A Literature Review of Time-Series Clustering Techniques and Machine Learning Techniques Used for Monitoring of Wind Turbines

Written by
Yohann Jacob Sandvik

Abstract

Here the abstract will be.

Contents

Chapter 1

Introduction

Clustering is a method used to categorize big amounts of data into groups known as *clusters*, when there is little, or no information available about the underlying groups. It is a popular choice for extracting patterns from large datasets, because clustering falls within the category of *unsupervised machine learning*, meaning that it does not require the dataset to be labelled. Clustering is also a common step in data mining algorithms, where the goal is to learn rules relating the different variables in a dataset. In the last decade clustering has started to become more common to use on time-series datasets as they are abundunt, and labeling is often costly and time-consuming. Time-series clustering has been applied on financial time series, medical time series and time series from a variety of other industries.

1.1 Motivation

As of 2018 wind power, together with solar power made up 7% of the worlds electricity production,[1] and has been referred to as "the fastest growing source of energy" by the Norwegian company Statkraft[2]. As the effects of climate change steadily are becoming a reality shifting to renewable energy sources is imperative, and wind power will certainly play a bigger part in meeting the worlds energy demand in the future.

To make wind power as a whole more lucrative, a good start would be to reduce the downtime, and improve the performance of turbines. The argument that time-series clustering may be a good approach for this is two-fold.

1. A single wind turbine can have several houndred sensors sampling up to every second, meaning that a wind farm can produce colossal amounts of time-series data. An unsupervised approach is usefull simply because labelling of all this data is cumbersome.
2. When wind farms become big enough it will become to costly to manually inspect every turbine to construct an effective model for condition monitoring, further automation is required [1]. Time-series clustering is then a good alternative for condition monitoring.

1.2 Objective

This literature review has three objectives in the form of questions.

[1] `https://www.iea.org/geco/electricity/`

[2] `https://www.statkraft.com/globalassets/old-contains-the-old-folder-structure/documents/wind-power-aug-2010-eng_tcm9-11473.pdf`

Objectives

1. What machine learning methods are currently being used to monitor the condition, and performance of wind turbines?
2. What are the different methods of model based time-series clustering currently used?
3. What time-series clustering methods (if any) are appropriate to test on time-series data produced by wind turbines?

The literature review is meant to be a preliminary work for a master thesis where select techniques will be evaluated on actual time series data produced by a wind farm in Norway. The project assigment is also a continuation of the master thesis written in the spring of 2019 by Espen Waaga. In his thesis he explored the effectiveness of clustering raw time series in regards to similarity in time, and shape. In his "Future work" section Espen Waaga suggests that a natural next step would be to look at clustering with regards to similarity in change, which is done by implementing a model-based approach.

1.3 Structure of Review

Chapter 2

Method

2.1 Search terms

To find the relevant literature on the subjects of interest the search engine Oria was used to search the university library of the NTNU. Oria was preferred over other search engines such as Google Scholar because Oria allows one to combine multiple search terms in unison using "AND" or "OR", and because it allows the user to choose whether the search term should be in the title, subject, or other parts of the articles. The review will only consider articles published in peer-reviewed journals. Table 2.1 summarizes the search results. The *Title* and *General content* columns show which terms were used in the different searches; which terms where required to be in the title, and which terms could be in the "general content", meaning any part of the article. Let "×" represent the AND operator between two search terms, and "∧" represent the OR operator. The "*" operator means that the search will include any word beginning with the word before the star, e.g *detect** includes *detection*, *detecting*, *detected*, etc. The N_f and N_i columns show how many results each search yielded and how many articles from each search were included in the review, respectively.

Nr.	Title terms	General terms	N_r	N_i
1	time × series × clustering	None	219	121
2	(monitor* ∧ detect*) × clustering	time × series	187	21
3	wind × turbine* × (monitor* ∧ detect*) × review	None	32	3
4	wind × turbine* × (monitor* ∧ detect*)	machine × learning	100	47
		Total number of articles included		193

Table 2.1: Search results

2.2 Screening method

To make sure that the articles used were relevant, the review is limited to articles published 2014 or later. Some older articles are included through backward snowballing for their historical importance. There were three levels of screening, screening of the title, abstract, and full article. Title-screening was primarily for seeding out duplicate articles returned from the search-engine. The screening of the abstract and full-article were to identify the articles that were not relevant for the review and exclude them.

It has been a challenge to include enough literature to get a good overview of the different methods within time series clustering, but also not more literature the author alone could handle in the time available. So although the objectives is to get an overview of the different time series clustering methods, and an overview of the current machine learning methods used for monitoring wind turbines, the author does not claim to have made a complete exhaustive summary of all the possible methods.

When screening articles from search one and two, articles meeting one (or more) of the following criteria were discarded:

- Primary goal is time-series prediction/forecasting. It is outside the scope of this assignment.
- Uses subsequence time-series clustering methods. It is outside the scope of this assignment.
- Time-series clustering is used only as a minor preprocessing step. Not considered relevant enough to the objectives of the review.
- Data used is not time series data. Not considered relevant enough to the objectives of the review.
- Paper does not actually use clustering algorithms. Not considered relevant enough to the objectives of the review.
- The data used consists of image time series. Data to different from data to be used in master thesis.
- The time-series clustering methods explored in the work are **not** model-based or feature-based. The raw-data-based approach has been somewhat covered in Espen Waaga's master thesis, so work using this appraoch will be emmitted in this review.
- The specific model-based approach used measures the tail dependence of time series. Method not found relevant enough for the data that will be used in the master thesis, more relevant for financial time series.

Search number three was used to find existing literature reviews on condition monitoring of wind turbines. Three good literature reviews on the subject where found, and one good review on machine learning methods used for condition monitoring of wind turbines was found in search four. So, when screening the remaining articles from search number four the focus was to find articles not included in the aforementioned reviews, to complement them as well as possible.

Chapter 3

Wind Turbine Monitoring

3.1 Wind turbine Components

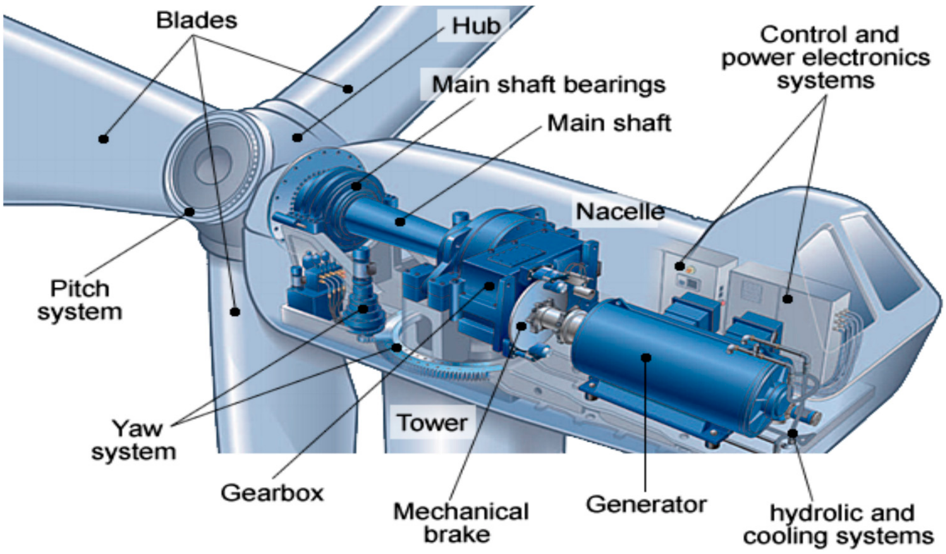

Figure 3.1: Illustration of the different parts of a wind turbine, taken from [2]

Figure 3.1 shows the main parts of a wind turbine which includes the rotor (blades and hub), shafts, gearbox and generator. Simplified a wind turbine works by wind pushing the blades, generating torque that makes the hub rotate. The hub is connected to the gearbox through the main shaft. The gearbox then gears down the torque and gears up the rotational speed to a level that the generator can use to induce current, that goes to a station that transforms the voltage to a level that can be used in the electrical grid.

3.2 Sensors and Data Acquisition

Information about a wind turbine can come from many sources, it can come from external sources such as images from a camera, or from internal sensors measuring operational data. The collective term for systems measuring operational data is supervisory control and data acquisition (SCADA) systems. To choose what algorithm to use, or what model to use, one must first consider what data one has available. From the literature considered, these where the most used forms of data used as input for the model.

