UNIVERSITY OF LJUBLJANA

MASTERS THESIS

Development and analysis of new activation based load profiles

Author:
Jakob JENKO

Supervisor:
Dr. Carolina FORTUNA and
Dr. Marko MEŽA and

A thesis submitted in fulfillment of the requirements for the degree of Masters of electrical engineering

in the

-
ICT

November 5, 2022

Declaration of Authorship

I, Jakob JENKO, declare that this thesis titled, "Development and analysis of new activation based load profiles" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

Date:

"In science, great oaks grow from little acorns."

D. Everett

UNIVERSITY OF LJUBLJANA

Abstract

faculty of electical engineering
ICT

Masters of electrical engineering

Development and analysis of new activation based load profiles

by Jakob JENKO

An increasing amount of energy data holds enormous amounts of untapped potential. The efficient presentation of energy data to humans and machines through load profiles is the constant narrative throughout the thesis.

Besides the amount of data, the consumption of electricity itself is increasing. A third of electrical energy in the EU is consumed in residential areas. Therefore, optimizing consumption would leave a significant impact on reducing the human footprint.

Another problem that the EU is facing is an aging population. Using the energy data, it would be possible to design an anomaly detection system that would detect accidents such as falls, strokes or dementia-induced altered behavior.

The idea of solving big issues using big data led us to perform a comprehensive review of existing publications and use cases. It was found that some publications have already tried to solve the mentioned issues. Using the related work publications, we have made a table of possible ways to present the data, which left us with a lot of gaps. These gaps present load profiles that were not yet researched or used. The question we tried to answer next was, are these profiles really useful? It could be, that they are impractical, which could be the reason they were not used before.

During the thesis, we utilized these previously unused load profiles. This was done first by looking at how data and load profiles are related in high dimensional space using t-SNE. Newly obtained knowledge aided us when designing and constructing the elderly care system. The successful implementation of newly proposed load profiles proved that the new profiles can be efficiently utilized. While we filled in few gaps, many are left to be researched.

Acknowledgements

I would like to thank the following people without which this thesis would not be possible. Dr. Carolina Fortuna and Dr. Marko Meža helped me with the research, by providing continuous feedback and support. The scientific and open-source community provided the data and the tools used throughout the thesis. Friends and colleagues, who through discussion helped induce the ideas that were put forward through the thesis. My family, who inspired, supported and encouraged me to pursue my studies in the first place. And my partner Nika, who supported me through all my years of studies in all aspects possible.

Contents

List of Figures

List of Tables

List of Abbreviations

LP	**L**oad **P**rofile
ALP	**A**activation **L**oad **P**rofile
SLP	**S**tandad **L**oad **P**rofile
TP	**T**able of load **P**rofiles
P	**P**ower (profile)
A	**A**ctivation (profile)
ZEB	**Z**ero **E**nergy **B**uilding
DR	**D**emand **R**esponse
AD	**A**nomaly **D**etection
EC	**E**lderly **C**are
DER	**D**istributed **E**nergy **R**esources
HVAC	**H**eating, **V**entilation and **A**ir **C**onditioning
EV	**E**lectric **V**ehiecle
PV	**P**hoto **V**oltaics
EU	**E**uropean **U**nion
NILM	**N**on **I**ntrusive **L**oad **M**onitoring
t-SNE	**t**-distributed **s**tochastic **n**eighbor **e**mbedding
PCA	**P**rincipal **C**omponent **A**nalysis

Chapter 1

Introduction

Climate change calls for a shift to renewable energy and restructuring of the electric power industry. Sources Eurostat, 2022 show as of the time of reading this paper, 44 % of produced electricity in Europe was from combustible sources such as gas, fuel, and coal. Even though that is a significant decrease of 10 % in the last 10 years, it is a significant Co2 emitter. The same source Eurostat, 2022 also states that a third of energy is consumed by the residential sector. It is estimated, that the human population will reach 10 billion inhabitants in the next 10 years, and ever-increasing ownership of electrical appliances such as smartphones, HVACs, and EVs will further elevate this issue. Acknowledging this, reducing consumption in the residential sector could leave a significant impact on the human footprint.

The EU aims to be climate neutral by 2050, therefore it seeks to improve the efficiency of every part of pollution contributors through The European Green Deal. A large part of these contributors is the Energy sector. A subpart of the energy sector is the residential sector, where many advancements could be made to help to reach the goal.

This could be achieved through various applications and methods that use load profiling as their core technology. Authors in Chuan, Rao, and Ukil, 2014 proposed a method to reduce peak loads by studying consumer appliance usage patterns. Paper Csoknyai et al., 2019 studied consumer usage patterns, and returned feedback that contributed to reducing consumption. Another notable way is the use of distributed energy resources and managing them in such a way as to decrease the net output of energy flow such as the authors describe in Moreno Jaramillo et al., 2021. All described methods would reduce and alleviate the load off the power grid.

Load profiling in building energy consumption is not a novelty and had been in research since the 1980s. While it was thought that aggregated LPs of households are relatively predictable, recent data obtained using smart meter data showed large deviance from user to user due to different lifestyles, as the author states in Proedrou, 2021. In recent years LPs have changed due to renewable energy accelerated development of distributed energy resources such as residential photovoltaic power plants, home wind energy, and using EVs with home batteries. Socioeconomic changes such as work-from-home, also drastically reshaped the LP curve.

The thesis aims to propose and develop new, previously unused LPs, that will contribute to mitigating the raised issues. Before we disclose our contributions, let us first have an overview of what load profiles are and in which other use cases they can be utilized, besides the energy saving that we just mentioned.

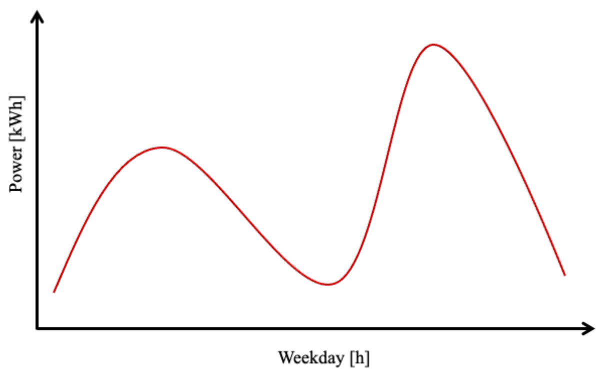
1.1 Definition and types of LPs

Author Proedrou, 2021 defines terms as following:

- Load: the electricity that all the electricity-powered devices in the household consume in unit time.
- Profile: a graph representing the significant features of the electricity load over time.

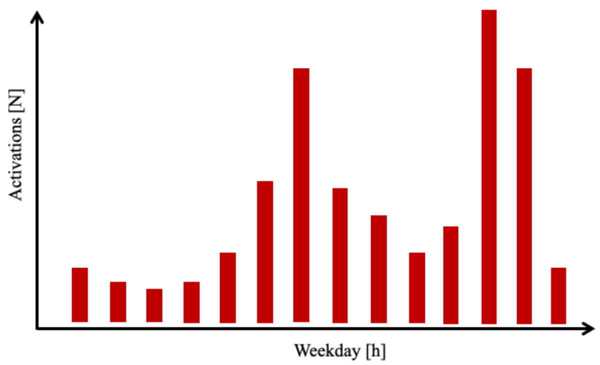
While the most commonly used feature is power, there are other derivatives such as the number of activations of an appliance or operating time. Usually, the LPs are presented as a daily power consumption profile such as shown in Figure 1.1. This profile is also known as the standard daily load profile. While the LP is a sketch, it still presents consumption trends in morning and evening peaks.

FIGURE 1.1: Average daily usage profile for an appliance or a building



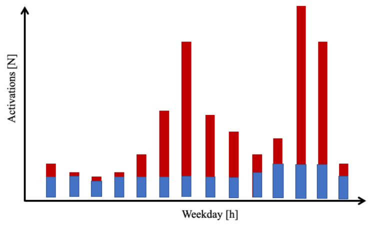
Alternatively, we can use a histogram-based presentation such as can be seen in Figure 1.2. While Figure 1.2 presents the same data as Figure 1.1, due to data processing, it could potentially reveal more relevant consumption patterns.

FIGURE 1.2: Histogram of daily activations profile for an appliance or a building



LPs can present whole-house usage as well as per-appliance usage, where each presentation has its advantages and disadvantages. To present more information, sub-meter data can be used to present whole-house usage with per-appliance contributions. Two of such examples can be seen on Figure 1.3a and 1.3b.

(A) Histogram of daily activations profile for appliance A and B



(B) Average weekday power consumption for appliances A, B and C

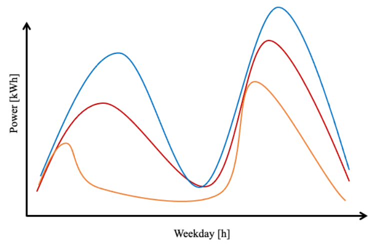


FIGURE 1.3: LPs with multiple appliances

To present as much information as possible, all the above-mentioned attributes can be presented in a multidimensional way such as shown in Figure 1.4 and 1.5.

FIGURE 1.4: Number of daily activations/power consumption of one appliance/house in one-month period

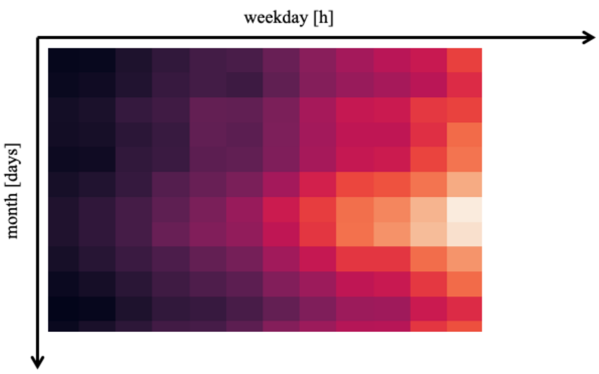
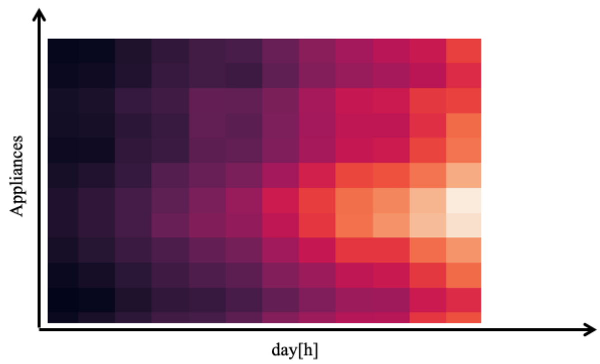


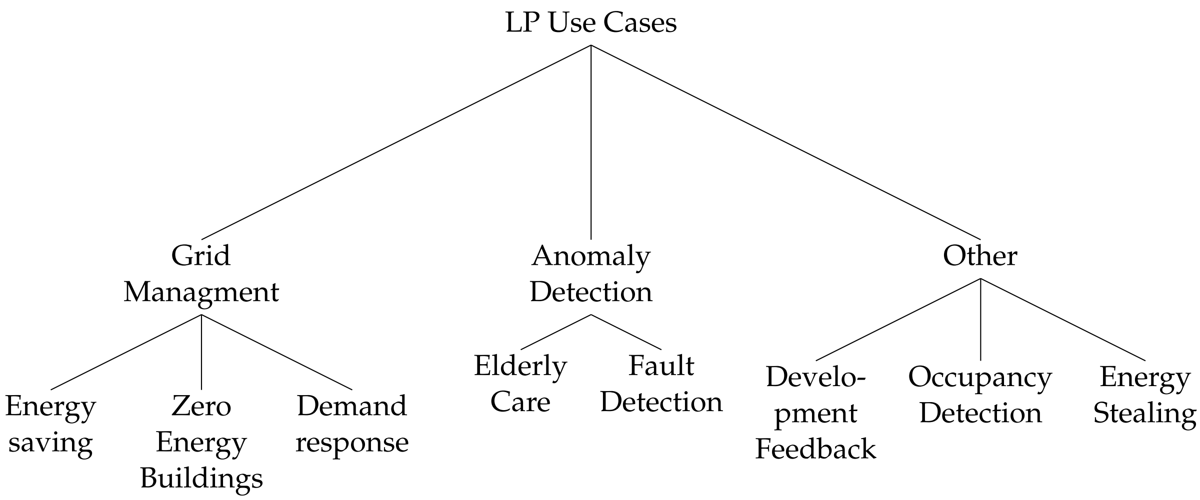
Figure 1.4 is a sketch, and it does not present real-world data. Even though, it is still possible to see the consumption throughout each day that the plot presents one month of data, where we can see the consumption throughout each day. The brightness presents the activity of the household or an appliance. The brighter the plot, the more activity for that hour of that day of the month. One other thing to keep in mind when reading a such profile is that the origin is placed in the upper left corner. This originates from image processing standards.

