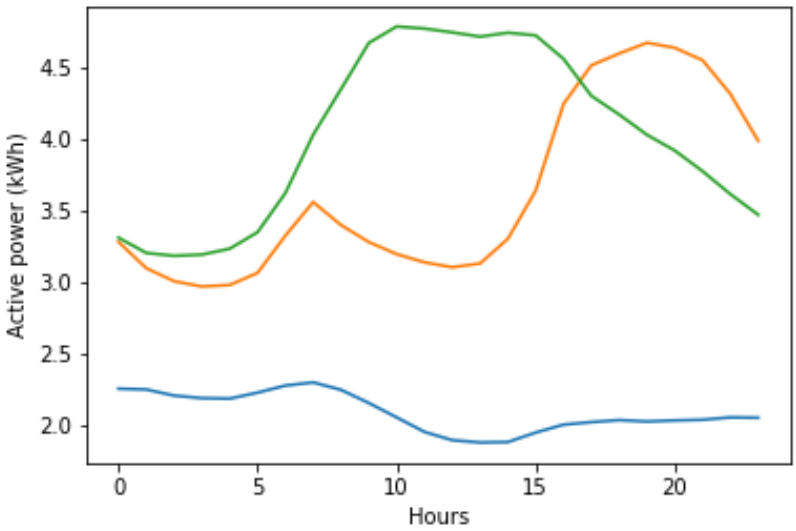
Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Electric Power Engineering

Stian Norheim

Clustering of AMS-data

Master's thesis in Energy and Environment
Supervisor: Eivind Solvang, Maren Istad, Karoline Ingebrigtsen
June 2020



NTNU
Norwegian University of Science and Technology

SINTEF

Stian Norheim

Clustering of AMS-data

Master's thesis in Energy and Environment
Supervisor: Eivind Solvang, Maren Istad, Karoline Ingebrigtsen
June 2020

Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Electric Power Engineering

NTNU
Norwegian University of
Science and Technology

Abstract

The worldwide rollout of advanced metering systems (AMS) seen today makes large amounts of AMS-data available. The AMS-data is suitable for Big-Data techniques, and clustering is one technique commonly used. Clustering is a technique to divide a dataset without external labelling information into groups with different characteristics. This master thesis investigates clustering on AMS-data, and presents a solid foundation of clustering theory, AMS and useful background information to use the clustering results most efficiently. The thesis presents a comprehensive clustering analysis on four datasets using 6 different clustering algorithms and 5 data representation techniques. Validation of the results is done with three cluster validation indexes (CVIs) and manual inspection to insure robustness of the results. A comparison with the practice used today and an analysis of capacity in the grid based on AMS-data are also performed.

The results show that the partitional algorithms are the preferred choice. The old, simple and robust K-Means performs well and has a low computational cost. The K-Shape got similar CVI scores as K-Means, despite the CVI scores calculated being biased towards algorithms using the euclidean distance (ED) measure. The hierarchical algorithm using single linkage shows interesting results, being able to isolate outliers. The data representation techniques show potential for a considerable reduction in time, and also clustering with different characteristics.

Sammendrag

Utrullingen av AMS (smarte strømmålere) man ser i dag, både i Norge og globalt, gjør store mengder strømdata (AMS-data) tilgjengelig. På disse AMS-dataene kan man bruke Big-Data teknikker, og clustering er en teknikk mye brukt. Clustering er en teknikk som deler et datasett uten ekstern gruppering av dataene inn i forskjellige clustre (grupper) med ulike egenskaper. Denne masteroppgaven undersøker clustering av AMS-data, og presenterer clustering teori, AMS og nyttig tilleggsinformasjon for å kunne bruke resultatene fra clusteringen på en best mulig måte. Masteroppgaven presenterer en grundig analyse av 6 clustering algoritmer og 5 data representasjonsteknikker på fire ulike datasett. Validering av resultatene er gjennomført ved hjelp av tre clustering validerings indekser (CVIs) og manuell analyse for å sikre robuste resultater. En sammenligning med dagens praksis og en analyse av kapasitet i nettet basert på AMS-data er også gjennomført.

Resultatene viser at oppdelingsalgoritmer (engelsk: partitional algorithms) er det beste valget. Den gamle, enkle og robuste algoritmen K-Means presterer bra, og er den raskeste algoritmen. K-Shape presterte omtrent likt som K-Means på CVI'ene, selv om CVI beregningene favoriserer algoritmer som bruker en distansemåler kalt euclidean distance (ED), noe K-Means gjør. Den hierarkiske algoritmen som bruker singel lenke kriterium viser interessante resultater ved at den klarer å isolere data med avvikende verdier. De ulike teknikkene for å representere AMS-dataene viser at det er et potensiale for en betydelig reduksjon i beregningstiden for algoritmene, i tillegg til at man kan endre egenskapene til clusteringen.

Acknowledgements

I would like to acknowledge everyone that has been helping and supporting me along the way in the process of making this master thesis.

Especially I would like to thank my three supervisors. My main supervisor, Eivind Solvang, for guiding me steadily through the working and writing process, as well as many useful advises, both for work and private situations. My co-supervisor Maren Istad for many useful suggestions and for giving me insight in the daily work-life. My co-supervisor Karoline Ingebrigtsen for good technical discussions as well always being ready to help, despite being a thousand kilometer away most of the time.

Many other people have also helped me along the way, who deserve to be acknowledged. I thank Volker Hoffmann and Katarzyna Michalowska from SINTEF, Marianne Blikø and Rune Paulsen from Tensio and Gunnar Berge, Gunnwald Olseth, Line Bergfjord and Johannes Gjesdal from BKK. I thank the students Kari Walstad and Håvard Refvik for useful discussions and mutual problem solving. I also want to thank the guys from the reading place, for making a great working-environment as well as a great time together. Last, but not least, I thank my amazing girlfriend for all her love and support.

Table of Contents

Abstract i

Sammendrag ii

Acknowledgements iii

Table of Contents vii

List of Tables ix

List of Figures xii

Abbreviations xiii

1 Introduction 1

2 Theory 3
- 2.1 Electricity system . . . 3
  - 2.1.1 History of electric systems . . . 3
  - 2.1.2 History of electricity markets . . . 4
- 2.2 Advanced Metering Systems . . . 5
  - 2.2.1 Research on AMS-data . . . 6
- 2.3 Network tariffs . . . 7
- 2.4 Clustering theory . . . 8
  - 2.4.1 Time-series representation . . . 9
  - 2.4.2 Distance measure . . . 10
  - 2.4.3 Centroid extraction . . . 20
- 2.5 Clustering algorithms . . . 21
  - 2.5.1 Partitional methods . . . 21
  - 2.5.2 Hierarchical clustering . . . 22
  - 2.5.3 Different clustering approaches . . . 23
- 2.6 Cluster validation indexes . . . 23

2.6.1 Silhoutte index . . . 24
2.6.2 Calinski-Harabasz . . . 25
2.6.3 Davies-Bouldin . . . 25
2.7 Estimating the number of clusters . . . 26
2.7.1 CVI scores . . . 26
2.7.2 Gap statistics . . . 26
2.8 Multiple step approaches . . . 27
2.9 Prediction of load profiles . . . 27
2.9.1 Prediction of load profiles today . . . 28
2.9.2 Future load profiles . . . 28
2.10 Temperature corrections . . . 28
2.10.1 Temperature corrections . . . 28

3 Methods **31**
3.1 Clustering components . . . 31
3.1.1 Time-Series representation . . . 31
3.1.2 Distance measures . . . 32
3.1.3 Cluster centroids . . . 32
3.1.4 Clustering algorithms . . . 32
3.1.5 Cluster centroid representation . . . 33
3.2 Experimental settings . . . 33
3.2.1 Datasets . . . 33
3.2.2 Time-series samples . . . 33
3.2.3 CVI scores . . . 34
3.2.4 Software packages used . . . 34
3.3 Tested clustering approaches . . . 34
3.3.1 Parameters . . . 34
3.3.2 Optimal number of clusters . . . 35
3.3.3 Presentation of results . . . 36
3.3.4 Practical application . . . 36
3.4 Comparison with FASIT . . . 37
3.5 Analyses of capacity in the grid . . . 37

4 Results **39**
4.1 Clustering approaches . . . 39
4.2 Clustering algorithms . . . 39
4.2.1 K-Means . . . 40
4.2.2 K-Shape . . . 40
4.2.3 K-Means with DTW . . . 42
4.2.4 Self Organizing Maps . . . 42
4.2.5 Hierarchical clustering . . . 43
4.3 Reduction techniques . . . 45
4.3.1 PCA . . . 45
4.3.2 Wavelets . . . 47
4.4 Transformation techniques . . . 48
4.5 Average load profiles . . . 49

4.6 Practical application . . . 50
4.7 Comparison with FASIT . . . 52
4.8 Transformer analyses . . . 53
4.8.1 Energy transformed . . . 53
4.8.2 Max power . . . 53
4.8.3 Time of use . . . 54
4.8.4 Max power aggregated by consumers . . . 54

5 Discussion **57**
5.1 Clustering approaches . . . 57
5.1.1 Clustering algorithms . . . 58
5.1.2 Data representation techniques . . . 60
5.1.3 Practical applications . . . 62
5.1.4 FASIT . . . 62
5.2 Analyses of capacity in the grid . . . 62
5.2.1 Temperature corrections . . . 63
5.2.2 Aggregation . . . 63
5.2.3 Usefulness of aggregation . . . 63
5.2.4 Assumptions and uncertainties . . . 63
5.3 Broader discussion . . . 64
5.3.1 Applications of clustering . . . 64
5.3.2 Assumptions and approaches . . . 67
5.3.3 Metering frequency . . . 68
5.4 The working process . . . 69

6 Conclusion **71**

7 Further work **73**

Bibliography **75**

List of Tables

2.1 Results from the Euclidean distance example 13
2.2 Elements compared in the orange warping path from figure 2.5 14
2.3 Values of the load profile used in the SBD example 18
2.4 cross-correlation values for different shifts 19

3.1 Clustering approaches tested. All approaches use the Z-transformed datasets, except for the approach with Min-Max normalization. 35

4.1 Results from clustering analyses 40
4.2 The optimal cluster numbers 40
4.3 Optimal cluster number using MM-normalization and Z-transformation . 48
4.4 Optimal cluster number with all load profiles and the load profiles averaged (A) and Z-transformation (no extra label). 50
4.5 Some key values from temperature corrections on max power 54

6.1 Summary of results 71

List of Figures

2.1 Four main components in clustering 8
2.2 The three load profiles used in the euclidean distance example 12
2.3 Comparison of load profiles 12
2.4 Comparison of normalized load profiles 13
2.5 The DTW paths from example 1 DTW 14
2.6 The two load profiles analysed in DTW example 2 15
2.7 Comparison of normalized load profiles 16
2.8 Comparison of DTW computation with two different restrictions 16
2.9 Comparison of normalized load profiles 17
2.10 Load profiles used in the SBD example with different shifts 18
2.11 Example of a two-step model. Figure taken from [1] 27

4.1 K-Means on dataset 1 with 3 clusters. Two different cluster representations. 41
4.2 K-Means on dataset 4 with 11 clusters. Two different cluster representations. 41
4.3 K-Means on dataset 4 with 2 clusters. Two different cluster representations. 41
4.4 K-Shape on dataset 1 with 3 clusters. Two different cluster representations. 42
4.5 K-Means with DTW on dataset 1 with 3 clusters. Two different cluster representations. 42
4.6 SOM on dataset 4 with 11 clusters. Two different cluster representations. 43
4.7 Hierarchical clustering with single linkage criteria on dataset 3 with 2 clusters. Two different cluster representations. 44
4.8 Hierarchical clustering with single linkage criteria on dataset 3 with 11 clusters. Two different cluster representations. The clusters not labelled contain only one sample. 44
4.9 Hierarchical clustering with ward linkage criteria on dataset 1 with 11 clusters. Two different cluster representations. 45
4.10 CVI-score for the reduction techniques 46
4.11 K-Means clustering with a PCA and Z-transformation of data on dataset 1. Two different cluster representations. 46

4.12 K-Means clustering with a wavelet transformation of dataset 1 with 3 clusters. Two different cluster representations. 47
4.13 K-Means clustering with a wavelet transformation on dataset 1 with 3 clusters. Two different levels in the wavelet algorithm. 47
4.14 K-Means clustering with two different transformation techniques on dataset 3 with 11 clusters 48
4.15 K-Means clustering with two different transformation techniques on dataset 2 with 10 clusters. 49
4.16 Clustering with all samples (dashed lines) and with average load profiles (full lines) for dataset 1 with 3 clusters. 49
4.17 Hierarchical clustering with single linkage criteria and average load profiles on dataset 1 with 3 clusters. Two different cluster representations. . . 50
4.18 Clustering in the practical application on streetlights. 51
4.19 Average consumption for one of the abnormal load profiles 51
4.20 Comparison of clustering and FASIT load profiles on households for weekdays. Two different cluster representations. 52
4.21 Comparison of clustering and FASIT load profiles on households for weekends. Two different cluster representations. 52
4.22 Energy consumed, with and without temperature correction, below transformers 53
4.23 Maximum power below transformers 53
4.24 Maximum power below transformer normalized 54
4.25 Difference between PMT and PT 55
4.26 Time of use for transformers 55
4.27 Measured peak and aggregated peaks for transformers 56
4.28 Measured peak and aggregated peaks for transformers 56
4.29 Differences between the aggregated PMT and PT 56

5.1 Example of visualisation for customer introducing a battery 65

Abbreviations

AMS	=	Advanced metering systems
CH	=	Calinsky-Harabasz
CVI	=	Cluster validation index
DB	=	Davies-Bouldin
DSO	=	Distribution System Operator
DTW	=	Dynamic time warping
ED	=	Euclidean distance
HI	=	Hierarchical clustering
KM	=	K-Means
KS	=	K-Shape
MM	=	Min-Max (normalization)
PCA	=	Principal component analysis
PMT	=	Maximum power measured temperature corrected
PT	=	Maximum temperature corrected consumption
SBD	=	Shaped-based distance
SI	=	Silhoutte
SOM	=	Self-organising maps

Chapter 1

Introduction

The introduction of advanced metering systems (AMS) is a part of the digitization of the power system. In Norway, 98.5% of the customers had AMS installed by the end of 2018. With AMS installed, the power market can be more efficient and the operation of the grid can be improved from the AMS-data. Until now, the main focus for the Norwegian distribution system operators (DSOs) has been to get the AMS up and running, and in the years to come it is expected that more analyses will be done with AMS-data. The AMS make large quantities of data available, which are suitable for Big-Data techniques. Clustering is a Big-Data technique which divides a dataset without external labelling information into clusters with different characteristics. These clusters might be valuable, as they provide information about patterns in the dataset analysed.

Clustering is a well known Big-Data technique and it is commonly used on AMS-data. There are many clustering methods and algorithms being used today, and improvements to the algorithms are constantly proposed. Transformation of the input data is also a common way to change the characteristics of the clustering. There are a lot of literature available on clustering of AMS-data, and some of the motivations for doing the clustering are demand side management (DSM), specialized network tariffs and as a pre-stage before prediction models.

This master thesis starts by presenting a solid fundament of clustering theory and AMS in chapter 2. This chapter also includes background information to use the clustering results most efficient. The methods and experimental setup in the clustering study performed in this master thesis are presented in chapter 3. The results are shown in chapter 4 and discussed in chapter 5. Chapter 6 draws some conclusions of the work done, and chapter 7 proposes further work.

Chapter 2

Theory

This chapter presents the theory which the analyses in this master thesis are based on. The chapter starts by a presentation of electric systems in section 2.1, included to explain the smart meters' position in the modern electricity systems we have today. Section 2.2 presents AMS and their requirements, which is important information to understand the possibilities as well as limitations regarding Big-Data analysis on AMS-data. Section 2.3 presents network tariffs, both the ones being used today as well as the ones proposed by the regulating authorities, which is important information to understand the changes of consumption in the future. Section 2.4 explains clustering as well as important techniques in clustering, before section 2.5 explains the different clustering algorithms used in this master thesis. Section 2.6 presents three cluster validation indexes (CVIs), which are being used to quantify the quality of the clustering, before section 2.7 and 2.8 present practical techniques used in clustering approaches. Section 2.9 presents how the average load profiles used today are calculated. The chapter ends by looking at temperature corrections of electric consumption in section 2.10, which are used in analyses of capacity in the grid.

2.1 Electricity system

This section presents the building blocks of a modern electricity system. This section is important background information to understand how AMS can contribute to a more efficient operation of the power grids and power markets. Readers with good insight in electricity systems might therefore skip this section.

2.1.1 History of electric systems

In 1831, Michael Faraday discovered how to produce electric energy from mechanical energy [2]. However, it took many decades before the electricity could be used for lightning purposes. In 1878, the inventor Thomas Edison managed to make the first functioning light bulb, and the first public presentation found place in 1879 [3]. The following decade

many cities installed electric street lights and since then the electric systems have increased to the large interconnected systems we see today [4].

Many things have changed since the discovery of electricity, but the physical laws still apply today the same way as they did 200 year ago. Therefore, some of the problems they struggled with in the late 19th century, we are still facing today [5]. Two of the problems are frequency and voltage quality. In AC systems, an increase in the active power consumed will contribute to reduce the frequency in the grid. This means that in order to keep the frequency stable, the production and consumption need to be synchronized. The electrical engineer also knows that that there is a close link between the voltage magnitude and the reactive power. Therefore, changing the reactive power consumed or produced might be an easy way to change the voltage magnitudes in the grid. Safety is another big problem, both for humans and buildings.

The electricity to power the first street light came from coal, and fossil fuels have always been an important source to make electricity [6], [7]. However, the last decades the electricity mix in Europe has been changing towards renewable energies. The renewable sources are growing fast, and in 2018, 26% of the electricity in the European Union came from renewable sources (including hydropower), an increase from 20% in 2009 and 12 % in 1990 [7], [8]. It has also been a decrease in the electricity from fossil fuels, from 56 % in 1990 to 51 % in 2009 and 46 % in 2018 [7], [8].

It is likely that the share of renewables will be even higher in the future, and therefore it is important to understand the different characteristics between fossil fuels and renewables. One of the differences is that generators using fossil fuel have inertia and renewable sources do not [9]. Loosely speaking, inertia means the energy necessary to change a motion, and systems with higher inertia are more stable. The higher inertia an electric system has, the lower the change in frequency will be if the load is increased or decreased. This makes a system with more renewables more susceptible to load changes, both increasing and decreasing load. The other difference is the planning of the generation. While fossil fuel plants are dependent on the fossil fuel, the renewables are dependent on its renewable sources. The fossil fuels can be extracted and stored for a long time, but that is not always the case for renewables. For solar and wind, the two most increasing renewables today, the energy from the sun and in the wind has to be used immediately [8].

2.1.2 History of electricity markets

An electric market is a system where selling and buying of electricity find place [10]. Most high-developed countries have today deregulated their electrical markets [10]. This deregulation means separating production of electricity, buying of electricity and building and maintenance of the electrical infrastructure [10]. These three parts are separated to allow competition in the market [10].

The three parts in the electricity markets have different aims and tasks, presented in the following. The distribution network operator (DSO) has the responsibility for building and maintaining the electrical infrastructure, which make it a natural monopoly (as it is not socioeconomic to have parallel transmission lines from different companies). The DSOs get their income from the network tariffs, which are explained in section 2.3. The electrical production companies produce electricity and get their income from selling electricity on the electricity market. The electricity producers are free to produce whenever they

want, and they try to maximize their profit by producing when the electricity prices are the highest. The electricity suppliers provide electricity to its customers, who pay the electricity supplier. The electricity supplier buys the electricity on the electricity market.

The leading power market in Europe is Nordpool [11]. This market consist of offers and bids which can be formulated under certain conditions [12]. The prices are calculated based on these offers and bids, and if there are not sufficient capacity between bidding areas, there might be different area prices. Norway is for example divided into five different areas, which might get different prices. Nordpool offers trading in a ”day-ahead” market and a ”intraday” market. The day-ahead market set the prices for the following day, and is the main arena for trading electrical power [13]. The deadline for the offers and bids for a given day in the day-ahead market is at 12:00 the previous day, and the prices get published at 12:42 or later [13]. The intraday market offers flexibility if there is a difference between the sell order or the purchase order and the actual consumption or generation. The intraday market secures the balance between the generation and consumption (plus losses) and helps to keep the power grid stable.

In average, the prices in the day-ahead market follow the demand [14]. However, the prices are set from the offers and bids in the market, and they might therefore not reflect the actual demand in the network. One of the reasons for this is the marginal prices for producing electricity, and the renewable energies have a marginal cost close to zero [15]. Another reason is the costly starting and stopping of production from the power plants. As the price of electricity does not follow the demand in the network, the price of electricity does not reflect the available capacity in the grid either. This is important to keep in mind when discussing network tariffs (section 2.3).

The increased digitization creates new possibilities. With the introduction of AMS, it is possible to bill customers according to their hourly (or even finer) consumption. Previously, customers were billed based on their energy use, typically over 1 month, but hourly billing is already in use today [16], [17]. AMS gives the customers the possibility to adapt to the prices in the market, by using energy when the electricity is at its cheapest. For most customers it probably feels overwhelming to check the electricity prices every day and use electricity when it is on its cheapest. Therefore automatic solutions seem to be the future, with a popular example of charging of electric cars [18]. Room and water heating, which contributes with a large share of the electricity consumption in Norway, might also be moved without affecting the customers life and habits, due to large thermal constants (the temperature in the room or the water stay stable after heating) [19].

2.2 Advanced Metering Systems

Advanced Metering Systems (AMS), also called smart meters, are devices with advanced metering abilities connected to a two-way communication network. The main contribution of AMS is to provide a more efficient power market and improve the operation of the grid [20]. The requirements of the AMS in Norway are stated in the Norwegian regulation no. 301 of 11 March 1999 [21]. For this master thesis, the most relevant requirements are the following:

- Metering with a minimum frequency of 60 minutes and possibility to increase the

frequency to 15 minutes.

- Metering of active and reactive power in both directions.

The last couple of years is has been a big rollout of AMS in Norway. The rollout came as a demand from the regulating authority, NVE (Norwegian water and Resources and Energy Directorate) in Norway [21]. The rollout should be finished by the end of 2018, but a few percent still missed AMS by the beginning of 2019 [22]. NVE is now aiming for 100% implementation by the beginning of 2021 [22]. In a proposal published by NVE in May 2020, an increase of the metering frequency to 15 by the second quarter of 2023 [23] is presented. The electric consumptions from all customers in Norway are gathered in a hub called Elhub, which became operational in the beginning of 2019 [24]. The customers can access their consumption data in this hub, and the hub also makes switching of electricity suppliers easier [16].

The AMS provide many improvements compared to the old meters [20]. One of these is, as mentioned in section 2.1.2, the opportunity of hourly billing (or even finer), compared to the monthly billing which was the practice in Norway before [25]. This is possible due to the two-way communication and the increased metering frequency. The AMS have also an increased accuracy compared to older electricity meters, which make the billing more accurate [26]. AMS can also cut or limit the maximum consumption of a customer and send out signals of faults in the grid [21]. Beneficial for the customer, AMS can receive information about electricity prices and tariffs [21]. In addition, the smart meters installed in Norway have the opportunity to meter the voltage [16].

There are, however, also challenges and consequences regarding the rollout of AMS. One consequence is the cost associated to the installation, which is 10 000 MNOK for Norway, and in average 3500 NOK for each household [27]. The installation cost will be paid through the grid fee, which for households will increase with around 300 NOK per year [27]. An Official Norwegian report published in 2015 looks at some vulnerabilities with AMS [28]. This report expresses a concern about the dependency of commercial communication infrastructure, the privacy policy and the security aspects of remote disconnection of customers [28]. The report also concludes that no risk or vulnerability analysis was conducted before it was decided to install AMS in Norway. More information about security aspects regarding the rollout in Norway can be found at [29], [30] and [20].