- Vibration measurement
- Acustic emission monitoring
- Temperature measurement
- Power signal measurement
- Oil debris monitoring
- Strain monitoring
- Optical fiber monitoring
- Ultrasonic testing
- Image analysis

Analysis of vibration signals is the most common form of condition monitoring used in industry for any form of rotating equipment [3]. By measuring the acustic emission generated by a component of a wind turbine, one can estimate how much damage it has obtained. The temperature of components in a wind turbine is closely correlated with the health of the component, and is therefore used often in condition monitoring applications [4]. The power signal can also say a lot about how well a wind turbine is performing, specifically the wind speed - power curve. When monitoring the debris in the oil of a wind turbine gearbox one is analysing the size, type and number of wear particles present in the lubricant, as they can indicate the degree of damage in the gearbox [5]. Strain monitoring, optical fiber monitoring, ultrasonic testing, and image analysis are all used to detect structural damage in different components of the wind turbine, usually the blades [6, 7, 8, 9, 10, 11], or tower [12]. The most common approach however was to use a combination of multiple sensors-values to make predictions about the condition about the wind turbine.

3.3 Machine learning techniques

A machine learning is a subset of artificial intelligence. Machine learning models that extract rules from data, which can then be applied to classify, or estimate components of another dataset. Machine learning algorithms are formally divided into *supervised learning*, *unsupervised learning* and *semi-supervised learning.* Supervised learning models require labelled datasets to extract information from the dataset, and are usually used to perform classification tasks, or to estimate a variable that is considered dependent on the input variables (regression). Unsupervised learning algorithms do not require labelled datasets. Semi-supervised learning uses a combination of labelled and unlabelled datasets.

3.3.1 Feature extraction

There are two central problems in condition monitoring that can be solved by feature extraction and selection. The first is the sheer volume of information being produced. A wind turbine with only 20 sensors, sampled at 100 Hz will produce 170 MB of information per day. Feature selection is used here to reduce the number of features to only those relevant for condition monitoring. The second problem is that for systems using only one signal such as vibration, there are many components that are superposed to create the measured signal, and noise is present. Feature extraction is used to separate the interesting components from each other, and cancel the noise. For some machine learning models feature extraction is not a neccesary

Extraction method	Articles
ARMA models	[13, 14, 6, 9, 15]
Discrete Wavelet Transform (DWT)	[14, 16, 17, 18]
Principal Component analysis (PCA)	[6, 19, 9, 18, 11, 20, 21]
Basic signal statistics	[22, 18, 23]

Table 3.1: Feature extraction methods

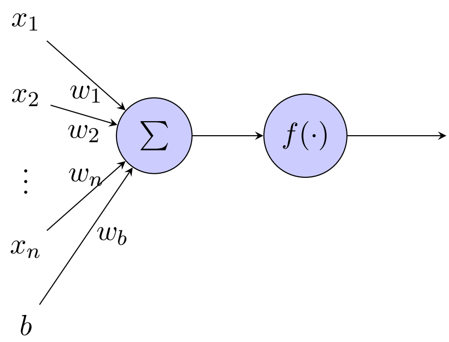

Figure 3.2: Perceptron

preprocessing step, for others careful though must be given as to how to extract features. Table 3.1 shows the most frequent methods found in the articles.
The DWT is a method used for decomposing time-dependent signals. It is a better alternative than other time-frequency decomposition tools, such as the Fourier transform (FT), for non-stationary signals. However, the FT is also used [14, 22]. Basic signal statistics refers to values easy to calculate over a fixed window size such as the root-mean-square (RMS) value, min/max value, etc. PCA is an unsupervised machine learning form used in multivariate systems. It produces the linear combinations of variables that has the highest variance, also called the *principal components*.

3.3.2 Regression-based models

A frequent approach is to use a supervised learning model to capture the normal behaviour of a wind turbine (or wind turbine component) by predicting the value of one time dependent variable, and then using the deviation between the prediction and the actual value (Estimation error E_e) to detect anomalous behaviour. What varies in these approaches is what machine learning model they use to model the wind turbine. The input data used in this approach is most often SCADA data. The target variable is either a particular signal which is excluded from the input signals, or the target is to recreate the input signal as accurately as possible. The latter approach is called an *auto-encoder*. Three curves in particular hold a lot of information about the performance of a wind turbine, namely the wind speed - active power curve, wind speed - rotor speed curve and wind speed - blade angle pitch curve. The majority of the regression based models, used a subset of these curves as target variables [24, 25, 26, 27, 28, 29, 30]. Another popular approach is to forecast the temperature in the gearbox, or generator windings [31, 30, 32, 33, 34, 4]. As mentioned before temperature is closely correlated with the health of a component, but since temperature changes so slowly it is hard to use temperature monitoring alone for fault prediction [12]. If one however can make a model capture the complex sources of temperature change, the deviation between predicted temperature, and actual temperature could be used for fault prediction. Table 3.2 shows the typical machine learning models used for regression.

Machine learning model	Articles
Neural networks (NNs)	[28, 35, 30, 32, 33, 34, 36]
Gaussian process (GP) regression	[24, 25]
Support vector regression (SVR)	[26, 27, 29]
PCA	[20]
K-nearest neighbours (KNN)	[26]
Random forest (RF)	[26]
Network of restricted Boltzman machines (RBMs)	[31, 4]

Table 3.2: Machine learning algorithms used by normal behaviour models

Straczkiewicz and Barszcz [35] train a NN to estimate the vibration in a gearbox using gearbox oil temperature, wind speed, rotor speed and active power as an input. By performing linear regression on the estimation error of the NN they are able to detect early states of damage in a gearbox, months before it is replaced. Rodriguez-Lopez et al. [32] compare different NN using SCADA data from 14 wind turbines monitored over several years, trained to estimate the temperature of the bearing on the non-drive end of a generator. They are able to detect failueres within 2 months of occurance. Mazidi et al. [30] propose a model of three different NN that estimate the rotor speed, gearbox temperature, and generator winding temperature using SCADA data. The estimation error for the NNs are then forwarded to a proportional hazard model which sets a dynamic threshold for anomalous behaviour. NN are by far one of the best models to capture complex non-linear relationships between input variables, and in contrast to the kernel methods such as GP and SVR they are less relient on data being preprocessed, and features being extracted from the data beforehand. Instead what features a NN is able to extract is decided by the size, and complexity of the archtitecture. One of the disadvantes of using NNs is that the require a lot of training data to be accurate compared to other. The amount of training data required by a NN is also decided by the complexity of it's architecture. Gonzalez et al. [26] chooses not to use NNs in their approach because they are prown to overfitting, instead they compare three other approaches of SVR, KNN and RF to estimate the active power. They found that the RF regressor trained with high frequency SCADA data produced the best results when trained on short periods. Tao et al. [27] use Grey correlation algorithm for eigenvector extraction and genetic algorithms for feature selection and an SVR to estimate all three performance curves.

Wang et al. [4] use a *Deep Belief Network* (DBN) that consists of two layers of RBMs, and an output layer. It uses the DBN to predict the gearbox main bearing temperature, and sets a threshold for E_e to detect anomalies. Zhao et al. [31] also uses a network of RBMs, but use them an auto-encoder, and then uses the reconstruction error R_e to detect an anomaly. DBN have a great potential for both regression and classification tasks, but as with NN their performance is highly dependent on their architecture, and the data used to train them. Yang, Huang, and Yang [36] use an auto-associative NN as an auto-encoder, and then use the Hotel T_2 statistic as a dynamic threshold for R_e. The use of dynamic thresholds is very

3.3.3 Supervised classification-based models

Machine learning model	Articles
Neural network based	[7, 8, 37, 10, 38, 39]
Tests several individual classifiers	[13, 6, 16, 40, 15]
Uses fusion / ensamble of classifiers	[14, 9, 41]
Support Vector Classifier (SVC)	[22, 42, 17, 43, 18, 23]
Hierarchical Extreme Learning Machine (H-ELM)	[44]
DBN	[45]
Hidden Markov model (HMM)	[46]

Table 3.3: Machine learning algorithms used by supervised classification models

Shihavuddin et al. [7], Qiu et al. [8], Reddy et al. [37], and Zhang, Lu, and Wang [10] use images as input, and different types of NN for classifying structural damage in the blades. Chen et al. [39] uses SCADA data as input, and deep NN for detecting ice on the blades. The testing, and comparing of several individual classifiers is also quite popular, however the most popular classifier by far is the SVC. The SVC has also been used for detecting damage in the blades [22], general fault detection in the turbine [43], but most for gearbox faults [42, 17, 18, 23]. Kim and Yun-Ho [46] use a statistical model based on HMMs. They combine multiple vibration signals measured by a condition monitoring system with thresholds, and set different alarm levels. These alarm levels act as the observations for the HMM. The states of the model are *Normal* and *Fault*. The model is trained with sequences of observations, which it uses to determine which state is most probable of producing said sequence of observations.

