Figure 1, it is divided into four steps: (1) data pre-processing, (2) data processing, (3) unsupervised rule mining and finally (4) ranking and interpretation of the rules. In the first step, the raw data are pre-processed and put into a usable form. The second step includes two distinct data mining techniques: (a) unsupervised k-means clustering and (b) supervised data labelling. Each technique is carried out independently with the same data set and is expected to influence the results differently. The output of either one of

them serves as input to the third step. Here, an association rule mining algorithm - called apriori ruling - is applied. The algorithm interconnects logically the environmental parameters and the component failures. The resulting rules, on the one hand, give an insight on the environmental conditions that have the highest impact on WT components failures; and, on the other hand, enable to evaluate for which input parameters which machine learning technique, (a) or (b), is more appropriate. The framework has been developed, so that in future works, parameters can be upgraded or changed easily.

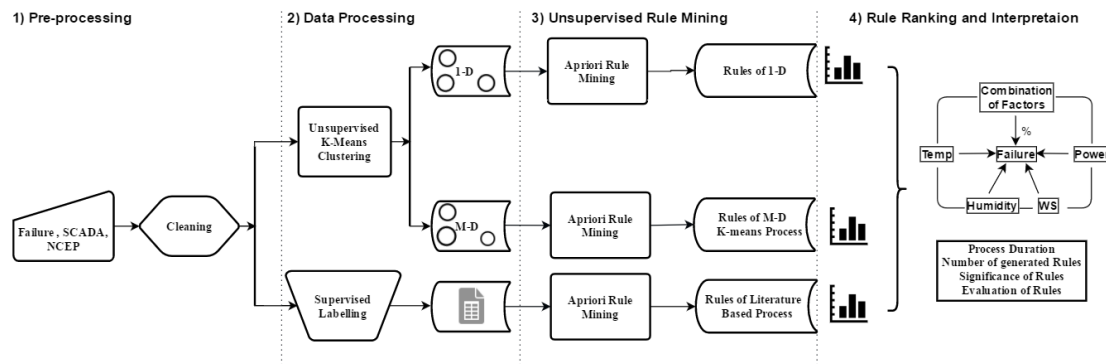


Figure 1: Visualisation of the developed framework and its sub-processes

3.1. Data Pre-Processing

In this step, the historical failure data were cleaned, in order to ensure that the indicated failure meets the herein used failure definition (causing a WT stop and intervention). Thus, remote wind turbine stops without any component failure or stops related to external factors, such as grid restrictions or annual inspections, etc. were excluded. Then the failures were assigned to the affected component, using a so called wind turbine taxonomy. A taxonomy is a component breakdown, which groups the assemblies in several stages according to their functionality and physical location. There are several sophisticated WT taxonomies available, such as the RDS-PP® [21], and the Reliawind taxonomy, e.g. Wilkinson et al. [22]. As the availability of the commercial taxonomies is restricted, many of the important players in wind energy, who do not belong to the VGB, have developed their own internal WT taxonomy.

In this study a recently presented modern WT taxonomy, Reder et al. [16], has been applied. This taxonomy was developed accounting for both modern and older WT technologies. It is based on information provided by big European wind

farm owners and previously presented WT taxonomies, in order to guarantee the comparability of the analysis to previous and future research projects. Further advantages of using this taxonomy are discussed in detail by Reder et al. [16].

The environmental parameters, as explained in Section 2, were extracted from the respective data sources. The analysis is very sensitive to missing values, which are usually represented computationally in the data as "not available" numbers, or *NAs*. Thus, only wind farms with complete SCADA history for the observation period were chosen for this analysis. The SCADA data were compared among closely located WTs to ensure the functionality of the SCADA systems. However, in the case of missing values, the latter could be imputed using one of the techniques described by Moritz et al. [23].

3.2. Data Processing

After having classified the failures regarding the WT taxonomy and having defined which conditions will be taken into account, this information needs to be processed. The aim of the data processing is to define thresholds for each variable and its values, in order to provide a suitable input to the apriori ruling process. As mentioned before, the data processing was carried out for each variable using two distinct learning based categorization techniques in parallel:

- (a) Supervised Labelling
- (b) Unsupervised K-Means Clustering
 - (b1) 1-Dimensional Input
 - (b2) Multi-Dimensional Input.

Supervised labelling (a) is a manual variable classification process that is expected to lead to highly biased results, as the classification is based on opinions and expert judgements, which in most cases are a result of costly experiments or profound experience in the field. The unsupervised k-means clustering (b) is a purely data-driven technique carried out with two different input dimensions: (b1) 1-dimensional input and (b2) multi-dimensional input, as explained in detail in Section 3.2.2. The unsupervised data mining certainly results in much higher computational effort, but is expected to provide better results than the supervised labelling. However, no expert judgements and/or experiments are needed as it is exclusively based on the input data, reducing the overall effort of this technique. A detailed description of both techniques is given in the following sub-sections 3.2.1 and 3.2.2 and their performance will be evaluated in Section 7.

3.2.1. Supervised Labelling

Supervised labelling is used to define thresholds and assign labels to the input parameters based on expert judgements and findings from literature. This process is conducted manually and requires profound knowledge about common parameter classification. In this work, for example the wind speed labelling was carried out using labels according to common literature findings and the typical cut-in and cut-out wind speeds of WTs. They were divided into calm, low, high and stormy wind conditions. Another example is the relative humidity (RH). The relative humidity can be labelled in terms of its resulting corrosiveness, as indicated for example in Jiang et al. [24], Leygraf et al. [25] and Xiang et al. [26]. Nonetheless, it has to be mentioned that this only represents the corrosiveness for one certain type of steel. As a WT system usually consists of many different types of materials, this is not entirely correct. Very costly experiments would be required in order to quantify the RH in terms of its corrosiveness for the entire system. Nonetheless, it is considered a good indicator and in the case study the RH will be described according to the resulting corrosiveness. Five labels, from dry air to highly corrosive and precipitation, were used to manually define the thresholds, one for each hourly value, as explained in Section 2.5. A detailed description of the assigned labels in the case study can be found in section 4.1. Also, the limitations of this technique will be discussed.

3.2.2. Unsupervised Clustering

Contrarily to the supervised labelling, the unsupervised clustering does not need or permit any manual intervention. It defines automatically clusters for each input parameter depending exclusively on their appearance in the data set. K-means is one of the most popular and commonly used clustering methods in unsupervised learning, and will be used in this work. As the outcome of this method depends on the input data, two input layouts were considered. This has been done in order to find the optimal way of analysing the correlation of failures and external conditions.

Multi-dimensional (Multi-D) input is the standard input format for k-means clustering, taking into account all time steps of the input parameters as one observation before a specific failure. This is done for each failure separately. This has been applied in a similar form in several studies, for example by Yildirim et al. [27], taking into account wind power and coordinates as one multi-dimensional input to k-means clustering.

One-dimensional (1-D) clustering takes into account all time steps of each of the parameters in the data set as one dimensional input vector. For example, all wind

speed values for each considered time step are included in one sequence, not distinguishing between the failures. This results in a one-dimensional grouping similar to the labelling, but without prior knowledge. Contrarily to Multi-D it takes each data entry as separate variable and allows to analyse them independently. This is illustrated in Figure 2, giving an example of the wind speed taken from a specific WT before failure. In this case, for example, the multi-dimensional input gathers all nine wind speed measurements as one observation to represent the wind conditions before failure. Thus, it is especially handy to analyse the general conditions over the observation period. Contrarily, one-dimensional inputs are representing one attribute, such as wind speed, with 9 observations (each 10 minute step) for all analysed turbines in one array. This can be used to analyse one of the time steps in more detail. In literature both forms are present, however, K-means algorithms are mostly used with multi-dimensional input data. The latter is especially useful for complex databases, but does not enable the analysis of a single time-step. Thus, in the presented framework, both techniques are included and their performance will be discussed in Section 4.