2.2.1 Research on AMS-data

The introduction of AMS provide many opportunities, and there is a lot of research which use AMS-data. The expected usage of AMS-data for up to five years ahead of time for Norwegian DSOs was presented in [20], published in 2019. This paper expects usage of clustering mainly to be from 1-4 year ahead of time by the writing of this thesis. ENERGYTICS is a project in SINTEF which has demonstrated Big Data technologies and data science on AMS-data [31]. This project looked among other things on prediction of faults, automatized investment decisions and analyses of voltage quality and consumption data. In Norway, ENOVA supports seven projects exploring opportunities with AMS carried out by DSOs [32]. These pilots look at technologies, services and business models to motivate customers to use less electricity. FME CINELDI aims to develop flexible, robust and intelligent electrical networks and uses AMS-data in many of their research areas [20].

Overview of research done, both in Norway and worldwide, with AMS can be found in [16] and [20].

2.3 Network tariffs

With the introduction of AMS comes, as described in the previous section, many opportunities, both to customers, companies and DSOs. As seen in the previous section, the technical qualities of AMS provide opportunities to utilize the grid more efficiently. One of the instruments for using the grid most efficient is through network tariffs. A network tariff is the price a customer must pay to the network utility for the electric infrastructure. Traditionally, the network tariffs for households in Norway have been based on a fixed amount for being connected to the grid plus a cost proportional to the amount of energy used [14]. Other customer groups, for example customers with a high power or energy consumption, have had a cost based on their maximum power consumed [14].

Even though the old tariffs are easy to understand for the customers, they do not reflect the cost in the electric grid [33]. Most of the cost associated to the electrical infrastructure is related to available capacity in the grid, not to how much energy is being consumed [33]. The change proposed by NVE is to include a cost related to the power consumed, as well as changing the cost for the energy consumed to the marginal cost of energy-consumption, so that it does not cover fixed costs [33]. This will make the network tariffs reflect the costs in the grid. The cost related to the power consumption is proposed to be calculated based on the highest peak consumption each day. The proposed changes will give incentives for customers to reduce their peaks [33].

However, not everyone agrees on how the network tariffs should look like [34]. There are many reasons why tariffs are difficult to set and agree on, discussed briefly in the following [33]. One is that the tariffs should be easy to understand. The concept of paying for the energy needed is natural for most people, while the reason behind paying more for using more power at the same time is harder to understand. Another reason is the division of costs. With changed network tariffs, the costs for the same load consumption patterns will inevitably change. This has led to people feeling that the authority (NVE) is aiming to target certain customers in the grid, like this article shows [35].

It is also important to remember that the new network tariffs are not the only incentives to change the consumption patterns for customers. The introduction of AMS gives electricity suppliers hourly data from the customers, and billing based on the hourly consumption is therefore increasingly popular [36]. This means that customers changing their consumption pattern can lower their electricity bill. A survey answered by 1200 Norwegian household customers showed that more than half of the customers were willing to change their consumption pattern to reduce their electricity bill [37]. As mentioned in section 2.1.2, the average spot-price follows the demand, so it is expected that the introduction of hourly billing will in average give lower peaks in the grid. However, hourly billing also can increase the peaks if the electricity price is low in a period with already high consumption. This is the reason why, as explained in [33], the network tariffs do not follow the spot-price.

The effect of network tariffs and hourly billing are important inputs when discussing changes in future load profiles. A presentation of the average load profiles the DSOs

use today for their customers, as well as a discussion of future load profiles, are given in section 2.9.

In literature, a common motivation for performing clustering is to make specialized and dedicated network tariffs [38], [39]. The customers in the different clusters are then given specialized tariffs, as these tariffs can be more specialized to the characteristics of the clusters [38].

2.4 Clustering theory

This section presents clustering theory necessary to understand the clustering algorithms and approaches used in this master thesis. Cluster analysis is in [40] defined as the ”the formal study of methods and algorithms for grouping”. The motivation behind clustering is to understand and learn about the data analysed [40]. Clustering groups similar object together, which gives a better base for understanding the different characteristics that exists in the dataset [40]. Clustering is also used to organize data and represent the data by prototypes. There are many clustering algorithms today, and they differ both in complexity and methodology [40].

Inspired by [41], the presentation of clustering is divided into 4 components, shown in figure 2.1. In the other chapters, the combination of one clustering algorithm and one distance measure is called an algorithm, in order to keep in line with the expressions commonly used in the literature of clustering of AMS-data. In this master thesis, the combination of the four components is called one clustering approach. The input of a clustering approach is the data that is to be clustered, and the output contains cluster labels (which groups the data belongs to) and cluster centroids (representations of the groups).

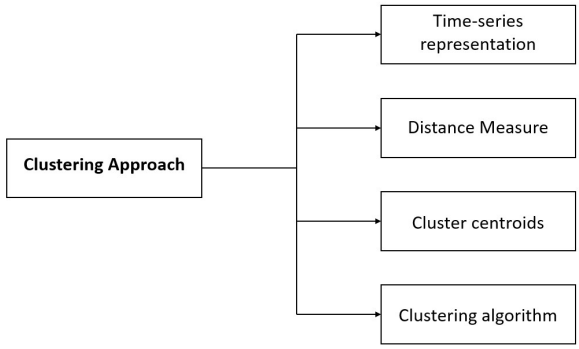


Figure 2.1: Four main components in clustering

In literature it is often emphasised that clustering is an exploratory tool [40], [42], [43]. These articles say that there is no way to find an optimal clustering approach in general, and therefore multiple clustering approaches should be tested for a problem. The literature also says that there is no single best cluster validation index (CVI) to evaluate the quality of cluster partition either [43], [39], [44].

Clustering is a big data technique used in many domains where a lot of data is available [41]. Some of the domains are biology, medicine, finance, voice recognition and energy [41], [45]. This means that the datasets, the objectives and the preferred algorithms can be different in the different domains [41]. This master thesis will focus on clustering theory and algorithms suitable for AMS-data.

To formalize the discussion of clustering, some terms are now presented. The dataset analysed, X, contains n samples which are divided into k clusters. A sample $\vec{x}$ is a vector $\vec{x} = [x_1, x_2, x_3, \ldots, x_m]$ consisting of m elements (scalars). Sample i is written as $\vec{x_i}$, and element a in a sample $\vec{x_i}$ is written x_{ia}, or just x_a, depending on the context. A cluster is defined as c_j with the corresponding cluster centroid $\vec{c_j}$. The total number of samples in a cluster is $|c_j|$, and the total number of samples in the dataset is N. A sample is also called a time-series or a load profile, and the words are used interchangeably in this thesis depending on the common practice in literature. All of the terms mentioned are commonly used in literature.

The rest of this subsection presents the three first components from figure 2.1, while section 2.5 is devoted to the clustering algorithms.

2.4.1 Time-series representation

The first step in a clustering approach is the representation of the data. The data is called time-series if it has values as a function of time, which is the case for AMS-data [41]. In the literature for clustering of AMS-data, three time-series representations have been found, presented in the following:

Reduction techniques

The time-series are transformed from the original m dimensional space to a lower dimensional space. The motivation behind dimensionality reduction is to reduce memory requirements, increase the computational speed on the clustering algorithm and reduce the noise [41]. Two commonly used reduction techniques on time-series clustering, discrete wavelet transformation and principal component analysis (PCA), are used in this master thesis [41].

PCA is a technique commonly used in AMS-data analyses, which compared to other reduction techniques have achieved good results [46], [47]. The PCA is a linear reduction method that uses eigenvalues of the covariance matrix to reduce the dimensionality of the dataset. The details of the PCA computation is out of the scope of this master thesis, and details can be found in [47].

Practical implementations of discrete wavelet transformation use filters to get representations of the input data [48]. The input data goes through a low pass and a high pass filter, and the output from the filters are approximation and detail coefficients. The low pass filter gives the approximation coefficients, which represent an averaging of the input data. For each time the data goes through the low pass filter, the size is halved, and the "level" is increased by one. By running the data one time through the low pass filter (level 1), the data is halved, and for level 2 the data is one quarter of the original size and so on. This way the desired size of the dataset can be obtained.

Feature extraction

Feature extraction can be done by representing a time-series with some features. Some features used in [49] and [50] to represent load profiles are average values and peak values. In [49], the feature extraction is claimed to improve the clustering. However, the CVI scores (described in section 2.7.1) are calculated based on different data representations, which do not make them a fair comparison.

Transformations

In many analyses, the shape of the time-series is more important than the actual values [41], [51]. The traditional clustering algorithms mainly cluster based on absolute values and most clustering on time-series use traditional algorithms [41], [38]. The most common way to cluster based on the shape of the time-series instead of their absolute values is to transform the data before doing the clustering. Some of the modifications proposed are:

- Min-max normalization defined as ([1]): $\frac{\vec{x}-x_{min}}{x_{max}-x_{min}}$, where $\vec{x}$ is a sample and x_{max} and x_{min} are the highest and lowest elements in the sample. This normalizes each element in the sample to lie in the range $[0, 1]$.
- Z-transformation defined as ([51]): $\frac{\vec{x}-\mu}{\sigma}$ where μ and σ are the average and the standard deviation. The transformed samples can be both negative and positive and has no limits, but an average of 0.

Average load profiles

With average load profiles, each customer is represented by only one sample in the dataset. For hourly readings, this sample contains 24 elements, which are the average consumption for that customer that hour. This is represented mathematically as:

$$\vec{x} = \sum_{i=1}^{24} \sum_{j=1}^{K} x(i, j) \tag{2.1}$$

where $x(i, j)$ is the consumption hour i on day j, and K is the number of days in the dataset.

2.4.2 Distance measure

The distance measure is the second component in a clustering approach and it provides a quantification of the similarity or dissimilarity between two samples. In literature, distance measures are also commonly called similarity measures. As clustering try to group similar samples together, it is crucial to quantify similarities between samples [51], [42].

First, time-series invariances will be presented before the three different distance measures used in this thesis will be described. For the reader not familiar with distance measures and their computation, examples with graphs are included for each distance measure. This is done to make it easier for the reader to gain more intuitive insight in the characteristics of each distance measure, as the mathematical expressions might look complex at

first sight. If the reader is comfortable with the mathematical expressions, or have previous knowledge about distance measures, the examples may be skipped.

time-series invariance

A good distance measure should be invariant to important distortions for the domain it is intended for [51], [52]. This means that one sample's similarity to other samples should not be affected by distortions considered irrelevant for that domain. Important time-series distortions for clustering on AMS-data are ([1], [53]):

- Scaling and translation invariance: One sample's similarity to other samples should not be affected by its amplitude (scaling) or offset (translation). A transformation of the sample $\vec{x}$ to $\vec{x'} = a\vec{x} + b$ should have the same similarity to other samples.
- Shift invariance: Samples that are similar in shape but shifted in time should be treated as similar.

Other time-series invariances can be found in [51], [52]. Both transformations shown in the previous subsection incorporate scaling and translation invariance. The following presents the three distance measures used in this master thesis.

Euclidean distance

This is the most used distance measure, and it is used in the popular K-Means and many other algorithms [40], [42]. The euclidean distance measure is fast and simple, but does not incorporate shift invariance. The euclidean distance (ED) of two samples, $\vec{x}$ and $\vec{y}$, is given as:

$$ED(\vec{x}, \vec{y}) = \sqrt{\sum_{k=1}^{m} (x_k - y_k)^2} \tag{2.2}$$

Written out by words, the euclidean distance is the square root of the sum of element-wise differences squared. A lower ED-value means more similar load profiles. A value of 0 means two identical load profiles, while the there is no upper limit of the ED.

Euclidean distance example

This example will illustrate the characteristics of the euclidean distance measure. We look at three load profiles and calculate the euclidean distance between them. The three load profiles can be seen in figure 2.2. The orange and green are load profiles made by SINTEF used for calculating the electric consumption not delivered to households during outages (explained in section 2.9.1), so they can be seen as household consumptions. The last one has constant power during the whole period.

First, we calculate the euclidean distance (ED) between load profile 1 and load profile 2 using equation (2.2), and we get an ED of 2.43. This is illustrated in figure 2.3a, where we see the distance between each element for the two load profiles.

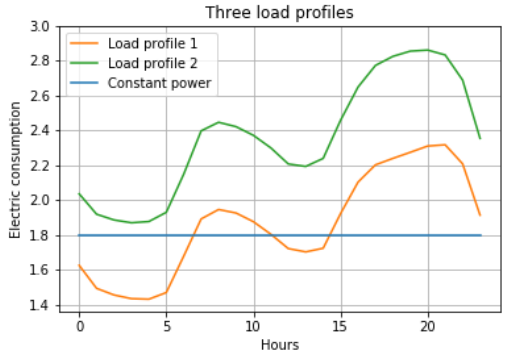
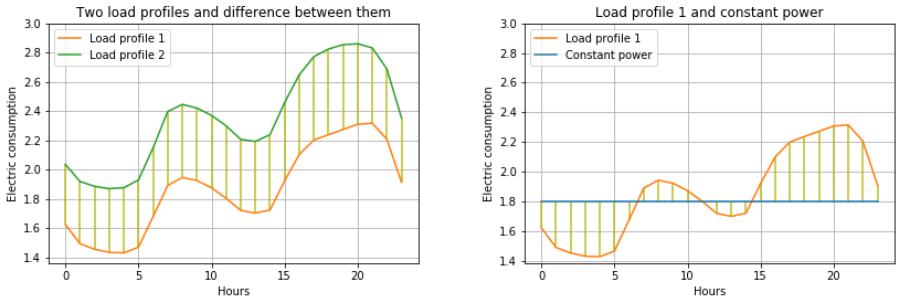


Figure 2.2: The three load profiles used in the euclidean distance example

Then we calculate the euclidean distance between load profile 1 (orange) and the load profile with constant power, and we get an ED of 1.45. This is a reduction of 40 percent, compared with the ED for load profile 1 and 2. We can see this difference by comparing figure 2.3a and 2.3b, which both have the same scale.



(a) Load profile 1 and 2

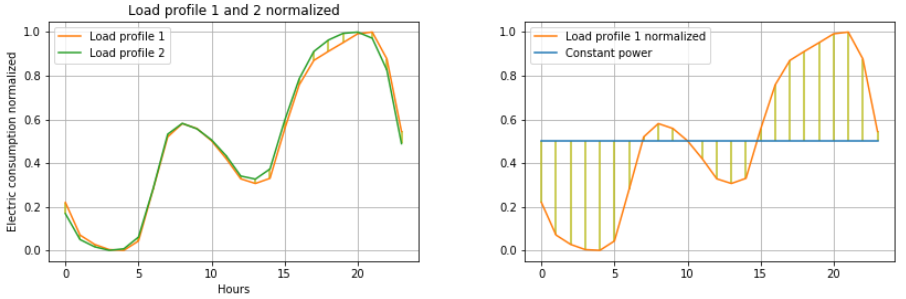
(b) load profile 1 and constant power

Figure 2.3: Comparison of load profiles

A way to incorporate scaling and translation invariance is by transforming the time-series. A comparison of the same load profiles normalized using the Min-Max technique (section 2.4.1) are shown in figure 2.4. We see that the distances between the first and second load profile are almost gone, while the distances between the constant and the first load profile (orange) stay roughly the same.

The results are summarized in table 2.1. With the raw data, the load profile 1 and constant power are most similar using the ED. However, when normalizing the load profiles, load profile 1 and 2 are by far the most similar.

As the two load profiles representing households share many characteristics, this example shows that it might be advantageous to transform the time-series before performing the clustering. This way the scaling and translation invariance is incorporated.



(a) Load profile 1 and 2 normalized

(b) load profile 1 and constant power normalized

Figure 2.4: Comparison of normalized load profiles

	Load profile 1 and constant power	Load profile 1 and 2
Raw data	1.45	2.43
Normalized	1.60	0.15

Table 2.1: Results from the Euclidean distance example

Dynamic Time Warping

The dynamic time warping (DTW) is known as the best distance measure for time-series that incorporates shift invariance [51]. The DTW is a quite old distance measure, presented by Sakoe and Chiba in 1978, for the purpose of voice recognition [54]. The DTW computes the distance between two time-series by stretching and compressing the time-series. This way the distance measure can detect similarity between two time-series even if their patterns are slightly shifted in time.

The DTW algorithm calculates a m by m similarity matrix containing all the square distances between any two points of the samples, $\vec{x}$ and $\vec{y}$. Then the algorithm calculates a warping path $W = [w_1, w_2, ..., w_k]$, where each element w_i contain a pair of elements, (a, b), one from each sample ($\vec{x}$ and $\vec{y}$). One pair can for example be the third element of $\vec{x}$ and the fourth element of $\vec{y}$, represented as $w_i = (3, 4)$. The warping path is contiguous, which means that for each step of the warping path, from w_i to w_{i+1}, there are three possibilities:

1. The x-value increases by one and the y-value stays the same
2. The y-value increases by one and the x-value stays the same
3. Both the x-value and the y-value increase by one

This means that if element i in the warping path is $w_i = (a, b)$, then w_{i+1} will be either:

- $w_{i+1} = (a + 1, b)$
- $w_{i+1} = (a, b + 1)$

- $w_{i+1} = (a+1, b+1)$

This warping path gets calculated to minimize the sum of the distances between the pairwise elements in the warping path, as shown in equation (2.3). The distance $d(w_i)$ is the square distance between the two elements, x_a and y_b, represented by w_i.

$$DTW(\vec{x}, \vec{y}) = min\sqrt{\sum d(w_i)} \tag{2.3}$$

The warping path can be recursively calculated by minimizing the following equation:

$$\gamma(i,j) = ED(i,j) + min\{\gamma(i-1, j-1), \gamma(i-1, j), \gamma(i, j-1)\} \tag{2.4}$$

Normally the warping path gets constricted to a subset of the similarity matrix M [51], [54]. One restriction is the radius of the DTW "band", explained in the following example:

Example 1 DTW

In this example we compare two time-series, $\vec{x}$ and $\vec{y}$, and the radius of the DTW band is r. Then one element in $\vec{x}$, say x_a, can be compared with all the elements in the range $[y_{a-r}, y_{a+r}]$. This means that one element can be compared with a maximum of $2r+1$ elements in the other time-series. It is important to also remember that the DTW path needs to be continuous.

An example with $r = 2$ and two different warping paths is shown in figure 2.5. The green path is comparing elements x_1 with y_1, x_2 with y_2, x_3 with y_3 and so on. However, the orange warping path compares different elements in the two time-series. For the orange warping path, the elements that get compared are shown in table 2.2.

Elements in time-series 1	1	2	2	2	3	4	4	5
Elements in time-series 2	1	1	2	3	4	4	5	5

Table 2.2: Elements compared in the orange warping path from figure 2.5

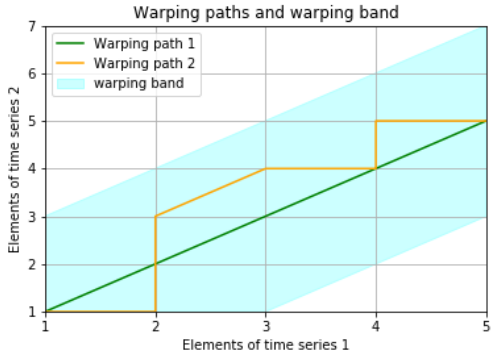


Figure 2.5: The DTW paths from example 1 DTW

In [55] the radius of the warping path for clustering of AMS-data with hourly resolution is set to 1. This means that the consumption in one hour can be compared with the previous hour, the same hour and one hour ahead in the other load profile.

Example 2 DTW

This example shows the DTW's ability to align two time-series with slightly different times on the spikes. Two load profiles with two identical spikes, but slightly shifted in time, are shown in figure 2.6. We compare these two load profiles using ED and DTW.

Figure 2.6: The two load profiles analysed in DTW example 2

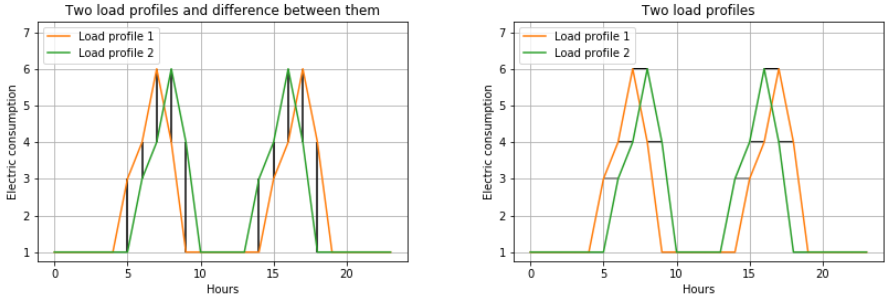
Figure 2.7 shows the graphs and the distances for ED and DTW. The distances between the time-series are the black lines, but be aware that the distances are only calculated in the y-axis. The DTW algorithm can therefore achieve lower distances than the ED algorithm, by comparing elements slightly shifted in time. The ED of the load profiles is 6.63, but the DTW is 0. The DTW obtains a perfect match because the elements in the time-series are the same, but slightly shifted in time. This shows that the DTW incorporates shift invariance.

This example is quite trivial, and in most cases a perfect fit is not obtained. The interested reader is encouraged to look for and do examples of computations to get more insight and hands-on experience with the dynamic time warping computations. Examples can be found online at [56] and [57].

Example 3 DTW

This example shows the importance of choosing the right restrictions on the DTW path. The DTW is a distance measure made for voice recognition, which means that uncritical implementation of the algorithm in other domains might not work so well.

This example uses the two load profiles from the example with ED and calculates the DTW score with no restrictions on the path, and a restriction of the radius of the DTW



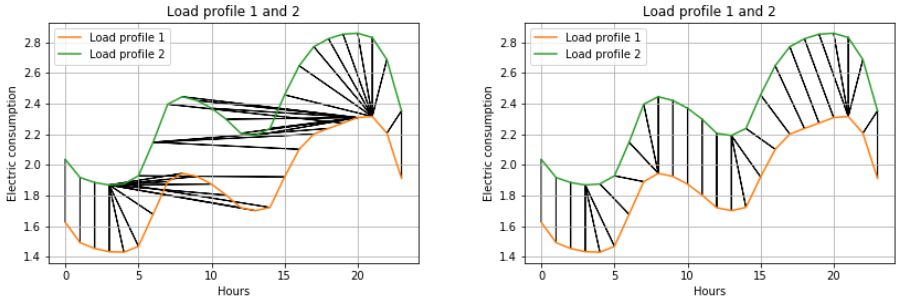
(a) Euclidean distances

(b) DTW distances

Figure 2.7: Comparison of normalized load profiles

band on 2.

The two computations are visualised in figure 2.8, where the black lines connect the elements that are compared in the computation. Without any restrictions of the dynamic warping path, we get the computation as in figure 2.8a, and with a restriction of the band on 2, we get the computation shown in figure 2.8b. The figures show the ”one to many” and ”many to one” principle that DTW is based on, that one element in one of the time-series can be compared to more than one element in the other time-series.



(a) DTW with no restriction

(b) DTW with restriction r = 2

Figure 2.8: Comparison of DTW computation with two different restrictions

The impact of the ”one to many” and ”many to one” effect can, as suggested in the literature, be minimized by adding restrictions to the warping path. By adding restrictions, single elements get less compared with elements of the other time-series, which reduces the effect of single elements. However, it can be still be seen in some degree that peak values get compared more to the other time-series.

If the data gets normalized before the DTW gets computed, the results gets very different. By maintaining no restrictions, we get the result as in figure 2.9. Here, only one element in each time-series gets compared with two elements in the other time-series. As the maximum ”shift” is only 1, the result with restrictions, even a bandwidth of 1, will be

the same as without any restrictions.