- Explain why so many articles are bunched together in ”several individual classifiers”
- Explain why H-ELM won’t be explained here.
- Explain fusion of classifiers.
- State that an explanation of HMMs will be given later.

3.3.4 Unsupervised classification-based models

The classification models have been split into supervised and unsupervised classifiers because, the unsupervised methods are of greater interest for this review. The use of unsupervised learning methods are not as widespread as the use of supervised learning methods. Stetco et al. [47] only included one article in their review that compared a regression model based on feed forward neural networks to two unsupervised models using gaussian mixture models and self organizing maps. It should be noted that the articles using an unsupervised learning approach that are included in this review, generally were published after [47].

Machine learning model	Articles
RBMs	[48]
K-means clustering	[21]
One-class SVM (OCSVM)	[11]
Multiway PCA	[19]

Table 3.4: Machine learning algorithms used by unsupervised classification models

Zhang and Ma [21] is the first implementation only implementation found using time-series clustering used for condition monitoring of wind turbines. They first use paralell factor analysis (PARAFAC) for dimensionality reduction, which is a generalisation of bilinear PCA, and then use K-means clustering on the reduced feature space. In their approach they are able to identify distinct operation modes of the wind turbines, and reduce data redundancy by PARAFAC. However they mention that there is still further work that can be done, especially in testing their model with data from turbines operating at different wind speeds. Yang, Liu, and Jiang [48] use a spatiotemporal pattern network for feature extraction, and then use unsupervised stacked RBMs for anomaly detection. Wang et al. [11] uses images as input, and the hidden layers of a convolutional NN trained on an unrelated dataset to extract features, and then compresses the features using PCA. The compressed features are then fed into an OCSVM which classifies faults. Pozo, Vidal, and Salgado [19] uses multiway PCA to set up a baseline, and then uses hypethis testing to determine whether the power curve of a wind turbine is an anomaly or not. A summary of the articles related to machine learning methods for condition monitoring of wind turbines can be found in 7.2.

3.4 Discussion

The supervised classification models seem to be the most popular ones, however regression based models are a close second. The supervised classification models show great promise in terms of accuracy, but are not of that much interest for this review, since the wind turbine data that will be used in the spring is not labelled with faults occurance, or anomalous behaviour. The regression based models don't require explicit labeling to model normal behaviour, so they could be used to detect some anomalous behaviour. What is problematic with using a complex regression based model such as a NN or DBN, is that without labelled data it might be hard to validate the performance of said model, because it often hard to determine **why** a particular observation is regarded as an anomaly. However, what can be transferred from these papers are the different methods for feature extraction, and selection. ARMA models, and HMMs are frequently used for time signal analysis, and will be expanded upon in section 4.3. PCA, and generalizations of PCA is also a valuable tool for dimensionality reduction, and will also be expanded upon, in section SECTION. What is preferable with ARMA models, HMMs, and PCA compared to NNs and DBNs is that they are easier to interpret, which is valuable when using unlabelled data.

It is promising to see that K-means clustering paired with PARAFAC showed such good results. Since the model developed by Zhang and Ma [21] still needs to be tested on wind turbines in more variable conditions (in terms of wind speed), there is still room for exploration on this topic. The use of RBMs and OCSVM for unsupervised anomaly detection is very interesting, but suffers the same problem as regression based models: without labelled data it is hard to validate the models, and interprate the anomalies they detect. The work done by Pozo, Vidal, and Salgado [19] strengthens the argument for exploring PCA, and generalizations of PCA for feature extraction, and selection.

Chapter 4

Time Series Clustering

Time-series clustering (TSC) can be divided into three components. How the time series is represented, which metric is used to measure similarity between time series, and what algorithm is used to cluster a set of time series. This chapter will give a short description of what specific type of TSC that this report will consider, it will go through the aforementioned steps of a TSC system, give an in-depth description of some time-series models and finally show some metrics used to evaluate a TSC system.

In this report we will deal with *discrete time series.* A time series is defined as a set of observations $\{x_t\}$ recorded at a specific time t. A discrete time series is a time series where the set of times when observations are made (T_0) is discrete [49]. A multivariate time series can be viewed as a set of vectors $\{\mathbf{x}_t\}$ where each set of vector elements $\{x_t^i\}$ is an individual time series. This means that the elements of the same vector $[x_t^1, x_t^2, ..., x_t^N]$ are separate observations made at the same time instance t. In a wind turbine, measurements of the temperature in the gear box made every 10 seconds can be considered a univariate time series. While the set of measurements made every 10 seconds of the temperature in the gear box, the power produced by the turbine, and the wind speed ahead of the blades can be considered a multivariate time series.

4.1 Types of Times-Series Clustering

There are three types of TSC, *whole-series TSC*, *subsequence TSC* and *time-point TSC*. Whole-series TSC is when multiple "whole" time series are clustered with respect to each other. Subsequence TSC comprises the clustering of subsequences of the same time series with respect to each other. The defining difference between whole-series and subsequence TSC is that whole-series TSC clusters multiple time series while subsequence TSC clusters clusters different subsequences of the same time series. When performing time-point TSC the goal is to cluster individual observations of a time series wrt. to each other. In this review we will only consider work using whole-series TSC, so when the phrase *time-series clustering* is used, one can assume that whole-series TSC is what is being refered to.

4.2 Representation Methods

There are numerous ways which a time series can be represented. Aghabozorgi, Shirkhorshidi, and Wah [50] define a time series representation given time series data $\{x_t\} = \{x_1, x_2, ..., x_T\}$ as transforming the time series into another vector $\{x_t\} = \{\hat{x}_1, \hat{x}_2, ..., \hat{x}_L\}$ where $L < T$. In theory $L = T$, but most often the point in transforming a time series is to reduce the amount of information present in the raw time series to more easily reveal patterns of interest. According to Aghabozorgi, Shirkhorshidi, and Wah [50] and Philippe Esling [51] representation methods can broadly be categorized into four categories:

- **Non-data adaptive**: Here the parameters of the transformation are the same for every time series [51]. Spectral transformations such as the Discrete Fourier Transform (DFT) and the Discrete Wavelet Transform (DWT) fall within this category.
- **Data adaptive**: The name here implies that the parameters of the transformation such as the length of the time series can change for every specific time series [51]. These methods focus on approximating a given time series in the best possible manner by reducing the global reconstruction error [50].
- **Model based**: These representations attempt to fit a stochastic model to the data [50]. Examples of models include Autoregressive Moving Average (ARMA) models, and Hidden Markov Models (HMM).
- **Data dictated**: This representation method differs from the aforementioned methods in as much that it automatically chooses the compression ratio based on the characteristics of the raw time series [50].

Data adaptive representations will be better at approximating time series than non-adaptive methods, but since transformation parameters can change for different time series it is harder to compare time series [50]. What is worth noting about model based representation methods is that they make assumptions about the underlying process that is generating the time series data [51]. Hence, it can be a good way of integrating a priori knowledge about the time series into the clustering system. In the sub-section below there is a more in-depth description of the two common time series models ARMA models, and HMM.

4.3 Time Series Models

To select suitable mathematical models for a dataset, we have to allow for the random nature of future observations. This is done by assuming that each observation in a time series x_t is a realization of a particular random variable X_t. The time series can then be modelled as a collection/set of random variables $\{X_t\}$, also known as a *stochastic process* [49].

To define an ARMA model, one needs to have a clear understanding of the terms white noise process, and stationary process. We say that the stochastic process $\{Z_t\}$ is "white noise" with zero mean, and variance σ^2 ($\{Z_t\} \sim WN(0, \sigma^2)$) if and only if $\{Z_t\}$ is zero mean, and every random variable contained in $\{Z_t\}$ is uncorrelated with every other random variable contained in $\{Z_t\}$. A stochastic process $\{X_t\}$ is said to be weakly wide-sense stationary if the mean, and variance are constant for all terms in the process. $\{X_t\}$ is said to be weakly short-term stationary if the mean and variance of terms are constant for distinct time periods within the duration of the process, but are not constant for all terms in the process. For brevity the term "stationary process" will be used when refering to a *weakly wide-sense stationary process*.