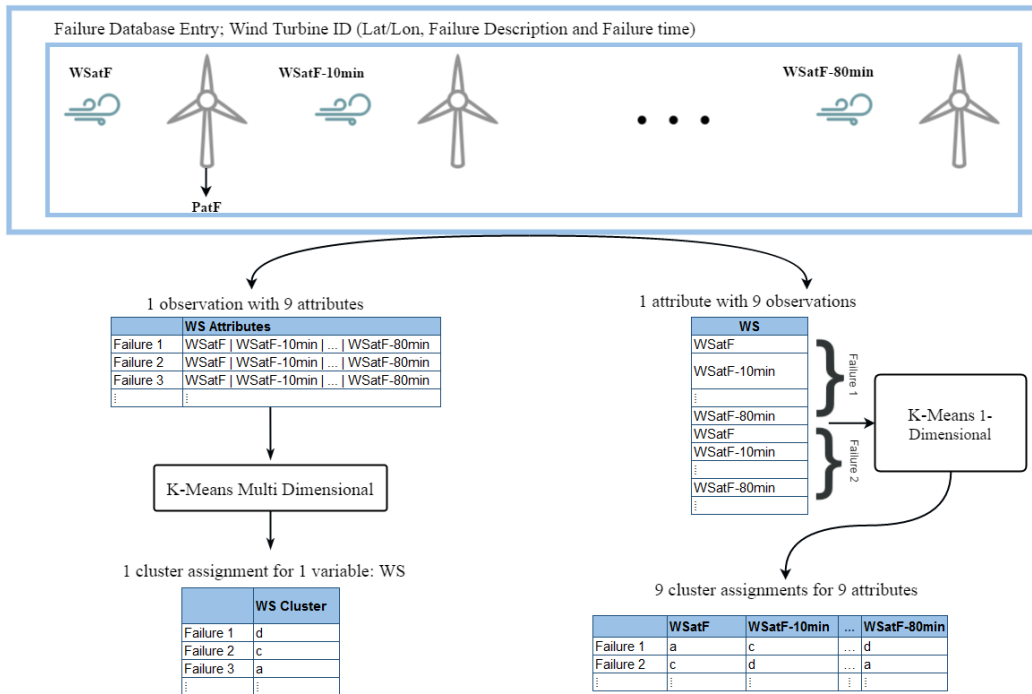


Figure 2: Multi-D and 1-D Data Pre-Processing, example for Wind Speed

A methodology developed by Yildirim et al. [28] using k-means clustering for apriori ruling in wind power analysis is adapted in this study for analysing the Multi-D and 1-D inputs. The main process starts with the following steps, which are carried out subsequently:

1. *Determine the Number of Clusters:*

Before the actual clustering, the number of desired clusters for each variable has to be defined. This is a common problem for partitioning methods such as k-means clustering, where k indicates the number of clusters for which the clustering is carried out. Thus, an algorithm has to be run firstly, in order to determine the optimal number of clusters for the given data set. There are around 30 methods and indices to define the optimal number of clusters, such as for example the elbow technique, silhouette and gap statistic methods. An overview over these techniques is given by Charrad et al. in [29] including an algorithm based on most of the 30 techniques that is capable of calculating the optimal grouping for a given data set, by varying all possible combinations of the number of clusters.

2. *K-Means for Clustering:*

Throughout the k-means clustering process, the data set is analysed with the aim of assigning each data point to one specific cluster. The clustering process consists of two steps that are repeated subsequently until the best result is obtained:

2.a. *Cluster assignment step:* After setting up the optimal number of clusters in the previous step, each data point is assigned to one of the cluster centroids, to which it is located closest.

2.b. *Move centroid step:* Then, the position of the centroids is moved to the mean location of all assigned data points.

3.3. *Association Rule Mining and Ranking*

Association Rule Mining (ARM) is a popular approach to find correlations, frequent patterns, or interesting relations among variables in data sets. The ARM algorithm chosen for this study is the apriori ruling algorithm. It uses prior knowledge to generate frequent patterns from the data set containing transactions, see e.g. Agrawal et al. [30].

3.3.1. Problem Definition and Terminology of the Association Rule Mining

The problem of association rule mining is defined as: finding all rules that are succeeding a pre-defined value of *support* and *confidence* in a given set of transactions. Thus, in a given database (set of transactions) the algorithm finds rules, which predict the occurrence of an item with respect to the occurrence of other items in one transaction. Further explanation can be found for example in [31]. For a better understanding the terminology used throughout this study will be discussed very briefly.

- *Group*: is indicating all data for one specific parameter, e.g. *WS*, *RH*, *T*, *Component*, etc.
- *Item (I)*: is a member of a group. This can be a measured value, such as $WS_{20} = calm$ and $WS_{80} = 3$; or a failed component, e.g. *Gearbox*.
- *Item Set*: is a collection of one or more items. A *Frequent Item Set* is an item set, which has a support value higher than the defined minimum support.
- *Transaction (t)*: is a combination of items of different groups. As stated in Yildirim et al. [28], associations are defined based on the transactions. Each transaction will generate at least one association rule.
- *Association Rule*: is an implication of the form *Item Set X* \rightarrow *Item Set Y*
- *Data Base (DB)*: is a set of transactions used as input to the case study, containing measure values for each time step as well as historical failures. For example: $DB = \{t_1, t_2, \dots, t_{146}\}$
- *Subset DB (SubDB)*: Is a part of the whole database. In this study the subsets are related to all the data recorded for a specific component, e.g. $SubDB_{Gearbox} \subset DB$ or $SubDB_{Generator} \subset DB$, etc.

In the following section the metrics, *support*, *confidence* and *lift* will be explained in more detail.

3.3.2. Process of the Association Rule Mining

The task of association rule mining consists of three main steps: (1) the generation of item sets, transactions and databases; (2) the counting phase (finding the combination of all frequent item-sets in the DB); and (3) the filtering phase (testing and generating the strongest rules among the item-sets). Subsequently

the final post-processing is carried out to evaluate and bench-mark the generated rules. The strength of each rule is rated by three metrics: *support*, *confidence* and *lift*. They are defined as shown in Table 2 for X (an item on the left side of the rule), Y (an item on the right side of the rule) and the total number of transactions N . *Freq* stands for the absolute frequency of appearance. The values of each metric have to be pre-defined, in order to determine the desired filter criteria for the strength of the rules.

Table 2: Input metrics and their formulations for the apriori ruling

Metric	Value	Formula
Support	0.3	$supp(X \rightarrow Y) = \frac{Freq(X \rightarrow Y)}{N} = Prob(X \wedge Y)$
Confidence	0.5	$conf(X \rightarrow Y) = \frac{supp(X \rightarrow Y)}{supp(X)} = Prob(X Y)$
Lift	-	$lift(Y \rightarrow X) = \frac{supp(X \rightarrow Y)}{supp(X)supp(Y)}$
Minimum Rule Length	2	-

The *support* metric is the primary measure in the evaluation of the rules in this paper. It can be seen as the frequency of occurrence for each association rule in relation to the total number of transactions. The logical conjunction *AND* (\wedge) is used to indicate the relation between items. At this point, it has to be underlined that there is a notation difference between the probability of an association indicating the *support* as ($Prob(Y \wedge X)$) and the conditional probability representing *confidence* as ($Prob(X | Y)$). This is explained in detail in [32].

Each rule has one support value, which must be greater than the pre-defined limit. In this manner the generated rules can be compared based on their *support* values. The *support* indicates the importance of the rule and depending on the input data, the number of analysed parameters and the desired results. It can hold any value between 0 and 1. For the verification of the framework with the presented case study, the minimal support was set to 0.3. This has shown to be the best compromise between the size of the input data, the number of resulting rules and the amount of information that can be drawn out of the latter. If two rules have the same *support* value, the rule with the higher confidence value will be taken as more important. Likewise is the application of the *lift* metrics in case that two rules have the same *support* and *confidence* values; see also Tan et al. [31]. Furthermore, the objective was to use the *support* metric as the only indicator for the

strength of the rules. This has several advantages and simplifies the evaluation process. In order to achieve this, the *confidence* and the *lift* of all rules needed to be equal to 1. With this, these two metrics lose their weight when comparing the rules among each other and the *support* is the only important comparison metric. As *lift* is frequently referred to as "interestingness", it indicates the correlation of different rules within a DB. A $lift > 1$ and $lift < 1$ indicate a positive or negative correlation between the rules, respectively. Whilst $lift = 1$ indicates independent rules and allows the other metrics to be more representative. The *confidence* metric behaves in a similar manner. This is achieved by dividing the *DB* into component related *SubDBs*. More information on how this has been realised can be found in the example given in Section 3.3.4, along with the methodology for calculating the ruling metrics.