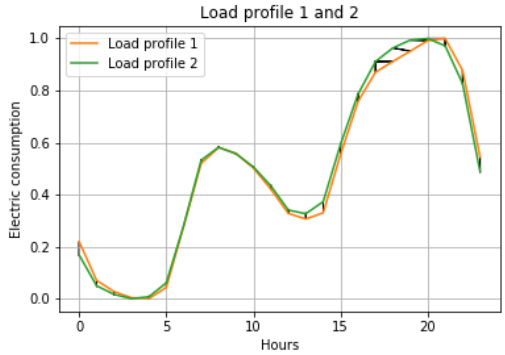


Figure 2.9: Comparison of normalized load profiles

Shape-Based Distance

This distance measure incorporates shift invariance and is computationally much less expensive than DTW [51], 2.4.2 .

The shape-based distance (SBD) uses the cross-correlation between two samples to incorporate the shift invariance. The cross-correlation is calculated by sliding one sample over the other and calculating the inner product for each shift. A shift of a sample is given as:

$$\vec{x}_{(s)} = \begin{cases} (0, 0, ..., 0, x_1, x_2, ..., x_{m-s}), & s \geq 0 \\ (x_{1-s}, ..., x_{m-s}, 0, ..., 0), & s < 0 \end{cases} \tag{2.5}$$

where s is the number of zeros in the beginning or end of the vector. The cross-correlation CC is calculated for all possible shifts $s \in [-m, m]$ as:

$$CC_w(\vec{x}, \vec{y}) = R_{w-m}(\vec{x}, \vec{y}), \quad w \in [1, 2..., 2m-1] \tag{2.6}$$

where

$$R_k(\vec{x}, \vec{y}) = \begin{cases} \sum_{l=1}^{m-k} x_{l+k} \cdot y_l, & k \geq 0 \\ R_{-k}(\vec{y}, \vec{x}) & k < 0 \end{cases} \tag{2.7}$$

The coefficient normalization is calculated as:

$$NCC_c = \frac{CC_w(\vec{x}, \vec{y})}{\sqrt{R_0(\vec{x}, \vec{x}) \cdot R_0(\vec{y}, \vec{y})}} \tag{2.8}$$

and takes values between -1 and 1. The SBD is calculated as:

$$SBD(\vec{x}, \vec{y}) = 1 - \max_{w} NCC_c \tag{2.9}$$

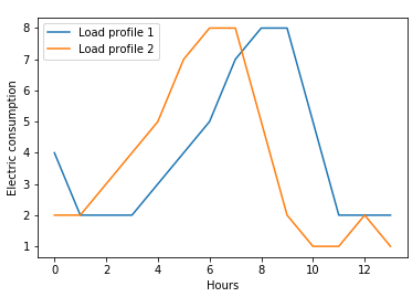
where $\max_w NCC_c$ is the NCC_w calculated with the w that gives the largest CC_w. Since NCC_c lies between -1 and 1, SBD will lie between 0 and 2. The lower the value is, the more similar the time-series are. An efficient computation of SBD is given in [51].

SBD example

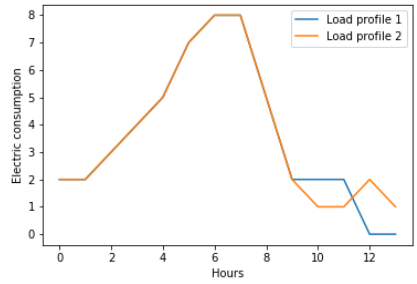
This example shows the computation of the SBD. Two load profiles, shown in figure 2.10a, are analysed. The values of the two load profiles are shown in table 2.3.

Hour	0	1	2	3	4	5	6	7	8	9	10	11	12	13
Load profile 1	4	2	2	2	3	4	5	7	8	8	5	2	2	2
Load profile 2	2	2	3	4	5	7	8	8	5	2	1	1	2	1

Table 2.3: Values of the load profile used in the SBD example



(a) Load profile 1 and 2 without any shifts



(b) Load profile 1 with a shift of s = -2 and load profile 2 with no shift

Figure 2.10: Load profiles used in the SBD example with different shifts

Now we start the computation of the shape-based distance. Load profile 1, called LP_1, is shifted for values $[-3, 3]$ in equation (2.10). (In the computation of SBD, all possible shifts are calculated, but it is sufficient to look at only a subset of the possible shifts to illustrate the computation). We calculate the shifts using using equation (2.5) and we get the shifts shown in equation (2.10).

Shift (on LP_1)	s = 3	s = 2	s = 1	s = 0	s = -1	s = -2	s = -3
cross-correlation value (CC)	146	167	195	234	259	268	253

Table 2.4: cross-correlation values for different shifts

$$LP_1 = \begin{cases} (0,0,0,1,4,2,2,2,3,4,5,7,8,8,5), & s = 3 \\ (0,0,1,4,2,2,2,3,4,5,7,8,8,5,2), & s = 2 \\ (0,1,4,2,2,2,3,4,5,7,8,8,5,2,2), & s = 1 \\ (1,4,2,2,2,3,4,5,7,8,8,5,2,2,2), & s = 0 \\ (4,2,2,2,3,4,5,7,8,8,5,2,2,2,0), & s = -1 \\ (2,2,2,3,4,5,7,8,8,5,2,2,2,0,0), & s = -2 \\ (2,2,3,4,5,7,8,8,5,2,2,2,0,0,0), & s = -3 \end{cases} \tag{2.10}$$

Then the cross-correlation CC is calculated for each shift according to equation (2.6) and (2.7). The computation of the cross-correlation for $s = 0$ is shown in equation (2.11), where LP_1 is load profile 1 and LP_2 is load profile 2.

$$\begin{aligned} CC = R_0(\vec{LP_1}, \vec{LP_2}) = \sum_{l=0}^{13} x_l \cdot y_l = 1 \cdot 2 + 4 \cdot 2 + 2 \cdot 2 + 2 \cdot 3 + 2 \cdot 4 + 3 \cdot 5 \quad ... \\ + 4 \cdot 7 + 5 \cdot 8 + 7 \cdot 8 + 8 \cdot 5 + 8 \cdot 2 + 5 \cdot 1 + 2 \cdot 2 + 2 \cdot 1 + 2 \cdot 1 = 234 \end{aligned} \tag{2.11}$$

Table 2.4 shows the cross-correlation value for all the shifts calculated in this example. It shows the highest value for s = -2. Figure 2.10b shows load profile 1 with a shift of $s = -2$ and load profile 2 with no shift.

The next step is to calculate the coefficient normalization. To do that we need the inner product for each of the samples, LP_1 and LP_2. By using equation (2.7), we calculate the inner products, $R_0(\vec{LP_1}, \vec{LP_1}) = 292$ and $R_0(\vec{LP_2}, \vec{LP_2}) = 271$. The coefficient normalization is then calculated using equation (2.12)

$$NCC_c = \frac{CC(\vec{x}, \vec{y})}{\sqrt{R_0(\vec{x}, \vec{x}) \cdot R_0(\vec{y}, \vec{y})}} = \frac{268}{\sqrt{292 \cdot 271}} = 0.95 \tag{2.12}$$

Then the value of the shape-based distance is calculated by using equation (2.9):

$$SBD(\vec{LP_1}, \vec{LP_2}) = 1 - \max_w NCC_c = 1 - 0.95 = 0.05 \tag{2.13}$$

A value of 0.05 indicates a high similarity between the load profiles. This can be confirmed from figure 2.10b, where the values for hour 0-8 are identical.

Comparison DTW and SBD

While both SBD and DTW incorporate shift invariance, there are some differences. As the DTW can be both compressed and stretched it can incorporate multiple shifts, while SBD

can only incorporate discrete shifts of the whole time-series. The SBD might ignore some elements in the time-series, but it compares every element in one time-series with only one element in the other time-series. On the other hand, the DTW uses all the elements, but might compare one element in one time-series to more elements in the other time-series. This might put more emphasis on single elements in DTW computation than SBD computation.

2.4.3 Centroid extraction

The third component of the clustering is the centroid extraction, which is often closely related to the distance measure.

A central question in cluster analysis is how to extract the cluster centroid. The most common objective function to minimize in a cluster partition is ([51]):

$$P = min \sum_{j=1}^{k} \sum_{\vec{x_i} \in c_j} dist(\vec{x_i}, \vec{c_j})^2 \tag{2.14}$$

Recall that k is the number of clusters, $\vec{x_i}$ is a sample, c_j is cluster j with the corresponding centroid $\vec{c_j}$. For each distance measure, the optimal centroids are found by inserting the distance measure in equation (2.14).

Euclidean distance

The euclidean distance is given by equation (2.2). By inserting this distance in equation (2.14), the optimal centroid is found to be the arithmetic mean of all the samples in the cluster, given by equation (2.15). This is proven mathematically in the appendix of the specialization project leading up to this master thesis [58].

$$\vec{c_j} = \frac{\sum_{\vec{x_i} \in c_j} \vec{x_i}}{|c_j|} \tag{2.15}$$

Recall that $|c_j|$ is the total number of samples in cluster j.

Dynamic time warping

In theory, by inserting the DTW distance (equation (2.3)) into equation (2.14) and minimize the equation, we get the optimal cluster centroids. However, due to the characteristics of the DTW, this minimization is not a trivial task [51], [59]. Many approaches to this minimization are proposed, and the most efficient and accurate seems to be the DBA computation [51], [59]. This technique is presented in [59], and is a heuristic method which iteratively refines the cluster centroid, initially picked as one sample in the cluster.

Shape-based distance

The cluster centroid extraction is a bit different for the SBD approach. As the SBD computation maximize the similarity, the centroid computation of SBD also aims to maximize

similarity. Each cluster centroid is maximized according to equation (2.16). An efficient computation of the cluster centroids is given in [51].

$$\vec{c_j} = \max \sum_{\vec{x_i} \in c_j} NCC_C(\vec{x_i}, \vec{c_j})^2 \tag{2.16}$$

2.5 Clustering algorithms

This section presents clustering algorithms, the last clustering component from the structure presented figure 2.1. Clustering algorithms are divided into methods, which describe the general structure to solve a problem [40]. The most tested clustering methods on time-series are [41], [38]:

- Partitional clustering
- Hierarchical clustering
- Other clustering approaches

The main focus of this thesis has been on partitional algorithms, because of their high computational speed and accuracy [51], [39]. However, the clustering literature emphasise that clustering is an exploratory tool and recommend testing multiple algorithms [40], [42], [43]. Therefore, some more algorithms have been tested and are described in the following subsections.

2.5.1 Partitional methods

A partitional method partitions the data in multiple clusters, usually by optimizing a criterion function. The criterion function is based on the distance measure, and the most common objective function was given in equation (2.14). Some distance measures are presented in section 2.4.2.

The literature shows that partitional methods are commonly used with clustering on time-series and AMS-data. Partitional methods got the best results in [39] and [51], and other articles ([46], [53], [47]) only use partitional methods. An advantage of partitional methods are the lower computational cost compared to other methods [51], [42].

The simplest partitional algorithm is the K-Means algorithm [40]. This is the most used clustering algorithm and despite being a very old algorithm, it is performing surprisingly well compared to newer algorithms [40], [42]. It is also among the fastest clustering algorithms [51], [42].

The following presents the partitional algorithm used in this master thesis:

K-Means

The algorithm starts with a random assignment of all the samples in the dataset to k-clusters and then reassign the samples to different clusters in an iterative process until convergence is achieved [42]. The reassignment process is based on the distance between the samples and the cluster centroids. Convergence is achieved when a certain number

of iterations is reached or when no samples are reassigned during one full iteration. The distance measure used in the original K-Means algorithm is the euclidean distance and the way to calculate the cluster centroid is to take the average of the samples in the cluster.

The algorithm can be summarized in four steps:

- Randomly assigning the samples to the clusters and calculating the cluster centroids.
- For each sample in the dataset, calculate the distance to all of the centroids and put the sample in the cluster with the least distance.
- Update the clusters centroids
- Repeat the two previous steps until convergence.

Note that the algorithm needs a prespecified number of cluster, k, as input. Ways to find the optimal number of clusters are given in chapter 2.7.

Many improvements to the old K-Means approach have been proposed [40]. The algorithm is frequently proposed with other distance measures, for example DTW and SBD (used in K-Shape). The K-Means algorithm is also proposed with a new centroid computation in K-medoid, where the clusters centroids are calculated as the median of the cluster instead of the average [40]. Other extensions of K-Means can be found in [40].

2.5.2 Hierarchical clustering

This clustering method creates a hierarchical structure of the data, either with a bottom-up approach or a top-down approach [42]. The bottom-up approach starts by assigning all the samples in their own clusters and merge clusters together based on their distance to other samples to make bigger clusters. The top-down approach does the opposite, it starts with all samples in one cluster and then divide the clusters into sub-clusters. The most common approach, and the fastest, is the bottom-up approach [42], [39].

The results of a hierarchical algorithm depend on the distance measure (section 2.4.2) as well as the linkage criterion [42]. The linkage criterion decides how clusters are merged. The result of the clustering partition is closely linked to the linkage criterion, where more complex criteria tend to make average size groups, while simple criteria tend to isolate outliers [42], [39], [38]. A simple linkage criterion, yet commonly used, is the single linkage criterion [42]. This criterion merges the clusters with the least distance between any pair of samples (one sample from one cluster and the other sample from the other cluster). A more complex criterion is the ward criterion [39]. This criterion uses something called within-cluster distance. This is the sum of all distances between the cluster centroid and all of the samples in that cluster. The criterion for each merging is to minimize the increase of the sum of the within-cluster distances.

A major drawback for hierarchical algorithms, compared to partitional algorithms, is the high computational time [51], [42]. The most promising results for hierarchical algorithms on time-series of AMS-data is for detecting outliers [39], [38]. In [39], the single linkage criterion clustered together streetlights, which were wrongly categorized as households. As wrong categorization is a problem for DSOs [39], hierarchical algorithms are interesting despite the high computational cost.

2.5.3 Different clustering approaches

Many clustering approaches have been tested on time-series and AMS-data [41], [38]. Although the most common approaches are partitional and hierarchical methods, most clustering approaches have also been tested on time-series and AMS-data [38], [41]. One of the most popular algorithms on electric load profiles which is not partitional or hierarchical is the self-organising maps algorithm (SOM), which is presented next. For a overview of other algorithms used on time-series and AMS-data, [41], [38] and [16] can be useful sources.

Self-organising maps

The self-organizing maps algorithm was presented by Kohonen in 1982 [60]. The foundation of the algorithm is the physical structure of memory in living organisms. The algorithm uses a neural structure to map the input data onto a map, often of two dimensions [61]. The aim of the algorithm is to fit the neural structure into the input data, which is done by assigning each neuron in the neural map a weight. The weight of each neuron is updated in an iterative training process, and the more training the better fit.

As the cluster centroids are closely linked to the SOM-algorithm, the cluster centroid representation is presented here. The centroids of the self-organizing maps are made based on the neurons made from the clustering [62]. The K-Means algorithm is used on the neurons, making neural centroids. Then, each sample is given a cluster label based on the nearest neural centroid.

The details of the computation of the algorithm is out of the scope for this master thesis. The interested reader can find details in the originally paper from 1982 ([60]), or newer publications on the topic, like the Springer series from 2012, also written by Kohonen [63].

2.6 Cluster validation indexes

An important question in clustering is how to evaluate the quality of a clustering partition. This can be done through clustering validation indexes (CVIs). In the literature, it is often distinguished between internal and external validation techniques [43], [39]. The external techniques compare external grouping information of the dataset with the clustering labels from the clustering to quantify the clustering quality. Internal techniques quantify the quality of the clustering based on the distances between the samples and clusters in the dataset. This master thesis will focus on internal CVIs, as customers from different customer groups might have same characteristics and therefore desired to be put in the same cluster. If the goal, however, is to see how well clustering manages to categorize the customers in the dataset compared to the already existing groups, external CVIs would be the preferred choice.

Most of the internal CVIs, and the three presented in the following, express the quality of the clustering as a combination of cohesion and separation of the clusters [43]. With cohesion means the tightness of a cluster, in other words the distance between the samples in the cluster. With separation means the distance between the clusters. In literature cohesion

is also called intra-variance or within-variance and separation is also called inter-variance or between-variance [43]. A good clustering should have the cohesion as low as possible and the separation as high as possible.

As explained in chapter 2.4, there is no best clustering algorithm for all situations. A consequence of this is that there is no best clustering validation technique for all situations either [43], [44]. In literature, a solution to this is to use multiple CVIs [39], [44]. The three indexes explained in the following subsections are all well-established [44].

Some general criteria or guidelines for validating a cluster partition by manual inspection can be found. In [39], the criteria that a cluster partition should fulfil are:

- Compactness: Clusters should be compact, which means consist of similar samples.
- Differentiable: Clusters should be possible to distinguish from each other. Clusters with indistinguishable characteristics should have been merged.
- Substantial: Clusters should be large enough to gain benefits from. This means to be able to use the cluster in later analyses or to gain knowledge of the group of samples in that cluster. A cluster containing very few samples might not fulfil this criterion.
- Stable: The clustering partition should be stable over time. This means that running the algorithm multiple times should give similar results.

These criteria might be helpful when analysing the clusters. Some of the criteria can be quantified through CVIs, for example the compactness and differentiable. But in the literature, papers show that the cluster partitions on AMS-data with the best CVI scores might not fulfil the differentiable or the substantial criteria [39], [38]. These papers recommend to also analyse the final cluster partition manually. Therefore, it can be important to have the criteria listed in mind when analysing the clusters.

Worth noting is also that the distance measure is included in the CVIs. This means that cluster partitions with different distance measures are not comparable. However, for algorithms using the same distance measure, a CVI quantify the quality of the partition, and the partitions with the best score should be the best partition. It is also worth noting that the CVI scores are calculated on a dataset using one (or no) transformation of the dataset. Therefore, the results using one transformation will not comparable to other transformations, as the differences between the elements in the transformed spaces are different.

In the following the three CVIs will be explained:

2.6.1 Silhoutte index

This CVI got the best results out of 30 CVIs in [43] when testing on a synthetic dataset and it was the CVI used in ENERGTYICS [64].

The CVI was presented in [65], and the formula for the silhouette index is:

$$Sil(C) = \frac{1}{N} \sum_{c_j \in X} \sum_{\vec{x_i} \in c_j} \frac{b(\vec{x_i}, c_j) - a(\vec{x_i}, c_j)}{max\{a(\vec{x_i}, c_j), b(\vec{x_i}, c_j)\}} \tag{2.17}$$

Where

$$a(\vec{x_i}, c_j) = \frac{1}{|c_j|} \sum_{\vec{x_j} \in c_j} d_e(\vec{x_i}, \vec{x_j})$$

$$b(\vec{x_i}, c_j) = \min_{c_l \in C \setminus c_j} \left\{ \frac{1}{|c_l|} \sum_{\vec{x_j} \in c_l} d_e(\vec{x_i}, \vec{x_j}) \right\}$$

where $c_l \in C \setminus c_j$ means for all clusters except for cluster j and N is the number of samples in the dataset. In equation (2.17), a is the average distance between all samples in cluster c_j, and b is the average distance to all samples in the closest cluster to $\vec{x_i}$. The cohesion is based on the average distance between all samples in the cluster and the dispersion is based on the average distance to the nearest cluster. The SI score is maximum 1 and minimum -1. The higher SI score, the better prediction.

2.6.2 Calinski-Harabasz

This index got the best results in [66] and the second-best result in [43] for synthetic datasets. The formula for the Calinski-Harabasz (CH) index is [43]:

$$CH(C) = \frac{N-k}{k-1} \cdot \frac{\sum_{c_j \in C} |c_j| d_e(\vec{c_j}, \vec{X})}{\sum_{c_j \in C} d_e(\vec{x_i}, \vec{c_j})} \tag{2.18}$$

where

$$\vec{X} = \frac{1}{N} \sum_{\vec{x_i} \in X} \vec{x_i}$$

where N is the number of samples in the dataset, k is the number of clusters, $|c_j|$ is the number of samples in cluster j and $\vec{X}$ is the average sample in the dataset.

The cohesion is based on the distance from all samples in a cluster to its cluster centroid. The separation is based on the distance from the cluster centroids to the global cluster ($\vec{X}$). The higher the CH score is, the better quality on the clustering.

2.6.3 Davies-Bouldin

This is one of the most used CVIs ([43]) and is used in [39], [44] and [38]. The formula for the Davies-Bouldin (DB) index is [43]:

$$DB(C) = \frac{1}{k} \sum_{c_j \in C} \max_{c_l \in C \setminus c_j} \left\{ \frac{S(c_j) + S(c_l)}{d_e(\vec{c_j}, \vec{c_l})} \right\} \tag{2.19}$$

Where

$$S(c_j) = \frac{1}{|c_j|} \sum_{x_i \in c_j} d_e(x_i, \vec{c_j})$$

The cohesion is based on the distance from all the samples in a cluster to its centroid and the separation is based on the distance between two cluster centroids. The lower the DB score is, the better the quality is.

2.7 Estimating the number of clusters

A central problem in cluster partitions is how many clusters is the optimal number in a partition [40]. Most clustering algorithms require a prespecified number of clusters before starting the clustering. Two options will be presented in the following:

2.7.1 CVI scores

This option is based on the CVI indexes presented in the previous section. The optimal cluster number is found by running the clustering algorithm multiple times with different number of clusters, calculate the CVI score each time and choose the clustering with the highest CVI score. This procedure is commonly done in literature, for example in [64], [49] and [47].

2.7.2 Gap statistics

The Gap Statistics is a method to estimate the number of clusters in a dataset and was proposed by Tibshirani in year 2001 [67]. The method calculates the average distance between all samples within the clusters in the whole dataset for different values of k, the number of clusters. Mathematically, this is represented as:

$$W_k = \sum_{j=1}^{k} \frac{1}{|c_j|} \sum_{i,l \in c_j} d(\vec{x_i}, \vec{x_l}) \tag{2.20}$$

where c_j is cluster number j and $|c_j|$ the number of samples in cluster j. The gap is then calculated as the difference between the expectation of $log(W_k)$ from an reference distribution and the actual $log(W_k)$ observed. This is represented mathematically as:

$$Gap_n(k) = E_n^*\{log(W_k)\} - log(W_k) \tag{2.21}$$

where $E_n^*\{log(W_k)\}$ is the expectation of $log(W_k)$ from the reference distribution. For the calculation of the reference distribution it is referred to [67].

The principle of the gap statistic is that for data well separated with K clusters, $log(W_k)$ will decrease faster than its expectation ($E_n^*\{log(W_k)\}$) for $k \leq K$ and it will decrease slower than its expectation for $k > K$. From equation (2.21) we can then draw the conclusion that the gap is the biggest for $k = K$. For more details about the reference distribution as well as the exact computation of the optimal cluster number k, see [67].

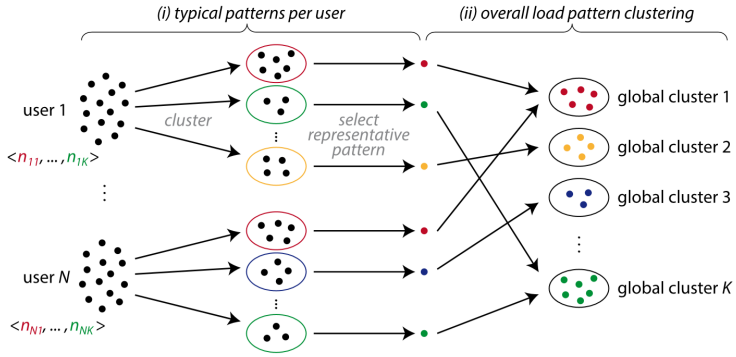


Figure 2.11: Example of a two-step model. Figure taken from [1]

2.8 Multiple step approaches

Cluster analysis on load profiles differ not only in the algorithms used, but also in the number of clustering steps used. The literature reveals both one-step approaches ([44], [39], [53]) as well two-step approaches ([1], [68], [47]). By a one-step approach only one clustering is performed on the dataset, while in a two-stage approach, the second clustering step uses the centroids from the first clustering step.