Alternatively, we can use the heatmap to present the activity of all appliances in a household over a period of time. Such example is Figure 1.5. In this case, we plot the consumption throughout the day, and it enables us to compare which appliances are being activated together throughout the day.

FIGURE 1.5: Consumption for each appliance in a day



1.2 LP use-cases



The load profiling method has a lot of different use cases across different fields. In our case, we will split use cases into three classes.

The first class is grid management. For example, it can be used to save energy by studying users' usage patterns and returning feedback, with suggestions on how to improve consumption. In cases where buildings have grid batteries and PV installed, the same feedback could be used to minimize the amount of energy being pulled from the grid. These are so-called zero-energy buildings (ZEB). Electrical energy providers could use demand response programs in combination with the LPs to optimize the management of the grid, with minimal impact on users' daily lives.

The second class is anomaly detection. The load profiles could be used to help the elderly in case of an accident or even help prevent one. They could be used to detect all kinds of early malfunctions in the operation of appliances, which would reduce service costs and save energy.

The last class is other, where occupancy detection, development feedback and energy stealing are all cases where LPs could be used.

A more detailed description of each use-case with publications will be addressed in the next chapter in section 2.2

1.3 Contributions

The main goal of the master's thesis is to propose suitable LPs for supporting residential building consumption optimization and elderly care management. To achieve this goal, we propose the following steps, where each step is a contribution to the scientific community.

1. Surveying the state-of-the-art LPs (2)
2. Development of multidimensional activation LPs (ALP's) (4)
3. Visual analysis of ALP's (5)
4. Propose a new anomaly detection method for elderly care (6)

The first contribution will be provided by taking a look at existing research and use-cases. Using the publications, a table of profiles will be constructed. The table will provide an overview of existing work, and reveal the gaps with LPs that were not yet utilized. Using the related work we will try to determine in what field each LP could be used. While we will fill some gaps, many will be left open for fellow researchers to pursue.

The second contribution will be provided in chapter 4. Here we will offer an in-depth look into the LPs, by presenting the profiles and showing how they present the consumption patterns. Each LP presents a different pattern and therefore has a different use case.

The third contribution will be provided in chapter 5. Here, we will use the t-SNE dimensionality reduction algorithm to show how samples are related. By doing that we will obtain an understanding of the content that the datasets hold.

This newly found knowledge should help us provide the last contribution. It will be provided in chapter 6. Here, we will design and construct elderly care assisted living system by utilizing one of the proposed profiles. The system will be able to detect anomalies in the daily routine of the elder. It should be simple, efficient and ready for real-world use. With that, we should be able to prove that the LP can be efficiently utilized, thus achieving the main goal of this thesis.

Chapter 2

Related work and table of profiles

In the first part of the chapter, we will review the existing work done and show possible use-cases for the load profiles. In the second part of the chapter most commonly used LP features will be presented. Using them, a table of profiles will be built. The table will be populated using the publications from the first part of the chapter. This will enable us an overview of existing work, and expose possible missing gaps in scientific research.

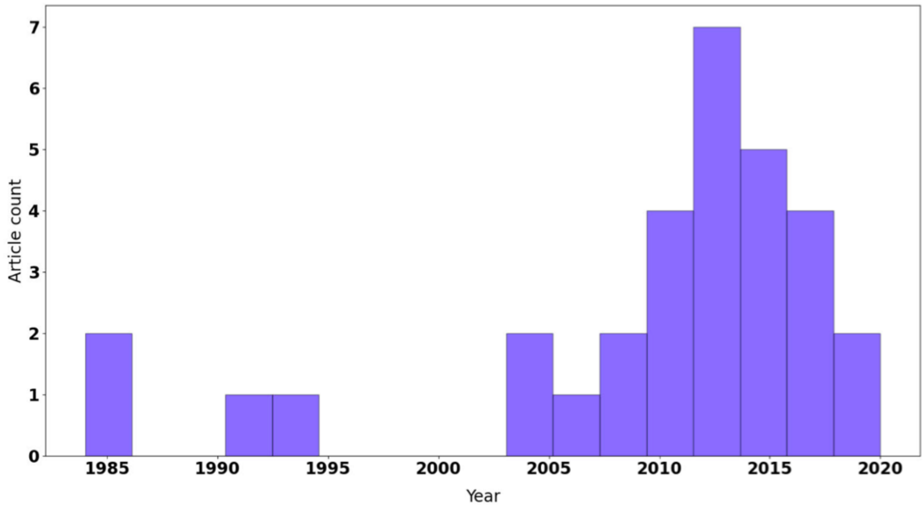
2.1 Related work

Work related to load profiling can be found in two research verticals. The first one is load profiling and load profile models, in most cases study the load profile curve of a building or appliance. The second vertical is anomaly detection in energy consumption data. There are quite a few connections between the two. For example, if one wants to do anomaly detection, one must first build some kind of "normal consumption profile", in other words, a load profile.

2.1.1 Load profiling

One of the first publications on load profiling was published by Train, Herriges, and Windle, 1985. They used a bottom-up approach using sub-meter data and other socioeconomic and demographic characteristics to create a load profile or statistically adjusted engineering (SAE) as they call it. They can adjust the curve based on weather, dwelling size, and income. In the same year, Walker and Pokoski, 1985 published a paper where they used a bottom-up approach with psychological factors to create probability models of when will an individual use an appliance. Since then there were two more in 1995. Research picked up the pace in 2005 with 7 publications in 2013 as Figure 2.1 shows.

FIGURE 2.1: Distribution of publications on load profiling from 1985 to 2020. The graph was published by Proedrou, 2021.



Load profiling can be performed in two ways: bottom-up and top-down. A bottom-up approach as Swan and Ugursal, 2009 states "calculates the individual dwelling energy or electricity consumption and extrapolates these results over a target area or region" Whereas with Top-down approach as Swan and Ugursal, 2009 states "uses the total energy or electricity consumption estimates to assign them to the characteristics of the building stock" In other more general words, Bottom-up uses sub-meter data, Top-down uses aggregated data. In our case, we take a deeper dive into the bottom-up approach.

Proedrou, 2021 did a comprehensive review on load profiling. The author defined various load-profile application subgroups such as demand-side management, planning and control design of energy systems, and residential load profiles. The author also grouped modeling techniques as probabilistic models, Markov chains, and Monte Carlo. The author first disclosed the current state of load profiling and issues with past work. They made a review of existing load profiling models and asses the-state-of-the art. Next, they pointed out future research directions and applications of load profiling models. Finally, the author exposes issues that researchers face and addresses possible solutions with conclusions.

Gerbec et al., 2005 tried to assign typical load profiles to a particular group of consumers based on their activity. To achieve that, they used probabilistic neural networks as a way of classification. Their methodology was tested in real-use scenarios.

Gao, Liu, and Zhu, 2018 makes use of the bottom-up method to build a forecasting framework for household load profiling, which takes into account the consumption patterns of residents. A model falls into the demand response use case. They have developed a "single-day extraction model", designed to select the same days by comparing environmental and household factors, which influence energy consumption. By using this approach, they have improved the accuracy of predicting the behavioral patterns of dwellers. Results show that their method successfully modeled daily usage.

Chuan, Rao, and Ukil, 2014 uses load profiling to optimize energy consumption distribution during the day. This reduces peak usage and alleviates load off the grid. The author used the bottom-up method, that is, using sub-meter data. Using this data, they made daily usage analyses on a one-hour basis. Using this information they optimized the daily activation of appliances so that peak usage was not as high. Results show that peak shedding was successful.

Csoknyai et al., 2019 analyzes energy consumption patterns and intervention strategies in residential buildings. Authors achieve this using a "serious game approach" with a combination of direct user feedback using smart meters. The application also provides advice, comparisons, savings, reduction goals, and monitoring. The approach takes into account almost all dimensions of residential energy usage. Their results show that their serious game was not able to induce energy-saving behavior.

Jeong et al., 2021 used extreme points in the appliance usage curve to cluster usage profiles. Usually, the first usage peak is in the morning, and the second one is in the evening. Additionally, they used demographic characteristics that are: region, area, age, salary, etc. to improve the results. Using collected data, they clustered profiles. They discovered 6 different usage profiles, where every cluster had a physical meaning such as energy-saving, morning heavy, evening heavy, etc.

Another clustering methodology was proposed by Park and Son, 2019, using load image profiles and image processing. They represented time series data as an image. The image is a grid of squares where the y-axis contains monthly data with a resolution of one day, x-axis contains daily data with a resolution of one hour. Grid is color filled with an algorithm that authors developed, where red means more activity and blue less. Using digital image filters they transformed the type-1 image to type-2 and from there used a threshold to obtain type-3. Using that information they clustered data based on images similarly. They used three different clustering methods: k-means, FCM, and EM algorithm. Using the Davies-Bouldin index, they were able to prove that image-based clustering performs better than non-image.

Abreu, Câmara Pereira, and Ferrão, 2012 clustered different load profiles using electricity consumption data and surveys using data from residential homes. They used PCA and k-means resulting in 5 clusters. Similar to other load profiling papers.

Whereas most of the above-mentioned papers focused on aggregated consumption of building to build a load profile, authors Issi and Kaplan, 2018 focused on appliance-level load profiling. Their main contribution was to create a realistic per-appliance load profile. They developed a wireless measurement system with smart plugs that enabled them to obtain power signatures for each appliance. They evaluated the data and based on observations they determined working cycles for each appliance. Furthermore, they concluded that 15 % of consumed power can be shifted, where they took tariffs into account.

2.1.2 Anomaly detection in building energy consumption data

A review on anomaly detection in building energy consumption data was written by Himeur et al., 2021. Here, the authors took a deep dive into detecting anomalies in energy consumption in buildings. The author first makes an overview of existing anomaly detection schemes and applications. Second, they perform a critical analysis and an in-depth discussion of the state-of-the-art. Next, they describe current trends such as NILM anomaly detection. Finally, they assemble a set of future research directions. Both reviews pointed out that NILM anomaly detection or NILM load profiling is a possible future research direction.

Rashid, Stankovic, et al., 2019 authors propose an algorithm that functions on top of existing state-of-the-art NILM algorithms Hidden Markov model, combinatorial optimization, Latent Bayesian Modeling, and Graph-based Signal Processing. They focus on three appliances, a fridge, freezer, and heater. Their metric was the number of operation cycles and energy used within those cycles. They implemented sigma variables to represent standard deviation and used rule-based anomaly detection. So if energy or counts are significantly larger than the mean then the day is considered anomalous. Their rule had only one manual setting and that was a number of standard deviations before the sample was considered anomalous. Their results show that sub-meter anomaly detection works decently whereas NILM-based anomaly does not work at all.

Rashid, Singh, et al., 2019 published another paper in the same year, where they took a similar approach, except that they used only compressor-based appliances such as fridges and air conditioners. They also added a rule to their existing rule-based anomaly detection algorithm, but the results still showed that NILM algorithms are not there yet.

Castangia et al., 2021 used disaggregated sub-meter data to detect anomalies in use consumption. They used a private dataset of 20 homes from northern Italy with no synthetic anomalies. Dataset included data from 2018 to 2020 meaning it included covid-induced anomalies. The authors first pre-processed the data by aggregating input load in hourly energy consumption, the second derived additional features, which are the time of use and duration of the activation. They use that data to detect single-point deviations for which they implemented the isolation Forest algorithm and anomalous trends for which to detect, they implemented Change Point Detection.

2.2 Use-cases

The general classification of use-cases was done in section 1.2. Here, we will focus on presenting these use-cases in great detail. This will be achieved by analyzing the use-case publications and in some cases providing additional solutions.

2.2.1 Grid management

Zero energy buildings and energy saving

As mentioned before many applications for load profiling could be used to reduce energy use and increase energy efficiency. With the emerging EV-market and ever-increasing installation of heat pumps, more and more energy is being used in form of electricity. This means, that most of the current power grids would have to be upgraded to keep up with demand.