Depending on the desired outcome, the ruling can lead to uni- or multi-conditional results. The metric that controls the outputs' conditionality is the rule length. The *Minimum Rule Length* indicates how many items have to be included in a rule, in order for it to be considered. The predefined minimum for the ruling process in this study is 2 - corresponding to at least one input and one output (in this case a failed component). For more input items the rule length increases and the rules become multi-conditional.

3.3.3. Example for the Interconnections and Transactions

Figures 3 and 4 show examples for three resulting rules using a rule length of 3 each. They are visualised in a similar manner as proposed in [33]. A rule length of 3 corresponds to two inputs and one output. The orange circles represent each association rule, the green ones stand for the input items. The blue circle is the output item for each rule - in the present case it is the *failed component*. Each figure contains more than one rule for the same output item. This was chosen in order to demonstrate that the same input item can contribute to various rules.

The arrows indicate the direction of the rule leading to the desired outcome. Thus, the construct green circle (*input items*) \rightarrow blue circle (*output item*) represents one rule. The numbers attached to the orange circles are related to the *support*. Figures 5 and 6 show examples for 10 resulting rules using a rule length of 2, indicating one input and one output item.

3.3.4. Example for the Evaluation Process - Ruling Metrics

In this section an example for the calculation of the evaluation metrics is given. This shall give an explanation on how the *support* value can be used as only rule metric, while setting *confidence* and *lift* equally to 1.

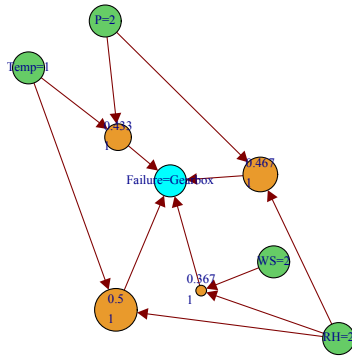


Figure 3: Multi-D K-means application to Gearbox Failures - rule length 3

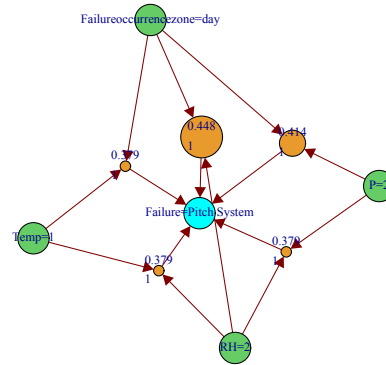


Figure 4: Multi-D K-means application to Pitch Failures - rule length 3

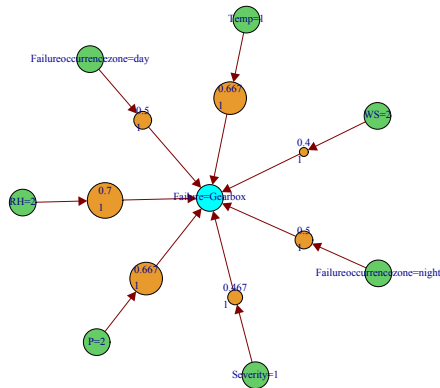


Figure 5: Multi-D K-means application to Gearbox Failures - rule length 2

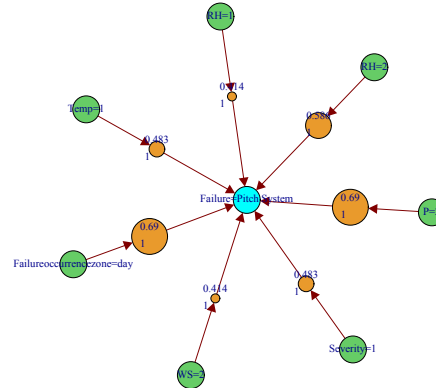


Figure 6: Multi-D K-means application to Pitch Failures - rule length 2

The evaluation of the rules (post-processing) is carried out in two steps: (a) find frequency of appearance of a combination of items, and (b) calculate and compare the evaluation metrics. Table 3 gives an example for the frequency of appearance of item sets in a DB with 149 failures of different components.

When analysing the whole DB, the item set (2) was found 21 times and item set (3) was found nine times. When considering only the part of *DB* that contains failures related to the gearbox, *SubDB_{GB}*, the ratio of appearance of item set (2) obviously changes. Thus, due to the smaller input it becomes equal to expression in the 6th row of Table 3. This effects the calculation of the evaluation metrics, as shown in Table 4. When using *SubDBs* as reference input, the *confidence* and *lift* metrics become equal to 1 and *support* can be used as sole indicator for the

Table 3: Counting appearances in DB(N=146) and $SubDB_{GB}$ (N=30)

Input	N	Item Set	Freq.
(1) DB	146	$Freq(Failure = GB)$	30
(2) DB	146	$Freq(severity = 3 \wedge T = 1)$	21
(3) DB	146	$Freq(severity = 3 \wedge T = 1 \wedge Failure = GB)$	9
(4) $SubDB_{GB}$	30	$Freq(Failure = GB)$	30
(5) $SubDB_{GB}$	30	$Freq(severity = 3 \wedge T = 1)$	9
(6) $SubDB_{GB}$	30	$Freq(severity = 3 \wedge T = 1 \wedge Failure = GB)$	9

strength of the rules.

Table 4: Calculating metrics for DB(N=146) and $SubDB_{GB}$ (N=30) Metrics

Input	Item	Supp.	Conf.	Lift
DB	$Prob((Failure = GB \wedge Sev = 3 \wedge T = 1) = 9/146)$	0.06	-	-
DB	$Prob((Failure = GB (Sev = 3 \wedge T = 1)) = \frac{(9/146)}{(21/146)})$	-	0.43	-
DB	$\frac{Prob((Failure=GB (Sev=3 \wedge T=1)) = \frac{(9/146)}{(21/146)})}{Prob(Failure=GB)=(30/146)}$	-	-	2.09
$SubDB_{GB}$	$Prob(Failure = GB \wedge Sev = 3 \wedge T = 1) = 9/30$	0.3	-	-
$SubDB_{GB}$	$Prob((Failure = GB (Sev = 3 \wedge T = 1)) = \frac{(9/30)}{(9/30)})$	-	1	-
$SubDB_{GB}$	$\frac{Prob((Failure=GB (Sev=3 \wedge T=1)) = \frac{(9/30)}{(9/30)})}{Prob(Failure=GB)=(30/30)}$	-	-	1

This enables a fast and very representative evaluation, as the support values can be converted to percentage of appearance directly. Furthermore, the pre-defined minimum confidence for the application of the framework of 0.5 is exceeded here-with. This is beneficial, especially for analysing different components and many environmental parameters.

4. Results and Discussions

The presented framework has been applied to a case study considering five main WT components. The results of each step of the framework will be presented in the following in order to evaluate its performance and analyse the weather conditions before component failures. The open source programming language R

[34] was used to perform the case study analysis as well as the comparison of the proposed methodologies.

4.1. Results of the Data Processing

Table 5 displays the thresholds for the manual labelling process, the minima and maxima for the clusters obtained for the one-dimensional (1-D) input data and the cluster centroids for the multi-dimensional (Multi-D) input data. It can be seen, that for some parameters, the number of defined labels differs highly from the number of clusters. Also, the thresholds are in some cases very different.

Table 5: Results for Labelling and Clustering with One-Dimensional (1-D) and Multi-Dimensional (Multi-D) Input

Minimum and Maximum Values for the Labelling									
Label	WS [m s ⁻¹]	Label	RH [%]	Label	P [-]	Label	Downtime [h]	Label	T [°C]
calm	< 3	dry air	20 - 40	consumption	$P_o < 0$	Minor	< 48	freezing	-10 - 0
low	3 - 10	moist air	40 - 60	not efficient (ne)	$0 \leq P_o < P_m$	Major	≥ 48	very cold	0 - 5
high	10 - 26	corrosive	60 - 80	efficient (e)	$P_o \geq P_m$			cold	5 - 10
storm	> 26	highly corr. precipitation	80 - 98 100					cool	10 - 15
								mild	15 - 20
								room temp.	20 - 25
								warm	25 - 30
								hot	30 - 35
								very hot	35 - 40

Minimum and Maximum Values for the Clusters obtained for the 1-D Input									
Cluster	WS [m s ⁻¹]	Cluster	RH [%]	Cluster	P [-]	Cluster	Downtime [h]	Cluster	T [°C]
1	0 - 6.568	1	32.46 - 62.3	1	0 - 0.27	1	27.5 - 144	1	3.34 - 10.65
2	6.62 - 12.98	2	62.84 - 79.19	2	0.3 - 0.7509	2	151.17 - 360	2	10.68 - 14.85
3	13 - 27.08	3	79.2 - 98.46	3	0.759 - 1	3	402 - 980.66	3	14.96 - 20.15
								4	20.44 - 30.13

Centroids of the Clusters obtained for Multi-D Input ¹									
Cluster	WS [m s ⁻¹]	Cluster	RH [%]	Cluster	P [-]	Cluster	Downtime [h]	Cluster	T [°C]
1	3.97	1	64.63	1	0.12	1	72.68	1	10.24
2	9.36	2	80.78	2	0.53	2	226.84	2	17.03
3	15.98			3		3	563.98	3	22.69

¹ As the multi-dimensional input also results in multi-dimensional clusters, the minima and maxima cannot be displayed. Thus, the cluster centroids are being displayed instead, giving an idea about the location of the clusters.