A two-step approach for clustering of AMS-data is normally done by first finding the typical load patterns for each customer and then cluster these to global clusters. This is shown in figure 2.11.

The two-step clustering approach have some advantages over the single step approach, listed in the following ([1], [68], [47]):

- Computational speed: The first stage can be carried out in parallel, which can be significantly faster on large datasets. And if the first step is already done, the second step will be considerable faster than doing the whole clustering in a one-step approach.

- Size reduction: Performing two steps instead of one significantly decreases the size of the datasets used in the clustering steps.

- Analysis of individual customers: By doing the clustering in two steps, it is possible to stop after one step to analyse individual customers.

2.9 Prediction of load profiles

This section presents the average load profiles being used today as well as a discussion of future load profiles.

2.9.1 Prediction of load profiles today

The current approach to estimate load profiles in Norway during outages is described in [69]. This method based on average load profiles was developed in the 90's, and is being used by Norwegian DSOs today [70]. The method divides customers into categories, and each category has four different load profiles. Two of the load profiles for each category are assigned for a high consumption period (from December to February) and the two other load profiles are assigned for a low consumption period (from March to November). For each consumption period, one load profile represents weekdays and the other represents weekends. Each load profile has two parameters, A and B. Parameter A is a temperature dependent variable and parameter B is temperature independent. For each load profile, the following equation is used to calculate the consumption for each hour h:

$$P_h = (A_h \cdot T + B_h) \cdot \frac{W}{W_N} \tag{2.22}$$

where T is the average temperature the given day, W is energy consumed the last measuring period for the customer and W_N is the calculated energy consumption the last measuring period corrected by temperature for the given customer. The procedure for calculating $\frac{W}{W_N}$ is given in [69].

2.9.2 Future load profiles

There are of course many factors that influence customers electricity consumption. Some of these factors were mentioned in section 2.3, where network tariffs and hourly billing were discussed. The new network tariffs proposed in Norway gives incentives for the customer to reduce its peaks. However, new electrical appliances, with electrical cars as a common example, is expected to increase the peaks in the grid [33]. Examples of other factors analysed are the building's energy efficiency and the rebound effect [71] and electricity reduction due to greater insight in own consumption [72].

In literature, demand side management (DSM) is a common motivation for clustering. DSM means actions designed to reduce the cost associated to electric consumption [73]. In [50], determinants for the sensitivity of DSM is analysed. This means how likely people are to change their consumption due to incentives. This is crucial to design demand response programs most efficiently [50]. By incorporating the sensitivity in the cluster analysis, the effects of the demand response programs can more accurately be analysed [50].

2.10 Temperature corrections

This sections presents the approach used by Norwegian DSOs to temperature correct energy and power consumption.

2.10.1 Temperature corrections

The temperature has in Norway a large impact on the electricity consumption in households [19]. This is because Norwegian households' main source for room heating is elec-

tricity [19]. In order to see trends on consumption affected by the temperature, it is important to do temperature corrections [74].

Common practice among the Norwegian DSOs today is to correct the energy as well as the maximum power for temperature [75]. The energy is corrected for by a method called "graddagsmetoden" in Norwegian, which compares the temperature in a period, normally a year, by the normal temperature for that period [74]. The method is translated as "degree method", which will be used in the following. A value called "degree days" is calculated as the the number of days where the temperature is lower than 17 degrees multiplied by the difference between the average temperature the given day and 17 degrees. If a day has an average temperature of 12.5 degrees, then that day will contribute with 17 - 12.5 = 4.5 "degree days". Then each day of the year is summarized, and we get the "degree days" for a given year. This value is compared to the normal "degree days" for a given area, found for the period 1981-2010 [75]. The values for the "degree days" for different areas from 2010 - 2019, as well as the normal degree days can be found on ENOVA's home page [76].

The "degree day" method assumes a temperature dependent and a temperature independent electricity consumption. Different building categories have different temperature dependent coefficients, and Enova publishes estimated coefficients, which can be found at [77]. The most common temperature dependent coefficient is 0.5, but different DSOs use different coefficients [78], [79], [80], [75]. The formula for correcting for temperature is:

$$E_{corrected} = E_{measured}\left\{(1-k) + k\frac{DD_{normal}}{DD_{year}}\right\} \tag{2.23}$$

where k is the temperature coefficient, DD_{normal} is the normal degree days and DD_{year} is the degree days for a given year.

The common practice is to correct the maximum power also, which is done in a simpler way than the energy [75], [78], [79]. The power is multiplied by a factor, which depends on the temperature measured compared to a reference temperature. The temperature measured is the average temperature over three days, the two previous days as well as the given day. The reference temperature is the minimal average temperature over three days the last 10 years, which is called the extreme minimum temperature. The factors used for some big electric utilities in Norway varies from 1 to 1.7 % for each degree in difference [78], [79], [80]. The following equation is used for calculating the power corrected by temperature:

$$P_{corrected} = P_{measured} + k \cdot diff \cdot P_{measured} \tag{2.24}$$

where k is the factor and $diff$ is the difference between the extreme minimum temperature and the temperature measured.

Be aware that the normal practice only is to temperature correct the maximum power measured [79]. Therefore, other times might have a higher temperature corrected consumption than the maximum power measured temperature corrected [79].

There are large uncertainties about the values of the temperature sensitivities [81]. Especially the last years the uncertainty has grown, due to new appliances and customer habits [81]. One of the biggest DSOs in Norway is using a temperature coefficient based on analyses performed in 1991 [80]. By analysing AMS-data, these temperature coefficients can be updated to better reflect the temperature sensitivities in the grid today.

Chapter 3

Methods

This chapter presents the methods used and the experimental settings in this master thesis. The methods are based on the theory presented in chapter 2, and the results are shown in chapter 4 and discussed in chapter 5. Section 3.1 starts the chapter by summarizing the techniques used in this thesis and section 3.2 describe the experimental settings. A description of the clustering approaches tested is given in section 3.3, and section 3.4 presents the comparison of the cluster results with the practice used today. The last section, section 3.5, presents the analyses of capacity in the grid, included to show other possibilities with AMS-data than clustering.

3.1 Clustering components

This section summarizes the four clustering components presented section 2.4 and 2.5.

3.1.1 Time-Series representation

The first component is the representation of the data, called time-series representation. The analyses in the master thesis are based on different time-series representations, which can be divided into three categories, with a total of 5 techniques.

- Dimentionality reduction: Transformation techniques to reduce the dimension of the dataset. Used to reduce the size of the dataset, thus reducing the computation time in the clustering algorithms. Two techniques are used in this thesis, Principal Component Analysis (PCA) and discrete wavelet transformations.
- Transformations: Transforming the data without reducing the dimensionality. Done to make a meaningful comparison of samples with different offset and amplitude. Two techniques are used in this thesis, the first one called Z-transformation and is defined as $\frac{\vec{x}-\mu}{\sigma}$ where μ and σ are the average and the standard deviation. The second one is called Min-Max (MM) normalization and is defined as: $\frac{\vec{x}-x_{min}}{x_{max}-x_{min}}$.

The second transformation technique transform each element in the range [0,1], while the first technique has no positive or negative limits.

- Average load profile representation: Each costumer is represented by one average load profile, which is the average consumption each hour for that customer. Equation (2.1) is used to get the average load profiles.

3.1.2 Distance measures

The second component is the distance measure, which quantify the similarity or dissimilarity between two samples. The analyses in the master thesis has been done with three different distance measures (described in detail in section 2.4.2):

- Euclidean Distance (ED): Defined as $ED(\vec{x}, \vec{y}) = \sqrt{\sum_{k=1}^{m}(x_k - y_k)^2}$
- Dynamic time Warping (DTW): Defined as $DTW(\vec{x}, \vec{y}) = min\sqrt{\sum w_i}$
- Shape-Based Distance (SBD): Defined as $SBD(\vec{x}, \vec{y}) = 1 - \max_w NCC_c$

3.1.3 Cluster centroids

The third component is the cluster centroids. The cluster centroid computation is explained for the three distance measures in section 2.4.3 and for SOM in section 2.5.3. As centroid computation is closely linked to the distance measure and algorithm, they are not summarized more in this section.

3.1.4 Clustering algorithms

The last component is the algorithm used in the clustering approach. The analyses in the master thesis are based on different clustering algorithms and different techniques. The algorithms are:

- K-Means: The classic and robust clustering algorithm. This algorithm will be tested with the distance measures ED, DTW and SBD (called K-Shape).
- Self-Organising Maps (SOM): Clustering approach which constructs an internal neural network which is self-organizing.
- Hierarchical clustering: Clustering approaches using a linkage criterion to merge clusters together. The hierarchical approaches tested are bottom-up algorithms with single and ward linkage criterion.

3.1.5 Cluster centroid representation

A central question with clustering is the centroid representations after the final clustering. As some of the techniques includes transformations, it could be easier to draw conclusions from a representation of the cluster centroids without the transformation. Therefore, in some cases, an additional representation of the centroids based on the ED centroid computation with the original dataset (not transformed) is provided. Recall that the ED centroid computation takes the average of all the samples in the cluster.

3.2 Experimental settings

This section describes the datasets, CVI computations and the software packages used in the analyses.

The analyses have been carried out on a shared SINTEF server with 24 physical cores, 512GB RAM and a NVIDIA Tesla K40c processor.

3.2.1 Datasets

The clustering analyses have been done on four different datasets. Three of the datasets contain the same customers, but have readings from different months. All the datasets contain hourly readings connected to a customer through a customer ID and a time stamp for the reading. The four datasets are presented in the following:

- Dataset 1: Contains 1227 customers and have readings over 2 months in the winter (December + January). Contain the same customers as dataset 3 and 4. The dataset contains 9 different tariff groups, and each customer is labeled with a tariff group. The two most frequent tariff groups are: Households, with around 60 % of the customers and holiday houses and cabins with around 20%.
- Dataset 2: Same dataset as dataset 1, for the summer (July + August).
- Dataset 3: Same dataset as dataset 1, for the fall (September, October and November).
- Dataset 4: Contain 4681 customers and have readings over 3 months (March - May). All customers are households.

Each reading is the consumption of the last 60 minutes. This means that a reading at 14:00 is the consumption between 13:00 and 13:59.

3.2.2 Time-series samples

As explained in section 2.4, clustering takes a dataset containing n samples and divide the samples into k clusters. For all the clustering in this master thesis, each sample consist of 24 elements, where each element in the sample represent the electric consumption (in kWh) for the corresponding hour. For all the approaches except for average load profiles, one customer has the number of samples equal to the number of days in the dataset. This means that one sample is the electric consumption over one day for one customer.

3.2.3 CVI scores

The CVI scores in this thesis are calculated using the Z-transformed datasets. This is chosen as 9 out of the 11 approaches presented in section 3.3 perform the clustering on the Z-transformed datasets. In the results, the silhouette index is abbreviated SI, Calinski-Harabasz abbreviated CH and Davies-Bouldin abbreviated DB. To make the comparison between the CVI scores easier, all scores will be divided by the corresponding CVI score for K-Means. This means that a DB score below 1 and a SI and CH above 1 indicate higher clustering quality, and a DB score above 1 and SI and CH score below 1 indicate lower quality.

3.2.4 Software packages used

The clustering performed in this thesis has been using publicly available software packages. The K-Means, hierarchical algorithms and PCA algorithms are imported from sklearn [82], K-Shape from tslearn [83], SOM from Sompy [62], and discrete wavelet transformation from pyWavelets [84].

3.3 Tested clustering approaches

This section presents the tested clustering approaches in this master thesis. Section 3.1 summarizes the clustering components presented in the theory. In the following of the thesis, the combination of a clustering algorithm and a distance measure will be called one clustering algorithm. This is done to keep in line with the expressions commonly used in the literature of clustering on AMS-data as well as make it easier to follow the results and discussions. One clustering approach is defined as the combination of a data representation technique and a clustering algorithm.

To overcome the large amount of clustering approaches possible, this master thesis follows an approach inspired by [41]. This article suggests to test every new approach against the well-known K-Means algorithm. Therefore, the K-Means algorithm using the Z-transformation is used as a base throughout the testing, and results are compared against the K-Means algorithm. The Z-transformation is chosen as a base time-series representation as it is more used in the literature than the MM-transformation.

In this master thesis the 11 approaches shown in table 3.1 are analysed. The table also shows the way the approach is written in the table in the results, as the descriptions get long for some approaches. All of the approaches use the Z-transformed datasets, except for the K-Means + Min-Max approach, which uses the Min-Max normalized datasets. All the approaches will be running on all of the datasets with the optimal cluster numbers found (explained in section 3.3.2). The average computation time will be noted, and the average CVI scores calculated for all approaches. There will also be a manual inspection of the cluster centroids for each approach.

3.3.1 Parameters

All the algorithms tested in this master thesis need input parameters. As this thesis focus mainly on the application and usefulness of the clustering approaches, the default parame-

Clustering approach	Shortened name in results
K-Means	K-Means
K-Shape	K-Shape
K-Means with Dynamic time warping	K-Means + DTW
Self-organising maps	SOM
Hierarchical clustering with single linkage criterion	Hierarchical + single
Hierarchical clustering with ward linkage criterion	Hierarchical + ward
K-Means with principal component analysis	K-Means + PCA
K-Means with discrete wavelet transformation	K-Means + wavelets
K-Means with Min-Max normalization	K-Means + Min-Max
K-Means with average load profiles	K-Means + AVG LP
Hierarchical clustering with single linkage criterion and average load profiles	HI + single + AVG LP

Table 3.1: Clustering approaches tested. All approaches use the Z-transformed datasets, except for the approach with Min-Max normalization.

ters have been used for the algorithms. The default parameters for K-Means can be found at [85], K-Shape at [86], hierarchical algorithms at [87], SOM at [62], PCA at [88] and discrete wavelet transformation at [89]. However, some of the algorithms require additional, non-default parameters. These algorithms are presented in the following:

- DTW: The K-Means algorithm with DTW uses a DTW band. Inspired by [1], the radius of the DTW band is set to 1. This is done to include dynamic time warping, as well as not making the time differences too big when comparing two elements in the load profiles.
- PCA: The PCA algorithm requires a number of components. The more components, the more accurate the representation gets, and the bigger the transformed dataset gets. The right number of components is therefore a trade-off between size (and thus computation time) and accuracy. For the computation of the CVI scores on all the datasets, the number of components for each sample is chosen as 6, which is a 75% reduction. In addition, an analysis with 3, 6 and 12 components on dataset 1 will be performed.
- Wavelet: The discrete wavelet transformation goes through a low pass and a high pass filter for each level of the algorithm. For each level, the size of the output is half of the input size. For the CVI computations on all datasets, level 2 is chosen to get 6 elements. As for PCA, the discrete wavelet transformation is also analysed having 3 and 12 elements on dataset 1. The wavelet used in the computation is chosen as the Haar wavelet, inspired by [1].

3.3.2 Optimal number of clusters

The CVI scores in this thesis are calculated with the optimal number of clusters, found by the approach described in section 2.7.1. The maximum number of clusters is set to be

11 in this thesis, chosen to not get way too many clusters. In literature, optimal cluster numbers are sometimes found to be very high, in some cases over 100 [38]. This is likely to not make the clusters differentiable, and much harder to visualise in a good way [38]. The clustering approach used to calculate the number optimal number of clusters is the K-Means with Z-transformation, as this is used as a basis in this thesis. The literature emphases that clustering is an exploratory tool and suggests using multiple CVIs to obtain a more robust result. Therefore, the optimal cluster number is found for all of the three CVIs presented in section 2.6.

3.3.3 Presentation of results

The results from the clustering analyses in this master thesis are evaluated based on the CVI scores, computation time and manual inspection of the clusters. The CVIs provide a quantification of the cluster quality, but as explained in section 2.6, they have some shortcomings. As the CVIs include the distance measure in their formula, a fair comparison of CVI scores from approaches using different distance measures is not possible. As the K-Means is used as a base, the CVI scores use the ED distance measure in their computation. The computation time is an objective measure of the clustering approaches, and is added as the computation time is a crucial factor when choosing clustering approaches [41].

The manual inspection is included because of the shortcomings of the CVI scores. The manual inspection is based on the criteria listed in section 2.6. The manual inspection makes it possible to compare the approaches with different distance measures. As some of the clustering algorithms have different distance measures, more emphasis has been put on the manual inspection of these approaches.

Some of the approaches have been studied a bit more deeply, due to their characteristics, and are explained in the following:

Reduction techniques

The CVI scores for both reduction techniques are calculated with 50%, 75% and 87.5% reduction, respectively 12, 6 and 3 elements for each sample (originally 24), on dataset 1. The reduction techniques reduce the dimension of the dataset, and therefore it is interesting to see how the CVI scores changes depending on how much the dataset is reduced. The characteristics of the reduction techniques will also be illustrated by showing the clustering centroids for different reduction sizes.

New optimal cluster numbers

Optimal cluster numbers are also found using the Min-Max normalization and average load profiles. As these transformed datasets give different CVI scores, different optimal cluster numbers might be obtained.

3.3.4 Practical application

A practical application of the clustering will be performed to show an example of how to use clustering on AMS-data. The practical application will be performed on streetlights

from dataset 1 for January. As streetlights are expected to have a similar max consumption day after day, the Min-Max normalization is chosen to best see when the maximum and minimum consumption find place. The practical application will look for outliers in the streetlight category, to see if there are any customers with an abnormal consumption pattern. By looking only at outliers, the manual inspection looking for wrongly categorized customers might be improved.

3.4 Comparison with FASIT

A comparison of clustering results with a current prediction technique for load profiles will also be done. A current practice among DSOs is to use FASIT, which is explained in section 2.9.1. As FASIT divides the load profiles for customers in weekdays and weekends (including public holidays), the data for January is divided into weekdays and weekends (all public holidays in January 2019 came on weekends in Norway). The variables, A and B from equation (2.22) are found in [69], and the average temperature for the weekdays and weekends in January for the given area analysed were found extracting data from the API provided by The Norwegian Meteorological Institute [90]. The 806 household customers in dataset 1 for January are used in the comparison. As FASIT represents average load profiles, the average values in the dataset will also be shown. The comparison will be by manual inspection, as it is not possible to make a CVI score from a single sample (the FASIT load profile).

3.5 Analyses of capacity in the grid

This section presents analyses of the capacity used in the grid, seen from the transformer side. As AMS contain large amounts of data about electric consumption, more analyses can be done than only clustering of AMS-data. This section shows one of the many possibilities with AMS-data. If the connections between a transformer and the customers are known, aggregated values can be calculated from the AMS-data. This can reveal information that is not available for the DSO today.

The 806 households in dataset 1 for January are used in the analyses, and an additional file containing which transformer each customer is connected to is used to divide the dataset into subgroups of transformers. The temperature data used is extracted from the API provided by The Norwegian Meteorological Institute [90].

The coefficients for temperature correction of the energy and power consumption are found using the data from ENOVA's home page ([76]) as well as the information from the DSO providing the data. The equations for temperature correction of the energy and power are replicated in this section for convenience. The k in equation (3.1) is found to be 0.5 and the $\frac{DD_{normal}}{DD_{year}}$ to be 0.94 for the given year. The k in equation (3.2) is found to be 0.0195. The differences between the extreme minimum temperature and the three day rolling mean temperatures, called $diff$ in equation (3.2), are found using temperature data for the nearest weather station extracted from [90].

$$E_{corrected} = E_{measured}\left\{(1-k) + k\frac{DD_{normal}}{DD_{year}}\right\} \tag{3.1}$$

$$P_{corrected} = P_{measured} + k \cdot diff \cdot P_{measured} \tag{3.2}$$

The following presents the values calculated in the analyses:

- Energy consumption: Total energy consumption for the transformer in the period analysed. Found by summarizing all the consumptions for the customers below the transformer.
- Energy consumption corrected by temperature: Total energy consumption temperature corrected. Found by using equation (3.1).
- Max power: The maximum consumption in the transformer. Found by summarizing the consumption each hour for all consumers and taking the highest consumption.
- Time of use: Energy consumption / (max power * time). Gives a relation between the energy consumption and the maximum power. The energy consumption is given in kWh, the max power in kW and the time in h, so the time of use will get a value in the range [0,1].
- Max power corrected by temperature: As explained in section 2.10, there are two ways to calculate the max power corrected by temperature. The one being used by the DSOs today is to temperature corrected the maximum power measured. This is called PMT in this thesis. The other way is to temperature correct every consumption, and then choose the highest value. This is called PT in the thesis. By calculating the PT instead of PMT, a better value of the maximum temperature corrected consumption is obtained. The PT and PMT are found by using equation (3.2).
- Aggregated max values: Found by summarizing the maximum values for all the customers below the transformer. The aggregated max power, aggregated PMT and PT are found.

Chapter 4

Results

This chapter presents the results from the analyses done in this master thesis. The methods are explained in chapter 3, which are based on the theory in chapter 2, and the results are discussed in chapter 5. The first section presents the results for the 11 clustering approaches analysed as well as the optimal cluster numbers for the different datasets. Section 4.2 to section 4.6 present the manual analyses of the approaches and additional analyses for some of the approaches. Section 4.7 presents the comparison with FASIT, before the chapter ends by presenting the results from the analyses of capacity in the grid.

4.1 Clustering approaches

The CVI scores and the computation time for the 11 clustering approaches analysed are shown in table 4.1. (The full description of the approaches is given in table 3.1.). All the values in table 4.1 are divided by the corresponding value for K-Means. The average computation time on all the datasets for K-Means was 11 seconds, so the computation time using the different approaches can be found by multiplying with the factor for that approach. Recall that DB scores lower than 1 and SI and CH scores above 1 indicate better clustering quality than the K-Means algorithm. The hierarchical algorithms did not have the required storage capacity to run on dataset 4, so the values for the hierarchical algorithms are only for dataset 1 to 3.

The optimal cluster numbers found are shown in table 4.2.

4.2 Clustering algorithms

This section presents the manual inspection of the clustering algorithms. The manual inspection criteria are given in section 2.6.

Clustering approach	DB	SI	CH	Computation time
K-Means	1	1	1	11s
K-Shape	1,008	1,017	1,000	x57
K-Means + DTW	1,131	0,811	0,899	x1012
SOM	2,003	-0,559	0,123	x5
Hierarchical + single	1,109	-2,744	0,012	x52
Hierarchical + ward	1,277	0,444	0,581	x273
K-Means + PCA	1,027	0,962	0,989	x0,61
K-Means + wavelets	1,079	0,897	0,949	x0,56
K-Means + Min-Max	–	–	–	x0,93
K-Means + AVG LP	–	–	–	x0,05
HI + Single + AVG LP	–	–	–	x0,03

Table 4.1: Results from clustering analyses

CVIs	D1	D2	D3	D4
DB	3	10	11	11
SI	2	2	2	2
CH	2	2	2	2

Table 4.2: The optimal cluster numbers

4.2.1 K-Means

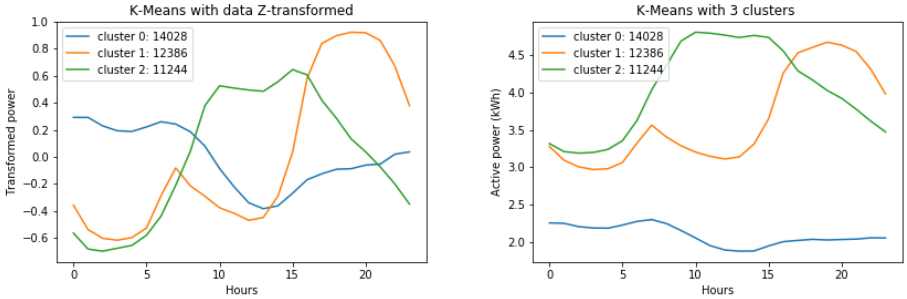
Some selected results from the K-Means algorithm are shown in figure 4.1 to figure 4.3. Figure 4.1 shows the results from dataset 1, where three distinct and substantial cluster centroids are shown. As the clustering of dataset 2, 3 and 4 got quite similar, only the results from dataset 1 and 4 are included.