On the other side, more and more photovoltaic systems are being installed, which is slowly shifting energy production towards end-users. Slowly energy grid is starting to shift towards so-called distributed energy resources or "DER" Moreno Jaramillo et al., 2021. DERs include all kinds of micro-energy sources such as PV, wind power, water power, and all kinds of energy accumulators that can store and release energy when needed such as heat pumps with hot water storage, home batteries, and EVs that can be used as a battery.

With smart management, these appliances could be used in a way that would reduce the net flow of energy and alleviate the load off the power grid. A way to achieve this is via load profiling and load modeling. To manage the appliances, a control system would have to be put in place Hledik and Lee, 2021. It would be enough to control a few appliances that consume most of the energy.

Since consumers take part in producing the energy, they are often called "prosumers" Parag and Sovacool, 2016. They will be an essential part of the European

Union's plan to reach zero-energy buildings and near-zero-energy buildings Parliament and Council of the European Union, 2021. The directive was accepted in 2010 and was recast in 2021. The plan is set to be realized in the next decade.

An actual use-case would be an EV owner with an installed PV system and heat pump, who works from home on occasion. In this case, two profiles would be developed. Normal workday and work-from-home day. Additional information would be obtained from the user's calendar. On a normal workday, the system would use PV energy to heat the water and store it, based on the user profile. On work-from-home days, the system would start charging the car with the morning sun, using only the PV energy. In the evening hours, when consumption rises and production falls, EVs could inject the power back into the house. Again using appliance load profiles to mitigate net energy flow as close to zero as possible (zero-energy building). With the ever-increasing power capacity and increasing range of EVs, more and more battery capacity could be used for mitigation. In the case of grid batteries, similar steps could be taken. This process is called vehicle-to-grid, and it is an important step towards zero-energy buildings Robledo et al., 2018 and Mehrjerdi and Hemmati, 2020.

One other way to use user load profiles is to optimally distribute the load by studying user's usage patterns as Chuan, Rao, and Ukil, 2014 and C. Li, Srinivasan, and Reindl, 2015 proposed in their papers. This could be further extended to neighborhoods connected into peer 2 peer energy distribution networks. As mentioned earlier, the way to save energy consumption is to distribute it as locally as possible. Knowing the usage patterns of all peers, the system could optimally distribute the energy using DERs across all homes without dwellers even noticing.

Another use-case could be using a heat pump and heat storage, where besides the user's usage patterns system would also obtain weather forecasts from the internet. Heat pumps that extract heat from the air are more efficient when temperature differences are smaller. The heat pump could store energy when warm and release the energy when cold. Based on the user usage profile, energy could be optimally distributed.

Many papers have been published, where authors explored ways to reduce the energy consumption of users by studying user consumption patterns, such as Spataru and Gauthier, 2014, Cellura et al., 2013, Verbong, Beemsterboer, and Sengers, 2013 and Spataru and Gauthier, 2014. Energy saving is done through instant feedback, reduction goals, rewards, and by comparing their user profile to the average user as the authors did in Csoknyai et al., 2019. Source Commission et al., 2006 states that as much as 20 % of energy could be saved by managing consumption.

Demand response

An increasing percentage of renewable resources is troubling energy distributors, due to the nature of renewable resources. In the prior chapter, it was mentioned how energy-saving measures would benefit users and their peers. One other use-case would be cooperation between end-user and energy distribution companies. Joint actions between them would benefit both as authors show in Albadi and El-Saadany, 2008 and Moslehi and Kumar, 2010

The electricity provider could control the main appliances so that load on the power grid is uniform, with as few peaks and valleys as possible. For this to function, users would have to allow the installation of energy meters and controllers on appliances that use the most electricity Shen, Jiang, and B. Li, 2015. One way to achieve this is to control the voltage of loads Zakariazadeh et al., 2014 the other way is to shift the loads in time C. Li, Srinivasan, and Reindl, 2015. This process is called direct load control Hledik and Lee, 2021, and it is part of demand response program Chen, 2018.

"DR program is a voluntary PJM program that compensates end-use (retail) customers for reducing their electricity use (load) when requested by PJM during periods of high power prices, or when the reliability of the grid is threatened." Chen, 2018

The benefit to the user would be the lower cost of charging EVs and heating the building. This is already done through so-called small and high tariffs. More detailed user load profiles would enable the electricity provider to introduce real-time tariffs.

The user would have three options. The first one would be that users can use the appliances as freely as they desire, this would result in a normal tariff. The second option would be to use the appliances as regularly as possible, this would lead to lower tariffs. The third option would be to leave the management of main appliances to the electricity provider via direct load control. The provider would combine the user appliance load profile and the real-time market price of energy to optimize the cost Graditi et al., 2015. This would lead to free or even negative prices of electricity since distribution companies have to keep the frequency of the grid as stable as possible.

For them to stabilize the frequency, they sometimes have to resort to load shedding. Load shedding is a process where a load is disconnected from the grid to keep the grid in sync Lopes, Moreira, and Madureira, 2006. Commonly whole neighborhoods are being disconnected, affecting their daily lives. Using user load profiles, distribution companies could disconnect the load in a way that would minimally affect the end user. When they would need to load the grid due to low demand, they could charge EVs free of charge or even pay to do so. This benefits the company as well since they do not need to lower energy production, which can be expensive.

2.2.2 Anomaly detection

One use-case of anomaly detection was already mentioned in the Elderly care chapter. One more thing that could be detected, using load profiling, would be the altered operation of appliances. In the case of a fridge, the system would detect that duty cycles are too long. The increased duty cycle can be caused by cooling liquid leakage, the fridge being open or compressor motor malfunction. Heat pumps work on the same basis as fridges, meaning the same anomalies could be detected. The malfunction could also be detected in heating element appliances such as toasters or boilers. Since mentioned appliances are one of the largest consumers in a household, early enough detection could lead to large energy-saving benefits Rashid, Stankovic, et al., 2019.

Elderly care

The aging population is an increasing socioeconomic issue. The elderly are facing many issues when staying at home alone for extended periods. Accidents such as falls or the inability to do choirs due to health-related issues or even dementia-induced issues such as leaving appliances on for long periods could all be detected, using sub-meter data such as authors Visconti et al., 2019 and Patrono, Rametta, and Meis, 2018 explore in their papers.

To detect falls or other issues a normal daily appliance use profile would be developed. It would involve routine behavior of users such as turning on the coffee machine in the morning, the stove and oven at the noon or using the toaster in the evening. All these routines could be measured and tracked. Using this data, a profile would be developed. The probability of an anomaly and a threshold would enable the system to detect an issue.

An example would be: the coffee machine not turning on in the morning or the stove and kitchen vent not being used at the noon. Another issue could be detected if the appliance would be used more frequently or for extended periods of time. This could indicate that the user forgot to turn off the stove, oven, or even a light. The same system could detect that a fridge or a freezer was left open since the duty cycles would be longer and more frequent. As soon as the issue would be detected it would notify the caregiver to check on the patient.

2.2.3 Other

Load profiling could also be used as feedback for the engineers and designers, of how a device is being used and if it is being used as designed. This would enable the manufacturers to improve their products according to user's needs, without unnecessary features.

Yip et al., 2018 uses anomaly detection algorithms and load profiling to detect energy lost due to non-technical losses. This occurs after the smart meter is exposed to cyber or mechanical attacks and its measurements are off.

One other use-case could be occupancy detection of buildings such as the authors explore in Kleiminger et al., 2013. Information about occupancy could be used as part of elderly care monitoring or in the case of building automation, to run certain tasks when a user enters or leaves the room or a building.

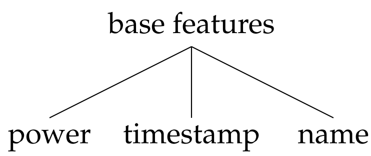
2.3 Table of profiles

While in related work we examined load profiling in general, this section focuses on how data is presented with load profiles. It can be portrayed in various shapes and forms, using all kinds of attributes and features to do so.

First, main load profiling features will be defined. Second, using these features a general load profile table will be constructed. Third, references from related work and use cases will be mapped to the table, from which main features will be selected. Fourth, using a reduced feature set a more detailed table will be formed. Again, the table will be populated using the same references as before. Finally, using this information a research direction will be formed.

2.3.1 Feature set

If we want to find the base features, we have to look at how consumption measurements are done in most buildings. The following three features enable us to know the amount of energy being consumed by the user.



If we translate these features to the time domain and observe them over a specified amount of time, new features emerge. The most notable example is the observation of electrical power over one hour. The result is energy E and it is one of the most common ways used to bill a customer for his power consumption.

We can also extract features such as the number of activations or time of operation for each activation. This can be done using sensors to detect activity or even deduct this from power consumption data. In cases where we are observing individual appliances, this can be done using simple signal processing techniques. In cases where we are observing buildings, this could be achieved using more complex disaggregation algorithms also known as NILM algorithms.

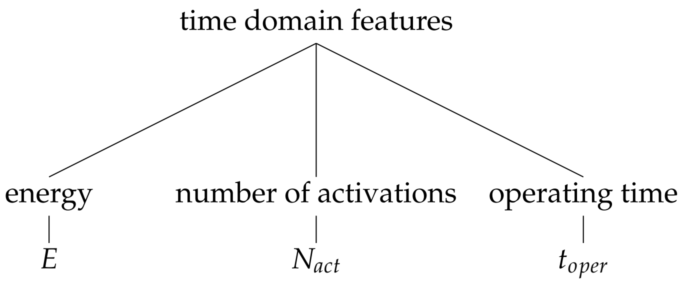
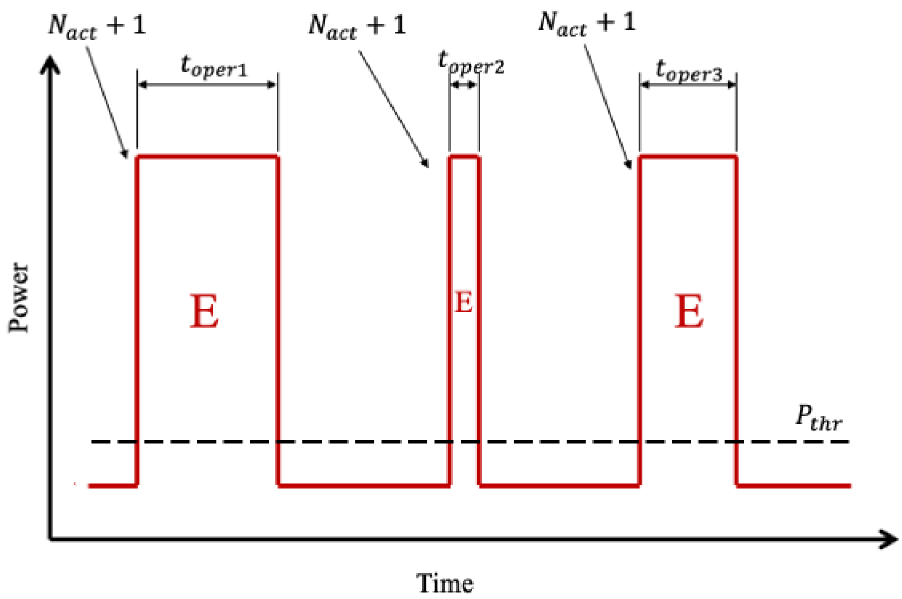


FIGURE 2.2: Simple signal processing of power consumption for a single appliance

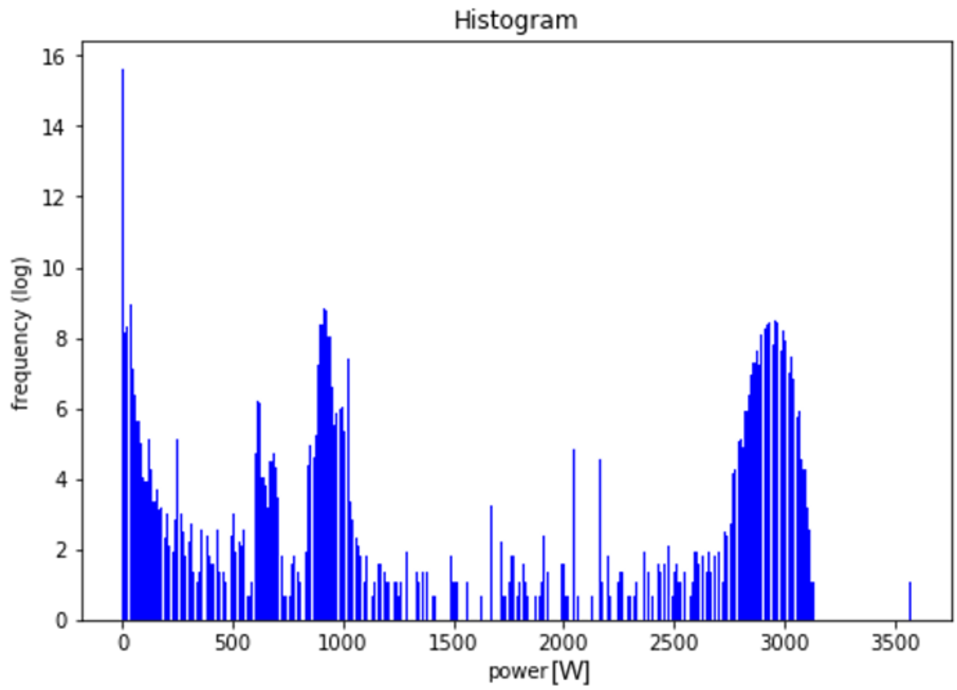


As we can see in Figure 2.2 all three-time domain features can be extracted from the graph. Energy E is equal to the area under the graph or in other words integral of power over time. N_{act} can be measured based on the number of times the power value exceeded some pre-defined threshold P_{thr}. The t_{oper} is the time between on and off events, where we use the same threshold as with N_{act}. While there are other features, such as time between activations, or total operational time that could be extracted, these were not commonly used in related work.

frequency domain features
|
Operation Modes

The same as we can present power in the time domain, the same can be done in the frequency domain. One actual example can be seen in Figure 2.3. Here, it is hard to extract more features, but one possibility could be detecting the number of operation modes based on the number of peaks, using signal processing algorithms.