4.3.1 Autoregressive Moving Average Models

An ARMA model descirbes a time series in terms of difference equations. It can be considered a combination of two smaller models, an autoregressive (AR) model and a moving average (MA) model. Let $\{X_t\}$ be a stationary process. An MA(q) model will describe every term X_t as a linear combination of q distinct white noise terms as in equation (4.1).

$$X_t = Z_t + \theta_1 Z_{t-1} + ... + \theta_q Z_{t-q} \tag{4.1}$$

Whereas an AR(p) model will describe every term X_t as a linear combination of p previous terms of $\{X_t\}$ as in equation (4.2)

$$X_t = \phi_1 X_{t-1} + ... + \phi_p X_{t-p} \tag{4.2}$$

Putting equations (4.2) and (4.1) together, an ARMA(p, q) model will describe every term X_t as a linear combination of p previous terms, and q white noise terms as in equation (4.3).

$$X_t - \phi_1 X_{t-1} - ... - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + ... + \theta_q Z_{t-q} \tag{4.3}$$

Given that the polynomials $1 + \theta_1 z + \theta_2 z^2 + ... + \theta_q z^q$ and $1 - \phi_1 z - \phi_2 z^2 - ... - \phi_p z^p$ have no common factors [52].

4.3.2 Hidden Markov Models

Let $\{X_n\}$ be a stochastic process where the random variables contained in $\{X_n\}$ only can take on a finite number of values which we will call states. Let X_n denote the state at time period n. The probability of X_n transitioning from state i to state j at the next time period $n + 1$ is denoted p_{ij}. It might seem natural that p_{ij} is conditional on what the state was in the last time period. $\{X_t\}$ is said to be a *Markov chain* if p_{ij} only is conditional on the last past state, as shown in equation (4.4).

$$\begin{aligned} p_{ij} &= P(X_n = i | X_{n-1} = i_{n-1}, X_{n-2} = i_{n-2}, ..., X_1 = i_1, X_0 = i_0) \\ &= P(X_n = i | X_{n-1} = i_{n-1}) \end{aligned} \tag{4.4}$$

Suppose now that the states that the process is in are hidden from the observer. Instead there exists a finite set of signals $\{S\}$ that are emmitted when the process enters a state. In addition, let the probability of emmitting signal s, at time period n, in state j ($P(S_n = s | X_n = j)$) be independent of previous states, and signals emmitted. A model of this type where the signals $S_1, S_2, ...$ are observed, and the underlying Markov states remain hidden is called a *hidden Markov chain model* [53].

4.4 Similarity Metrics

One similarity metric is euclidean distance. Another similarity metric is dynamic time warping.

4.5 Clustering Algorithms

As mentioned before clustering is a form of unsupervised machine learning. The goal is to divide the dataset into clusters, by maximizing some similarity metric for members of the same cluster, and minimizing the same metric for members of different clusters.

4.6 Evaluation Indices

Chapter 5

Discussion

Chapter 6

Conclusion

Chapter 7

Appendices

7.1 Machine learning for wind turbine condition monitoring

Ref.	Input	Feature extraction method	Machine learning model
[13]	Vibration signal.	ARMA model and J48 decision tree	Tests a set of (38) meta-, misc-, rule- and tree-based classifiers for fault detection in blades.
[31]	SCADA data.		Deep auto-encoder made of Restricted Boltzman Machines (RBMs) to model normal behaviour of SCADA variables (gearbox and generator temperature). Uses E_e for anomaly detection, with adaptive threshold set using extreme function theory (EFT).
[48]	SCADA data.	Spatiotemporal pattern network	Unsupervised use of RBMs for anomaly detection.
[14]	SCADA data.	Time-frequency domain analysis, DWT and ARMA model	Uses fusion of several classifiers for fault detection in a wind turbine.
[6]	Ultrasonic testing.	Tests linear and non-linear PCA and ARMA models	Neighbourhood component analysis for feature selection. Tests 20 different supervised classifiers for detecting ice on blades.
[24]	Wind-Power curve.		Uses Gaussian Process regression with EFT to determine whether a particular power curve is an outlier.
[22]	Acustic emission.	FFT.	Uses Distinguishability Measure for feature selection, and logistic regression and SVC for binary blade fault classification.
[44]	Power signal, wind speed and ambient temperature.		Hierarchical Extreme Learning Machine (H-ELM) for detection of anomolous behaviour.
[25]	SCADA data.		Gaussian processes regression to estimate wind-power curve
[26]	SCADA data.		Tests KNN, random forest, and SVR to estimate power curve. Detects anomalies by E_e.
[54]	Vibration signals.		Uses unsupervised dictionary learning extracting features which are then used to determine fault in drivetrain bearings.
[35]	Oil temperature, wind speed, rotor speed and active power.		Trains ANN to estimate vibration signal, uses E_e for anomaly detection.
[19]	SCADA data.	PCA.	Sets up baseline model using multiway PCA, then finds outliers by hypothesis testing whether multivariate distribution is equal to baseline.

[16]	FAST wind turbine simulator.	Image texture analysis tools.	KNN, Linear Discriminant Model, decision trees, bag-tree, linear SVC.
[36]	SCADA data.	K-means for outlier elimination.	Uses Auto-Associative Neural Networks as an auto-encoder, and the Hotel T2 statistic as a dynamic threshold for the R_e.
[27]	SCADA data.	Grey correlation algorithm for eigenvector extraction.	Use genetic algorithm for feature selection, and SVR for estimating performance curves (active power, rotor speed and blade pitch angle).
[30]	SCADA data.		Uses three NN for normal behaviour modelling of rotor speed, gearbox temperature and generator temperature. E_e sendt to proportional hazard model which sets dynamic threshold.
[40]	SCADA data.		Uses Inductive Transfer Learning and five differen ML classifiers for ice detection on blades.
[42]	Vibration signals.	Variational mode decomposition (VMD).	Uses multi-scale permutation entropy (MPE) used for feature selection, COVAL for domain normalization and an SVC for binary fault classification.
[29]	SCADA data.		Uses bins and SVR to estimate blade angle pitch curve.
[32]	SCADA data.		Tests different architectures of ANNs to for estimating temperature of non-drive end bearing. Uses E_e for anomaly detection.
[7]	Images taken by drones		Recurrent Convolutional Neural Network to classify structural damage in blades.
[8]	Uses images taken from ground level		Convolutional Neural Network, and YOLO-based small object detection approach (YSODA) for damage detection in blades and hub.
[17]	Vibration signals, acustic emission and oil particle analysis	DWT.	Uses decision tree for feature selection, and SVC for assessing fault severity in gearbox
[37]	Uses images taken from ground level		Convolutional Neural Network to detect cracks & damage in blades.
[9]	Ultrasonic testing	PCA and ARMA models.	Neighbourhood Component analysis for feature selection and an ensamble of KNN, linear SVC, decision trees, LDA and subspace discriminant to estimate amount of dirt and mud on blades.

[43]	Uses pitch position, rotor speed and generator speed.		Detects faults with an SVC with parameters optimized by Cuckoo-swarm optimization.
[15]	Vibration signals	ARMA model.	Dominating features selected with J48 decision tree, fault classification done with Bayesian- and lazy classifiers.
[18]	Vibration signals, acustic emission and oil particle analysis	DWT and PCA.	Dominating features selected with decision tree, fault detection done with SVC.
[10]	Images taken from imaging array		Uses a deep neural network for binary classification of blade defects.
[11]	Uses images taken by drone	Uses a CNN trained on an unrelated image dataset to extract general features.	Compress features with PCA, and pass them to a unsupervised one-class SVM.
[41]	Uses the FAST wind turbine simulator to get SCADA data.	Random forest.	Uses XGBoost to train an ensamble of classifiers for specific faults.
[33]	SCADA data.		Uses several ANN to build a normal behaviour model of temperature in gearbox and high speed shaft, then uses E_e together with the age of the age of the turbine to predict anomolous behaviour.
[28]	SCADA data.		Uses Pearson product-moment rank correlation to select features, and applies different ANN structures to predict the active power.
[45]	SCADA data from a simulink model of a wind turbine.		Uses DBN for detecting anomolous behaviour. First traines individual RBMs to recreate input, and then uses labeled data to fine tune DBN to detect faults.
[34]	SCADA data.		Uses kolmogorov-smirnov test to compare different turbines at same moment in time combined with the E_e of the gearbox bearing temperature of an ANN to detect anomalies.
[46]	Vibration signals.	Approximates vibration distributions at different rotor speeds with Weibull distribution.	Uses a HMM for statistical fault detection.
[38]	SCADA data.		Compares linear models, ANNs and state-dependent parameter models for fault detection.
[21]	SCADA data.	parallel factor analysis (PARAFAC) as a decomposition method	uses K-means clustering after decomposition for fault detection.