Figure 4.2 shows the results from dataset 4 with 11 clusters. The results on dataset 2, 3 and 4 with 10 and 11 clusters showed 7 to 8 clusters with low transferred values and one steep peak, as figure 4.2a shows. The 2 to 3 other clusters have different shapes, where one has a Z-value close to 0 for all the hours (green load profile in figure 4.2a). Although many clusters, all of clusters from all the datasets fulfil the distinguishable and substantial criteria. In figure 4.2 all the clusters sizes lie between 22 967 and 64 317.

Figure 4.3 shows the results from the clustering on dataset 4 with 2 clusters. By comparing this figure with figure 4.2, it seems like clusters with peaks to the left are merged and the clusters with peaks on the right are merged. The clustering with 2 clusters also manages to make distinguishable and substantial clusters.

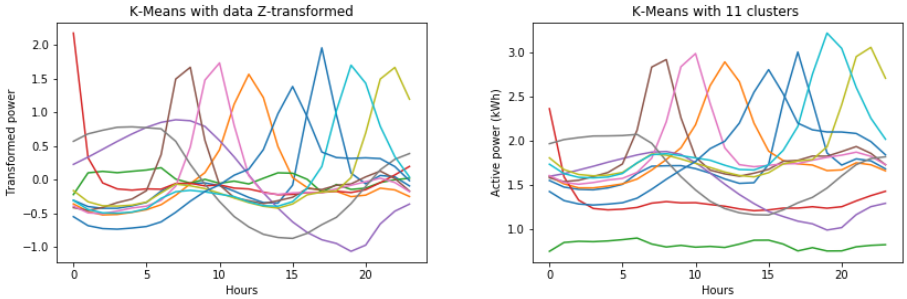
4.2.2 K-Shape

The results from the K-shape algorithm are very similar to the results from K-Means. This means that K-Shape also made distinguishable and substantial clusters. Figure 4.4 shows the results from K-Shape on dataset 1, which is almost identical to the results from K-Means. As the K-Shape got so similar results to K-Means, more results are not included.



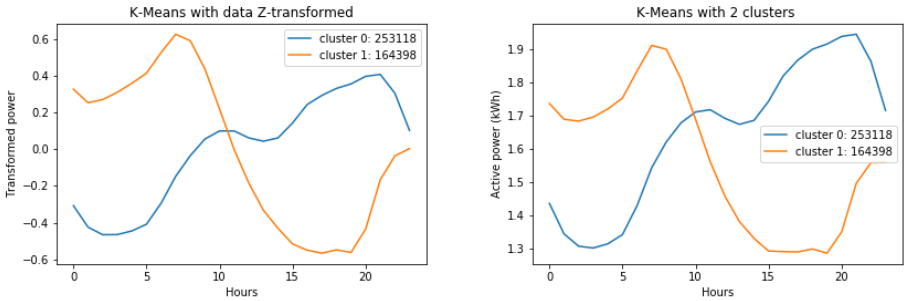
(a) Z-transformation representation **(b)** Original data representation

Figure 4.1: K-Means on dataset 1 with 3 clusters. Two different cluster representations.



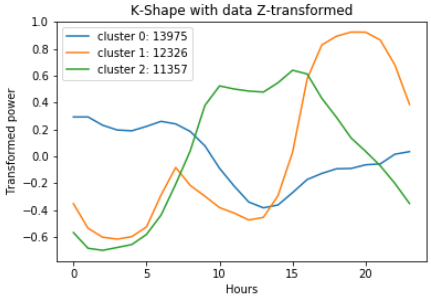
(a) Z-transformation representation **(b)** Original data representation

Figure 4.2: K-Means on dataset 4 with 11 clusters. Two different cluster representations.

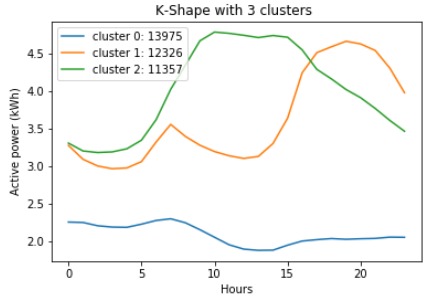


(a) Z-transformation representation **(b)** Original data representation

Figure 4.3: K-Means on dataset 4 with 2 clusters. Two different cluster representations.

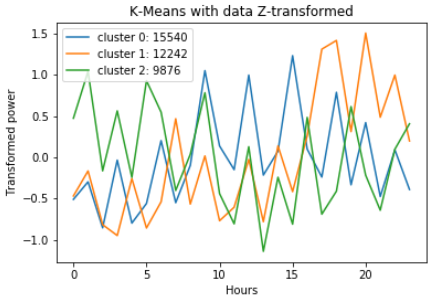


(a) Z-transformation representation

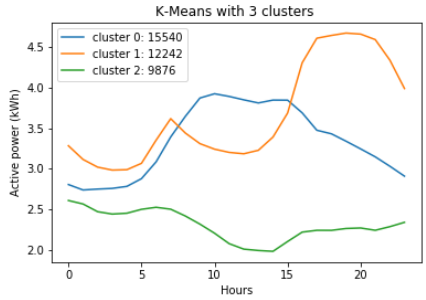


(b) Original data representation

Figure 4.4: K-Shape on dataset 1 with 3 clusters. Two different cluster representations.



(a) Z-transformation representation



(b) Original data representation

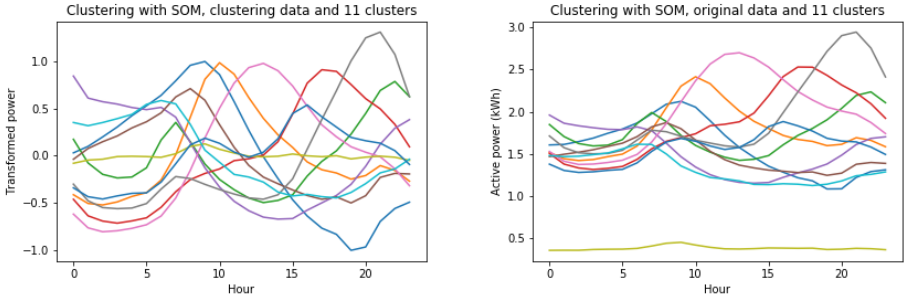
Figure 4.5: K-Means with DTW on dataset 1 with 3 clusters. Two different cluster representations.

4.2.3 K-Means with DTW

The clustering results from K-Means with DTW were different from the results from the original K-Means algorithm. The most different results are the shapes of the Z-transformed centroids, which are very fuzzy. The shapes of the transformed centroids do not represent the average values of the clusters. This is shown in figure 4.5. The cluster sizes and cluster centroids are in general different from the clusters from K-Means. This can be seen by comparing figure 4.5b and 4.1b. The cluster have different sizes, and the centroids have similar shapes, but with a different offset. Based on the manual inspection, the clustering with K-Means with DTW gives distinguishable and substantial clusters, but they might be a little bit less compact compared to K-Means, as they are less smooth.

4.2.4 Self Organizing Maps

This approach shows more different results than the previous approaches. The cluster centroids seem to be less distinguishable and less compact. This can be seen from the results on dataset 4 with 11 clusters, shown in figure 4.6. The SOM algorithm also has a



(a) Z-transformation representation

(b) Original data representation

Figure 4.6: SOM on dataset 4 with 11 clusters. Two different cluster representations.

tendency in making some of the clusters smaller, exemplified for dataset 4 with 11 clusters where the SOM approach has the smallest cluster as 7631, compared to 22 967 for K-Means.

4.2.5 Hierarchical clustering

This subsection presents the results from the two hierarchical clustering algorithms. Both of the hierarchical clustering algorithms did not have enough space requirements to run the algorithm on the biggest dataset, dataset 4.

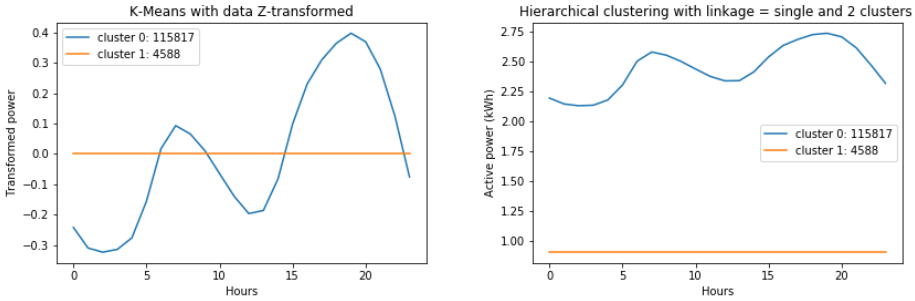
Single linkage criteria

The results from the hierarchical clustering with single linkage criteria are quite different from the results from the other clustering algorithms. On dataset 1, 2 and 3 with 2 clusters, the hierarchical clustering with single linkage divides the samples in one cluster with Z-values equal zero, and the rest of the samples in the other cluster. The clustering results from dataset 3 with 2 clusters are shown in figure 4.7. None of the other algorithms got similar results.

The results on dataset 1, 2 and 3 with higher number of clusters (respectively 3, 10 and 11), show two relatively big clusters and the rest of the clusters contain only one sample. The results from hierarchical clustering with single linkage on dataset 3 with 11 clusters are shown in figure 4.8. By studying figure 4.7 and 4.8, we see that the two clusters from the clustering with only two clusters are still there with 11 clusters, but in addition 9 clusters containing only one sample.

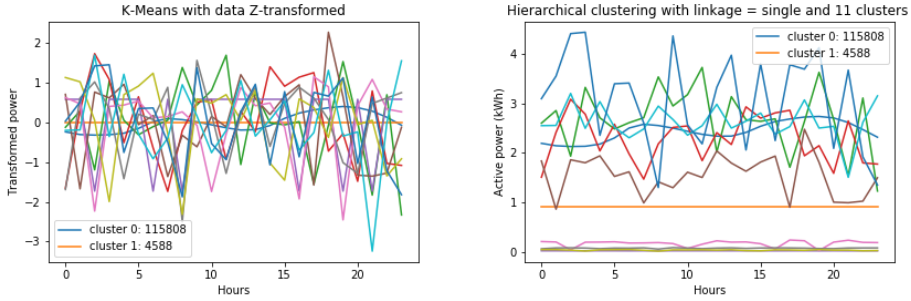
Ward criteria

The results from the hierarchical clustering with ward linkage on dataset 1 with 3 clusters is shown in figure 4.9. The results from the manual analysis is that this approach seems to give less smooth curves and the clusters are not as clearly defined as with the K-Means



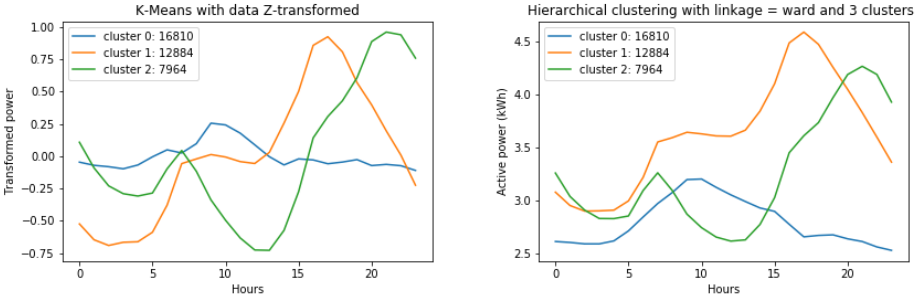
(a) Z-transformation representation **(b)** Original data representation

Figure 4.7: Hierarchical clustering with single linkage criteria on dataset 3 with 2 clusters. Two different cluster representations.



(a) Z-transformation representation **(b)** Original data representation

Figure 4.8: Hierarchical clustering with single linkage criteria on dataset 3 with 11 clusters. Two different cluster representations. The clusters not labelled contain only one sample.



(a) Z-transformation representation

(b) Original data representation

Figure 4.9: Hierarchical clustering with ward linkage criteria on dataset 1 with 11 clusters. Two different cluster representations.

approach. This suggests less distinguishable clusters, which is also confirmed by the CVI scores in table 4.1.

4.3 Reduction techniques

This section presents the results from the additional analysis with the reduction techniques, PCA and wavelets. As explained in section 3.3.3, the PCA and wavelets were tested with three different reductions, 50%, 75% and 87.5%, on dataset 1. The chosen number of clusters is 3, as this illustrates differences better than two clusters, the other optimal cluster number on dataset 1.

Figure 4.10 shows CVI scores on dataset 1 when reducing the dataset with 50%, 75% and 87.5%. In this figure, the dashed lines represent the clustering with wavelets while the full lines represent the PCA. The reductions correspond with 12, 6 and 3 elements, as the original size of the samples are 24 elements. The red line is one and represents the KM-scores. The figure shows that the wavelet scores get worse the more reduction, while the PCA score is not affected much of the size of the reduction on this dataset. The silhouette score is actually higher with 75% and 87.5% reduction, compared to 50% reduction of the dataset for PCA.

In the following subsections, visualisations for each reduction technique are presented to illustrate the characteristics of each reduction technique.

4.3.1 PCA

Figure 4.11 shows the results from the K-Means algorithm using PCA transformation with 6 principal components. By looking at the two first components (figure 4.11a), the clusters centroids are clearly distinguishable from each other. For component three to six, the components are almost 0, which indicate that 2 components would be enough to distinguish the clusters from each other. This is confirmed by the results from 3 and 12 components, shown in figure 4.10, where similar CVI scores are achieved.

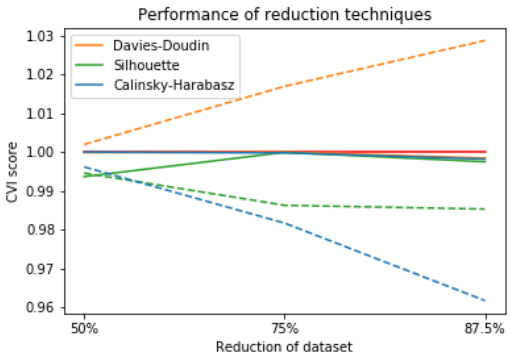
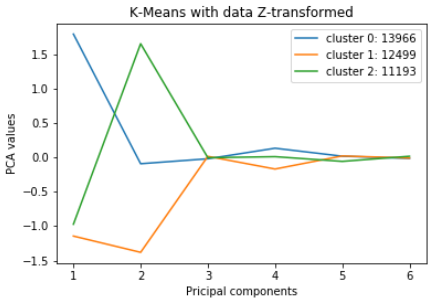
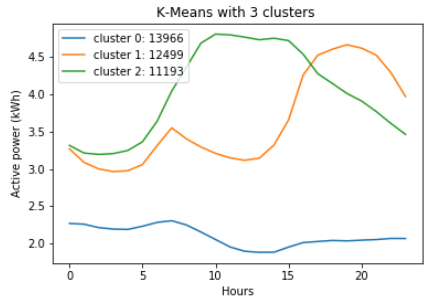


Figure 4.10: CVI-score for the reduction techniques

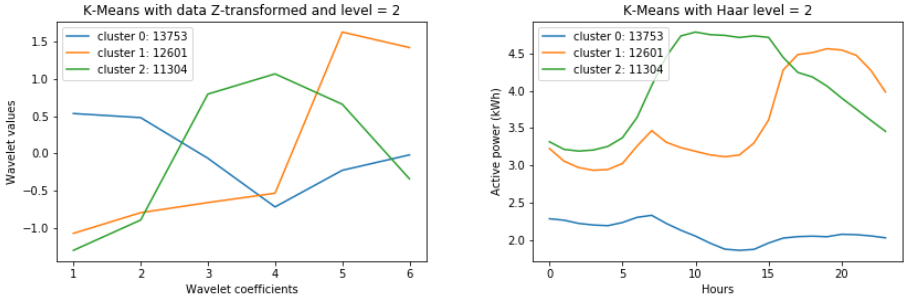


(a) PCA representation



(b) Original data representation

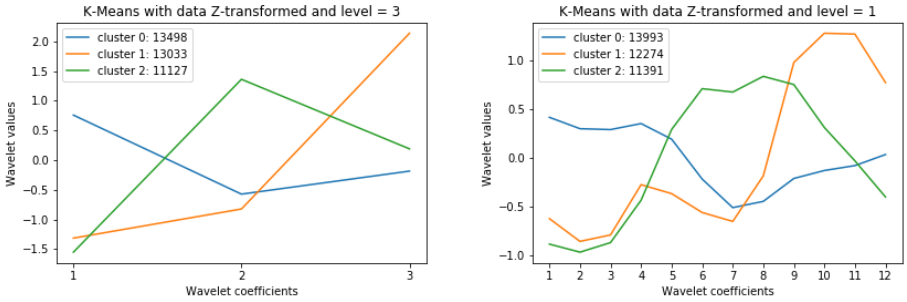
Figure 4.11: K-Means clustering with a PCA and Z-transformation of data on dataset 1. Two different cluster representations.



(a) Wavelet representation

(b) Original data representation

Figure 4.12: K-Means clustering with a wavelet transformation of dataset 1 with 3 clusters. Two different cluster representations.



(a) Wavelet level 3

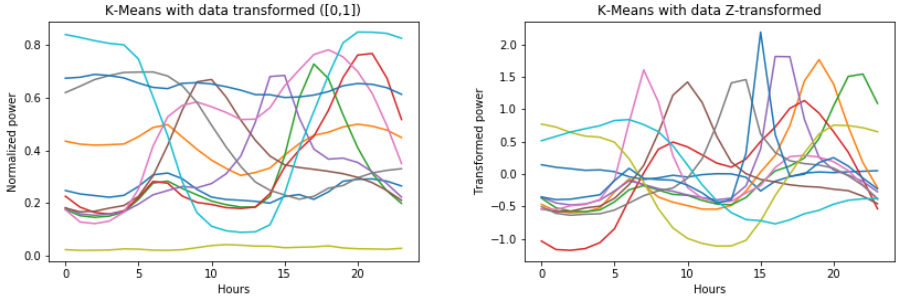
(b) Wavelet level 1

Figure 4.13: K-Means clustering with a wavelet transformation on dataset 1 with 3 clusters. Two different levels in the wavelet algorithm.

4.3.2 Wavelets

Figure 4.12 and 4.13 show the results from the K-Means algorithm using wavelet transformation. Figure 4.12 shows the cluster centroids using the data after the wavelet transform (figure 4.12a) as well as the representation using the ED representation on the original data (figure 4.12b). To show how the wavelet transformation changes by changing the number of levels (each level halves the number of elements), figure 4.13 is added. In figure 4.12, the level of the wavelet transformation is 2, to get the desired number of 6 components.

In figure 4.12, the cluster centroids are similar in shape, but the cluster sizes are a bit different, compared to K-Means. When decreasing the levels used in the wavelet algorithm (which means less reduction), the cluster sizes become more similar to K-Means. This can be seen in figure 4.13b. The figures show that the wavelet transformation try to approximate the shape of the original data, and the higher the level is (which means fewer elements), the worse the approximation gets. By comparing figure 4.12b and figure 4.13b, the shapes are quite similar, while figure 4.13a is less similar due to fewer elements.



(a) Min-Max normalization

(b) Z-transformation

Figure 4.14: K-Means clustering with two different transformation techniques on dataset 3 with 11 clusters

CVIs	D1+Z	D1+MM	D2+Z	D2+MM	D3+Z	D3+MM	D4+Z	D4+MM
DB	3	6	10	3	11	5	11	5
SI	2	2	2	2	2	2	2	2
CH	2	2	2	2	2	2	2	2

Table 4.3: Optimal cluster number using MM-normalization and Z-transformation

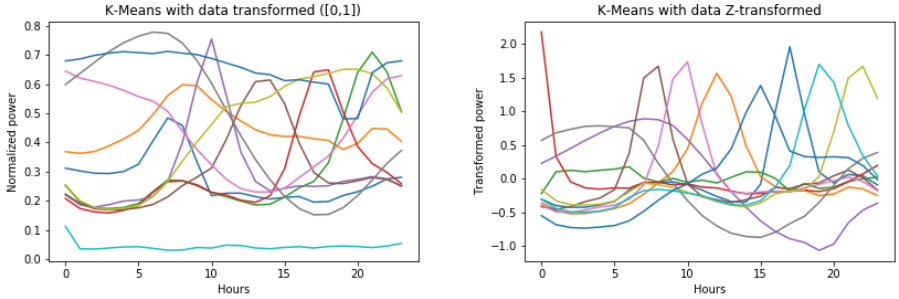
4.4 Transformation techniques

This section presents the other transformation technique used in this master thesis, the Min-Max normalization. Combined with the Z-transformation, these transformation techniques do not reduce the size of the dataset.

The results show that the Min-Max (MM) normalization tend to cluster more based on the consumption relative to the maximum and minimum consumption. This is well illustrated for dataset 3 with 11 clusters, shown in figure 4.14a for the MM-normalization and in figure 4.14b for the Z-transformation. While the Z-transformation only have one load profile with stable load during the day (the dark blue line around 0 in figure 4.14b), the MM-normalization has three load profiles with stable load (the yellow line and the two blue lines in figure 4.14a).

A trend is that with the MM-normalization the clustering is less affected by the peaks. This can be seen for dataset 4 with 11 clusters, shown in figure 4.15. Figure 4.15a shows the clustering with the MM-normalization, and figure 4.15b shows the clustering with Z-transformation. The cluster centroids using the MM-normalization do not have so distinct peaks, compared to the Z-transformation.

The optimal number of clusters using Z-transformation and MM-normalization are shown in table 4.3. The results are identical for the SI and CH index, while for the DB index, it is different for all the datasets.



(a) Min-Max normalization

(b) Z-transformation

Figure 4.15: K-Means clustering with two different transformation techniques on dataset 2 with 10 clusters.

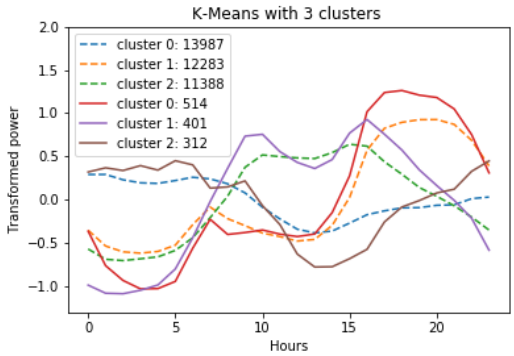


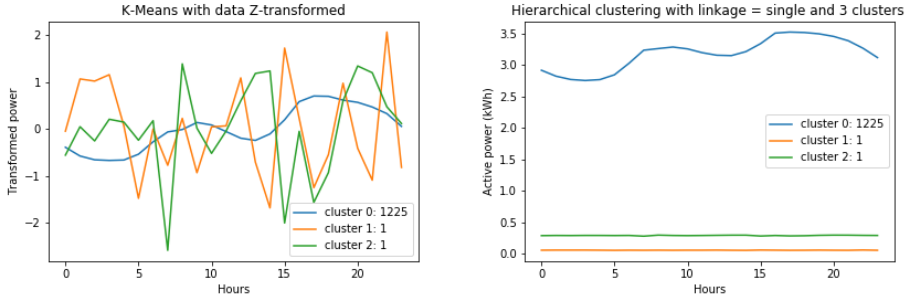
Figure 4.16: Clustering with all samples (dashed lines) and with average load profiles (full lines) for dataset 1 with 3 clusters.