For example the severity measurement, downtime, was clustered by the k-means algorithms into three categories, whilst the labelling only took two thresholds into account: major (≥ 48 hours) and minor (< 48 hours) interventions. Thus, k means actually provides more information and should be used in this case. Temperature data was labelled into eight categories, while the k-means algorithms only defined three and four clusters. This means, that the labels might give too much information, which is not very handy for the interpretation of the results. For categorising power production efficiency, labelling and 1-D clustering have three categories, while Multi-D takes into account only two. For relative humidity, the

dimensions were reduced significantly by clustering. Wind speed was labelled according to the cut-in and cut-out WS of a common WT. Four labels were assigned, including one for stormy conditions. This number was slightly reduced by both clustering techniques. As only one storm condition was represented in the data set, k-means was not able to assign this to a separate cluster.

The characteristics for the maintenance strategy are given in Table 6. No clustering was applied to this parameter. After consulting expert opinions and taking into account the typical working hours of the close-by located maintenance personnel, it has been divided into two segments: day and night time.

Table 6: Maintenance Labels

Working hours	Label	Characteristics
8 AM to 6 PM	Day Shift	Failure detection is more likely and maintenance personnel is available.
6 PM to 8 AM	Night Shift	Failure detection is less likely and availability of maintenance personnel is limited.

The aim of this comparison is to find the best technique to process the input data for further use in the apriori ruling algorithms. As the input parameters are all clustered according to their appearance in the data, k-means clustering was expected to show better results than the highly biased labelling. Depending on the amount of available information for each parameter, the use of a combination of both techniques showed the best results, though. For parameters with highly available information the use of labelling outperformed clustering and *vice versa*.

4.1.1. Performance of 1-D and Multi-D K-means Clustering versus Labelling

Table 7 displays the characteristics that have been observed for each technique by applying the framework to the case study data set. It is shown clearly that the techniques perform differently for each of the desired purposes.

Data Analysis: Labelling represents the most complex and a highly time consuming pre-processing. As reliable expert judgements are often hard to obtain and evaluate, this involves a considerable amount of manual work. Nonetheless, after the data thresholds are defined, the data processing is much faster than for the other two techniques. For the multi-dimensional clustering the data did not have to be pre-processed extensively, however, the actual data processing is more time consuming. The effort for one-dimensional clustering is slightly higher than for the multi-dimensional data pre-processing, as the data have to be converted to

Table 7: Characteristics of the used Techniques

	Labelling	1-D Clustering	Multi-D Clustering
ANALYSIS			
Time Consumption Data Pre-Processing	high	medium	low
Time Consumption Data Processing	low	high	medium
Expert judgement needed	✓	✗	✗
POSSIBLE RESULTS			
Behaviour over Time	✓	✓	✗
Punctual Analysis	✓	✓	✗
General Conditions	✗	✗	✓
POST-PROCESSING			
Interpretation of Results	medium	difficult	easy
Number of Rules per Component	very high	very high	low
Number of obtained Rules in the Case Study	15783	12814	127

a one-dimensional array. The data processing for 1-D was the most time consuming, as the computational effort increases with the one-dimensional input.

Possible Results:. Also the way of extracting the possible results differs highly for each of the three techniques. Labelling performs similar to 1-D clustering and both can be used to analyse one specific time step (e.g. WS_{atF} , RH_8 , etc.) or to detect e.g. alternations by considering subsequent time steps over the observation period. The multi-dimensional clustering is especially handy for analysing the general conditions over the observation period. It treats the input as a whole and cannot detect efficiently one specific time step nor variations over time.

Post-Processing:. The post processing and the interpretation of the results depend strongly on the amount of obtained rules. Whilst the Multi-D clustering results in a fairly low amount of rules and interpretation is quite easy, the 1-D clustering showed the opposite behaviour. Labelling also resulted in a very high amount of rules, however, these are quite easy to interpret due to the manually defined thresholds. The quality of the results, has to be evaluated for each input parameter separately. For this, the pre-processing restrictions, discussed in Section 4.1, should be taken into account. Generally speaking, for fast analyses where the accuracy of the outcome does not play a significant role, Multi-D is a very efficient technique. It avoids the time and resource consuming labelling and does not require as much effort in data processing as 1-D clustering. Nonetheless, a general problem of the k-means clustering is that it only considers data that serves as input to the algorithm. Thus, the input has to be carefully chosen and the results must be analysed and interpreted thoroughly. Labelling, however, is in some cases too

biased and can result in a loss of information.

4.2. Results and Discussion of the Weather Effects on Component Failures

In this section, the most relevant results derived from the application of the presented framework to the available case study are discussed. As stated, the primary importance metric for the classification of the rules is the support (*supp.*) value. Multi-D labelling was used for discovering the general conditions for wind speed (*WS*), relative humidity (*RH*), power production (*P*), severity of failure and ambient temperature (*T*) throughout the whole observation period before each failure. Labelling and 1-D clustering were used to define the conditions at a certain point in time, that occurred with high frequency. These can be for example WS_{20} or WS_{atF} , etc. Also the variations of a parameter over time (alternations) were analysed using these two techniques.

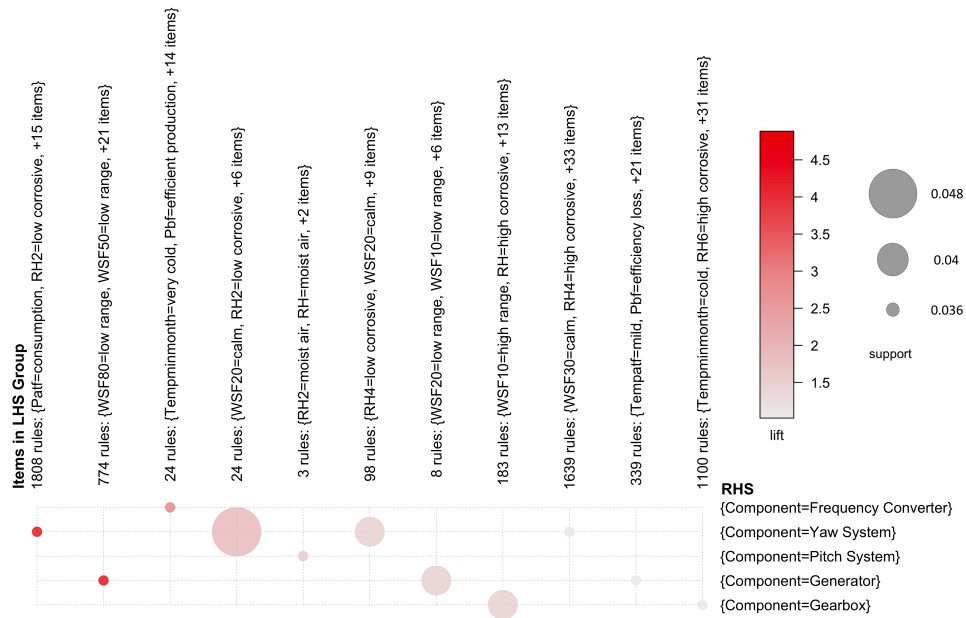


Figure 7: Grouped matrix for labelling rules with a minimum *support* of 0.03

Due to the overwhelming amount of obtained rules it is fairly hard to present or visualise all of them in a limited space. Figure 7 shows an example for a grouped rule matrix for the labelling results, which displays the cases of rule length ≥ 2 and a single input database for all observations. The grouped rule matrices for

1-D and Multi-D clustering are shown in Appendices A1-1 and A1-2. Since all the observations are taken into account in one single database, the lift metric has a meaningful scale in these figures.