FIGURE 2.3: Frequency of power values for the toaster. Actual data from the REFIT dataset.



In the case of Figure 2.3 we are observing a toaster over one year. Toasters are usually simple appliances using a heating element and a thermostat, meaning that the power consumption should be constant and set by the resistance of the heating element. The Figure shows 2.3 a nice normal distribution of power values around 3 kW, which we can assume is the heating element. We can notice two other peaks one at roughly 1 kW and the other at 0.7 kW. Since toasters usually do not have operation modes, we could assume that there are other appliances plugged into the metering device, meaning this could be a use-case for this kind of load profile.

2.3.2 General table

Using these features we can form a Table with all possible combinations. Table 2.1 is then populated with references from previous chapters. To understand the table more clearly, let's imagine that each feature is used as an axis label when plotting.

TABLE 2.1: General Table of load profiles

	power	number of activations
time	[18] [21] [10] [36] [70] [26] [25] [34] [1] [38] [57] [56] [33] [4] [15] [39] [18] [13] [50] [37] [25]	[14] [37]
operation time	[4]	[57] [56] [4]

Table 2.1 shows a combination of base features of power and time with 21 publications. One example of such profile can be seen on Figure 2.2 or 1.1 and is also known as standard load profile (SLP).

As we have seen in the previous section, the two other features, operation time and the number of activations are a derivation of the base features. A combination of the two has been used in three other papers. It shows how many times the appliance was activated for a certain amount of time. This load profile is commonly used for anomaly detection.

Derived features can be used in a combination with the base features. The combination between power and operation time LP shows us how long did an appliance operate for a certain amount of time. Only one publication used this set of features. Combining the time and number of activations LP could for example present at what time of the day appliance is being used the most. We have sourced only two publications that used this set of features.

Based on Table 2.1 it is possible to see that the most commonly published feature combination is time and power. This combination will be used as a baseline when making a more detailed table. Although the operating time feature was explored in a few publications, we are focusing on activation-based histogram representation. Based on Table 2.1 it is possible to see that not much attention was given to it.

There are many more ways to present the data. An extended Table can be found in Appendix B.

2.3.3 Detailed table

This section will focus on exploring possible activation-based load profiles, while using the power load profiles as a baseline. Features from 2.1 will be explored in higher detail. They will be split and arranged in a way that all 21 publications using power-based presentations will be divided into as many groups as possible. This should expose possible activation-based profiles as well as unpublished power-based profiles.

Sub-features

General features were already described in section 2.3.1. It is possible to further divide them into smaller so-called sub-features. These are reshaped and grouped as follows:

- Way of presenting a profile
  - Per-building
  - Per-appliance
  - Per-building and per appliance
- By time range of profile
  - Daily
  - Weekly
  - Monthly
  - Yearly
- Way of measuring usage
  - Average power use
  - Number of activations

2.3.4 Table of combinations or detailed table

The above-shown profiles can be combined, yielding a new way of displaying the data. Bellow, a map with combinations of the above-mentioned profiles is presented. The purpose of Table 2.4 to generate and show possible combinations. Some combinations that had similar output were grouped, and some that could not be sketched were discarded.

The profiles and figure graphics used in the Table 2.4 were sourced from section 1.1.

FIGURE 2.4: Table of combinations

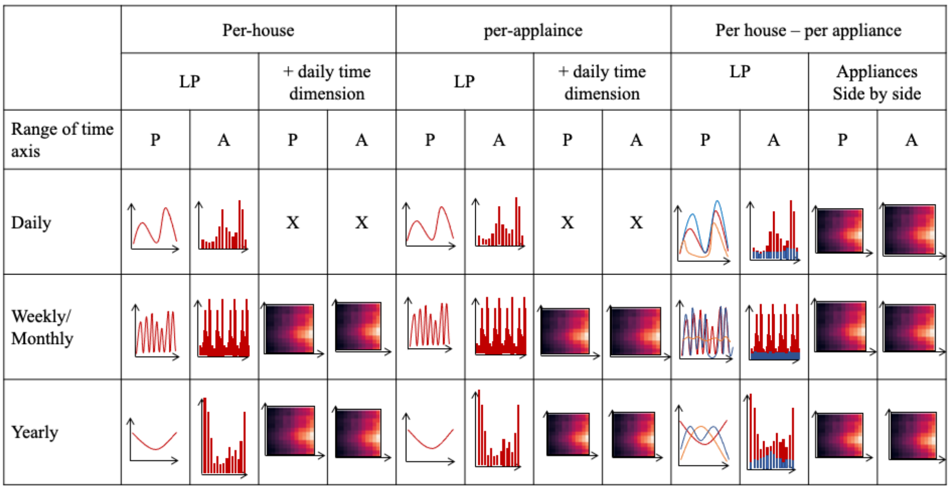


Table 2.4, uses features from previous subsection 2.3.3. In general, the Table 2.4 is formatted in a way that features from columns (time range) are used in the x-axis of a plot, and rows (consumption data) are used in the y or z-axis of a plot.

The column of the Table 2.4 presents the time domain. "Daily" means that the load profile presents average usage for one day and "Weekly" means it presents usage for a week. To be clear, for one to construct a decent daily profile, one needs a few weeks of data. The same goes for yearly profiles, in that case, one needs many years' worth of data.

The top row of the table is composed of 3 main groups. The first group focuses on per-building energy consumption. The second group examines the energy consumption of each appliance in a house separately. Third group analyses all appliances in a building.

The next row of the Table is further divided into two groups. First is the LP group which presents the given usage unit on the y-axis and time on the x-axis. Next is an LP with an additional time axis. In this case, we present the given usage unit on the z-axis and then time on the x and y-axis. Here, the second-time dimension can be anything from a week to a year. In the case of the per-building, the subgroup includes appliances instead of time. Example for this is Figure 1.5.

The last row presents the usage unit, that is power (P) or the number of activations (A).

In cases where the feature combination does not make sense, it is marked with an X.

2.3.5 Mapping references to the table of profiles

To find useful load profiles, references from the related work section 2.1 must be mapped.

TABLE 2.2: Table presents previously mentioned load profiles

Description	Per-building				Per-appliance				Per-building per-appliance			
	LP		+ daily time dimension		LP		+ daily time dimension		LP		Appliances side by side	
Range of time axis	P	A	P	A	P	A	P	A	P	A	P	A
Daily	[37] [18] [21] [10] [36] [13] [70] [26] [25] [34] [1] [38]		X	X	[57] [56] [33] [4] [15] [39]	[37]	X	X	[18] [13] [25]	[37]		
Weekly/ Monthly	[21] [10] [36]		[61] [50] [38]			[14]			[12]	[14]		
Yearly	[21] [10] [36]											

As can be seen from Table 2.2, most of the work (14 publications) has been done with standard daily load profiles with per-building power usage such as Figure 1.1. Quite a lot of work (6 publications), has been done with per-appliance daily power profiles. A few publications were based on weekly and yearly load profiles and a few used two-dimensional time and power presentations. Only one publication found used activation and time-based histogram such as shown in Figure 1.2. During the research we focused on publications from minority classes, meaning not all existing publications for standard load profiles are included. The purpose of Table 2.2 is to present missing scientific contributions and patterns of publications.

2.3.6 Mapping use-cases to the table of profiles

Table 2.3 includes arranged publications from the use-cases section 2.2. Similar pattern emerged as in Table 2.2.

TABLE 2.3: Table presents references mentioned in use-cases chapter

Description	Per-building				Per-appliance				Per-building per-appliance			
	LP		+ daily time dimension		LP		+ daily time dimension		LP		Appliances side by side	
Range of time axis	P	A	P	A	P	A	P	A	P	A	P	A
Daily	[16] [62] [46] [72] [42] [27] [73]		X	X	[46] [69] [52] [39]		X	X	[18]			
Weekly/ Monthly	[62] [36]											
Yearly	[62]											

2.3.7 Table of use-case groups

The Table 2.4 presents same publications as Table 2.3, but only group names are shown. The Table 2.4 indicates how groups are arranged. Where anomaly detection and elderly care are dominating in the per-appliance part of the table, zero energy buildings and demand response are dominating in a per-building part of the table.

TABLE 2.4: Table presents references mentioned in use-cases chapter

ZEB - zero energy buldings DR - demand response AD - anomaly detection EC - elderly care X - unfeasible	Per-building				Per-appliance				Per-building per-appliance			
	LP		+ daily time dimension		LP		+ daily time dimension		LP		Appliances side by side	
Range of time axis	P	A	P	A	P	A	P	A	P	A	P	A
Daily	ZEB, DR		X	X	AD, EC, ZEB		X	X	DR			
Weekly/ Monthly	ZEB											
Yearly	ZEB											

The figures listed above clearly depict the void not filled by publications. Although they may not be published, they still have a possible use case. In Table 2.5 empty spaces are filled with possible use-cases for given load profiles.

TABLE 2.5: Proposed use-cases for profiles

ZEB - zero energy buildings DR - demand response AD - anomaly detection EC - elderly care X - unfeasible	Per-building				Per-appliance				Per-building per-appliance			
	LP		+ daily time dimension		LP		+ daily time dimension		LP		Appliances side by side	
Range of time axis	P	A	P	A	P	A	P	A	P	A	P	A
Daily	AD, ZEB, DR,	AD, ZEB, DR,	X	X	AD, EC, ZEB, DR	AD, EC, ZEB, DR	X	X	AD, EC, ZEB, DR	AD, EC, ZEB, DR	AD, EC, ZEB, DR	AD, EC, ZEB, DR
Weekly/ Monthly	AD, ZEB, DR	AD, ZEB, DR,	ZEB, DR	ZEB, DR	AD, ZEB, DR	AD, ZEB, DR	AD, ZEB, DR	AD, ZEB, DR	AD, ZEB, DR	AD, ZEB, DR	AD, ZEB, DR	AD, ZEB, DR
Yearly	ZEB, DR	ZEB, DR	ZEB, DR	ZEB, DR	AD, ZEB, DR	AD, ZEB, DR	ZEB, DR	ZEB, DR	ZEB,DR	AD, ZEB, DR	AD, ZEB, DR	AD, ZEB, DR

2.3.8 Table of load profile potentials

Some combinations are indeed illogical and again others are less useful in a practical sense. The next Table 2.6 will try to rate the utilization potential of the profiles based on two characteristics. First is how well data is presented to the user, meaning that the load profile is clear about what it is presenting. The second is the effectiveness when being used in an algorithm, or in other words, how well data is presented to a machine.

These characteristics can not be easily measured, but it is possible to extract them based on the pattern of publications. To do that, we have to make two assumptions. The first one would be, that the larger the number of publications, the larger the effect of presenting the data to a human. The second would be, that the larger the number of use cases, the better the effectiveness of presenting the data to a machine. Using these two assumptions, we propose the following table. The Table has four possible classes.