4.5 Average load profiles

This section presents the results using the average load profiles as the input data for the clustering algorithm. As explained in section 2.9.1, with average load profiles one customer has one load profile, and each hour in this load profile is the average consumption this hour in the whole period.

Figure 4.16 shows the results of the clustering with all samples and with average load profiles on dataset 1 with 3 clusters. The clustering with average load profiles are the full lines, while the clustering with all samples are the dashed lines. The cluster centroids from the clustering with average load profiles have similar shapes, but more extreme values. The centroids from clustering with all samples seem like smoothings of the centroids from the clustering with average load profiles.

The results of hierarchical clustering with single linkage and average load profiles are shown in figure 4.17. These results show one large cluster and two clusters including only



(a) Z-transformation representation

(b) Original data representation

Figure 4.17: Hierarchical clustering with single linkage criteria and average load profiles on dataset 1 with 3 clusters. Two different cluster representations.

CVIs	D1	D1+A	D2	D2+A	D3	D3+A	D4	D4+A
DB	3	3	10	3	11	3	11	2
SI	2	2	2	2	2	2	2	2
CH	2	2	2	2	2	2	2	2

Table 4.4: Optimal cluster number with all load profiles and the load profiles averaged (A) and Z-transformation (no extra label).

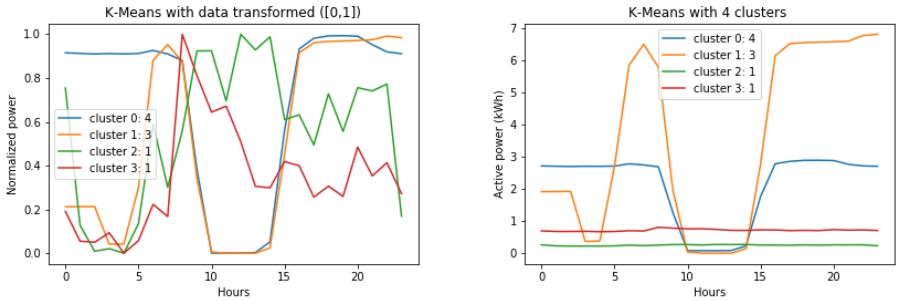
one customer. These two outliers have quite fuzzy consumption (shown in figure 4.17a), but they have a very low consumption (shown in figure 4.17b).

The results of the optimal cluster number computation using average load profiles and Z-transformation are shown in table 4.4. The results are identical for SI and CH, but lower with the DB index.

4.6 Practical application

This section presents the results from the practical application explained in section 3.3.4. The average profiles of the customers with categorization streetlights are taken as input to the K-Means algorithm with Min-Max normalization using 4 clusters. The number of clusters were chosen such that both outliers got isolated, while the other, and seemingly normal customers, got clustered together. The results are shown in figure 4.18, where we see two load profiles having a different shape of what we think of typical load profiles for streetlights. However, as shown in figure 4.18b, the average consumption of these two customers are quite low, below 1 kWh in average.

We analyse the red load profile a bit more, and its consumption is shown in figure 4.19. By analysing more we find that this customer has a minimum consumption this month of 0.212 kWh and maximum consumption of 0.470 kWh. From figure 4.18b we see that cluster 0 and cluster 1 go almost to zero when the lights are not turned on. This is not the case for cluster 2 and 3.



(a) Z-transformation representation

(b) Original data representation

Figure 4.18: Clustering in the practical application on streetlights.

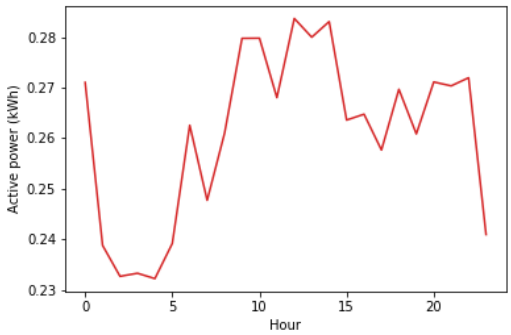
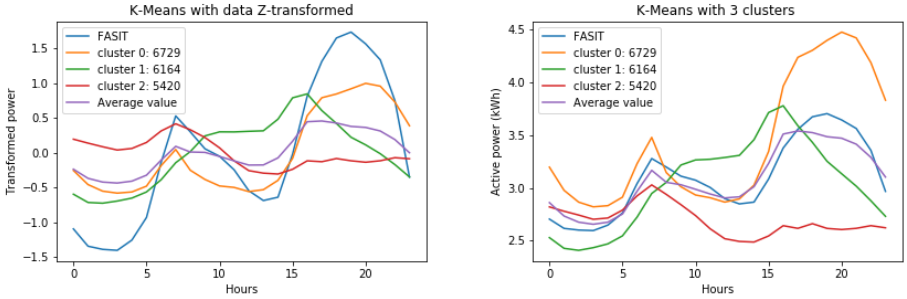


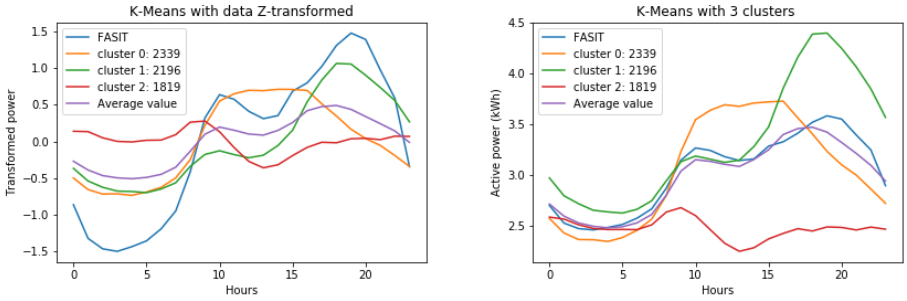
Figure 4.19: Average consumption for one of the abnormal load profiles



(a) Z-transformation representation

(b) Original data representation

Figure 4.20: Comparison of clustering and FASIT load profiles on households for weekdays. Two different cluster representations.



(a) Z-transformation representation

(b) Original data representation

Figure 4.21: Comparison of clustering and FASIT load profiles on households for weekends. Two different cluster representations.

4.7 Comparison with FASIT

This section presents the comparison of clustering results on dataset 1 and FASIT load profiles. As explained in section 3.4, the households in dataset 1 are extracted and divided into weekends and weekdays. Figure 4.20 shows the clustering results on dataset 1 with weekdays and the corresponding FASIT load profile. The transformed values of the FASIT profile and the average value differ quite much (shown in figure 4.20a), while with active power (figure 4.20b) they almost align. The same can be seen in figure 4.21, which shows the same results for the weekends. For both of the transformed FASIT load profiles, it seems like a scaling factor could make them quite similar to the average value profile.

While the average values and FASIT load profiles are quite similar in the active power representation, the datasets contain load profiles with shapes very different from the average. This is captured by the three clusters, which all have different shapes than the average load profile for both weekdays and weekends.

4.8 Transformer analyses

This section presents the results from the methods explained in section 3.5.

4.8.1 Energy transformed

The energy transformed for January is shown in figure 4.22, both with and without temperature corrections.

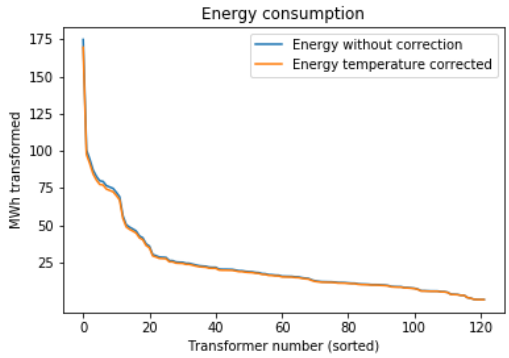


Figure 4.22: Energy consumed, with and without temperature correction, below transformers

4.8.2 Max power

The maximum power for the transformers in January, both with and without temperature corrections, are shown in figure 4.23. As the transformers have a high difference in the maximum power, it is hard to see the differences accurately between the maximum power with and without temperature correction.

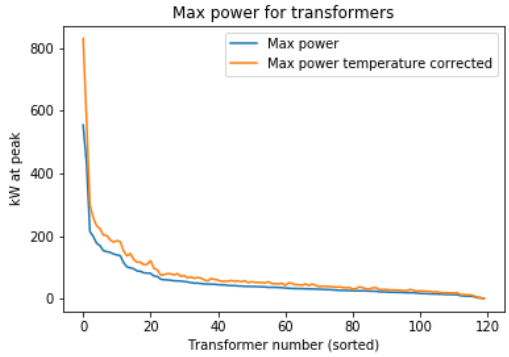


Figure 4.23: Maximum power below transformers

Figure 4.24 shows the maximum power corrected by temperature, both PMT and PT, divided by the maximum power measured. A value of, for example 1.2, means that the maximum power corrected by temperature is 20% higher than the maximum power measured. Recall from section 3.5 that PMT is an abbreviation for maximum power measured corrected by temperature and PT for maximum power corrected by temperature. The difference between PMT and PT is shown in figure 4.25, where the values are normalized by dividing by the maximum power measured for the transformer. In a bit more than half of the cases they coincide, and in other cases it is a big deviation, up to 20 %. Some values are summarized in table 4.5. The table shows a difference of 3.7 % in average between the PMT and PT. The table also shows that the PMT and PT coincide for 47 % of the transformers.

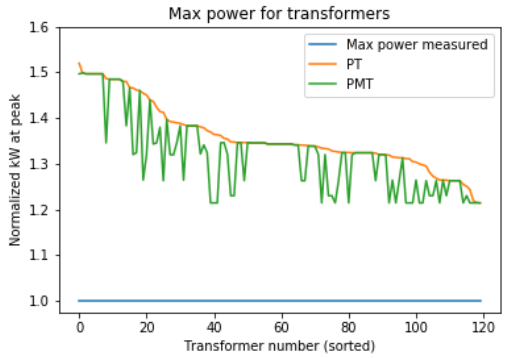


Figure 4.24: Maximum power below transformer normalized

Variables	Increase PT	Increase PMT	Coinciding peaks
Values	35.8 %	32.1 %	47 %

Table 4.5: Some key values from temperature corrections on max power

4.8.3 Time of use

The time of use for each transformer, is shown in figure 4.26. The figure shows that some transformers have a low time of use, and that most of the transformers lie between 0.4 and 0.7.

4.8.4 Max power aggregated by consumers

The transformers have max consumption when the sum of the consumption of the customers below the transformer is the highest. However, an interesting value is the aggregated maximum for each of the customers below a transformer. This value is given for each transformer in figure 4.27a. The figure shows peak values for the transformers, both

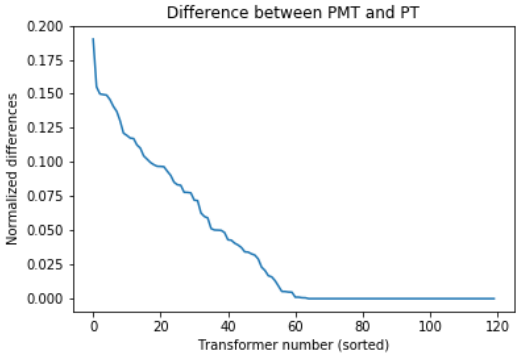


Figure 4.25: Difference between PMT and PT

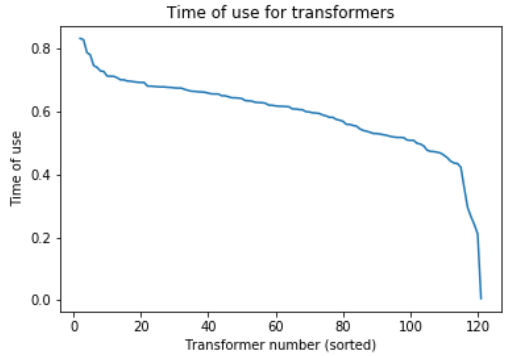
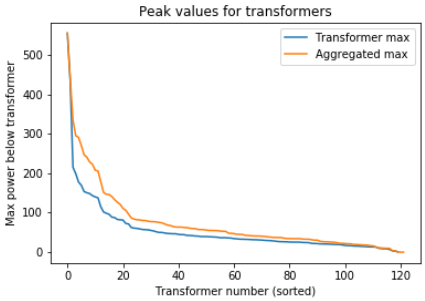


Figure 4.26: Time of use for transformers

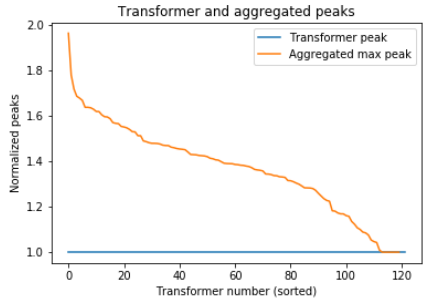
the measured peak value in the transformer and the aggregated peak values for all customers below the transformer. Figure 4.27b shows the same values normalized (divided by the peak value measured in the transformer). The average increase for the aggregated peak values is found to be 37%.

Figure 4.28a shows the aggregated maximum consumption measured and aggregated PMT. Figure 4.28b shows PMT and PT normalized by diving by the aggregated maximum consumption measured. The average increases for the normalized PMT and PT are found to be 36.3% and 39.8%. The normalized difference between PMT and PT is shown in 4.29. The average difference between PMT and PT is 3.5%.

The average increase between the measured maximum consumption in the transformers compared to the aggregated PT is 106%. This means that the average increase in the maximum value would be 106% higher if all the customers would have their maximum consumption at the same time and the temperature would be at the extreme minimum.

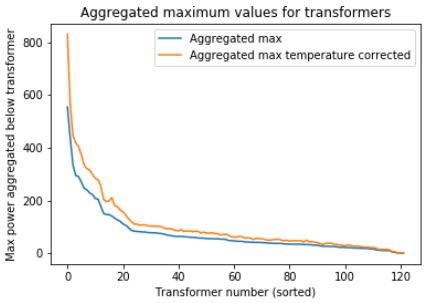


(a) Transformer and aggregated peaks

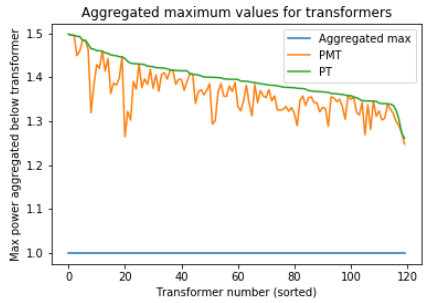


(b) Normalized peaks

Figure 4.27: Measured peak and aggregated peaks for transformers



(a) Transformer max and MPT



(b) Normalized peaks

Figure 4.28: Measured peak and aggregated peaks for transformers

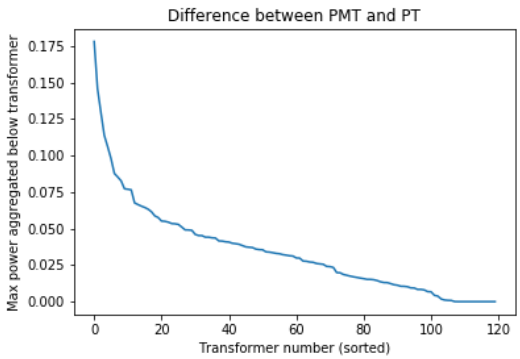


Figure 4.29: Differences between the aggregated PMT and PT

Chapter 5

Discussion

This chapter presents the discussion of the analyses performed in the master thesis, described in chapter 3 and presented in chapter 4. The discussion is based on the theory, presented in chapter 2, and the combination of the results and discussion will be summarized in the conclusion, chapter 6.

The discussion will take a similar structure as the results presented in chapter 4. The clustering approaches will be discussed first, followed by a discussion of the analyses of capacity in the grid. Then, a broader discussion including assumptions usefulness of clustering is presented. The chapter ends by a short description of the working process and challenges along the way. The discussion about the clustering approaches starts by looking at the clustering algorithms before the data representation techniques are discussed. Then, the practical application and the comparison with FASIT load profiles will be discussed.

5.1 Clustering approaches

As suggested in the literature, multiple clustering approaches have been tested in this master thesis. The clustering algorithms have been tested with multiple CVIs, also suggested in literature, to make the results more robust. However, as discussed in this section, there are some shortcomings of the CVIs.

This section begins with a short discussion of the CVIs and computation times before the clustering algorithms are discussed. Then the data representation techniques are discussed before the practical application and the comparison with the FASIT load profiles end the section.

CVIs

The CVIs are used to test the quality of the clustering, but as pointed out in the theory, the CVIs include a distance measure. This means that a comparison of two clustering methods using different distance measures is not a fair comparison. The CVIs scores calculated in this master thesis all use the ED, which favours the algorithms using the ED as distance

measure. For the algorithms that use the ED distance measure, the CVIs seem to be a good indicator of the quality. This gives us, combined with the execution time and the manual analyses, a basis to conclude that some of the approaches are not preferred.

5.1.1 Clustering algorithms

This master thesis tested 6 different clustering algorithms. Two of the algorithms use a different distance measure (K-Shape and DTW), which make a bigger emphasis on manual analyses. This is an obvious disadvantage, as subjectivity is influencing the results in a higher degree. In the following, all the algorithms will be discussed.

K-Means

This is the base approach in this master thesis. It performs pretty good compared to the other approaches, even though it is worth to keep in mind that the CVIs favours the ED distance measure. However, it achieves surprisingly similar results as the K-Shape, which suggest good quality, even though K-Means does not incorporate shift invariance. It is the fastest of the clustering algorithms, and the results show that by using a reduction technique the computation time can be lowered without lowering the clustering quality significantly. The results suggest that the simple K-Means with Z-transformation is a good choice for clustering load profiles, especially if the computation time is of important character.

K-Shape

The K-shape algorithm performs surprisingly similar to the K-Means algorithm, both in cluster sizes, centroid shapes and cluster quality measured from the CVIs. It should anyway be pointed out that the CVIs use the ED, which favours the K-Means algorithm. The similar results with CVIs favouring algorithms using ED might suggest an increased clustering quality using K-Shape, as the ED does not incorporate shift invariance. By deciding whether to use K-Shape or K-Means, one should evaluate the trade-off between increased computation time and a possible increase in clustering quality.

DTW

The results of the K-Means using the DTW distance measure are different from K-Means and K-Shape. The cluster centroids from the DTW computation are very fuzzy, and the centroids are not representative for the average shape of the samples in the cluster. This is probably because the centroids incorporate dynamic time warping, which lets each sample have local shifts to align the best way possible with the centroid. This give us additional information which indicates that the individual samples are quite fuzzy and chaotic, even though the average values are smooth. This is in line with other findings in literature ([49], [50]). The long trends of the DTW centroids, found by taking the rolling mean, seems to give similar shapes as the centroids from K-Means.

The clustering quality is a bit hard to make conclusions from. The CVI scores show a significant lower quality than K-Means and K-Shape, but as mentioned before, the CVI

scores are calculated using the ED distance measure. The sizes of the clusters are also somewhat different, except for with 2 clusters, where they are quite similar. The results indicate that using the DTW distance measure instead of ED give a quite different clustering. The computation time is a big disadvantage, and the DTW clustering has a computation time over 1000 times higher than K-Means, a factor which increases with the size of the dataset and the number of clusters. Based on the findings in this thesis, the K-Means with DTW cluster differently than K-Means and K-Shape, but the results raise a question whether or not the increased computation time is worth the possible increase in clustering quality.

Different results might be achieved when the frequency of the readings is increased to 15min (proposed to start in 2023 in Norway), as local time shifts might be more important with increased metering frequency.

SOM

The SOM algorithm seems to perform badly compared to the K-Means algorithm. The CVI scores are considerably worse than the K-Means, and the manual analyses seem to reveal less compact and differentiable clusters and the computation time is higher than K-Means. By running the algorithm multiple times, the results differ considerably more than the K-Means, which indicate a less stable method. The SOM-algorithm is quite complex compared to the other algorithms tested in this master thesis. Therefore, more insight in the algorithm and better setup of parameters and training might increase the performance of the algorithm. However, based on the bad results, it is unlikely that increased insight can make obtain similar CVI scores as the K-Means. More training will also increase the computation time of the algorithm. Therefore, based on the results in this master thesis, the SOM algorithm does not seem like a preferred choice for clustering on AMS-data.

Hierarchical clustering with single linkage

The hierarchical clustering with single linkage gives some interesting results, even though it has among the worst CVI scores. The results show a tendency to make a cluster which have constant power during the day (their values Z-transformed are 0). None of the other clustering approaches extracted this cluster, which tells us that the algorithm can detect clusters the others fail to detect. However, the algorithm tends to make clusters containing only one sample. This might in many cases not be desired, as the clusters might not fulfil the substantial criteria. The hierarchical algorithm with single linkage is considered good if the goal is to get additional information about outliers in the dataset.

Hierarchical clustering with ward linkage

The results with hierarchical clustering with ward linkage show higher CVI scores compared to the single linkage, but the results from ward linkage are not so interesting. The CVI scores are significantly worse than the K-Means, K-Shape and K-Means with DTW, even though the ward linkage uses the ED distance measure. By manual inspection the clusters seem to be less distinguishable. The computational time is also relatively high, around 270 times higher than K-Means on dataset 1 with 3 clusters.

Both of the hierarchical algorithms did not manage to run the biggest dataset, due to insufficient memory. From the theory we know that the computational time (and thus memory requirements) increases more than partitional algorithms, which especially favours the partitional algorithms on large datasets.

Stability criterion

The stability criterion was the fourth manual inspection criterion mentioned in the theory. This was not tested systematically in this thesis, but some tendencies were found when running the algorithms. Three of the algorithms showed a tendency to be pretty stable, while the other three showed more differences in cluster sizes on different runs. In general, the most stable algorithms got higher CVI scores, except for hierarchical clustering with single linkage. The stable algorithms seemed to be K-Means, K-Shape and hierarchical clustering with single linkage. The less stable algorithms seemed to be SOM, K-Means with DTW and hierarchical clustering with ward linkage.

5.1.2 Data representation techniques

This subsection discusses the results from the data representation techniques used in this master thesis. The data representation techniques can be divided to reduction techniques, which reduce size of the dataset, transformation techniques, which transform the dataset while keeping the size constant and the average load profile representation, which represent each customer with only one sample.

Reduction techniques

The PCA achieves good results on the CVI scores and a considerable reduction in the computation time. By reducing the size of the dataset with 75%, the computational time decreases with 39%, and reducing the CVI scores with 2.5%. On dataset 1 with 3 clusters, similar CVI scores were achieved with only 3 elements (87.5% reduction), so the reduction technique might perform well using few elements. The clustering with PCA uses a little bit more time than the with wavelets, but as the CVI scores are significantly higher, the PCA seems to be the preferred reduction technique.

The wavelet reduction technique also achieves relatively good CVI scores. At least compared with the SOM algorithm and the hierarchical algorithms. One point of interest is that the CVI scores with wavelets drops considerably with the increased reduction. The computation time is a bit lower than for PCA, but not enough to weight up for the lower CVI scores.

These reduction techniques were tested on load profiles with hourly readings, but as pointed out in the theory, 15-min readings are proposed in Norway from 2023. Therefore, reduction techniques might be even more important in the future, as the reduction in quality might be relatively lower compared to the increased computational speed. The results on clustering of time-series with higher resolution for the two reduction techniques might be different than the results found in this master thesis. It should also be pointed out that other reduction techniques might perform as good or better than the PCA, but that PCA anyways seems like a good choice.

Transformation techniques

There are multiple transformation techniques which incorporate scaling and translation invariance. The Z-transformation was chosen as the base transformation in this thesis because it is most used in the literature. As can be seen in the results in this master thesis, the two different transformation techniques used give clusters with different characteristics. This is reasonable, as they provide a different transformation in the same space. With the MM normalization, we saw a tendency of clustering based on their relative consumption compared to the maximum and minimum daily consumption. While the Z-transformation typically clustered maximum one cluster with constant consumption, the MM normalization gave multiple clusters with constant consumption. The clusters in the MM normalization had a different relative consumption, some low and some higher.