These matrices are a compact form of displaying the results and shall be discussed at this point using the example of Figure 7. The rules on the right hand side (RHS) of the figure suggest the following interpretations:

- For the frequency converter the rules indicate with small support and medium lift values that it commonly fails during months with very low monthly minimum temperatures and that the turbine shows efficient production values just before the failure.
- The rules for the yaw system show strong support and lift values for failures under calm wind speed conditions 20 minutes before the failure occurrences and low corrosive relative humidity values 2 hours before the failure. The other two rules show a similar but weaker relation between failures and calm wind speeds. However, these weaker rules show association with high corrosive relative humidity 4 hours before failure.
- The pitch system tends to fail for the moist air at the failure and moist air 2 hours before the failure occurrences, indicated by rules with relatively small support and weak lift values.
- The rules for the generator failures show medium support and medium lift values for failures under low wind speed conditions 10 minutes and 20 minutes before failure. The association between temperature, power efficiency and the generator failures is observed with weak support and weak lift.
- The gearbox failures are associated to rules with with medium support and significant lift values for failures under high wind speed conditions 10 minutes before the failure and high corrosive relative humidity. The association between monthly minimum temperature, relative humidity 6 hours before failure and generator failure is given with weak support and weak lift.

Appendices A1-1 and A1-2 can be interpreted in a similar manner as discussed above. Nonetheless, these grouped matrices do not permit a quantitative and item based interpretation. Thus the results are presented in Table 8, which gives the most frequently obtained rule for each environmental parameter with each technique presented in percentage of their appearance in *SubDB*, as discussed in Section 3.3.4. The table can be interpreted in three ways:

1. *Technique-based*: For analysing the possible results that can be extracted from each of the three techniques it is possible to compare the results row-wise for labelling, 1-D and Multi-D clustering for each component.
2. *Component/Item-based*: For inspecting one environmental parameter and comparing its influence on different components, the table can be interpreted column-wise. This allows extracting the general parameter conditions (Multi-D) and the condition at a specific time step (Labelling and 1-D) that most frequently appeared in the data.
3. *Combinations*: If the combinations of several parameters before component failure are of interest, all rows related to the component can be combined and analysed.

Table 8: Results for the three Approaches - Rules with highest Support Value (*supp.*)

	WS	supp.	RH	supp.	Power	supp.	Severity	supp.	Temp.	supp.	
Gearbox	Label	$WS_{atF} = \text{low}$	46%	$RH_{atF} = \text{highly corr.}$	60%	$P_{atF} = \text{ne}$ $P_{bf} = \text{ne}$	57%- 46%	Major	96%	$T_m = \text{cool}$	60%
	1-D	$WS_{60} = 2 \text{ (max: 3)}$	50%	$RH_2 = 3 \text{ (max: 3)}$	63%	$P_{atF} = 1 \text{ (max: 3)}$ $P_{bf} = 3 \text{ (max: 3)}$	43%- 67%	1	46%	$T_{atF} = 1 \text{ (max: 4)}$	53%
	Multi-D	$WS = 2 \text{ (max: 3)}$	40%	$RH = 2 \text{ (max: 2)}$	70%	$P = 2 \text{ (max: 2)}$	67%	1	46%	$T = 1 \text{ (max: 3)}$	66%
Generator	Label	$WS_{atF} = \text{low}$	53%	$RH_8 = \text{highly corr.}$	70%	$P_{atF} = \text{ne}$ $P_{bf} = \text{e}$	67%- 50%	Major	96%	$T_{min} = \text{cold}$	46%
	1-D	$WS_{30} = 2 \text{ (max: 3)}$	50%	$RH_8 = 3 \text{ (max: 3)}$	70%	$P_{atF} = 1 \text{ (max: 3)}$ $P_{bf} = 3 \text{ (max: 3)}$	40%- 80%	2	50%	$T_{atF} = 2 \text{ (max: 4)}$	30%
	Multi-D	$WS = 2 \text{ (max: 3)}$	46%	$RH = 2 \text{ (max: 2)}$	73%	$P = 2 \text{ (max: 2)}$	90%	2	50%	$T = 1 \text{ (max: 3)}$	46%
Converter	Label	$WS_{30} = \text{low}$	55%	$RH_{atF} = \text{highly corr.}$	48%	$P_{atF} = \text{ne}$ $P_{bf} = \text{e}$	78%- 48%	Major	81%	$T_{max} = \text{mild}$	48%
	1-D	$WS_{atF} = 1 \text{ (max: 3)}$	52%	$RH_{atF} = 3 \text{ (max: 3)}$	48%	$P_{atF} = 2 \text{ (max: 3)}$ $P_{bf} = 3 \text{ (max: 3)}$	48%- 81%	1	81%	$T_{atF} = 1 \text{ (max: 4)}$	63%
	Multi-D	$WS = 1 \text{ (max: 3)}$	51%	$RH = 2 \text{ (max: 2)}$	55%	$P = 2 \text{ (max: 2)}$	85%	1	81%	$T = 1 \text{ (max: 3)}$	51%
Pitch	Label	$WS_{atF} = \text{high}$	58%	$RH_8 = \text{highly corr.}$	62%	$P_{atF} = \text{ne}$ $P_{bf} = \text{e}$	79%- 35%	Major	86%	$T_m = \text{cool}$	37.9%
	1-D	$WS_{20} = 2 \text{ (max: 3)}$	51.7%	$RH_8 = 3 \text{ (max: 3)}$	62.1%	$P_{atF} = 1 \text{ (max: 3)}$ $P_{bf} = 3 \text{ (max: 3)}$	48%- 69%	1	48.3%	$T_{atF} = 4 \text{ (max: 4)}$	34%
	Multi-D	$WS = 2 \text{ (max: 3)}$	41%	$RH = 2 \text{ (max: 2)}$	58.6%	$P = 2 \text{ (max: 2)}$	69%	1	48.3%	$T = 1 \text{ (max: 3)}$	48.3%
Yaw	Label	$WS_{atF} = \text{low}$	63%	$RH_2 = \text{low corr.}$	60%	$P_{atF} = \text{ne}$ $P_{bf} = \text{e}$	53%- 43%	Major	70%	$T_{max} = \text{mild}$	47%
	1-D	$WS_{20} = 1 \text{ (max: 3)}$	70%	$RH_{atF} = 2 \text{ (max: 3)}$ $RH_{10} = 3 \text{ (max: 3)}$	56% 56%	$P_{atF} = 2 \text{ (max: 3)}$ $P_{bf} = 3 \text{ (max: 3)}$	33%- 63%	1	90%	$T_{atF} = 1 \text{ (max: 4)}$	36%
	Multi-D	$WS = 1 \text{ (max: 3)}$	60%	$RH = 2 \text{ (max: 2)}$	70%	$P = 2 \text{ (max: 2)}$	70%	1	90%	$T = 1 \text{ (max: 3)}$	53%

It should be noted, though, that much more information could be drawn out of the results. However, due to limited space, not everything can be discussed here.

Additionally, there are more environmental conditions, such as ice and snow that can be responsible for certain failures. These have not been considered directly in this analysis, but enter indirectly considering the temperatures and seasons. For easier interpretation, the results for 1-D and Multi-D clustering in Table 8 are always showing the maximum number of clusters for each parameter in brackets.

4.2.1. Results for the Labelling and 1-D Clustering - Environmental Conditions at Time of Failure

In this section the environmental conditions at the exact time of failure (punctual analysis) are discussed. As discussed before, they were obtained with the labelling and 1-D k-means clustering algorithms. Figures 8 (a) to (c) show the obtained results for supervised labelling and apriori ruling for the parameters relative humidity, wind speed and temperature at the time of failure occurrence. Figures 8 (d) to (f) represent the same results obtained from the 1-dimensional clustering and apriori ruling. These plots show the failure frequency, which indicates the total number of failures assigned to each of the clusters or labels for the different components. It can be seen that the number of clusters differs in some cases significantly from the numbers of assigned labels. If the amount of input data is rather modest, the use of both techniques on parallel can provide the best results. As shown in Table 7 the effort for 1-D clustering is lower than for labelling. This can lead to significant advantages when using very large databases.