- 1 - The load profile satisfies both assumptions and has a high utility rate and was already researched (very useful, but with low research potential).
- 2 - The load profile satisfies only one of the above-mentioned assumptions (has mid-research potential).
- 3 - The load profile does not suffice any of the above-mentioned assumptions and was not yet researched or practically used (high research potential, could be hard to utilize).
- X - The load profile is inexplicable (does not make any sense).

TABLE 2.6: Proposed classification of profiles

	Per-building				Per-appliance				Per-building per-appliance			
	LP		+ daily time dimension		LP		+ daily time dimension		LP		Appliances side by side	
Range of time axis	P	A	P	A	P	A	P	A	P	A	P	A
Daily	1	3	X	X	1	2	X	X	1	2	3	3
Weekly/ Monthly	1	3	2	3	3	2	3	3	2	2	3	3
Yearly	1	3	3	3	3	3	3	3	3	3	3	3

2.3.9 Table of possible future research directions

To find future research directions we must look into profiles that were least researched, such profiles are marked with the number 3 on Table 2.6. Some profiles were not researched because they may not present data as well and some were simply overlooked. This is why we have built the following Table 2.7. The Table was populated as follows:

- (1) - The load profile has high potential.
- (2) - The load profile has mid-potential.
- Empty - The load profile has low potential or was already researched.
- X - load profile is inexplicable

The process of evaluation was a bit complicated, but it can be summed down to the following rules.

If the load profile was used as a power profile, can it be used as an activation profile? Here, we must use common sense. For example. If we follow this rule for per-building power load profiles, it turns out that activation load profiles are not as useful since they are based on per-appliance load profiles. In other words, to build per-building activation load profiles we need per-appliance (sub-meter) data anyway. That is why we have assigned them to the second class.

The second rule was applied to 3D profiles. In the case where one dimension was commonly used, it is probably worth investigating it with a combination of additional dimensions.

Following these rules, the Table 2.7 was constructed.

TABLE 2.7: Possible future research contributions

	Per-building				Per-appliance				Per-building per-appliance			
	LP		+ daily time dimension		LP		+ daily time dimension		LP		Appliances side by side	
Range of time axis	P	A	P	A	P	A	P	A	P	A	P	A
Daily		(2)	X	X			X	X			(1)	(1)
Weekly/ Monthly		(2)		(1)	(1)		(1)	(1)			(2)	(2)
Yearly					(2)	(2)					(2)	(2)

Table 2.7 presents the possible future research directions. While some LPs have mid-research potential according to our rules, they are still worth investigating. In science, it often happens that use-cases change over time and research that seemed inapplicable suddenly finds its place.

We will focus on profiles with high research potential and use the number of activations as a unit of measure. When the aforementioned parameters are applied, the result is Table 2.8.

TABLE 2.8: Load profiles to be pursued

	Per-building				Per-appliance				Per-building per-appliance			
	LP		+ daily time dimension		LP		+ daily time dimension		LP		Appliances side by side	
Range of time axis	P	A	P	A	P	A	P	A	P	A	P	A
Daily			X	X			X	X				(1)
Weekly/ Monthly				(1)				(1)				
Yearly												

The profiles shown in Table 2.8 are our direction of research. In the next part of the thesis, we will try to utilize and present these load profiles. This will be done as follows. In chapter 5 we will use

- Per-building daily-weekly load profile
- Per-appliance daily-weekly load profile

with a t-SNE neighboring algorithm to find how they are related in high dimensional space. In chapter 6 we will use

- Per-building Per-appliance daily load profiles with appliances side by side

To build assisted living system for the elderly.

Chapter 3

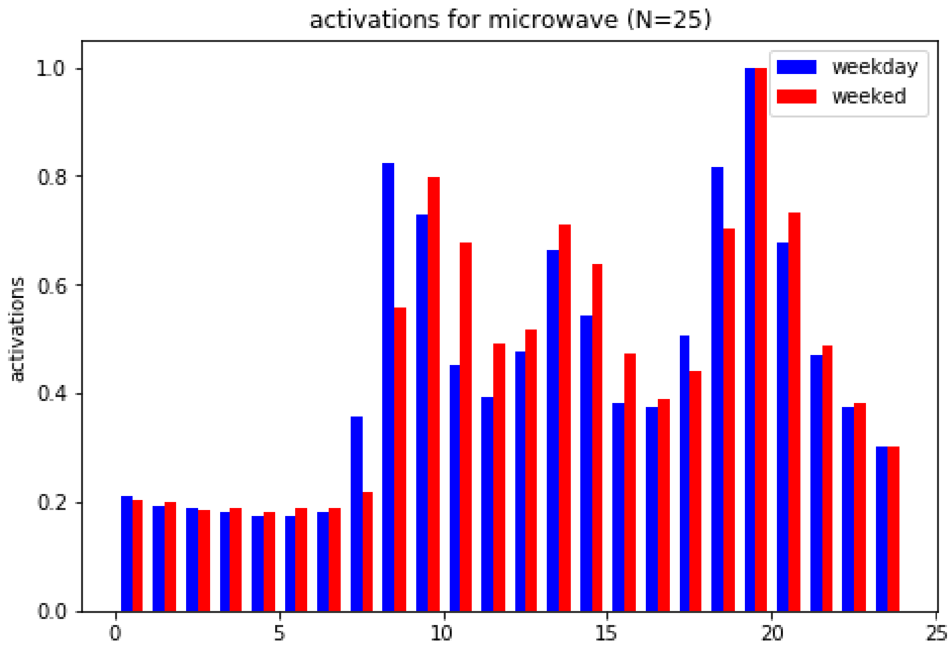
Methodology

The following chapter includes methodological procedures that are common for all chapters. More detailed methodological procedures will be described in each chapter separately.

3.1 Data

We used UK-DALE Kelly and Knottenbelt, 2015, REFIT Rashid, 2019, ECO Beckel et al., 2014, REDD Kolter and Johnson, 2011, and iAWE Batra et al., 2013. All datasets measured electrical energy consumption for residential buildings. They include main smart meter data, as well as sub-meter data for each appliance in a dwelling. For easier handling datasets will be sliced into 1-hour intervals.

FIGURE 3.1: Universal normalized daily usage profile for weekend and weekday for a microwave. Superposition of data from 25 homes.



Data will be then used to generate various load profiles. One such example can be seen in Figure 3.1. The histogram shows normalized daily activation for microwaves. It consists of data from 25 homes from 4 different datasets.

3.2 Tools used

To process the data and to obtain the results the environment and virtual machines from Google Colab (Bisong, 2019) were used. They offer access to Google GPU-accelerated compute machines with 12 GB of RAM. Colab also offers access to Drive cloud storage, where the dataset and results were stored. While running the experiments, we made use of Drives 100 TB pooled cloud storage, which is available to students of the University of Ljubljana. For development and version control, GitHub was used.

Within the Colab which uses a Jupyter (Kluyver et al., 2016) environment at its core, various python libraries were used. To store and read the datasets in hdf5 format we used h5py (Collette, 2013) and Pickle (Van Rossum, 2020). To load datasets into RAM and then handle them, the pandas (McKinney et al., 2010) library was used. For handling the large matrices and calculating we used NumPy (Harris et al., 2020). To present the data with graphs we have used Matplotlib (Hunter, 2007) and to present data with heatmap Seaborn (Waskom, 2021). For easier implementation, such as of the t-SNE, a Scikit (Pedregosa et al., 2011) and SciPy (Virtanen et al., 2020) libraries were used.

Chapter 4

Presenting Proposed LPs

Previously defined profiles will be presented in-depth. In general, each profile has its use-case already assigned in Table 2.3. Here, we will focus on exposing the main features, issues, and use cases.

Data for profiles in this chapter was used from building 2 from the REFIT dataset. Data was collected from 2013-09-18 to 2015-05-28.

4.1 Time ranges

One important thing to mention is the use cases for different time ranges of load profiles. Based on Table 2.2 it is possible to see that most publications and 2.3 use daily time range.

Generally, daily profiles are easier to build since they do not need as much data as others do. To build a decent profile one needs enough data. A sufficient amount of data is the amount that covers major events. For a daily profile, a few weeks of data is enough, weekly load profiles need a few months of data, monthly few months, and yearly few years. And this is the main issue, there is rarely enough data to build such profiles. Even then, usage patterns could change over a long period such as a decade. Combining that with a smaller number of use cases for such profiles, reveals why such profiles were not looked into as much in Table 2.2.

One more thing about time ranges that need to be mentioned are patterns that they present. Daily profiles present daily usage and enable us to extract contextual events such as waking up, cooking, leisure time, etc. The weekly pattern is also repetitive, and it enables us to see how appliance usage changes over the weekdays and weekends. The monthly profile has none of the above. It is not repetitive since each day of the month can be a different day of the week, and the period is too short to capture seasonal patterns. Alternatively, it could be presented as a week in a month, but there is no significant usage pattern to be revealed. The yearly profile on the other hand presents the seasonal effects on usage such as increased daylight and temperature.

4.2 Per-building

The section will be focused on per-building profiles, meaning whole building usage is presented as a single load profile. This kind of presentation is useful for observing general activation trends in a building. Possible use cases for per-building load profiles are grid management and energy saving.

When it comes to activation load profiles there is one issue compared to power load profiles. To build per-building power load profiles it is possible to use the main power meter, whereas, at activation load profiles, sub-meter data is needed. This can be solved using NILM algorithms, but they are not in a state of practical use yet.

The daily per-building load profile is also known as the standard load profile. According to Table 2.2 this is the most commonly used power profile. Figures 4.1a and 4.1b present usage patterns on different time ranges. The two profiles, therefore present different contextual cases.

(A) Daily (B) Yearly

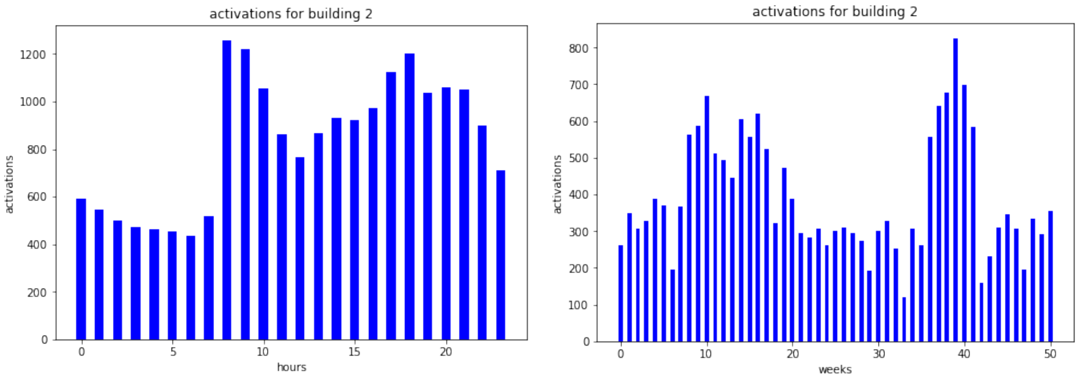


FIGURE 4.1: Per-building load profiles

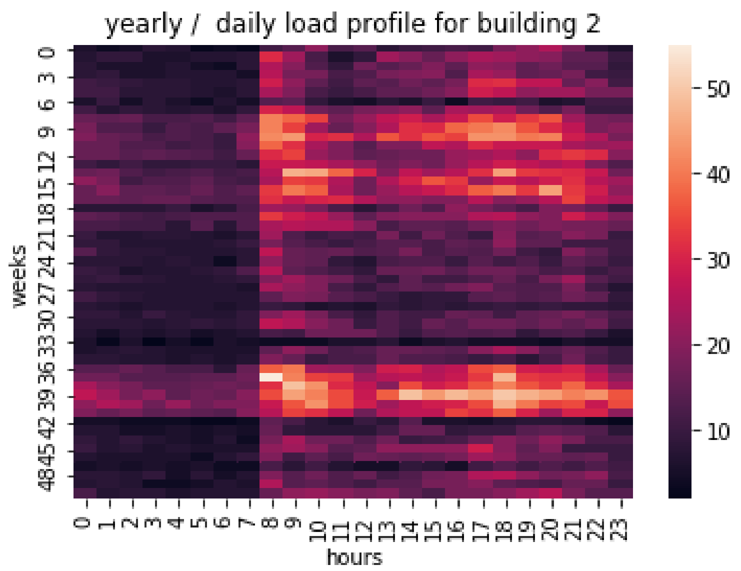
The first Figure 4.1a shows how activations change throughout the day. It is possible to see that there is some activity even throughout the night and early morning. These can mostly be related to fridges or other appliances that are not directly activated by users. At around 8 in the morning, it is possible to detect the first peak. These can be related to morning choirs. Then, at around noon, a dip occurs. The reason behind it is probably, that the dwellers are not home. In the afternoon, the rate of activations slowly increases until it peaks at around 19 o'clock. This slow rise could be a contribution of each dweller arriving home at different parts of the day.