In the context of demand response, the clusters from the MM-normalization could be valuable information. The customers with a low consumption relative to their maximum consumption might be able to shift their consumption if desired, while customers with high relative consumption, might not have this opportunity. Customers with medium constant consumption might also be valuable, if they are able to switch their consumption both up and down. However, customer's possibilities and willingness to change their consumption might differ greatly even within clusters with similar characteristics. But clustering with MM-normalized data might be a first step to reveal customers with constant consumption.

The optimal cluster number is also tested with the MM-normalization. For the DB index, different optimal cluster numbers were achieved for all datasets, while the cluster numbers were the same for the SI and CH index. It is hard to draw conclusions from the results, but they show that different optimal cluster number can be achieved with different transformation techniques.

Average load profiles

This data representation technique is a combination of reduction techniques and transformation techniques. It reduces the dataset that is being clustered, and it makes new samples to be clustered, with the same length. The intention behind average load profiles was to see if this reduction could maintain the characteristics of all the samples in the dataset, while drastically reduce the computation time. The computation time on K-Means is only 5% of the computation time with all the samples.

The result on the first dataset shows that the average load profile makes clusters with similar shape, but actually more extreme values. This was not expected, as the average load profile method averages the hourly consumption for each customer before the clustering is performed. This shows that clustering with average load profiles could work as a good approximation for early analyses on big datasets. Using the average load profile representation might make other and more time consuming algorithms, like K-Means with DTW, possible to carry out in early analyses of the dataset.

The results from the optimal cluster number calculations with average load profiles show that in general a lower clustering number is found. This might suggest that finding the optimal cluster number using the average load profile approach might be a good first approach to find the optimal cluster number in a large dataset, as the computation of optimal cluster numbers using all the samples in the dataset was quite time consuming.

Stability criterion

There were seen some tendencies for the different data representation techniques regarding the stability. The MM-normalization showed similar stability as Z-transformation and the reduction techniques showed an increasing stability with increasing CVI scores. The average load profiles showed a high degree of stability, but this clustering included significantly lower samples to cluster than the other clustering approaches.

5.1.3 Practical applications

The practical application shown in this master thesis presents an example of how clustering can be used to analyse a group of customers. The practical application carried out exemplifies one way to look for outliers. As the clustering can group customers, having a small dataset of customers or using an approach which naturally cluster outliers (for example hierarchical clustering with single linkage), some possible customers can be detected. By only inspecting these candidates, the detection of outliers, possibly with wrong categorization, can be much more efficient than inspecting all the customers.

5.1.4 FASIT

The results from this thesis seem to confirm that the common practice today for predicting load profiles is a good approximation of the average consumption. The second peak of the FASIT profiles is however a bit later, around 2 hours, than the peak found in the dataset for both weekdays and weekends. The comparison with FASIT is only done for one month, so one should be careful to draw too strong conclusions. The intention behind testing the clustering against the FASIT load profiles was to show that the current approach only consider average values, and not the variety in the dataset. This is an advantage with clustering, as the different characteristics in the dataset can be found. The next step to improve the common practice today is to find out how clustering, or other techniques, can be used to improve the predictions.

5.2 Analyses of capacity in the grid

The usage of clustering on AMS-data in Norway is expected to mainly find place from 1 year and head (section 2.2.1). Therefore, having analyses that are closer to the calculations and analyses performed today can be a bridge towards familiarizing oneself with analyses on AMS-data, and especially AMS-consumption data (active or reactive power) as used in this thesis. By being familiarised with analyses on AMS-data, clustering on AMS-data might be a reasonable next step in the analyses of AMS-data. The analyses of the capacity in the grid do also give additional and useful information which can be used immediately in the operation and planning of the grid. This is compared to the clustering, which is a great technique for grouping, but maybe a bit hard to see exactly how it can be used in the operation and planning of the grid.

The results from the analyses of capacity in the grid show many interesting and useful results, discussed in the following:

5.2.1 Temperature corrections

As mentioned in the theory, the difference between the maximum power measured temperature corrected (PMT) and the maximum power temperature corrected (PT) might not be the same. Using the sensitivity provided by the DSO in the area analysed, only 47% of the transformers have coinciding peaks for the PMT and PT. This means that in 53% of the transformers, the current approach does not represent the maximum consumption with temperature correction sufficiently. The average difference between PMT and PT is found to be 3.7% in the dataset analysed. For some transformers the difference between the PMT and the PT can be up to 20%. Using the PT instead of PMT might therefore be a big improvement for the DSOs.

5.2.2 Aggregation

The results using the aggregated PMT and PT show similar results to the non-aggregated PMT and PT, but the number of coinciding peaks is lower. The lower number of coinciding peaks is probably because a much higher number of consumption points are analysed. The aggregated PMT and PT use the consumption of all the customers below each transformer, while the non-aggregated PMT and PT only use the consumption in the transformer.

5.2.3 Usefulness of aggregation

The results from the aggregated maximum power is included to focus a bit on what could happen if customers change their consumption patterns. The last months with the corona virus have shown a shift in the consumption patterns in Norway ([91]), which tells us that previous consumption might not always reflect future consumption. This is important to keep in mind when using the results from AMS-data, and it is important to reflect about the consequences of a wrong prediction. When installing electrical infrastructure, the cost of predicting too low electric consumption might be very high, due to later upgrades of the lines or cables. On the other hand, the consequences of predicting too high or low losses in the grid might not have so big consequences.

The average increase of the aggregated peaks compared to the transformer peaks shows an increase of 37% in the dataset analysed. The results are ranging from zero 0% to 100% increase in consumption with aggregated peaks. These values could be quite different using different datasets, but the results show a big difference in the aggregated peaks and transformer peaks.

The average increase of the measured peak in the transformer compared to the aggregated PT is found to be 106% in the dataset analysed. This is a high increase, but a high increase is also expected due to aggregation of customer peaks and as well as temperature correction. It shows however the effect of extreme minimum temperature and aggregation, which is worth considering in some degree in the planning phase.

5.2.4 Assumptions and uncertainties

The temperature sensitivity used for this area by the DSO is quite high compared to commonly used temperature sensitivities. This might give a higher impact to the temperature

corrections than using a smaller temperature sensitivity. The analyses were first performed with a 1% temperature sensitivity (as this was the value used by the DSO in 2018), which also revealed a high percentage (42%) of transformers with non-coinciding PMT and PT peaks. As mentioned in the theory, there are uncertainties around the temperature sensitivities. By using AMS-data, these temperature coefficients can be updated to reflect the actual temperature sensitivities today. The temperature sensitivities can also be different during the year, for example each month or seasons in the year, which might reflect the true temperature sensitivities even better.

The values found in the analyses of capacity do not contain the losses in the grid, neither in the lines from the transformer to the customers nor the losses in the transformer. The values are simply aggregated values from the customers. However, if the characteristics and the topology of the grid were known, this could be included in the analyses.

5.3 Broader discussion

This section will discuss things that do not naturally fit under the discussion of the results.

5.3.1 Applications of clustering

This subsection presents a discussion of the applications and usefulness of clustering. As mentioned in the theory, clustering is the study of methods and algorithms for grouping. We have seen for the results in this master thesis that clustering provide a grouping of the data, and the results from the clustering depend on the components chosen in the clustering approach. When doing clustering, it is important to know what clustering can do, but it might be equally important to know the limitations and shortcomings of clustering. If clustering is an unfamiliar technique, it might be hard to see these limitations.

Additional information

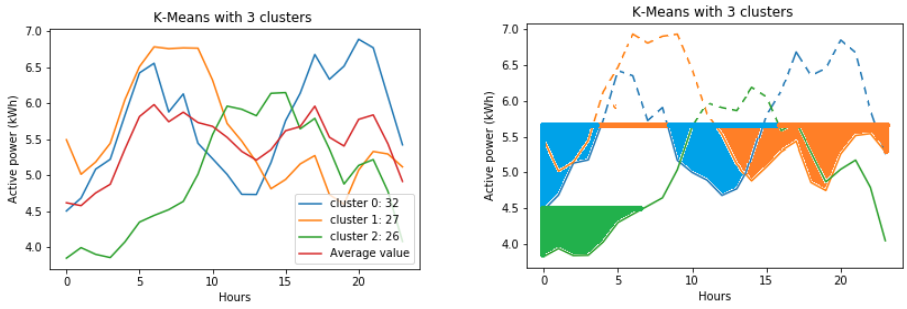
During the work with this master thesis, conversations with employees in Norwegian DSOs have revealed a desire to have more information about the customers, and to have improved load profiles on these customers. Some examples are customers with electric cars, solar panels, heating technology used, number of persons at home and working time. This is useful for DSOs, as more accurate load profiles can be made, and more accurate predictions about the customers consumption can be made. Many people see clustering as a technique which can solve this. However, as mentioned in the theory, clustering is mainly used in practical applications without external labeling information. If we already have a good grouping, why bother to use clustering? The external information can be valuable when using clustering, but it is mainly used as a validation technique, as it shows how well the clustering is performing.

A common question that employees in the Norwegian DSO encounter is whether it is available capacity in the grid for connecting one or multiple customers a given place in the grid. Although clustering can show and visualise characteristics of the dataset analysed, it seems not to be the best technique to answer the question above. Clustering is mainly a tool

for grouping, organizing and visualisation of characteristics in a dataset, not a prediction technique.

Visualisation

One of the strengths with clustering is the possibility to organize and visualise the data by centroids. An example became clear during a conversation with an electric supplier, where guidance of customers that wanted solar panels or batteries was discussed. When approaching and talking to customers, it is important to represent the findings in an understandable way for the customer. By visualising the customers consumption, the effect of including batteries can be added. An example is shown in figure 5.1, a customer from dataset 4. This customer has an average consumption relatively constant from 5:00 to 20:00, and the average consumption never goes above 6 kWh. As the network tariffs proposed (section 2.3) include a cost associated to the daily peak consumption, the average consumption is not the only interesting value. The clusters show that the consumption for all the clusters goes above 6 kWh, and most of the days up to 7 kWh, which means an increased cost. If the customer could install a battery, the maximum consumption in average consumption for each cluster could be decreased. An example of a decrease is shown in figure 5.1b, where the maximum consumption for each cluster centroid is around 5.7 kWh. In this figure, the dashed lines represent the previous consumption, while the filled areas show the new consumption. The actual figures shown to customers would be made nicer, but these simple figures show one way to visualise a customer's consumption.



(a) Clustering of customer

(b) Customer clusters with battery

Figure 5.1: Example of visualisation for customer introducing a battery

Centroid representation

The clustering centroids shown are the centroids which minimize the objective functions, and in many cases they are close to the average values of samples in the cluster. As many samples are gathered in a cluster, the centroids show the typical consumption pattern in that cluster. The samples in the clusters differ therefore probably significantly from the cluster centroids, at least the DTW algorithms give some indications of this. This is important to keep in mind when using the clustering results, that individual samples differ from the

cluster centroid. The usage of the results from AMS-data is briefly discussed in section 5.2.3.

Temperature independence

A strength with clustering is that temperature data might be irrelevant, as the consumption can be transformed to incorporate scaling and translation invariance. From the theory we know that the electric consumption is tightly related to the temperature. But if the relation is to multiply the consumption with a factor and add an offset, this will be cancelled with a transformation which incorporates scaling and translation invariance.

The two FASIT load profiles used in the ED example (shown in figure 2.4a) are the estimated consumption for households on weekdays in a high load period for two days with a temperature difference of $10\,^{\circ}\mathrm{C}$. By normalizing these load profiles they almost align, which indicate that these load profiles are almost temperature independent (that they align when they are transformed). However, it is not known if transformed load profiles are temperature independent in general. If the consumption is temperature independent, there is no reason to include temperature correction before the clustering. This makes the analyses simpler, which make them easier to understand.

Prediction models

Clustering as a first step before performing prediction algorithms might increase the accuracy of the predictions. This is because clustering can group customers with different characteristics, and the different characteristics might make them more sensible to some factors. A common approach in literature is to input the different clusters separately to the prediction models, which in most cases are shown to increase the accuracy of the predictions [44], [53], [92].

Clustering on other datasets

Performing clustering on other datasets might reveal different characteristics. The datasets in addition can be divided based on the seasons during the year, as chosen in this thesis, or based on the days in the week, as with the comparison with FASIT. Other divisions on the datasets, as for example customer groups, might also be done.

Clustering on reactive data

As mentioned in the specialization thesis leading up to this master thesis, clustering on reactive power consumption could give interesting results. The reactive power consumption is known to fluctuate in a higher degree than active power consumption, which could make it hard to see patterns. These patterns might be visible after clustering. Reactive power consumption is much less researched than active power consumption, so clustering could reveal useful information about the reactive power consumption. Especially interesting could be analyses of the reactive power from customers with solar panels, as these customers are known to have more voltage problems.

DSM and network tariffs

The clustering can give information about the characteristics of the customers in the grid. As clustering divide the dataset into groups, which have different characteristics, specialized programs can be performed on the different clusters. Two programs mentioned in the theory are demand side management (DSM) and network tariffs. The clustering might contribute with useful information for deciding these specific programs, but it is important to be aware that much work still lies ahead with specifying the programs after receiving the clustering results. The clustering results should be seen as additional useful information, and not as a solution to how to make the programs proposed.

Clustering as a toolbox

Clustering is a technique used in many fields where a lot of information is available, and it is commonly used today. With the introduction of AMS in the power system, there is also a lot of information available. This data can contribute to improve the operation of the grid as well as provide a more efficient power market. Clustering can be one of the tools to improve the operation of the grid, as the technique can give us information about the customers in the grid. It might be hard sometimes to see exactly what clustering can be used to, but it is advised that the clustering technique can be added to the toolbox of the electrical power engineer having access to or working with AMS-data. By understanding the methods and algorithms, when the engineer encounters a problem, clustering might be one way to solve the problem.

5.3.2 Assumptions and approaches

This subsection discusses the assumptions and approaches taken in this master thesis, and their consequences.

Accuracy of AMS-data

This master thesis analyses datasets of AMS-data, which might include errors. The AMS-data for the months from February to June 2019 for the customers in dataset 1, 2 and 3 contained errors, according to the DSO providing the AMS-data for these datasets. This shows that AMS-data do contain errors, even though the transfer and data apparently look fine. According to the DSO, the values for these months were three times as high as they should have been. Therefore, it could be beneficial to do some basic analyses to look for errors in the datasets. For dataset 4, the daily consumption was plotted against temperature, which showed high correlation. This suggests that the scaling and amplitude of the dataset was constant in the period analysed.

Loss of data

Another source of error in the AMS-data might be loss of data. The samples used as input to the clustering algorithms in this thesis were hourly readings over 1 day. Each sample containing at least one ”Not a Number” (NaN) were discarded before performing the clustering. Another approach could be to fill in the NaN values with 0 or another

number, if the sample contains one or few NaN values. By knowing the reasons behind the loss of data, it would be possible to fill the elements with reasonable values. This has not been the main focus in the thesis, so all samples containing NaN elements were discarded before the clustering.

Clustering algorithms

The clustering algorithms used in this master are provided from software packages available on the internet. The parameters used by each algorithm is chosen as the default values, unless otherwise stated in the text. By changing some of the parameters, different results might be obtained. For most of the algorithms, it is assumed that changing the parameters would only give slightly different results, as most of the parameters deal with things like initialization, tolerances and number of runs. These parameters do not change the way the algorithm works, but rather the how close the local minimum of the objective function in average gets is to the global minimum. It takes longer time to get closer to the global minimum, so there is a trade-off between computational time and accuracy.

Feature extraction, as mentioned in theory, is another data representation technique. By extracting features instead of using the raw data, the computation time and the noise might be reduced. The feature extraction technique can work in the same way as the average load profiles representation, as a fast first analysis. These results might be an aid to decide the next step to solve a given problem.

Optimal cluster numbers

In the theory, two methods to calculate the optimal cluster number were presented. In the analyses, only the method with CVI scores was performed. This was chosen due to the complexity of the Gap statistics compared to the CVI scores. Further research can find optimal cluster numbers using ”Gap Statistics”, and these cluster numbers can be compared with optimal cluster numbers found using the CVI scores.

Computation time

The analyses are performed on a shared server, as described in section 3.2. This is however assumed to have little importance in the computation times found, as multiple runs on the same same algorithms on the same dataset give roughly the same computation time. In addition, the starting point of the algorithms differs each time the algorithm runs, which leads to a different computation, and in many cases a different computation time.

5.3.3 Metering frequency

The clustering performed in this master thesis is done on hourly readings, while the regulating authorities in Norway proposes to increase the metering frequency to 15-min from 2023. Clustering with 4 times higher time resolution might give a bit different results. The reduction techniques might be important to keep the computational time low on large datasets, and the reduction techniques might decrease the clustering quality less since the

resolution already is high. The DTW distance measure might do better, as multiple local shifts might be more desired with increased metering frequency.

5.4 The working process

When I started the work with the specialization project ([58]) leading up to this master thesis in the fall of 2019, I was thinking clustering had many obvious applications in the power system. However, when diving into the vast sea of literature, I found no direct usage of clustering. Specialized programs could be made for the different clusters, but a lot of work making and testing the specialized programs still lies ahead after performing the clustering. The specialization project included cluster analyses using K-Means and K-Shape, but without any applications of the results. The last chapter, ”Further work”, in the specialization project explained what would be the focus of this master thesis.

I started the master thesis following the plan described in the specialization project and used the first period to read and implement CVIs and more algorithms. I was still looking for more direct usage of the clustering results than the specialized programs, but I did not find much. Therefore, I changed the focus from the plan proposed, and started working with something which was closer to the operation of the grid today, the analyses of capacity in the grid. The biggest problem of the thesis has without doubt been to see the applications of clustering on AMS-data.

Chapter 6

Conclusion

This section presents the conclusions from this master thesis. The conclusions are based on the results in chapter 4, discussions in chapter 5 and the theory in chapter 2.

The theory gives the reader good insight in clustering, as well as other background information useful when analysing and using clustering results. The theory makes the reader able to understand the clustering performed in this master thesis, as well as a basis to read scientific articles about clustering.

In this master thesis, 6 different clustering algorithms and 5 data representation techniques on four different datasets have been tested and analysed. The datasets are from four different seasons and contain from 1227 to 4681 customers. Representations of cluster centroids from the approaches tested are shown, both with the dataset transformed as well as on the raw data. As suggested in the literature, the different clustering approaches have been tested against the commonly used K-Means algorithm. The performance of the clustering is evaluated using the CVI indexes Davies-Bouldin (DB), Silhoutte (SI) and Calinsky-Harabasz (CH). Recall that lower DB-score and higher SI and CH-score means better clustering quality. The results are summarized in table 6.1, and all the CVI scores are divided by the CVI scores from K-Means. The computation time for each approach is the average computation time compared to K-Means.

Clustering approach	DB	SI	CH	Computation time
K-Means	1	1	1	11s
K-Shape	1,008	1,017	1,000	x57
K-Means + DTW	1,131	0,811	0,899	x1012
SOM	2,003	-0,559	0,123	x5
Hierarchical + single	1,109	-2,744	0,012	x52
Hierarchical + ward	1,277	0,444	0,581	x273
K-Means + PCA	1,027	0,962	0,989	x0,61
K-Means + wavelets	1,079	0,897	0,949	x0,56

Table 6.1: Summary of results

The results show that the partitional algorithms are superior. The hierarchical and SOM algorithms have a higher computational time than K-Means, and the clustering quality is lower. The hierarchical algorithm using single linkage shows some interesting results, as it detects outliers in a much higher degree than all the other clustering algorithms tested.

The three partitional algorithms tested are K-Means, K-Shape and K-Means with DTW. The clusters from the K-Shape algorithm are quite similar to the clusters from the K-Means algorithm, both in size, shape and CVI scores. The computation time is around 60 times higher than the K-Means algorithm on the datasets tested. As the CVI indexes are biased towards K-Means, the performance of K-Shape is considered slightly better. If increased accuracy is preferred, K-Shape might be better than K-Means, with the drawback of increased computation time. The K-Means with DTW provide a clustering with different sizes, shapes and CVI scores than the K-Means and K-Shape. The CVI scores are considerably worse, but it is hard to draw conclusion based on the CVI scores, as the algorithm is punished for its local time shifts. The computation time is much higher for K-Means with DTW than K-Means, a factor of around 1000, and increasing with the size of the dataset. The K-Means with DTW is commonly highlighted as the best algorithm incorporating shift invariance, but the results from this thesis raise the question whether the shift invariance incorporated weights up for the highly increased computational time.

The master thesis analyses 5 different data representation techniques. The Z-transformation is used as a base representation as this is the most common in literature. The results show that the PCA is the preferred reduction technique compared to discrete wavelet transformation. By reducing the dataset with 75%, the clustering quality drops by 3% using PCA, compared to 8% using discrete wavelet transformation. Reduction techniques might be more important as the metering frequency increases. The Min-Max normalization provides a different clustering than the Z-transformation, and the algorithm tends to cluster more based on the consumption relative to the maximum and minimum consumption. The representation using average load profiles is a fast approach, reducing the computation time with 95% in the datasets analysed. Although fast, the representation managed to keep the characteristics of the dataset analysed.

The comparison with FASIT shows that the current practice of predicting load profiles for households gives a good approximation to the average consumption of the dataset analysed. The results also show that while it is a good approximation to the average consumption, there are large differences within the dataset, which could be exploited further.

The results from the analyses of capacity in the grid can be used by grid operators to improve the planning and operation of the grid. The most interesting result was the difference on 3.7% between the peak consumption temperature corrected and the highest temperature corrected consumption. As the common practice is only to temperature correct the peak consumption, that value might not be representative for the highest temperature corrected consumption.

Clustering on AMS-data divides the dataset into clusters with different characteristics, which can be beneficial for visualisation or applications. Some applications of clustering on AMS-data are DSM, network tariffs and as a pre-stage before prediction models. By understanding the principles and components used in clustering, other applications for clustering on AMS-data might be found.

Chapter 7

Further work

This master thesis showed a comprehensive study of clustering algorithms and data representation techniques on datasets from four different seasons. The thesis showed important results from the clustering, but more research can of course be done to make the results even more robust. This especially applies for K-Shape and K-Means with DTW, as the CVI indexes used are biased towards the ED distance measure. One way to test the performance of these algorithms is using external labelling information, for example customer groups. If using customer groups as the external labelling information, one should keep in mind that customers from different groups might have similar consumption patterns, which the algorithms try to cluster together.

This master thesis presents two ways to calculate the optimal cluster number, but only one technique is used in the analyses. Further research can use the ”Gap Statistics” to find the optimal cluster numbers, which can be compared with optimal cluster numbers found using CVI scores, as done in this thesis.

Another interesting study area is clustering on reactive power data. The reactive power consumption is much less studied than active power consumption and is known to be quite fuzzy. Clustering might reveal some unknown patterns, which might improve the grid operation, especially in areas with voltage problems.

As much research lies ahead before finding the optimal clustering algorithms on AMS-data, maybe the most interesting and useful further work is to see what practical applications clustering can contribute in. Some applications are mentioned in the thesis: DSM, specialized network tariffs and as a pre-stage to prediction models. However, there might be many more areas where clustering can improve today’s practice. The ones that are working daily with AMS-data related problems are the ones best suited to see what improvements clustering can give. In order to see these improvements, insight in clustering principles as well as the clustering components is necessary.

Bibliography

[1] K.Mets, F.Depuydt, and C.Develder, "Two-stage load pattern clustering using fast wavelet transformation," *IEEE Transactions on Smart Grid*, vol. 7, no. 5, pp. 2250–2259, Sep. 2016.

[2] P.Jerstad, B.Sletbak, A.A.Grimenes, and R.Renstrøm, *Rom Stoff Tid Fysikk 2*, 1st ed. Cappelen Damm AS, 2008.