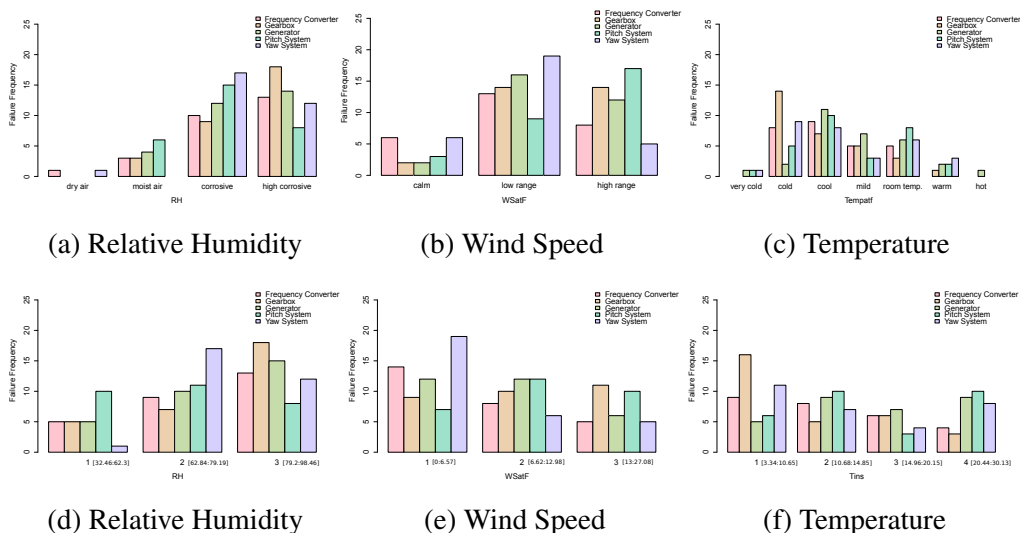


Figure 8: Conditions at Failure obtained for Labelling (a)-(c) and 1-D clustering (d)-(f)

All components showed an increased number of failure occurrence with higher relative humidities. Especially the frequency converter, the generator and the gearbox showed significantly higher numbers of failures in corrosive environments. Labelling and 1-D clustering showed similar results for this parameter, although the 1-D clustering only assigned 3 clusters. Nonetheless, the output is consistent when combining the labelling results for dry and moist air. It should be noted, that the case study data contained very few failure observations in dry air. Thus, the unsupervised clustering could not assign these to a separate cluster. Only analysing the clustering results, might lead to the assumption that quite a high number of failures occurred at low RH (cluster 1). However, it is important to compare this to the labelling results, where dry air is listed as a separate class. Hence, it can be seen that most failures occurred at least in moist air or higher relative humidities.

From an engineering perspective high wind speeds represent highly distributed loads on the WTs. Naturally, failure occurrences are expected to be in correlation with high and critical loads, hence, high wind speeds. Albeit, the results demonstrate a different trend. High wind speeds did not seem to affect all components. This is also connected to the design specific operational wind speed range of the respective wind turbine technology. In the results for the labelling technique it can be seen that only the pitch system suffered from a higher number of failures in the presence of high wind speeds. The other components showed more failures with wind speeds in the lower range. This has to be combined with the results of the long term wind speed conditions in Section 4.2.2. This seems contradictory to previous studies where failures are frequently related to high wind speeds. It has to be stated though, that for this case study only 80 minutes of wind speed data ahead of failure were considered. It is supposed that failures occur with a certain delay and this short time span is not representative enough, as the peak wind speed is not always reached within this observation. In further application of the framework the observation period should be extended as the wind speed conditions immediately before failure seem to be lower. Nonetheless, it states clearly that the pitch system is more likely to suffer immediately from failures in the presence of high winds.

Also, when comparing the labelling and 1-D clustering results, it can be seen that the failure distribution in Figure 8e is way more balanced among the different wind speed categories than in Figure 8b. This is, e.g. due to the nature of the unsupervised clustering, which tries to balance the number of data points per cluster, see also Section 4.1. Thus, the labelling showed more detailed results.

The temperature at failure time showed to be mainly in the range from cold to mild for most components. Especially the gearbox showed many failures in the pres-

ence of cool and cold temperatures, This was the case for both, Labelling and 1-D clustering. Using the labelling technique, provides the possibility to distinguish further between the different temperature thresholds.

4.2.2. Results for the Multi-D Clustering

Using Multi-Dimensional clustering allows finding general environmental conditions over a longer period of time and throughout the whole observation period ahead of failure, see Table 8.

Corrosive conditions were recorded before the failures of all components.

Gearbox failures showed high support values for cool mean monthly temperatures T_m . The failures occurred mostly during the winter months and high relative humidities were recorded. The wind speeds at failure were not significantly high.

The generator showed failure occurrences at mostly low wind speeds and temperatures at failure in the lower range (cluster 1) and a mean monthly temperature in the cold range. However, before failure occurrences the main wind speed conditions were in cluster 2 - being elevated velocities.

The converter showed failures mostly with mild monthly mean temperatures. and in highly corrosive conditions at failure. Mostly low wind speeds were recorded,, both, at failure and prior to failure (WS was in cluster 1).

Pitch failures mainly occurred in the presence of higher wind speeds at failure, and cool monthly mean temperatures, as well as highly corrosive surroundings. In high and especially changing wind speed conditions the pitch system has to intervene more often, and is thus, more likely to fail.

Yaw system failures occurred mainly at low wind speeds and cool monthly mean temperatures. The WTs use the yaw system to find the main leading wind direction in order to place the rotor in a favourable position for the energy production. Especially in low wind speed this is a quite difficult task and the yaw system is constantly employed, and hence, more likely to fail.

4.2.3. Analysing Changing Environmental Conditions over Time using 1-D Clustering and Labelling

After having analysed the weather conditions directly at failure occurrence (Section 4.2.1) and the general ones during the observation period before the failure (Section 4.2.2), now their behaviour throughout the observation period will be analysed. As it is not very handy to simply plot the time series before each failure, for this purpose, the same framework was applied with the condition of extracting the probability of occurrence of a certain label or cluster for each avail-

able time step. The results have been evaluated and combined in order to test if the conditions were alternating or steady throughout the observation period. As an example of the capabilities of the proposed framework, the wind speed and relative humidity alternations will be presented here. As discussed before, the Multi-D clustering did not perform well for punctual analyses and will not be used in this part. The labelling and the 1-D clustering based frameworks will be used. However, as mentioned in Section 4.1 both perform differently for different parameters. In order to analyse the wind speed, for which extensive previous information is available, labelling was used as data processing technique. For analysing the relative humidity a mixture of labelling and the 1-D-clustering algorithm was employed. In this case, previous knowledge about the harmful RH thresholds for WT's is limited and can only be assigned manually based on findings from other areas. Nonetheless, clustering reduced the number of categories too much and needs to be cross-checked with the assigned labels in order to draw useful conclusions. The temperature analysis is carried out using 1-D clustering, as it showed to reduce the number of input categories to a reasonable extend. Figure 9 summarizes the percentages of appearance for the characteristics of the wind speed time series before each component failure. It is distinguished between steady, indicating no significant changes in the thresholds; alternation between different thresholds; as well as descending and ascending behaviour towards the time of failure. Especially the pitch system and the gearbox showed an elevated number of failures in the presence of wind speeds that were steadily in the low or high range. Alternations in wind speed seemed to play a role for the generator and yaw system. The frequency converter rather showed higher numbers of failures exposed to wind speeds with steadily low or ascending behaviour. It has to be mentioned that 10-minute mean wind speeds are used for this analysis and wind speed alternations in higher frequency ranges are not captured, but could be affecting the components drastically. This will be subject to further studies.