The second Figure 4.1b shows how activations change over the year. Again, it is possible to observe two peaks. One in the spring and the other one in autumn. It is hard to correlate the activity with the seasonal effect since it seems like the activity is about the same in mid-winter as in mid-summer. The exact reason behind this pattern is unknown.

4.2.1 Per-building two-dimensional time

Alternatively, it is possible to combine Figures 4.1a and 4.1b and present activations as a heat map. The result is a Figure 4.2 showing more complex activation patterns.

FIGURE 4.2: Two-time-dimensional per-building load profile



By combining the Figures and presenting them with a heat map, additional features are revealed. For example, the black lines are the ones that probably present the vacation or other events where dwellers are away from home.

When analyzing Figure 4.1b it seems like dips in activity are for a similar reason, but Figure 4.2 shows these two dips from a different perspective. The peak activity in Figure 4.2 shows a routine or a pattern similar to what was seen in 4.1a, one peak in the morning and one in the evening. The same pattern can be observed in winter dip, even though the pattern is less clear it is present. The same cannot be said for the summer dip in the middle of the plot. Here, while the morning peak is visible, the evening one can barely be detected.

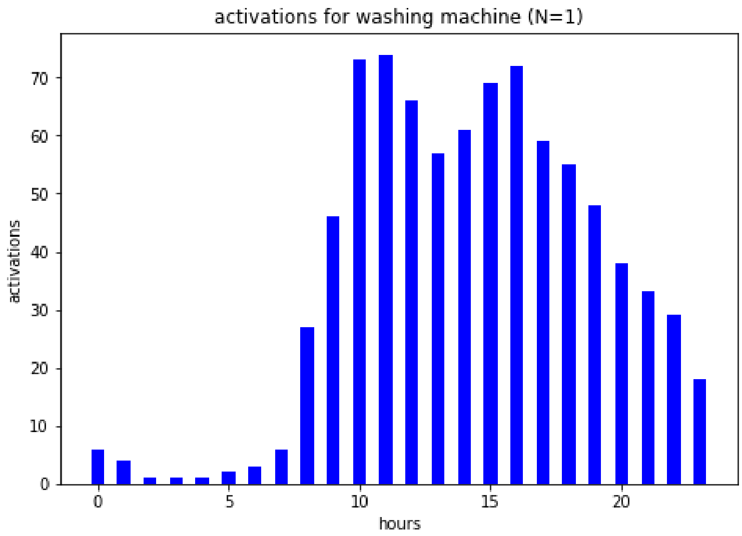
One more thing to mention is that the increased activity at the start of the fall increased activity throughout the night and day. This could point to that some new appliance was installed, which increased the number of activations.

Previously it was mentioned that these kinds of profiles are the most applicable in grid management. One such example could be load shedding. Using the Load profile above, electrical energy providers could find buildings with the least activity at that time of day. Combining that with power data, it could disconnect the buildings with the least activity and most power consumption.

4.3 Per-appliance

Per appliance load profiles offer a look into the consumption of each appliance. In the case of activation load profiles, this is an elemental load profile, since all other activation profiles are built on top of it. This also means that it is one of the most universal profiles since it can be used in all previously defined use cases.

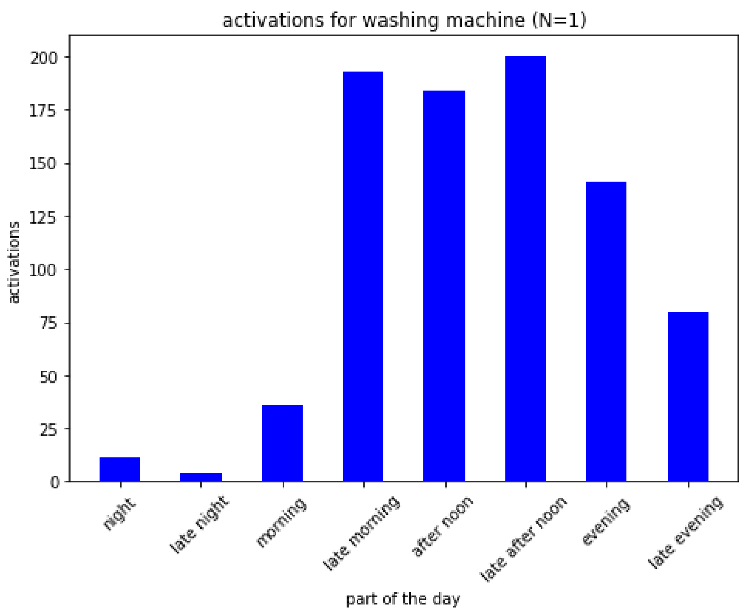
FIGURE 4.3: Daily per-appliance load profile



Looking at Figure 4.3, we can detect a similar pattern as in per-building Figure 4.1a. While the peaks are closer together, the pattern remains. One thing to notice here is, that the washing machine is used only throughout the day. This means that this household does not use the cheaper nighttime tariffs.

Another parameter that was not explicitly mentioned before, is the resolution of load profiles. Histograms can be presented using various resolutions or numbers of buckets. An optimal number of buckets is a number that clearly presents the usage pattern. 3-hour bucket size in Figure 4.4 does a good job at presenting the appliance usage at the main parts of the day. This offers a better contextual presentation that is easier to process using algorithms. As we can see in Figure 4.4, by increasing the extent of buckets, the two peaks join together into one larger peak. This coincides with the point of the presentation, where we want to present a more general pattern in key parts of the day.

FIGURE 4.4: Daily per-appliance load profile with larger buckets sizes

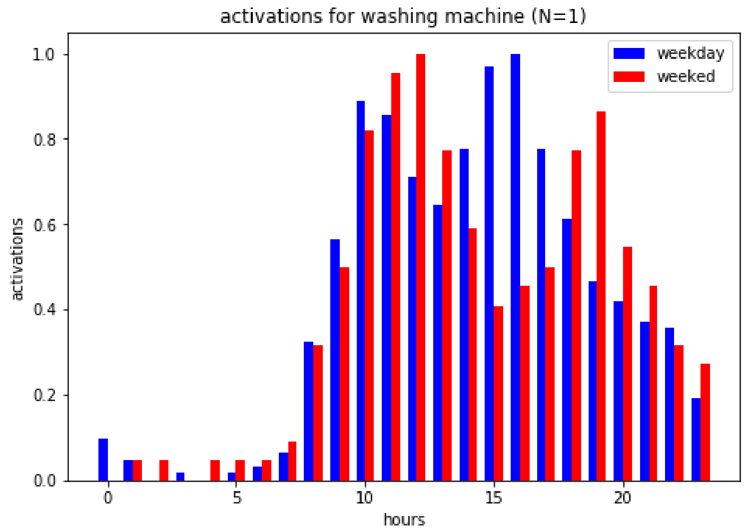


While the low resolution is useful for contextual presentation, high resolution is needed for time-sensitive applications such as elderly care, where we have to detect an accident as soon as possible. The hourly resolution would mean that in case of an accident system would need at least an hour to detect it. While this is sufficient for demonstrating the capabilities, real implementation would need to use lower-resolution data.

In the case where dwellers have different usage patterns during the weekends, two profiles would have to be developed. It is possible to present them both at once such as in Figure 4.5. This is essentially a variation of the weekly load profile that maintains high resolution. Since there are more weekdays than weekend days, activations had to be normalized accordingly.

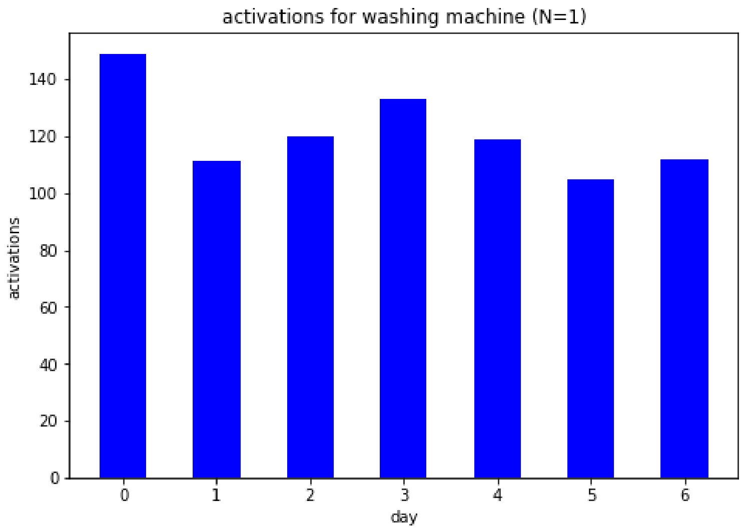
The Figure 4.5 again shows the same pattern as on Figure 4.3. What can we observe here is how these two patterns are the same but are shifted in time. On weekdays the first peak occurs at around 10 AM and the second at around 3 PM. On weekends the first peak does not occur until 12 AM and the second at around 6 PM. This shift in the pattern shows that while there is a change in behavior between weekends and weekdays it is not a drastic one, at least in this case.

FIGURE 4.5: Normalized daily per-appliance with weekday and weekend load profiles.



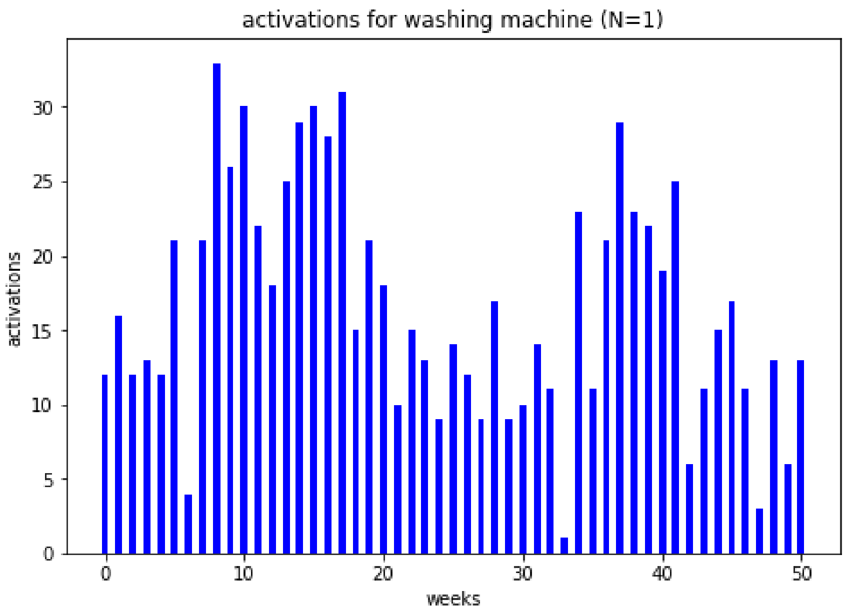
Another way to present weekly data is shown in Figure 4.6. In this case, weekdays are numbered, where 0 stands for Monday and 6 for Sunday. This resolution offers a look into how consumption pattern changes over the week. This is useful for applications such as grid management or energy saving. In this particular case, it is possible to see that the user most commonly uses the washing machine on Mondays and Wednesdays. Using a weekly weather report that would indicate high energy production on Wednesday, the electricity provider could offer a low cost for energy for that day. This kind of presentation could also be used to detect daily anomalies.

FIGURE 4.6: Weekly per-appliance load profile



In Section 4.1 we mentioned that the monthly presentation does not show any significant usage pattern, so it was not shown here. The yearly presentation again shows the more broad usage pattern, which can be seen in Figure 4.7. This is useful for grid management and energy saving, where one could detect seasonal changes in the usage of an appliance.