[20]	Uses a wind turbine simulator for SCADA data.		Multiple PCA models are as a statistical reference reflecting the data variability in local zones and used in parallel for online fault detection.
[23]	Vibration signals.	Variational mode decomposition	Uses Fisher score and ReliefF algorithm for feature selection. Feeds selected signals into a multi-class SVC for bearing fault detection.
[39]	SCADA data.		Uses deep neural networks for detection of icing on the blades.
[55]	SCADA data.		Combines NN with alarms generated by SCADA system to reduce false alarm rate.
[4]	SCADA data.		Uses K-means clustering to partition turbines into different operating states, and a specific DBN of RBMs for each cluster to forecast the gearbox main bearing temperature. Uses E_e to detect anomalies, threshold set by Mahalanobis distance.
[56]			This is a literary review of vibration based condition monitoring and fault diagnosis of planetary gearboxes in wind turbines.
[47]			This is a literary review of machine learning methods used for condition monitoring of wind turbines.

Table 7.1: Summary of machine learning methods for wind turbine condition monitoring

7.2 Time-Series Clustering

Ref.	Representation	Similarity measure	Clustering Algorithms	Evaluation
[57]	Mixture Gaussian hidden Markov model (MGHMM).		Expectation-maximization (EM).	Bayesian information criterion (BIC).
[58]	Variance ratio statistics.	Euclidean distance.	Hierarchical clustering mainly, and K-means.	Duda-Hart $Je(2)/Je(1)$ indices.
[59]	HMM. States correspond to concentration regimes.	Which state each HMM is in.	Cluster together TS with corresponding HMMs in the same state.	
[50]	This is a review of time series clustering.			
[60]	Raw TS and some extracted statistics: variance, covariance, spread and differences.		Growing hierarchical self-organizing map (GHSOM).	
[61]	Linear combinations of spline basis functions.	Euclidean distance.	State space modelling, K-means and complete-linkage hierarchical clustering.	L-curve and gap statistic.
[62]	EMD for filtering out stochastic components, then extract topological features.	Euclidean distance.	K-means.	Precision, recall, F1-score and Matthews correlation coefficient.
[63]	Construct network between time series using dissimilarity matrix. Use KNN, and ϵ-NN to create networks from matrix.	Test a multitude of different distance functions. DTW performs best.	Test many community detection algorithms to sort network into clusters.	Rand index.
[64]	Symbolic Aggregate ApproXimation (SAX).	Approximate distance (APXDIST), Euclidean distance, and DTW.	Custom three step algorithm (MTC), with preclustering, subclustering, and merging to form final clusters.	Evaluated on datasets with known labels, with accuracy, precision, recall and F-measure.

[65]	Generalized autoregressive conditional heteroskedasticity (GARCH) model.	Tests different metrics based on squared Euclidean distance between unconditional volitility and time varying volitility.	Tests different variations of fuzzy C-medoids.	Xie-Beni index, Fuzzy Rand index, and prototype definition.
[66]	Bispectral Smoothed Localized Complex EXponential (BSLEX) algorithm.	Aggragated quasi-distance between smoothed bispectra across blocks.	Agglomorative hierarchical clustering with Ward's linkage.	Silhuette index as stopping criterion, and Rand Index, entropy and purity to evaluate cluster effectiveness.
[67]	HMMs.	Kulback-Leibler distance between the likelihood of observation sequency T given HMM λ.	Partitioning around medoids (PAM).	Silhouette index, Bavies-Bouldin index and Dunn index.
[68]	Copula-based model for time series.	P-norm of difference between copula of two points, and upper bound copula.	Fuzzy PAM.	Fuzzy Silhouette (FS) index, adjusted Rand index (ARI), fuzzy Rand index (FRI).
[69]	DWT and ARMA models.	Test multiple similarity measures.	Agglomerative Hierarchical clustering with Ward linkage.	Uses the clusters produced to perform regional frequency analysis, and then evaluates model using bias, RMSE, relative RMSE (RRMSE) and Nash criterion.
[70]	Independent component analysis (ICA).	Not specified.	Hierarchical clustering with complete linkage.	
[71]	Symbolic Aggregate Approximation (SAX).	Aproximate distance between symbolic representations of TS.	Extended version of K-modes.	SSE for stopping criteria, Rand index, NMI, Purity, Jacard, F-measure, FM and entropy for evaulating clusters.
[72]	Tranforms the covariance matricesof the time series into a tangent space.	Euclidean distance.	Hierarchical clustering with average linkage.	
[73]	Normalized spectral densities.	Total variation distence.	Agglomorative hierarchical clustering with complete and average linkage.	Dunn's index.

[74]	Self-exciting threshold autoregressive model (SETAR).	Primarily tests Euclidean distance, Haussdorf distance and DTW, but, tests 22 different ones.	Primary method is spectral clustering, but also tests PAM, and fuzzy C-means.	Measures accuracy of method on clustering simulated data, and uses Gap statistic as stopping criterion.
[75]	AR model.	A type of exponential euclidean distance.	Fuzzy C-medoids.	Fuzzy Silhouette index.
[76]	Continous wavelet transform (CWT).	Multi-scale PCA similarity matric.	Fuzzy C-means.	Precision and recall of classification according to labels, and silhouette index.
[77]	Preprocessing using Hodrick-Prescott (HP) filter, feature extraction using state space models or regression trees.		Self-organizing map (SOM).	Silhouette index as stopping criterion.
[78]	ARIMA model.	Euclidean distance between AR weights.	Trimmed fuzzy C-medoids.	Decides number of clusters by looking at the rate of decrease, and second derivative of an objective function with regard to a trimming factor α.
[79]	Permutation based coding of time series.	Use four distance metrics based on mutual information, entropy and Cramer's V association measure.	Hierarchical clustering with single, complete and average linkage.	
[80]	Mixture of autoregressions (MoAR) models.	Expectation-maximization (EM).		
[81]	Use PCA and custom ICA algorithm for feature extraction.	Euclidean distance between extracted features.	Hierarchical clustering with average, single, complete and Ward linkage, and K-means.	CH, Friedman, C-index, Dunn's, SDbw and Silhouette index.
[82]	Extracts various signal statistics, and performs feature extraction using R package "Psych".	Euclidean distance.	Hierarchical clustering with complete linkage.	

[83]	DWT with the Haar wavelet, and a globel sensitivity analysis.	Euclidean distance, to minimize variance.	K-means.	
[84]	Multi-relational network in topological domain, static (time-invariant), and dynamic (time-varying).		Multi-nonnegative matrix factorization. Compares their approach to three other community detection algorithms.	Rand index, adjusted Rand index and purity.
[85]	Uses Singular value decomposition (SVD) to represent the cluster centroids.	Pearson correlation coefficient.	Fuzzy C-means with particle swarm optimization.	Objective function, precision, and F-measure.
[86]	Multivariate time series are transformed into 3-order hysteresis tensors, then multilinear PCA is used to reduce dimensionality.	Tensor distance metric. Cluster centers initialized based on cycle feature variation.	Tensor K-means (CTK-means).	RI, ARI, Jaccard coefficient and Folks and Mallows (FM) index.
[87]	Compares ten model-based clustering methods. GMM and Markov-switching model perform best.			Misclassification rates.
[88]	Vari-segmented DWT.	Euclidean distance.	K-means, hierarchical agglomerative clustering and SOM.	
[89]	Discrete Fourier transform (DFT).	Euclidean distance.	Delaunay Triangulation method.	Purity and F-measure.
[90]	Piecewise SVD, and piecewise aggregate approximation (PAA).	Euclidean Distance.	Balanced Iterative Reducing and Clustering using Hierarchies (BIRCH).	

[91]	Extracts signal statistics, and uses PCA for feature selection.	Euclidean distance.	K-means.	Analyses the correlations of specific features with different clusters.
[92]	Fragmented periodogram.	Euclidean distance.	Spectral clustering.	
[93]	Extract statistical features of time series, then use a convolutional auto-encoder for further feature extraction.	Mainly Euclidean distance for the model-based approach.	Test many different clustering algorithms.	Adjusted Rand index.
[94]	Compares a feature-based approach using PCA, with a model-based approach using state-space models for the individual time series.	Inverse exponential Euclidean distance for feature based approach, and Euclidean distance for model-based approach.		Silhuette index as stopping criterion and inertia.
[95]	AR model.	Euclidean distance.	Fuzzy C-means.	
[96]	Flexible space-time autoregressive (FSTAR) models.	Use Wald statistic to compare model parameters of univariate STAR models, and p-value as a similarity metric.	Hierarchical agglomorative clustering.	ARI.
[97]	Common PCA (CPCA).	Cluster centroids represented by common projection axis of all TS in a specific cluster, then reconstruction error of time series using cluster centroid used as similarity metric.	Custom algorithm, similar to K-means.	Precision.
[98]	Map the time series to multiple high-dimensional tensors using multiple kernals.	Matrix L^p-norms.	Self-developed multi kernal clustering algorithm (MKC).	Normalized Mutual Information (NMI) and Rand Index (RI).