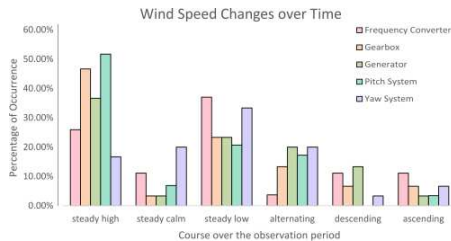


Figure 9: WS conditions over time obtained with supervised labelling

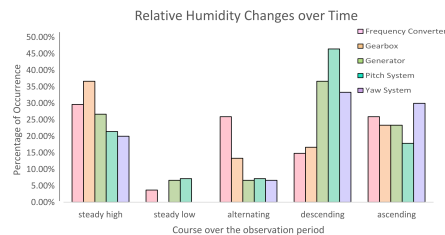


Figure 10: RH conditions over time obtained from 1-D clustering

Figure 10 shows the behaviour of the RH parameter before component failures. In general, most failures occurred in steady corrosive conditions or those conditions where the relative humidity was ascending towards high corrosive states. Descending relative humidity might be misleading, as it still takes into account the RH being in the (highly) corrosive range at least 10 hours before the failure occurrence. Thus, most components are affected by high relative humidity. The temporal changes in RH are not as significant as they are for WS.

4.2.4. Seasonal Failure Occurrence

Figure 11 presents the failure events within each season. All components showed an increase in failure frequencies during the winter months. This is probably due to higher wind speeds, lower temperatures, and other factors like snow and icing. Especially the gearbox seemed to be affected during the winter, which could be related to poor performance with lower lubrication oil temperatures.

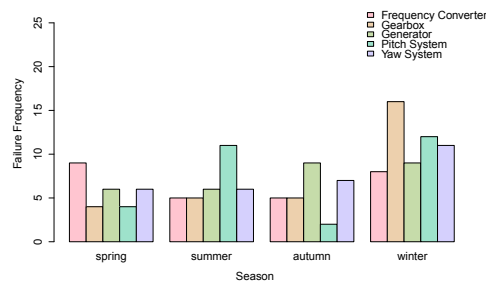


Figure 11: Number of failures per season and component

4.2.5. Energy Production Losses

As expected, in most cases the power production shortly before failure did decrease below the reference value for efficient production at the specific wind speed. This is especially interesting for failure detection, as the power production is a parameter, which is easy to measure in real time. The mean monthly power production at wind speed in the same wind farm during the month in which the failure was detected, was taken as reference to determine if the production is efficient (e) or not efficient (ne). The manufacturer's power curve did not represent well the on-site conditions and monthly variance and could not be used as reliable representation of the expected power production of the specific WTs. In general, labelling performed well to indicate, if the power production was efficient or not. 1-D clustering performed better in terms of stating to what extend the

power production was inefficient. Here, the production was assigned to three clusters according to the percentage of expected power output. Nonetheless, 1-D clustering did not distinguish between productions of 76%-100% in cluster 3, mixing efficient and slightly inefficient production. Multi-D clustering only showed that the general power production conditions were in cluster 2. The power production is directly related to wind speed. A power production in cluster 2 means that most failures occur around an efficiency of 50 % corresponding to a wind speed in low to medium range. When looking at the wind speed distributions this is actually the threshold with the highest probability of appearance. Hence, this result makes sense when comparing a longer period of time before failure. Nonetheless, it does not provide any specific information on the power production conditions before the failure occurrence. Table 8 shows an inefficient power production during the last 10 minutes for all components. The converter and pitch system showed the highest percentage of failures with an anticipated loss in energy production. The gearbox was the only component that showed an inefficient production also in P_{bf} . As indicated by 1-D clustering, the pitch system, gearbox and generator showed especially low efficiencies in the last 10 minute interval before failure. This could be indicating a gradual degradation before the failure.

To sum it all up, for analysing the efficiency of power production a mixture of labelling and 1-D clustering gives the best results. Labelling can be used for obtaining if the power production is efficient or not; 1-D clustering can be used to analyse then to what extend it was not efficient.

4.2.6. Downtimes and Availability of Maintenance Personnel

The failure severity in this study was categorised by the hours of downtime they have caused. As expected, the framework identifies the gearbox and generator as the most severe component in terms of downtime. The downtimes are strongly affected by the applied maintenance strategy and the availability of maintenance personnel at failure occurrence. Pitch system failures mainly occurred during the day shift (in 79% of the cases). As the occurrence of pitch system failures is strongly correlated to high wind speeds, this is most likely due to the diurnal wind speed pattern and its peak during the day. The frequency converter failures occurred in 65% of the cases during the night shift. By looking at the diurnal pattern of the relative humidity, this could be explained due to a rise in relative humidity during the late evening and night time. Also lower temperatures during the night time might be affecting this behaviour. For the generator and gearbox the failures occurred equally distributed between day and night shift.

5. Conclusions

A framework has been presented that is capable of correlating and analysing failure data and environmental conditions ahead of wind turbine component failures. It has been shown that it can be applied to several environmental parameters and for various analysis purposes. Three different data pre-processing techniques are employed. Either one is being followed by an apriori ruling process, where rules are established that are interconnecting the different parameters and calculating the probability of occurrence of their combinations.

The pre-processing techniques, (a) supervised labelling, (b1) unsupervised 1-D clustering and (b2) unsupervised Multi-D clustering performed differently for the different input parameters and output requirements. For obtaining a quick insight on the general conditions over the observation period, the Multi-D clustering based technique has shown very good results. It is fast and the interpretation of the results is quite easy, although, the number of resulting rules and their accuracy is lower compared to the other two techniques. Supervised labelling and unsupervised 1-D clustering are capable of analysing the environmental conditions before failures in more detail. Thus, the conditions at a certain time-step within the observation period, such as at the exact time of failure, can be extracted (punctual analysis). Also, the behaviour of each weather parameter over the whole observation period before failure can be observed by applying these two techniques. With this, it is possible to detect if the parameters alternate or are rather stable throughout an extended time before failure. Nonetheless, certain differences between these two approaches have been highlighted. Whilst 1-D clustering requires less pre-processing, labelling needs time extensive and costly expert judgements to establish the parameter thresholds accurately. Labelling performs best, when extensive knowledge on the parameters behaviour is available, such e.g. for wind speed. However, it can introduce a bias if the expert judgements are limited. If there are very few information for defining the parameter thresholds, as for example for RH, the 1-D clustering shows a better performance.

The performance of the framework was very satisfying and the spectrum of possible results is big. It can be applied to a variety of tasks in failure analysis, detection and prevention; and could significantly decrease O&M and condition monitoring expenses.

Acknowledgements

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agree-

ment No. 642108.

References

- [1] S. Martorell, A. Sanchez, V. Serradell, Age-dependent reliability model considering effects of maintenance and working conditions, *Reliability Engineering & System Safety* 64 (1) (1999) 19–31.
- [2] S. Martorell, M. Villamizar, S. Carlos, A. Sánchez, Maintenance modeling and optimization integrating human and material resources, *Reliability Engineering and System Safety* 95 (12) (2010) 1293–1299.
- [3] S. Carlos, A. Sánchez, S. Martorell, I. Marton, Onshore wind farms maintenance optimization using a stochastic model, *Mathematical and Computer Modelling* 57 (7-8) (2013) 1884–1890.
- [4] F. Ding, Z. Tian, Opportunistic maintenance for wind farms considering multi-level imperfect maintenance thresholds, *Renewable Energy* 45 (2012) 175–182.
- [5] L. XiaoFei, L. Min, Hazard rate function in dynamic environment, *Reliability Engineering & System Safety* 130 (2014) 50–60.
- [6] G. Van Kuik, J. Peinke, R. Nijssen, D. Lekou, J. Mann, C. Ferreira, J. W. V. Wingerden, D. Schlipf, P. Gebraad, H. Polinder, A. Abrahamsen, G. J. W. V. Bussel, P. Tavner, C. L. Bottasso, M. Muskulus, D. Matha, H. J. Lindeboom, S. Degraer, O. Kramer, S. Lehnhoff, M. Sonnenschein, P. E. Morthorst, K. Skytte, Long-term research challenges in wind energy a research agenda by the European Academy of, *Wind Energy Science* 1 (1) (2016) 1–39.
- [7] B. Hahn, Zeitlicher Zusammenhang von Schadenshäufigkeit und Windgeschwindigkeit, in: *FGW-Workshop Einfluss der Witterung auf Windenergieanlagen*, 1997.
- [8] P. Tavner, C. Edwards, A. Brinkman, F. Spinato, Influence of Wind Speed on Wind Turbine Reliability, *Wind Engineering* 30 (1) (2006) 55–72.
- [9] S. Faulstich, P. Lyding, P. Tavner, Effects of wind speed on wind turbine availability, in: *European Wind Energy Conference*, Brussels, Belgium, 2011.