[3] E. Palermo, "Who invented the light bulb?" https://www.livescience.com/43424-who-invented-the-light-bulb.html, accessed: 2020-04-10.

[4] Wikipedia, "Street light," https://en.wikipedia.org/wiki/Street_light, accessed: 2020-04-10.

[5] ——, "History of electric power transmission," https://en.wikipedia.org/wiki/History_of_electric_power_transmission, accessed: 2020-04-10.

[6] E. Gambrel, "Sources of energy from the 1800s," https://sciencing.com/sources-energy-1800s-8126819.html, accessed: 2020-04-10.

[7] Eurostat, "Electricity generation statistics – first results," https://ec.europa.eu/eurostat/statistics-explained/index.php/Electricity_generation_statistics_%E2%80%93_first_results#Production_of_electricity, accessed: 2020-04-10.

[8] European Environment Agency, "Electricity production by fuel," https://www.eea.europa.eu/data-and-maps/indicators/electricity-production-by-fuel-1/electricity-production-by-fuel-assessment-3, accessed: 2020-04-10.

[9] S. Garvey, A. Hoskin, and J. Rouse, "Fully renewable electricity could destabilise the grid – but old coal plants may have a solution," https://theconversation.com/fully-renewable-electricity-could-destabilise-the-grid-but-old-coal-plants-may-have-a-solution-1333 accessed: 2020-04-10.

[10] Wikipedia, "Electrcity Market," https://en.wikipedia.org/wiki/Electricity_market, accessed: 2020-04-17.

[11] Nordpool, “Simple, efficient & secure,” https://www.nordpoolgroup.com/, accessed: 2020-04-17.

[12] ——, “Order types,” https://www.nordpoolgroup.com/trading/Day-ahead-trading/Order-types/, accessed: 2020-04-10.

[13] ——, “Day-ahead market,” https://www.nordpoolgroup.com/the-power-market/Day-ahead-market/, accessed: 2020-04-10.

[14] A. Eriksen, H.Hansen, J. .Hole, T.Jonassen, V.Mook, S.Steinnes, and L.Varden, “RME HØRINGSDOKUMENT,” NVE, Oslo, Tech. Rep. 1, 02 2020, acessible at http://publikasjoner.nve.no/rme_hoeringsdokument/2020/rme_hoeringsdokument2020_01.pdf.

[15] B. Murray, “The Paradox of Declining Renewable Costs and Rising Electricity Prices,” https://www.forbes.com/sites/brianmurray1/2019/06/17/the-paradox-of-declining-renewable-costs-and-rising-electricity-prices/#2a44215f61d5, accessed: 2020-04-17.

[16] K.Ingebrigtsen, V.Hoffmann and A.J.Berre, “Applications of big data and data science in the electricity distribution grid,” SINTEF, Trondheim, Tech. Rep., 2020.

[17] Tibber, “Lad elbilen smartere.” https://tibber.com/no/smart-stromavtale, accessed: 2020-04-17.

[18] ——, “En strømavtale som matcher 2020,” https://tibber.com/no/losninger/smartlading-elbil, accessed: 2020-04-10.

[19] NELFO, “Effekteffektivisering i husholdningene – et alternativ til nettinvesteringer,” NELFO, Oslo, Tech. Rep., 2018.

[20] NVE, “Fremtidens Avanserte Måle og Styringssystem (AMS),” NVE, Oslo, Tech. Rep. 36, 08 2019, acessible at http://publikasjoner.nve.no/rapport/2019/rapport2019_34.pdf.

[21] ——, “FOR-1999-03-11-301: Forskrift om måling, avregning, fakturering av nettjenester og elektrisk energi, nettselskapets nøytralitet mv.” https://lovdata.no/dokument/SF/forskrift/1999-03-11-301, accessed: 2019-10-22.

[22] ——, “Smarte målarar (AMS),” NVE, Oslo, Tech. Rep. 24, 04 2019, acessible at http://publikasjoner.nve.no/rapport/2019/rapport2019_24.pdf.

[23] ——, “Høring: Modell for innføring av 15 minutters balanseavregning samt forslag til nye funksjonskrav og krav til sikkerhet for AMS,” https://www.nve.no/reguleringsmyndigheten/nytt-fra-rme/nyheter-reguleringsmyndigheten-for-energi/horing-modell-for-innforing-av-15-minutters-balanseavregning-samt-forslag-til-nye-funksjonskrav-c accessed: 2020-04-21.

[24] Elhub, “Om elhub,” https://elhub.no/om-elhub/, accessed: 2020-06-01.

[25] S. Mathisen, "Smarte målere leser av strømmen for deg," https://blogg.fortum.no/smart-strom-avanserte-male-og-styringssystemer-ams, accessed: 2020-06-01.

[26] NVE, "FOR-2007-12-28-1753: Forskrift om krav til elektrisitetsmålere," https://lovdata.no/dokument/SF/forskrift/2007-12-28-1753, accessed: 2020-04-22.

[27] ——, "Smarte strømmålere (AMS)," https://www.nve.no/stromkunde/smarte-strommalere-ams/, accessed: 2019-10-22.

[28] O. Lysne, "Norges offentlige utredninger - Digital sårbarhet – sikkert samfunn," Departementenes sikkerhets- og serviceorganisasjon Informasjonsforvaltning, Oslo, Tech. Rep. 13, 2015, acessible at https://www.regjeringen.no/contentassets/fe88e9ea8a354bd1b63bc0022469f644/no/pdfs/nou201520150013000dddpdfs.pdf.

[29] C.Frøystad, M.G.Jaatun, K.Bernsmed, and M.Moe, "Risiko- og sårbarhetsanalyse for økt integrasjon av AMS-DMS-SCADA," NVE, Oslo, Tech. Rep. 15, 2018, acessible at http://publikasjoner.nve.no/eksternrapport/2018/eksternrapport2018_15.pdf.

[30] F.Skapalen and B.Jonassen, "Veileder til sikkerhet i avanserte måle- og styringssystem," NVE, Oslo, Tech. Rep. 7, 2012, acessible at https://www.nve.no/Media/5525/veiledertil-sikkerhet-i-ams.pdf.

[31] SINTEF, "Energy analytics," https://www.sintef.no/projectweb/energytics/, accessed: 2019-10-27.

[32] ENOVA, "Pilotene – disse går foran," https://www.enova.no/privat/smarte-strommalere-ams/pilotene--disse-gar-foran/, accessed: 2020-06-01.

[33] A.B.Eriksen, H. Hansen, J. Hole, T. Jonassen, V. Mook, S. Steinnes, and L. Varden, "Effekteffektivisering i husholdningene – et alternativ til nettinvesteringer," NVE, Oslo, Tech. Rep., 2020.

[34] H. Løvik, "Nytt system for strømpriser og nettleie slaktes av en samlet energibransje," https://www.tu.no/artikler/nytt-system-for-strompriser-og-nettleie-slaktes-av-en-samlet-energibransje/432162, accessed: 2020-04-21.

[35] A. Torsheim, "Det nye forslaget til nettariff vil også straffe alle med varmepumpe og isolasjon," https://www.tu.no/artikler/det-nye-forslaget-til-nettariff-vil-ogsa-straffe-alle-med-varmepumpe-og-isolasjon/480229?utm_source=tu.no&utm_medium=feed&utm_campaign=feed_most_read, accessed: 2020-04-21.

[36] E. Lindvoll, "Pris på strøm - Så mye kan du spare på strøm til timepris," https://www.dinside.no/okonomi/sa-mye-kan-du-spare-pa-strom-til-timepris/70659242, accessed: 2020-04-21.

[37] Sentio Research Norge AS, "Forbrukerens tilpasning i strømmarkedet 2017," NVE, Oslo, Tech. Rep. 8, 01 2018, acessible at http://publikasjoner.nve.no/rapport/2018/rapport2018_08.pdf.

[38] G. Chicco, "Overview and performance assessment of the clustering methods for electrical load pattern grouping," *Energy*, vol. 42, p. 68–80, 06 2012.

[39] T.H.Dang-Ha, R.Olsson, and H.Wang, "Clustering methods for electricity consumers: An empirical study in hvaler -norway," 11 2016.

[40] A.K.Jain, "Data clustering: 50 years beyond k-means," *Pattern Recognition Letters*, vol. 31, no. 8, pp. 651 – 666, 2010, award winning papers from the 19th International Conference on Pattern Recognition (ICPR). [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167865509002323

[41] S. Aghabozorgi, A. S. Shirkhorshidi, and T. Y. Wah, "Time-series clustering – a decade review," *Information Systems*, vol. 53, pp. 16 – 38, 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0306437915000733

[42] M.Kantardzic, *Data Mining: Concepts, Models, Methods, and Algorithms*. John Wiley & Sons, Inc., 2011.

[43] O. Arbelaitz, I. Gurrutxaga, J. Muguerza, J. M. Pérez, and I. Perona, "An extensive comparative study of cluster validity indices," *Pattern Recognition*, vol. 46, no. 1, pp. 243 – 256, 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S003132031200338X

[44] J.Yang, C.Ning, F.Zhang, C.Deb, K.W.D.Cheong, S.Lee, C.Sekhar, and K.Tham, "K-shape clustering algorithm for building energy usage patterns analysis and forecasting model accuracy improvement," *Energy and Buildings*, vol. 146, 03 2017.

[45] Wikipedia, "Cluster analysis," https://en.wikipedia.org/wiki/Cluster_analysis, accessed: 2019-11-02.

[46] A. Aleshinloye, A. Bais, and I. Al-Anbagi, "Performance analysis of dimensionality reduction techniques for demand side management," in *2017 IEEE Electrical Power and Energy Conference (EPEC)*, Oct 2017, pp. 1–6.

[47] Y.Lu, T.Zhang, and Z.Zeng, "Adaptive weighted fuzzy clustering algorithm for load profiling of smart grid customers," in *2016 IEEE/CIC International Conference on Communications in China (ICCC)*, July 2016, pp. 1–6.

[48] A.Taspinar, "A guide for using the wavelet transform in machine learning," http://ataspinar.com/2018/12/21/a-guide-for-using-the-wavelet-transform-in-machine-learning/, accessed: 2020-05-29.

[49] S. Yilmaz, J. Chambers, and M. Patel, "Comparison of clustering approaches for domestic electricity load profile characterisation - implications for demand side management," *Energy*, vol. 180, pp. 665 – 677, 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0360544219310060

[50] S. Yilmaz, S. Weber, and M. Patel, "Who is sensitive to dsm? understanding the determinants of the shape of electricity load curves and demand shifting: Socio-demographic characteristics, appliance use and attitudes," *Energy Policy*, vol. 133, p. 110909, 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0301421519304872

[51] J.Paparrizos and L.Gravano, "k-shape: Efficient and accurate clustering of time series," *ACM SIGMOD Record*, vol. 45, pp. 69–76, 06 2016.

[52] G.Batista, E.Keogh, O.Tataw, and S.A.S.Vinícius, "Cid: An efficient complexity-invariant distance for time series," *Data Mining and Knowledge Discovery*, vol. 28, 04 2013.

[53] F.Fahiman, S.M.Erfani, S.Rajasegarar, M.Palaniswami, and C. Leckie, "Improving load forecasting based on deep learning and k-shape clustering," in *2017 International Joint Conference on Neural Networks (IJCNN)*, May 2017, pp. 4134–4141.

[54] H. Sakoe and S. Chiba, "Dynamic programming algorithm optimization for spoken word recognition," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 26, no. 1, pp. 43–49, February 1978.

[55] T. Teeraratkul, D. O'Neill, and S. Lall, "Shape-based approach to household load curve clustering and prediction," *IEEE Transactions on Smart Grid*, vol. PP, 02 2017.

[56] J. Coleman, "Dynamic programming example (dynamic time warping)," http://www.phon.ox.ac.uk/jcoleman/old_SLP/Lecture_5/DTW_explanation.html, accessed: 2020-03-30.

[57] B.Hasan, "Introduction To Dynamic Time Warping," https://riptutorial.com/dynamic-programming/example/25780/introduction-to-dynamic-time-warping, accessed: 2020-03-30.

[58] S. Norheim, "Clustering of AMS-data," Department of Electric Power Engineering, NTNU – Norwegian University of Science and Technology, Project report in TET4520, Dec. 2019.

[59] F. Petitjean, A. Ketterlin, and P. Gançarski, "A global averaging method for dynamic time warping, with applications to clustering," *Pattern Recognition*, vol. 44, no. 3, pp. 678 – 693, 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S003132031000453X

[60] T. Kohonen, "Self-organized formation of topologically correct feature maps." *Biol. Cybern.*, pp. 44–69, 1982.

[61] Wikipedia, "Self-organizing map," https://en.wikipedia.org/wiki/Self-organizing_map, accessed: 2020-04-21.

[62] V. Moosavi, S. Packmann, and I. Vallés, "Sompy: A python library for self organizing map (som)," 2014, gitHub.[Online]. Available: https://github.com/sevamoo/SOMPY.

[63] T. Kohonen, *Self-Organizationg and Associative Memory*, 3rd ed. Springer Verlag, 2012.

[64] SINTEF, "Unpublished work, ENERGYTICS."

[65] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *Journal of Computational and Applied Mathematics*, vol. 20, pp. 53 – 65, 1987. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0377042787901257

[66] G. W. Milligan and M. C. Cooper, "An examination of procedures for determining the number of clusters in a data set," *Psychometrika*, vol. 50, no. 2, pp. 159–179, Jun 1985. [Online]. Available: https://doi.org/10.1007/BF02294245

[67] R. Tibshirani, G. Walther, and T. Hastie, "Estimating the number of clusters in a data set via the gap statistic," *Journal of the Royal Statistical Society Series B*, vol. 63, pp. 411–423, 02 2001.

[68] Z.Jiang, R.Lin, and F.Yang, "A hybrid machine learning model for electricity consumer categorization using smart meter data," *Energies*, vol. 11, p. 2235, 08 2018.

[69] A.O.Eggen and H.Vefsnmo, "FASIT kravspesifikasjon," SINTEF, Trondheim, Tech. Rep., 11 2018.

[70] SINTEF, "Clustering of customers based on ams-data," https://www.sintef.no/projectweb/energytics/focus-areas/analysis-of-voltage-quality-and-consumption-data/clustering-of-customers-based-on-ams-data/, accessed: 2019-10-27.

[71] N.Sandberg, I.Sartori, M.Vestrum, and H.Brattebø, "Using a segmented dynamic dwelling stock model for scenario analysis of future energy demand: The dwelling stock of norway 2016–2050," *Energy and Buildings*, vol. 146, 04 2017.

[72] VaasaETT, "Smarte målere (AMS) og feedback," NVE, Oslo, Tech. Rep. 72, 10 2014, acessible at http://publikasjoner.nve.no/rapport/2014/rapport2014_72.pdf.

[73] Enel X, "What is demand side management?" https://www.enelx.com/en/questions-and-answers/eindustry/what-is-demand-side-management, accessed: 2020-05-30.

[74] S. Rahman, "Temperature correction of energy statistics," https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/65732/2089-ons-paper-temp-correction-of-energy-stat.pdf, accessed: 2020-04-21.

[75] Pöyry Management Consulting AS , "Evaluering av modeller for klimajustering av energibruk," NVE, Oslo, Tech. Rep. 7, 2014, acessible at http://publikasjoner.nve.no/rapport/2014/rapport2014_07.pdf.

[76] ENOVA, "Graddagstall ," https://www.enova.no/om-enova/drift/graddagstall/, accessed: 2019-10-22.

[77] ——, " Enovas byggstatistikk 2017."

[78] BKK, "Regional Kraftsystemutredning for BKK-området og indre Hardanger 2018 - 2038," https://nett.bkk.no/aktuelt?dtmSource=%2Fnett%2Fkraftsystem, accessed: 2020-04-13.

[79] NTE, "Kraftsystemutredning for tidligere Nord-Trøndelag 2018-2038," https://tn.tensio.no/nyhetsarkiv/nyhet/kraftsystemutredning-for-tidligere-nord-trondelag-2018-2038, accessed: 2020-04-13.

[80] Lyse, "Kraftsystemutredning for Sør-Rogaland 2018 - 2038," https://www.lysenett.no/getfile.php/131413-1571838924/Dokumenter/Utredninger%20og%20rapporter%20%28KSU%29/Hovedrapport_-_Kraftsystemutredningen_for_S\OT1\or-Rogaland_2018-2038.pdf, accessed: 2020-04-13.

[81] Hafslund, "Oslo Akershus og Østfold - Kraftsystemutredning 2018–2038 ," https://assets.ctfassets.net/p14w1ia3hr36/5rEIzlSbNS86agkkwuEoaG/38d5eaf946852b9c18ff3cfe868ee49c/kraftsystemutredning_HN_Hovedrapport_2018_2038.pdf, accessed: 2020-04-13.

[82] F.Pedregosa, G.Varoquaux, G.A.Gramfort, V.Michel, B.Thirion, O.Grisel, M.Blondel, P.Prettenhofer, R.Weiss, V.Dubourg, J.Vanderplas, A.Passos, D.Cournapeau, M.Brucher, M.Perrot, and E.Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[83] R. Tavenard, J. Faouzi, G. Vandewiele, F. Divo, G. Androz, C. Holtz, M. Payne, R. Yurchak, M. Rußwurm, K. Kolar, and E. Woods, "tslearn: A machine learning toolkit dedicated to time-series data," 2017, https://github.com/tslearn-team/tslearn.

[84] G. Lee, R. Gommers, F. Waselewski, K. Wohlfahrt, and A. O'Leary, "Pywavelets: A python package for wavelet analysis," *Journal of Open Source Software*, vol. 4, no. 36, p. 1237, 2019. [Online]. Available: https://doi.org/10.21105/joss.01237

[85] F.Pedregosa, G.Varoquaux, G.A.Gramfort, V.Michel, B.Thirion, O.Grisel, M.Blondel, P.Prettenhofer, R.Weiss, V.Dubourg, J.Vanderplas, A.Passos, D.Cournapeau, M.Brucher, M.Perrot, and E.Duchesnay, "sklearn.cluster.KMeans," https://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html, accessed: 2020-04-15.

[86] R. Tavenard, J. Faouzi, G. Vandewiele, F. Divo, G. Androz, C. Holtz, M. Payne, R. Yurchak, M. Rußwurm, K. Kolar, and E. Woods, "tslearn.clustering.KShape," https://tslearn.readthedocs.io/en/stable/gen_modules/clustering/tslearn.clustering.KShape.html, accessed: 2020-04-15.

[87] F.Pedregosa, G.Varoquaux, G.A.Gramfort, V.Michel, B.Thirion, O.Grisel, M.Blondel, P.Prettenhofer, R.Weiss, V.Dubourg, J.Vanderplas, A.Passos, D.Cournapeau, M.Brucher, M.Perrot, and E.Duchesnay,

"sklearn.cluster.AgglomerativeClustering," https://scikit-learn.org/stable/modules/generated/sklearn.cluster.AgglomerativeClustering.html, accessed: 2020-04-15.

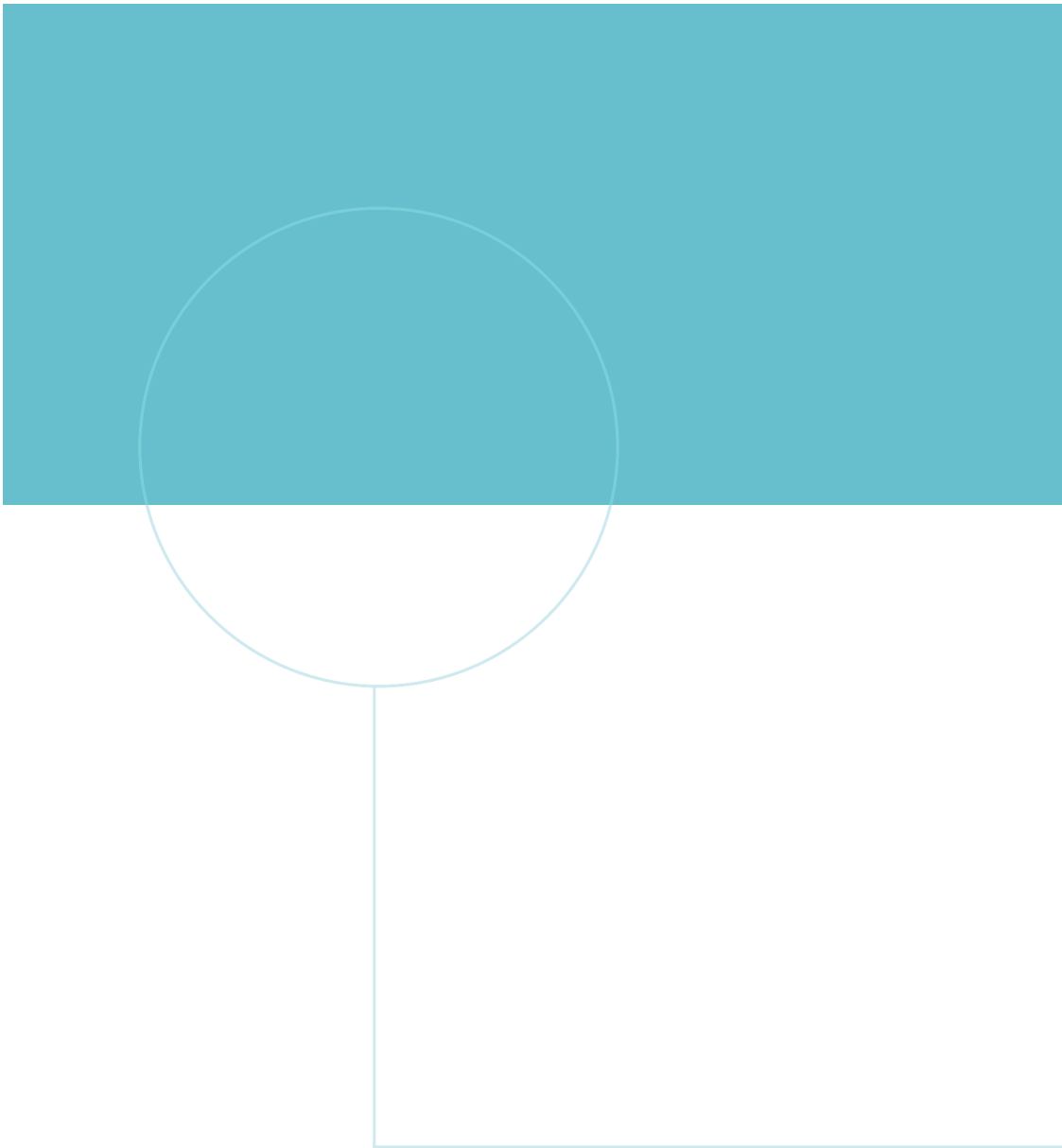
[88] ——, "sklearn.decomposition.PCA," https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.PCA.html, accessed: 2020-04-15.

[89] G.R.Lee, R.Gommers, F.Wasilewski, K.Wohlfahrt, and A.O'Leary, "Discrete Wavelet Transform (DWT)," https://pywavelets.readthedocs.io/en/latest/ref/dwt-discrete-wavelet-transform.html, accessed: 2020-04-15.

[90] The Norwegian Meteorological Institute, "What is frost?" https://frost.met.no/index.html, accessed: 2020-04-20.

[91] E.S.Viseth, "Hjemmekontor og karantene: Strømforbruket hjemme har endret seg," https://www.tu.no/artikler/hjemmekontor-og-karantene-stromforbruket-hjemme-har-endret-seg/488176, accessed: 2020-06-06.

[92] J. Cugliari, Y. Goude, and J. Poggi, "Disaggregated electricity forecasting using wavelet-based clustering of individual consumers," in *2016 IEEE International Energy Conference (ENERGYCON)*, 2016, pp. 1–6.



NTNU
Norwegian University of
Science and Technology

SINTEF