[99]	Vector autoregressive (VAR) models.	Euclidean distance between model parameters, then perform statistical test of whether TS come from same DGN.	Test two algorithms. One with fixed number of clusters, k similar to K-means, and one with an upper limit to the number of clusters k_{max}.	Purity index.

Table 7.2: Summary of model based time-series clustering methods

7.3 Code

```
import pandas as pd
import numpy as np
import matplotlib as mpl
import matplotlib.pyplot as plt
import seaborn as sns
import statsmodels.api as sm
```

Listing 1: Module fetching and data cleaning

Bibliography

[1] Espen Waaga. *Machine Learning for Automatic Classification of Wind Turbines*. English. Master thesis. 2019.

[2] Ml Hossain, A Abu-Siada, and SM Muyeen. "Methods for Advanced Wind Turbine Condition Monitoring and Early Diagnosis: A Literature Review". English. In: *Energies* 11.5 (2018). ISSN: 1996-1073.

[3] Henrique Dias Machado de Azevedo, Alex Maurício Araújo, and Nadège Bouchonneau. "A review of wind turbine bearing condition monitoring: State of the art and challenges". English. In: *Renewable and Sustainable Energy Reviews* 56 (2016), pp. 368–379. ISSN: 1364-0321.

[4] H. Wang et al. "Early fault detection of wind turbines based on operational condition clustering and optimized deep belief network modeling". In: *Energies* 12.6 (2019). ISSN: 19961073.

[5] P. Qian et al. "A novel condition monitoring method of wind turbines based on long short-term memory neural network". English. In: *Energies* 12.18 (2019). ISSN: 19961073.

[6] Alfredo Arcos Jiménez et al. "Linear and nonlinear features and machine learning for wind turbine blade ice detection and diagnosis". English. In: *Renewable Energy* 132 (2019), pp. 1034–1048. ISSN: 0960-1481.

[7] ASM Shihavuddin et al. "Wind Turbine Surface Damage Detection by Deep Learning Aided Drone Inspection Analysis". English. In: *Energies* 12.4 (2019), p. 676. ISSN: 1996-1073.

[8] Zifeng Qiu et al. "Automatic visual defects inspection of wind turbine blades via YOLO-based small object detection approach". English. In: *Journal of Electronic Imaging* 28.4 (2019), pp. 043023–043023. ISSN: 1017-9909.

[9] Alfredo Arcos Jiménez, Carlos Quiterio Gómez Muñoz, and Fausto Pedro García Márquez. "Dirt and mud detection and diagnosis on a wind turbine blade employing guided waves and supervised learning classifiers". English. In: *Reliability Engineering and System Safety* 184 (2019), pp. 2–12. ISSN: 0951-8320.

[10] Ningning Zhang, Chengzhi Lu, and Anmin Wang. "Study on wind turbine blade defect detection system based on imaging array". English. In: *E3S Web of Conferences* 118 (2019). ISSN: 25550403. URL: `http://search.proquest.com/docview/2301959230/`.

[11] Yinan Wang et al. "Unsupervised anomaly detection with compact deep features for wind turbine blade images taken by a drone". English. In: *IPSJ Transactions on Computer Vision and Applications* 11.1 (2019), pp. 1–7. ISSN: 1882-6695.

[12] Pierre Tchakoua et al. "Wind Turbine Condition Monitoring: State-of-the-Art Review, New Trends, and Future Challenges". English. In: *Energies* 7.4 (2014), pp. 2595–2630. ISSN: 19961073. URL: `http://search.proquest.com/docview/1537086932/`.

[13] A. Joshuva; V. Sugumaran. “A machine learning approach for condition monitoring of wind turbine blade using autoregressive moving average (ARMA) features through vibration signals: a comparative study”. English. In: *Progress in Industrial Ecology, An Int. J.* 12.1/2 (2018). ISSN: 1476-8917. URL: http://www.inderscience.com/link.php?id=95867.

[14] Vahid Pashazadeh, Farzad R Salmasi, and Babak N Araabi. “Data driven sensor and actuator fault detection and isolation in wind turbine using classifier fusion”. English. In: *Renewable Energy* 116.PB (2018), pp. 99–106. ISSN: 0960-1481.

[15] A. Joshuva and V. Sugumaran. “Improvement in wind energy production through condition monitoring of wind turbine blades using vibration signatures and ARMA features: a data-driven approach”. English. In: *Progress in Industrial Ecology* 13.3 (2019), p. 207. ISSN: 1476-8917.

[16] Magda Ruiz et al. “Wind turbine fault detection and classification by means of image texture analysis”. English. In: *Mechanical Systems and Signal Processing* 107.C (2018), pp. 149–167. ISSN: 0888-3270.

[17] Inturi Vamsi, G.R Sabareesh, and P.K Penumakala. “Comparison of condition monitoring techniques in assessing fault severity for a wind turbine gearbox under non-stationary loading”. English. In: *Mechanical Systems and Signal Processing* 124 (2019), pp. 1–20. ISSN: 0888-3270.

[18] Vamsi Inturi et al. “Integrated condition monitoring scheme for bearing fault diagnosis of a wind turbine gearbox”. English. In: *Journal of Vibration and Control* 25.12 (2019), pp. 1852–1865. ISSN: 1077-5463.

[19] Francesc Pozo, Yolanda Vidal, and Óscar Salgado. “Wind Turbine Condition Monitoring Strategy through Multiway PCA and Multivariate Inference”. English. In: *Energies* 11.4 (2018), p. 749. ISSN: 19961073. URL: http://search.proquest.com/docview/2041094406/.

[20] Azzeddine Bakdi, Abdelmalek Kouadri, and Saad Mekhilef. “A data-driven algorithm for online detection of component and system faults in modern wind turbines at different operating zones”. English. In: *Renewable and Sustainable Energy Reviews* 103 (2019), pp. 546–555. ISSN: 1364-0321.

[21] Wenna Zhang and Xiandong Ma. “Simultaneous Fault Detection and Sensor Selection for Condition Monitoring of Wind Turbines”. English. In: *Energies* 9.4 (2016), p. 280. ISSN: 19961073. URL: http://search.proquest.com/docview/1780819136/.

[22] T. Regan, C. Beale, and M. Inalpolat. “Wind Turbine Blade Damage Detection Using Supervised Machine Learning Algorithms”. In: *Journal of Vibration and Acoustics, Transactions of the ASME* 139.6 (2017). ISSN: 10489002.

[23] L. Fu et al. “Condition monitoring for the roller bearings of wind turbines under variable working conditions based on the Fisher score and permutation entropy.” In: *Energies* 12.16 (2019). ISSN: 19961073.

[24] Evangelos Papatheou et al. “Performance monitoring of a wind turbine using extreme function theory”. English. In: *Renewable Energy* 113.C (2017), pp. 1490–1502. ISSN: 0960-1481.

[25] Ravi Pandit and David Infield. “Gaussian Process Operational Curves for Wind Turbine Condition Monitoring”. English. In: *Energies* 11.7 (2018), p. 1631. ISSN: 19961073. URL: http://search.proquest.com/docview/2108516073/.

[26] Elena Gonzalez et al. “Using high-frequency SCADA data for wind turbine performance monitoring: A sensitivity study”. English. In: *Renewable Energy* 131 (2019), pp. 841–853. ISSN: 0960-1481.

[27] Liang Tao et al. “Abnormal Detection of Wind Turbine Based on SCADA Data Mining”. In: *Mathematical Problems in Engineering* 2019 (2019). ISSN: 1024-123X.

[28] Majid Morshedizadeh et al. “Improved power curve monitoring of wind turbines”. In: *Wind Engineering* 41.4 (2017), pp. 260–271. ISSN: 0309-524X.

[29] Ravi Pandit and David Infield. “Comparative assessments of binned and support vector regression-based blade pitch curve of a wind turbine for the purpose of condition monitoring”. English. In: *International Journal of Energy and Environmental Engineering* 10.2 (2019), pp. 181–188. ISSN: 2008-9163.

[30] Peyman Mazidi et al. “A health condition model for wind turbine monitoring through neural networks and proportional hazard models”. English. In: *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability* 231.5 (2017), pp. 481–494. ISSN: 1748-006X.