FIGURE 4.7: Yearly per-appliance load profile

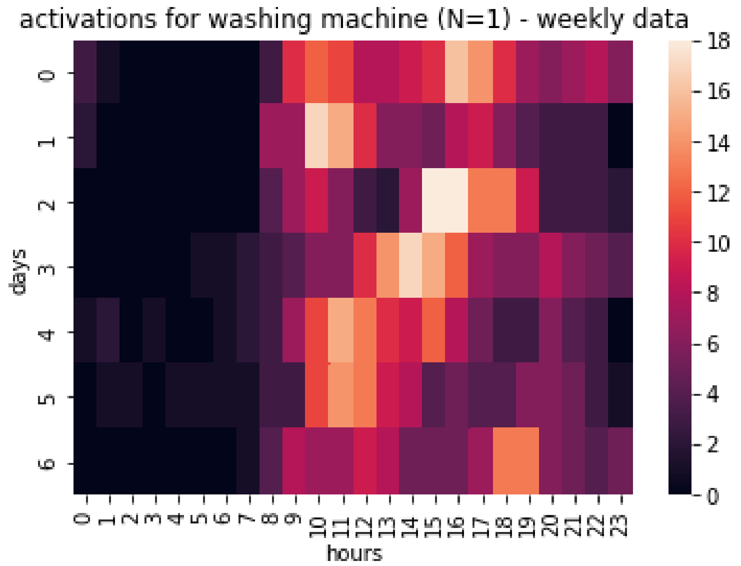


When comparing the pattern from Figure 4.7 to pattern from Figure 4.1b it is possible to see the very same pattern. When making a quick comparison, they seem like the same image, only when taking a closer look it is possible to see that differences do exists. We can make a similar conclusion here, as we did for Figure 4.1b. It is hard to do any deeper analysis without the metadata.

4.3.1 Two-dimensional time per-appliance load profiles

Using a combination of Figures 4.3 and 4.6, it is possible to generate Figure 4.8.

FIGURE 4.8: Two-dimensional time per-appliance load profile



In this case, a similar use-case could be fitted as in the first example in subsection 4.2.1. The first example used load shedding when the demand is too high. On the contrary, it can also occur if the grid demand is too low. There are two solutions to this issue. The first one is to decrease production, which can be slow and expensive. The second option is to load the grid, which can be done in many ways. One of the ways is to turn on appliances using a direct load control system or notify users to turn on appliances that they have commonly used at that time in the past. Due to the increasing percentage of renewable energy sources, more and more energy peaks will be weather dependent. By combining weekly wind forecasts, weekly cloud coverage, and user consumption profiles energy providers could notify users to turn on their appliances at peak usage times.

By analyzing Figure 4.8 it is possible to see that the user uses a washing machine, on Wednesdays from 15 to 16 o'clock quite commonly. Should weather reports indicate high production peaks, the electrical provider could offer low-cost energy for that time of day for all users with similar usage patterns. This could all be automated for appliances such as home grid batteries, water heaters, EVs, or even fridges with a control system. This would mean that the grid operator could regulate the demand instantly. By using load profiles it could prioritize appliances that would be used anyway, which would leave minimal impact on users' routines. While renewable energy is cheap to produce, it is expensive to store. Increased adoption of such resources will require a large amount of energy to be stored and released, this process is at best 80 % efficient. If that energy is optimally distributed, less energy would be lost due to conversion.

Other two-dimensional presentations

Figures 4.9 show how some appliances have a constant usage pattern over a year, whereas again others change it. Examples below are randomly picked appliances from UK-DALE and REFIT.

The Sub-figure 4.9a shows how computer usage changes over the year. In the first quarter, the computer was used from 10:00 a.m to 8:00 p.m. In week 18 it is possible to observe that the computer is less and less used throughout the day. Starting week 40 it is again possible to see that the computer is getting more and more use in the morning hours. This is a good example of how can a usage pattern slowly change through the year. Since the pattern seems to bounce back, it could be seasonally correlated.

The second example is Sub-figure 4.9b. It shows how TV usage changes over the year. Compared to the computer, it is possible to see that the pattern looks a lot more persistent with slight changes. Interestingly enough, when a close-up observation is made, it is possible to see that at the time when the computer was at its peak the TV was at it's low. And when the usage of computers decreased, the usage of TV increased. Due to the lack of metadata, it is hard to know the exact reason behind it.

The good thing about this change is, that it takes a few weeks before it changes. This will be important later in Chapter 6, when we will be designing an elderly care system, that will be based on periodical user behavior. This slow change gives the system time to adapt.

One observation of quick behavior change can be made in weeks 8-11 and weeks 38-37, where we can see a black row on all three sub-Figures in Figure 4.9. The instant decrease in activity is probably a vacation.

The last Figure is 4.9c where the load profile portrays the yearly use of the washing machine. In this case, the seasonal pattern is much clearer. It seems like the appliance was used in the early morning hours of the winter and early spring. This practice suddenly stops at week 13, until it appeared back in week 36.

(A) Computer (B) TV (C) Washing machine

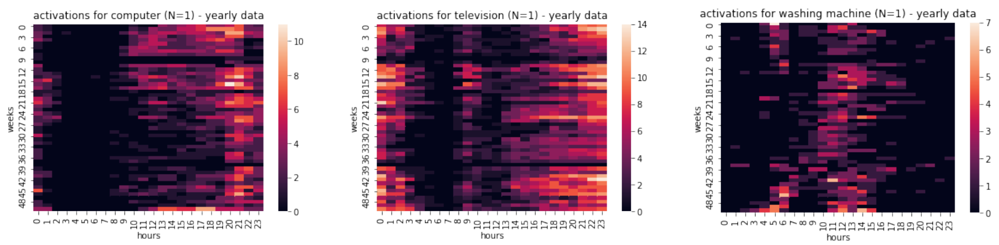
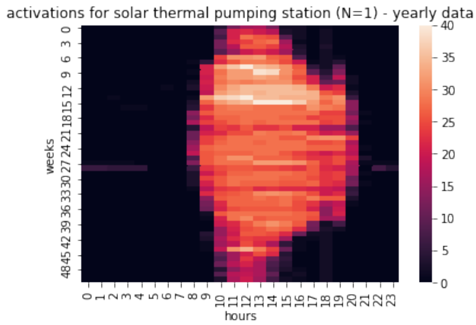


FIGURE 4.9: Various yearly two-dimensional load profiles for building 4 from REFIT.

Another example worth mentioning is Figure 4.10 from UK-DALE building 1, where data was collected from 2012-11-09 to 2017-04-26. Roughly 5 years of data mean that it is possible to build a decent profile.

(A) Solar thermal pumping station



(B) Light

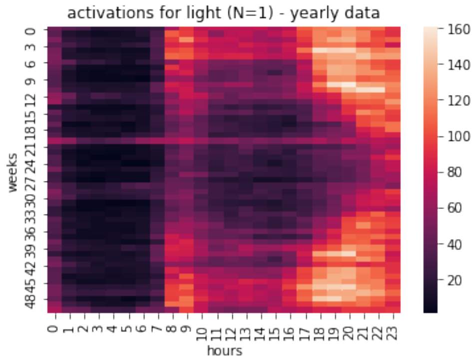


FIGURE 4.10: Effect of seasonal changes on load profiles

Appliance on Figure 4.10a activates when water in solar collectors heats up to a certain threshold. Since water heats up based on the strength of solar radiation, we can observe the change in solar radiation throughout the year for the UK.

Appliance on Figure 4.10b on the other hand works quite the opposite. We usually turn on the light when the solar radiation falls below a certain threshold, and turn it off when we sleep. The Figure is one of the best examples, where we can observe the combined effect of user behavior, in this case sleeping, and the seasonal effect of changing solar radiation on users' behavior.

Combining the Figures 4.10a and 4.10b enables us to differentiate between the two.

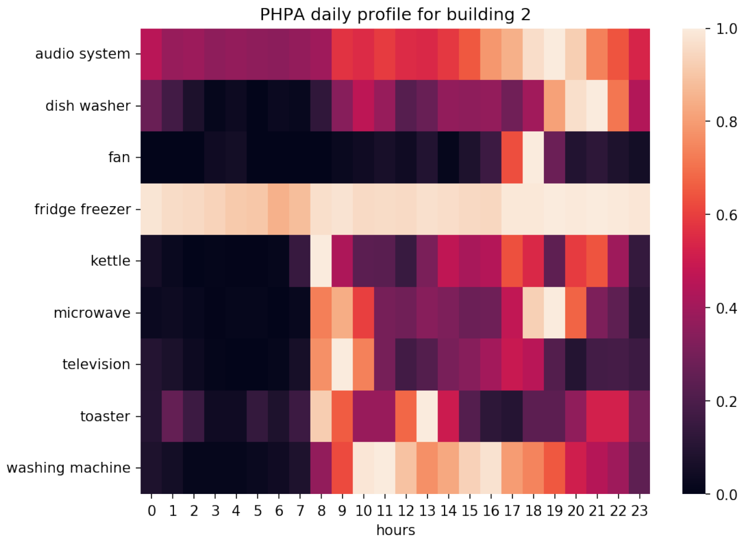
4.4 Per-building per-appliance

The last group of profiles is a combination of per-building and per-appliance load profiles. Observing the usage pattern of many appliances offers a better look into users' usage patterns.

In the case of elderly care, the goal is to observe a group of appliances. Activation of a group of appliances would yield a contextual event. If a stove and kettle are commonly used together each morning this use could translate to an event such as breakfast. To achieve this, one needs to observe all appliances at once such as shown in Figure 4.11.

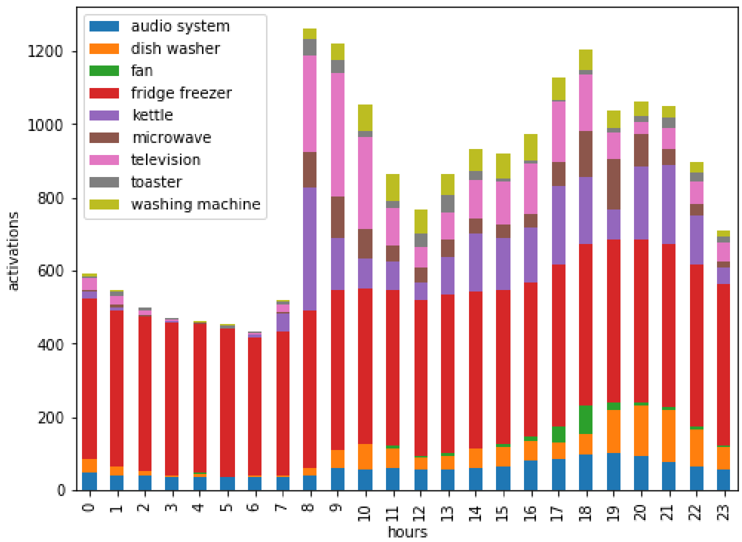
Figure 4.11 is also a good example of the elderly care system, that would detect an anomaly such as a fall, or a person unable to get up from the bed in the morning. This profile shows that the first thing in the morning used are a kettle and toaster, and with a delay of one hour, microwave and TV. This enables us to construct time thresholds in which appliances should be used. If none of these appliances are activated between set thresholds, morning would be considered anomalous. Although less likely, issues could also occur during the use of appliances. In case an elder falls during cooking, toasting bread or opening the fridge the duty cycle would increase, which would also be considered an anomaly. In case any of these anomalies are detected, the caregiver would be notified to check on the elder.

FIGURE 4.11: Daily per-appliance per-building building load profile



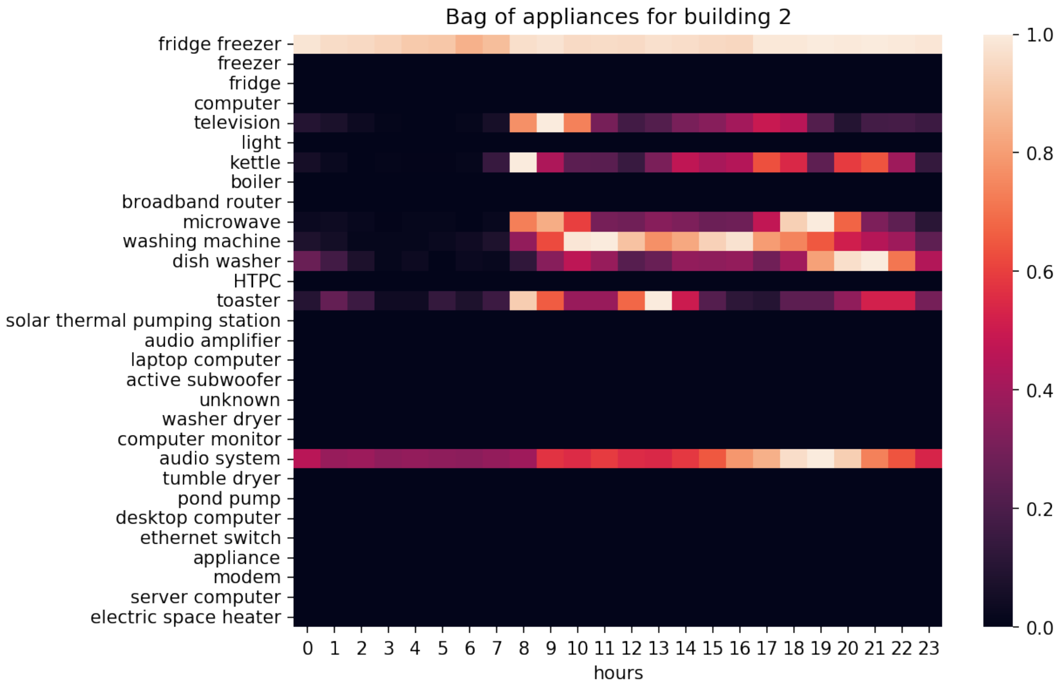
The very same data can be presented in an alternative way, such as shown in Figure 4.12. The usage pattern is the same as on 4.1a, except that it is possible to see the contribution of each appliance.