- [10] Á. M. Costa, G. Roshan, J. A. Orosa, Á. Rodríguez-Fernández, Case Study of Weather Maintenance in Wind Power Generation, *Arabian Journal for Science and Engineering* 39 (7) (2014) 5615–5624.
- [11] P. Tavner, D. M. Greenwood, M. W. G. Whittle, R. Gindele, S. Faulstich, B. Hahn, Study of weather and location effects on wind turbine failure rates, *Wind Energy* 16 (2) (2013) 175–187.
- [12] M. Wilkinson, T. Van Delft, K. Harman, The Effect of Environmental Parameters on Wind Turbine Reliability, in: *European Wind Energy Conference*, Copenhagen, Denmark, 2012.
- [13] G. Wilson, D. McMillan, Assessing Wind Farm Reliability Using Weather Dependent Failure Rates, *Journal of Physics: Conference Series* 524 (1) (2014) 1–10.
- [14] A. Kusiak, A. Verma, A Data-Driven Approach for Monitoring Blade Pitch Faults in Wind Turbines, *IEEE Transactions on Sustainable Energy* 2 (1) (2010) 87–96.
- [15] M. D. Reder, J. J. Melero, Assessing wind speed effects on wind turbine reliability, in: *Wind Europe Conference*, Hamburg, Germany, 2016.
- [16] M. Reder, E. Gonzalez, J. J. Melero, Wind Turbine Failure Analysis - Targeting current problems in Failure Data Analysis, *Journal of Physics: Conference Series - The Science of Making Torque from Wind* 753 (2016) 1–11.
- [17] G. Wilson, D. Mcmillan, G. Ault, D. Bank, Modelling the Effects of the Environment on Wind Turbine Failure Modes using Neural Networks, *UK Chinanet SuperGen Conference*.
- [18] M. Reder, J. Melero, Time series data mining for analysing the effects of wind speed on wind turbine reliability, in: *Safety and Reliability Theory and Applications*, CRC Press Taylor & Francis Group, 6000 Broken Sound Parkway NW, Suite 300 Boca Raton, Florida USA, 2017, pp. 109–117.
- [19] M. Kanamitsu, W. Ebisuzaki, J. Woollen, S.-K. Yang, J. J. Hnillo, M. Fiorion, G. L. Potter, Ncep/doe amip-ii reanalysis (r-2), *Bull. Amer. Meteor. Soc.* 83 (2002) –12.

- [20] M. Kemp, E. van Loon, J. Shamoun-Baranes, W. Bouten, Rncp: global weather and climate data at your fingertips, 2012, *Methods in Ecology and Evolution* 3 65–70.
- [21] VGB-PowerTech, VGB-Standard RDS-PP Application specification Part 32: Wind energy, Tech. rep., VGB-PowerTech, Essen, Germany (2014).
- [22] M. Wilkinson, K. Harman, B. Hendriks, F. Spinato, T. Van Delft, Measuring wind turbine reliability, results of the reliawind project, in: *Measuring wind turbine reliability, results of the reliawind project*, European Wind Energy Association Conference, 2011, pp. 1 – 8.
- [23] S. Moritz, A. Sardá, T. Bartz-beielstein, M. Zaefferer, J. Stork, Comparison of different Methods for Univariate Time Series Imputation in R.
- [24] G. Jiang, J. Keller, P. L. Bond, Determining the long-term effects of H₂S concentration, relative humidity and air temperature on concrete sewer corrosion, *Water Research* 65 (2014) 157–169.
- [25] G. O. Lloyd, Atmospheric corrosion, in: M. Radojevic, R. Harrison (Eds.), *ACIDITY - Sources, Consequences and Abatement*, Elsevier Science Publishers Ltd, 2000, pp. 1–8.
- [26] Y. Xiang, Z. Wang, Z. Li, W. Ni, Long term corrosion of X70 steel and iron in humid supercritical CO₂ with SO₂ and O₂ impurities, *Corrosion Engineering, Science and Technology* 48 (5) (2013) 395–398.
- [27] N. Yıldırım, B. Uzunoğlu, Spatial clustering for temporal power ramp balance and wind power estimation, in: *International Conference on Cyberworlds (CW)*, Gotland, Sweden, 2015, pp. 224–228.
- [28] N. Yıldırım, B. Uzunoğlu, Data Mining via Association Rules for Power Ramps Detected by Clustering or Optimization, in: L. M. Gavrilova, K. C. J. Tan, A. Sourin (Eds.), *Transactions on Computational Science XXVIII: Special Issue on Cyberworlds and Cybersecurity*, Springer Berlin Heidelberg, 2016, pp. 163–176.
- [29] M. Charrad, N. Ghazzali, V. Boiteau, A. Niknafs, NbClust: An R package for determining the relevant number of clusters in a data set, 2014, *Journal of Statistical Software* 61 (6) 1–36.

- [30] R. Agrawal, R. Srikant, Fast Algorithms for Mining Association Rules, in: Proceedings of the 20th International Conference on Very Large Databases, 1994, pp. 487 – 499.
- [31] P.-N. Tan, M. Steinbach, V. Kumar, Association Analysis: Basic Concepts and Algorithms, in: Introduction to Data mining, Addison-Wesley, 2005, Ch. 6, pp. 327–414.
- [32] M. Hahsler, A Probabilistic Comparison of Commonly Used Interest Measures for Association Rules, 2015, accessed on 15/12/2016.
URL http://michael.hahsler.net/research/association_rules/measures.html
- [33] M. Hahsler, S. Chelluboina, arulesViz: Visualizing Association Rules and Frequent Itemsets, 2015, r package version 1.1-0, accessed on 15/12/2016.
URL <https://CRAN.R-project.org/package=arulesViz>
- [34] R Core Team, R: A Language and Environment for Statistical Computing, 2015, R Foundation for Statistical Computing, Vienna, Austria, accessed on 15/12/2016.
URL <https://www.R-project.org/>

Appendix A. Results visualised as Ruling Plots

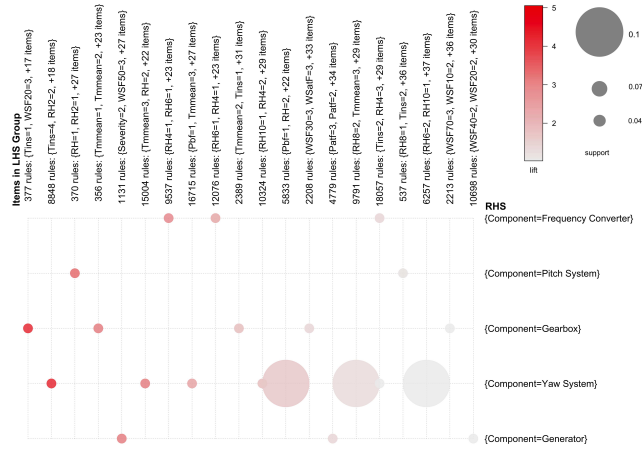


Figure A.12: Grouped matrix for 1-D clustering rules with a min. *support* of 0.03

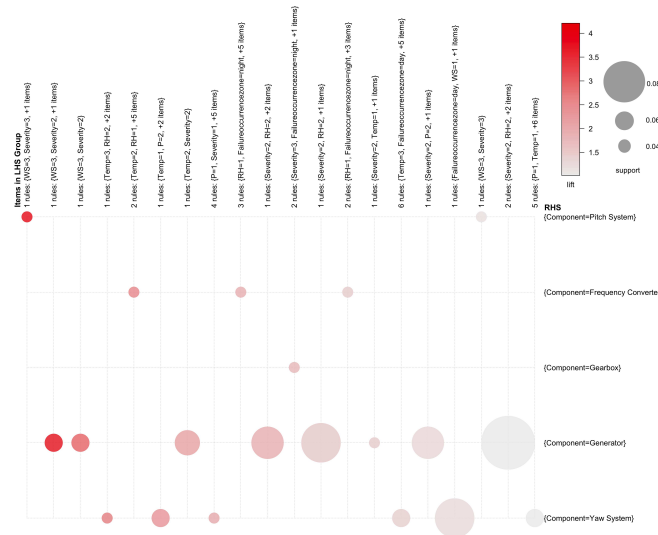


Figure A.13: Grouped matrix for Multi-D rules with a min. *support* of 0.03