[31] Hongshan Zhao et al. “Anomaly detection and fault analysis of wind turbine components based on deep learning network”. English. In: *Renewable Energy* 127 (2018), pp. 825–834. ISSN: 0960-1481.

[32] MA Rodriguez-Lopez et al. “Methodology for Detecting Malfunctions and Evaluating the Maintenance Effectiveness in Wind Turbine Generator Bearings Using Generic versus Specific Models from SCADA Data”. English. In: *Energies* 11.4 (2018). ISSN: 1996-1073.

[33] P Bangalore and M Patriksson. “Analysis of SCADA data for early fault detection, with application to the maintenance management of wind turbines”. English. In: *Renewable Energy* 115 (2018), pp. 521–532. ISSN: 0960-1481.

[34] P. Guo, J. Fu, and X. Yang. “Condition monitoring and fault diagnosis of wind turbines gearbox bearing temperature based on kolmogorov-smirnov test and convolutional neural network model”. In: *Energies* 11.9 (2018). ISSN: 19961073.

[35] Marcin Straczkiewicz and Tomasz Barszcz. “Application of Artificial Neural Network for Damage Detection in Planetary Gearbox of Wind Turbine”. In: *Shock and Vibration* 2016 (2016). ISSN: 1070-9622.

[36] Hsu-Hao Yang, Mei-Ling Huang, and Shih-Wei Yang. “Integrating Auto-Associative Neural Networks with Hotelling T2 Control Charts for Wind Turbine Fault Detection”. English. In: *Energies* 8.10 (2015), pp. 12100–12115. ISSN: 19961073. URL: `http://search.proquest.com/docview/1732948122/`.

[37] Abhishek Reddy et al. “Detection of Cracks and damage in wind turbine blades using artificial intelligence-based image analytics”. English. In: *Measurement* 147 (2019), p. 106823. ISSN: 0263-2241.

[38] Philip Cross and Xiandong Ma. “Model-based and fuzzy logic approaches to condition monitoring of operational wind turbines”. English. In: *International Journal of Automation and Computing* 12.1 (2015), pp. 25–34. ISSN: 1476-8186.

[39] Longting Chen et al. “Learning deep representation of imbalanced SCADA data for fault detection of wind turbines”. English. In: *Measurement* 139 (2019), pp. 370–379. ISSN: 0263-2241.

[40] Hongguang Yun et al. “An Adaptive Approach for Ice Detection in Wind Turbine With Inductive Transfer Learning”. English. In: *IEEE Access* 7.99 (2019), pp. 122205–122213. ISSN: 2169-3536.

[41] Dahai Zhang et al. “A Data-Driven Design for Fault Detection of Wind Turbines Using Random Forests and XGboost”. English. In: *IEEE Access* 6 (2018), pp. 21020–21031. ISSN: 2169-3536.

[42] H. Ren et al. "A new wind turbine health condition monitoring method based on VMD-MPE and feature-based transfer learning". In: *Measurement: Journal of the International Measurement Confederation* 148 (2019). ISSN: 02632241.

[43] A. Agasthian, Rajendra Pamula, and L. Kumaraswamidhas. "Fault classification and detection in wind turbine using Cuckoo-optimized support vector machine". English. In: *Neural Computing and Applications* 31.5 (2019), pp. 1503–1511. ISSN: 0941-0643.

[44] Peng Qian et al. "A novel wind turbine condition monitoring method based on cloud computing". English. In: *Renewable Energy* 135 (2019), pp. 390–398. ISSN: 0960-1481.

[45] D Yu et al. "A radically data-driven method for fault detection and diagnosis in wind turbines". English. In: *International Journal of Electrical Power and Energy Systems* 99 (2018), pp. 577–584. ISSN: 0142-0615.

[46] Sangryul Kim and Seo Yun-Ho. "Development of a Fault Monitoring Technique for Wind Turbines Using a Hidden Markov Model". English. In: *Sensors* 18.6 (2018), p. 1790. ISSN: 14248220. URL: http://search.proquest.com/docview/2108718625/.

[47] Adrian Stetco et al. "Machine learning methods for wind turbine condition monitoring: A review". English. In: *Renewable Energy* 133 (2019), pp. 620–635. ISSN: 0960-1481.

[48] Wenguang Yang, Chao Liu, and Dongxiang Jiang. "An unsupervised spatiotemporal graphical modeling approach for wind turbine condition monitoring". English. In: *Renewable Energy* 127 (2018), pp. 230–241. ISSN: 0960-1481.

[49] Peter J. Brockwell and Richard A. Davis. *Time Series: Theory and Methods Second Edition*. Springer-Verlag New York, Inc., 1991.

[50] Saeed Aghabozorgi, Ali Seyed Shirkhorshidi, and Teh Ying Wah. "Time-series clustering - A decade review". eng. In: *Information Systems* 53 (2015), p. 16. ISSN: 0306-4379.

[51] Carlos Agon Philippe Esling. "Time-series data mining". In: *ACM Cumputing Surveys* 45 (2012).

[52] Peter J. Brockwell and Richard A. Davis. *Introduction to Time Series and Forecasting Third Edition*. Springer International Publishing Switzerland, 2016.

[53] Sheldon M. Ross. *Introduction to Probability Models 11th Edition*. Academic Press Elsevier Inc., 2014.

[54] Sergio Martin-del-Campo, Fredrik Sandin, and Daniel Strömbergsson. "Dictionary learning approach to monitoring of wind turbine drivetrain bearings". In: (2019).

[55] Alberto Pliego Marugan, Ana Maria Peco Chacon, and Fausto Pedro Garcia Marquez. "Reliability analysis of detecting false alarms that employ neural networks: A real case study on wind turbines". English. In: *Reliability Engineering and System Safety* 191 (2019), p. 106574. ISSN: 0951-8320.

[56] Tianyang Wang et al. "Vibration based condition monitoring and fault diagnosis of wind turbine planetary gearbox: A review". English. In: *Mechanical Systems and Signal Processing* 126 (2019), pp. 662–685. ISSN: 0888-3270.

[57] Jose Dias, Jeroen Vermunt, and Sofia Ramos. "Clustering financial time series: New insights from an extended hidden Markov model". eng. In: *European Journal of Operational Research* (2015), p. 852. ISSN: 03772217. URL: http://search.proquest.com/docview/1664477966/.

[58] Joao A Bastos and Jorge Caiado. "Clustering financial time series with variance ratio statistics". eng. In: *Quantitative Finance* 14.12 (2014), pp. 2121–2133. ISSN: 1469-7688. URL: http://www.tandfonline.com/doi/abs/10.1080/14697688.2012.726736.

[59] Alvaro Gomez-Losada, Jose Carlos M Pires, and Rafael Pino-Mejias. "Time series clustering for estimating particulate matter contributions and its use in quantifying impacts from deserts". eng. In: *Atmospheric Environment* 117.C (2015), pp. 271–281. ISSN: 1352-2310.

[60] Yu-Chia Hsu and An-Pin Chen. "A clustering time series model for the optimal hedge ratio decision making". eng. In: *Neurocomputing* 138.C (2014), pp. 358–370. ISSN: 0925-2312.

[61] Francesco Finazzi et al. "A comparison of clustering approaches for the study of the temporal coherence of multiple time series". eng. In: *Stochastic Environmental Research and Risk Assessment* 29.2 (2015), pp. 463–475. ISSN: 1436-3240.

[62] Cassio M.M Pereira and Rodrigo F de Mello. "Persistent homology for time series and spatial data clustering". eng. In: *Expert Systems With Applications* 42.15-16 (2015), pp. 6026–6038. ISSN: 0957-4174.

[63] Leonardo N Ferreira and Liang Zhao. "Time series clustering via community detection in networks". eng. In: *Information Sciences* 326.C (2016), pp. 227–242. ISSN: 0020-0255.

[64] Saeed Aghabozorgi and Teh Wah. "Clustering of large time series datasets". eng. In: *Intelligent Data Analysis* 18.5 (2014), pp. 793–817. ISSN: 1088-467X. URL: http://search.proquest.com/docview/1620092812/.

[65] Pierpaolo D'Urso, Livia De Giovanni, and Riccardo Massari. "GARCH-based robust clustering of time series". eng. In: *Fuzzy Sets and Systems* 305 (2016), pp. 1–28. ISSN: 0165-0114.

[66] Jane Harvill, Priya Kohli, and Nalini Ravishanker. "Clustering Nonlinear, Nonstationary Time Series Using BSLEX". eng. In: *Methodology and Computing in Applied Probability* 19.3 (2017), pp. 935–955. ISSN: 1387-5841.

[67] Shima Ghassempour, Federico Girosi, and Anthony Maeder. "Clustering multivariate time series using Hidden Markov Models." eng. In: *International journal of environmental research and public health* 11.3 (2014), pp. 2741–2763. ISSN: 1660-4601. URL: http://search.proquest.com/docview/1510403562/.

[68] Marta Disegna, Pierpaolo D'urso, and Fabrizio Durante. "Copula-based fuzzy clustering of spatial time series". eng. In: *Spatial Statistics* 21 (2017), pp. 209–225. ISSN: 2211-6753.

[69] H Zaifoglu, B Akintug, and A. M Yanmaz. "Regional Frequency Analysis of Precipitation Using Time Series Clustering Approaches". eng. In: *Journal of Hydrologic Engineering* 23.6 (2018). ISSN: 1084-0699.

[70] Thelma Safadi. "Using independent component for clustering of time series data". eng. In: *Applied Mathematics and Computation* 243.C (2014), pp. 522–527. ISSN: 0096-3003.

[71] Saeed Aghabozorgi and Teh Wah. "Approximate Clustering of Time-Series Datasets using k-Modes Partitioning". eng. In: *Journal of Information Science and Engineering* 31.1 (2015), pp. 207–228. ISSN: 1016-2364. URL: http://search.proquest.com/docview/1686443066/.

[72] Jiancheng Sun. "Clustering multivariate time series based on Riemannian manifold". eng. In: *Electronics Letters* 52.19 (2016), pp. 1607–1609. ISSN: 0013-5194. URL: http://search.proquest.com/docview/1845816549/.

[73] Pedro C. Alvarez-Esteban, Carolina Euan, and Joaquin Ortega. "Time series clustering using the total variation distance with applications in oceanography". In: *Environmetrics* 27.6 (2016), pp. 355–369. ISSN: 1180-4009.

[74] Sipan Aslan, Ceylan Yozgatligil, and Cem Iyigun. "Temporal clustering of time series via threshold autoregressive models: application to commodity prices". eng. In: *Annals of Operations Research* 260.1-2 (2018), pp. 51–77. ISSN: 0254-5330.

[75] Pierpaolo D'Urso, Livia De Giovanni, and Riccardo Massari. "Time series clustering by a robust autoregressive metric with application to air pollution". eng. In: *Chemometrics and Intelligent Laboratory Systems* 141 (2015), pp. 107–124. ISSN: 0169-7439.

[76] Joao Francisco Barragan, Cristiano Hora Fontes, and Marcelo Embirucu. "A wavelet-based clustering of multivariate time series using a Multiscale SPCA approach". eng. In: *Computers and Industrial Engineering* 95 (2016), pp. 144–155. ISSN: 0360-8352.

[77] Binoy B Nair et al. "Clustering stock price time series data to generate stock trading recommendations: An empirical study". eng. In: *Expert Systems With Applications* 70 (2017), pp. 20–36. ISSN: 0957-4174.

[78] Pierpaolo D'Urso et al. "Autoregressive metric-based trimmed fuzzy clustering with an application to PM10 time series". eng. In: *Chemometrics and Intelligent Laboratory Systems* 161 (2017), pp. 15–26. ISSN: 0169-7439.

[79] Maria Ruiz-Abellon, Antonio Gabaldon, and Antonio Guillamon. "Dependency-Aware Clustering of Time Series and Its Application on Energy Markets". eng. In: *Energies* 9.10 (2016), p. 809. ISSN: 19961073. URL: `http://search.proquest.com/docview/1831861660/`.

[80] Hien Nguyen et al. "Maximum Pseudolikelihood Estimation for Model-Based Clustering of Time Series Data". eng. In: *Neural Computation* (2017), p. 990. ISSN: 08997667. URL: `http://search.proquest.com/docview/1884823978/`.

[81] Jafar Rahmanishamsi, Ali Dolati, and Masoudreza Aghabozorgi. "A Copula Based ICA Algorithm and Its Application to Time Series Clustering". eng. In: *Journal of Classification* 35.2 (2018), pp. 230–249. ISSN: 0176-4268.

[82] B Hulsegge and K.H de Greef. "A time-series approach for clustering farms based on slaughterhouse health aberration data". eng. In: *Preventive Veterinary Medicine* 153 (2018), pp. 64–70. ISSN: 0167-5877.

[83] Yongkyu Lee, Jonggeol Na, and Won Bo Lee. "Robust design of ambient-air vaporizer based on time-series clustering". eng. In: *Computers and Chemical Engineering* 118 (2018), pp. 236–247. ISSN: 0098-1354.

[84] Lihua Zhou et al. "Clustering Multivariate Time Series Data via Multi-Nonnegative Matrix Factorization in Multi-Relational Networks". eng. In: *IEEE Access* 6 (2018), pp. 74747–74761. ISSN: 2169-3536.

[85] Z. Izakian and M. Mesgari. "Fuzzy clustering of time series data: A particle swarm optimization approach". eng. In: *Journal of Artificial Intelligence and Data Mining* 3.1 (2015), pp. 39–46. ISSN: 2322-5211.

[86] H. He and Y. Tan. "Pattern Clustering of Hysteresis Time Series with Multivalued Mapping Using Tensor Decomposition". In: *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 48.6 (2018), pp. 993–1004. ISSN: 21682216.

[87] Karen Kazor and Amanda Hering. "Assessing the Performance of Model-Based Clustering Methods in Multivariate Time Series with Application to Identifying Regional Wind Regimes". eng. In: *Journal of Agricultural, Biological, and Environmental Statistics* 20.2 (2015), pp. 192–217. ISSN: 1085-7117.

[88] D. Muruga Radha Devi and P. Thambidurai. "Similarity measurement in recent biased time series databases using different clustering methods". In: *Indian Journal of Science and Technology* 7.2 (2014), pp. 189–198. ISSN: 09746846.

[89] Narges Shafieian. "A Novel Method for Transforming XML Documents to Time Series and Clustering Them Based on Delaunay Triangulation". eng. In: *Applied Mathematics* 6.6 (2015), pp. 1076–1076. ISSN: 2152-7385. URL: `http://search.proquest.com/docview/1718957579/`.

[90] Ibgtc Bowala and Mgnas Fernando. "A novel model for Time-Series Data Clustering Based on piecewise SVD and BIRCH for Stock Data Analysis on Hadoop Platform". eng. In: *Advances in Science, Technology and Engineering Systems* 2.3 (2017), pp. 855–864. ISSN: 2415-6698.

[91] Jiechen Wang et al. "Relationship Between Urban Road Traffic Characteristics and Road Grade Based on a Time Series Clustering Model: A Case Study in Nanjing, China". eng. In: *Chinese Geographical Science* 28.6 (2018), pp. 1048–1060. ISSN: 1002-0063.

[92] Jorge Caiado, Nuno Crato, and Pilar Poncela. "A fragmented-periodogram approach for clustering big data time series". eng. In: *Advances in Data Analysis and Classification* (2019), pp. 1–30. ISSN: 18625347. URL: `http://search.proquest.com/docview/2239964127/`.

[93] Gerrit Bode et al. "A time series clustering approach for Building Automation and Control Systems". eng. In: *Applied Energy* 238 (2019), pp. 1337–1345. ISSN: 0306-2619.

[94] Omid Motlagh, Adam Berry, and Lachlan O'Neil. "Clustering of residential electricity customers using load time series". eng. In: *Applied Energy* 237 (2019), pp. 11–24. ISSN: 0306-2619.

[95] Yongping Zeng et al. "Fuzzy clustering of time-series model to damage identification of structures". eng. In: *Advances in Structural Engineering* 22.4 (2019), pp. 868–881. ISSN: 1369-4332.

[96] Edoardo Otranto and Massimo Mucciardi. "Clustering space-time series: FSTAR as a flexible STAR approach". eng. In: *Advances in Data Analysis and Classification* 13.1 (), pp. 175–199. ISSN: 1862-5347.

[97] Hailin Li. "Multivariate time series clustering based on common principal component analysis". eng. In: *Neurocomputing* 349 (2019), pp. 239–247. ISSN: 0925-2312.

[98] Yongqiang Tang et al. "Tensor Multi-Elastic Kernel Self-Paced Learning for Time Series Clustering". eng. In: *IEEE Transactions on Knowledge and Data Engineering* (2019), pp. 1–1. ISSN: 1041-4347.

[99] S. Deb. "VAR Model Based Clustering Method for Multivariate Time Series Data". eng. In: *Journal of Mathematical Sciences* 237.6 (2019), pp. 754–765. ISSN: 1072-3374.