FIGURE 4.12: Stacked daily per-appliance per-building building load profile



These load profiles are useful when it comes to analyzing the usage pattern in one building. To be able to process the load profiles across many buildings a new profile, seen in Figure 4.13, must be introduced. The idea is derived from the bag-of-words method used in text processing, where a list of the most commonly used words is formed, and then used to process the text. Here, It is possible to use the activation data from all five datasets. A list of appliances is sorted by the number of activations and then only the top 30 appliances are selected. Using this list it is possible to present the usage of each building universally. This solves the issue of different appliances in different buildings.

FIGURE 4.13: Universal presentation of per-building per-appliance load profile



While analyzing Figure 4.13 we can see that the fridge freezer is most commonly activated. Since there is no pattern, and it is activated randomly, the pattern is presented as a white line. For the graph to be balanced, we have normalized the activations. If we had not done this, we could observe only the fridge freezer, due to its activation dominance.

Other, more dynamic appliances have a much clearer presentation of their activity. One other thing that we can notice is that there are a lot of empty load profiles for certain appliances. This is because we have no data for these appliances for this household. Probably, this is one of the biggest weak points of this load profile.

The Bag of appliances was not shown on the Table of profiles 2.8, since it is a special case of the per-building per-appliance profile shown in Figure 4.11.

4.5 Conclusion

This chapter showed how some activation profiles from Table 2.2 present real-world data, analyzed the presentations and further elaborated on their use-cases.

It was possible to see how each load profile presents its unique user activation pattern. Figure 4.1a offered us a unique look into how users behave on daily basis and Figure 4.1b how this behavior changes over a year. Next, with Figure 4.2, we presented how combining these figures presents new features, that were otherwise hidden. Further on it was shown how the very same presentations can be used on appliance data. For example, Figure 4.10 showed how this yearly change could be affected by the seasons. Finally, we have shown how more detailed profiles 4.11 could be used for practical applications such as elderly care.

Chapter 5

Visual analysis of LP's using t-SNE

5.1 Introduction

The load profiles can present big data in a way, that can enable users or algorithms to extract activation patterns. The one thing they do not offer is a comparison between activation patterns. One such example would be when checking how similar are certain usage patterns. This could be usage patterns within a household, where we are looking for appliances that are used similarly, or when observing multiple buildings and how their consumption differs. To measure the similarity of activation profiles, the t-SNE dimensionality reduction algorithm will be used.

The clustering of load profiles was researched many times before, as it was described in related work Chapter 2. We will be working with dimensionality reduction, where clusters are usually formed as a side product. The following clustering publications are worth mentioning. We have seen that authors Gerbec et al., 2005, Jeong et al., 2021 and Abreu, Câmara Pereira, and Ferrão, 2012 have clustered regular one-dimensional load profiles, as well as with 2D image-based load profiling in publications published by Park and Son, 2019.

The publication by Aréchiga et al., 2016 compared various dimensionality reduction techniques for clustering and visualization of load profiles. Their goal was to compare Principal Component Analysis, Isometric Feature Mapping, Sammon Mapping, Locally Linear Embedding and Stochastic Neighbor Embedding. They used power daily load profiles from residential and industrial areas. This publication was of the closest resemblance to our goals, that we were able to find.

In all cases, work has been done with the power load profile, whereas in this case, we will try to find similarities between activation profiles using a t-SNE algorithm. Most of the publications used single-time dimensions, whereas we will use two-time dimensions.

Although the use-cases were presented in-depth in Chapter 2, it is worth mentioning one specific use case. The increasing price of energy resources, could lead to over-saving and living in cool homes. By using similarity metrics between profiles across different buildings, it would be possible to detect outliers when it comes to heating. With this approach, it would be possible to detect users, that are living in below-average cool homes and offer them cheaper plans.

5.2 Goals

The chapter will demonstrate the application of previously unused load profiles, and show the practical use case using a t-SNE neighboring algorithm.

Using this algorithm we will show how user data is related in high dimensional space. For this, we will use per-building, per-appliance load profiles with many combinations, where each one enables us a different view of load profiles.

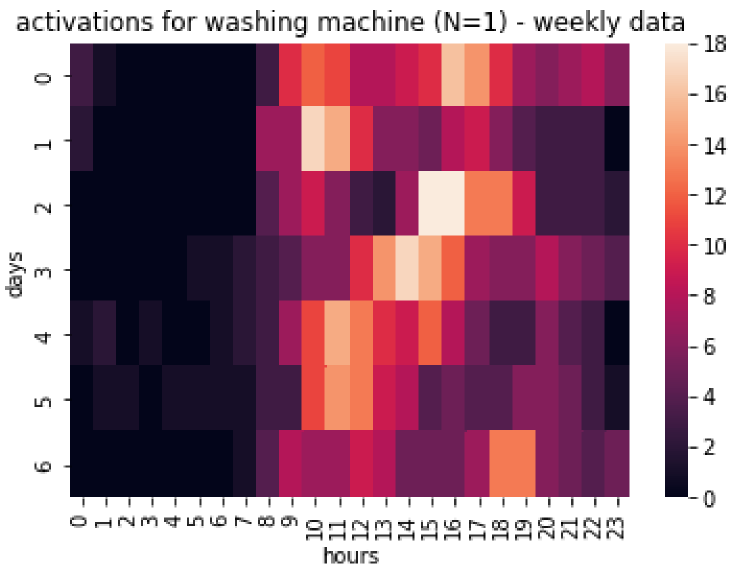
5.3 Methodology

5.3.1 Load profiles

Weekly-daily load profile

During testing, a weekly-daily load profile constructed from a month of data will be used. Y-axis will present the days in a week and X-axis presents the hours in a day. Weekdays are labeled from 0 to 6, and hours from 0 to 23. Since we are working with images, the origin is placed in the upper-left corner. This means that a pixel in the upper-left corner presents the first hour of a week, this would be a Monday from midnight to one o'clock. The lower-right corner presents the last hour of the week. Since there are roughly 4 weeks in each month, each pixel will present 4 samples. One such example of profiles that we will use, was already presented in chapter 4 with Figure 4.8. For practical reasons, we are presenting it again here with Figure 5.1.

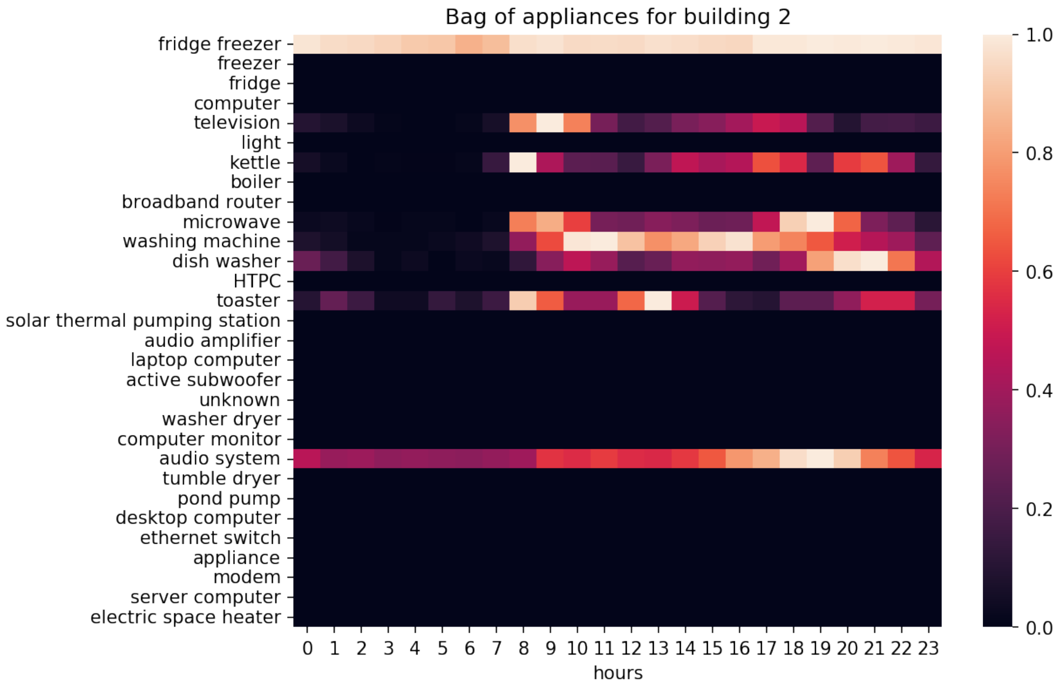
FIGURE 5.1: Weekly per-appliance load profile



Bag of appliances load profile

Another load profile that will be used at the end of this Chapter will be the bag-of-appliances load profile. The profile was presented and analyzed in depth in chapter 4 and was presented in Figure 4.13. But again, for ease-of-use purposes, we will summarize the profile here.

FIGURE 5.2: Universal presentation of per-building per-appliance load profile



To build the profile seen in Figure 5.2, we used the data from all 5 datasets and made a list of the most commonly used appliances. Only the top 30 appliances were selected. This enables us to have the same load profile for all buildings, and thus enables us to see how the usage differs across them. One problem that arises here is the missing appliances. These appliances present themselves as a black line. A lot of missing appliances may cause the image to be primarily black, which could cause trouble for the algorithm processing this as an image.

Data

We have on average roughly one year of data per building. In some cases few weeks and in others up to 5 years for some appliances. By slicing this data into 1-month-long intervals and converting them to load profiles we were able to obtain 5218 samples.

5.3.2 Datasets

To present the results REFIT (Rashid, 2019), UK-DALE (Kelly and Knottenbelt, 2015), ECO (Beckel et al., 2014), REDD (Kolter and Johnson, 2011) and iAWE (Batra et al., 2013) datasets will be used. Combined datasets include data for over 25 homes, where some have up to 5 years of data. The structure of datasets will be analyzed in larger depth in the next chapter 6.

5.3.3 t-SNE algorithm

The t-SNE Maaten and G.E. Hinton, 2008 or t-distribution stochastic neighboring embedding is a method for portraying high dimensional data in low dimensional space. This process is also known as dimensionality reduction.

One of the well-known dimensionality reduction algorithms is PCA. The key difference between the two is that one is linear, and the other is non-linear. PCA, linear, projects data in new space and finds the one with the least variance between data points. SNE Geoffrey Hinton and Roweis, 2003, non-linear, is composed of two main parts. The first one is converting the high-dimensional Euclidean distances between data points into conditional probabilities that represent similarities. The pairs with high similarity have a high probability, and pairs with lower a low probability. Second, it uses Kullback-Leibler divergence to minimize it with respect to a location on a map. To achieve this it uses gradient descent to minimize the cost function. Over many iterations, similar data points should be close together and far away from dissimilar objects. Similar data points usually form clusters.

t-SNE uses SNE as a basis, except that it uses t-student distribution instead of normal to calculate the similarity.

In our case, two dimensions will be used. Since this is a non-linear dimensionality reduction, the axis usually presents dimensions that are hard to comprehend by the brain. Since the algorithm uses similarity at the base of the algorithm, it is possible to see which samples are more similar to each other.

5.4 Results

The results will be presented in three subsections

- Per-building load profile
- Per-appliance load profile
- Per-building per-appliance load profile

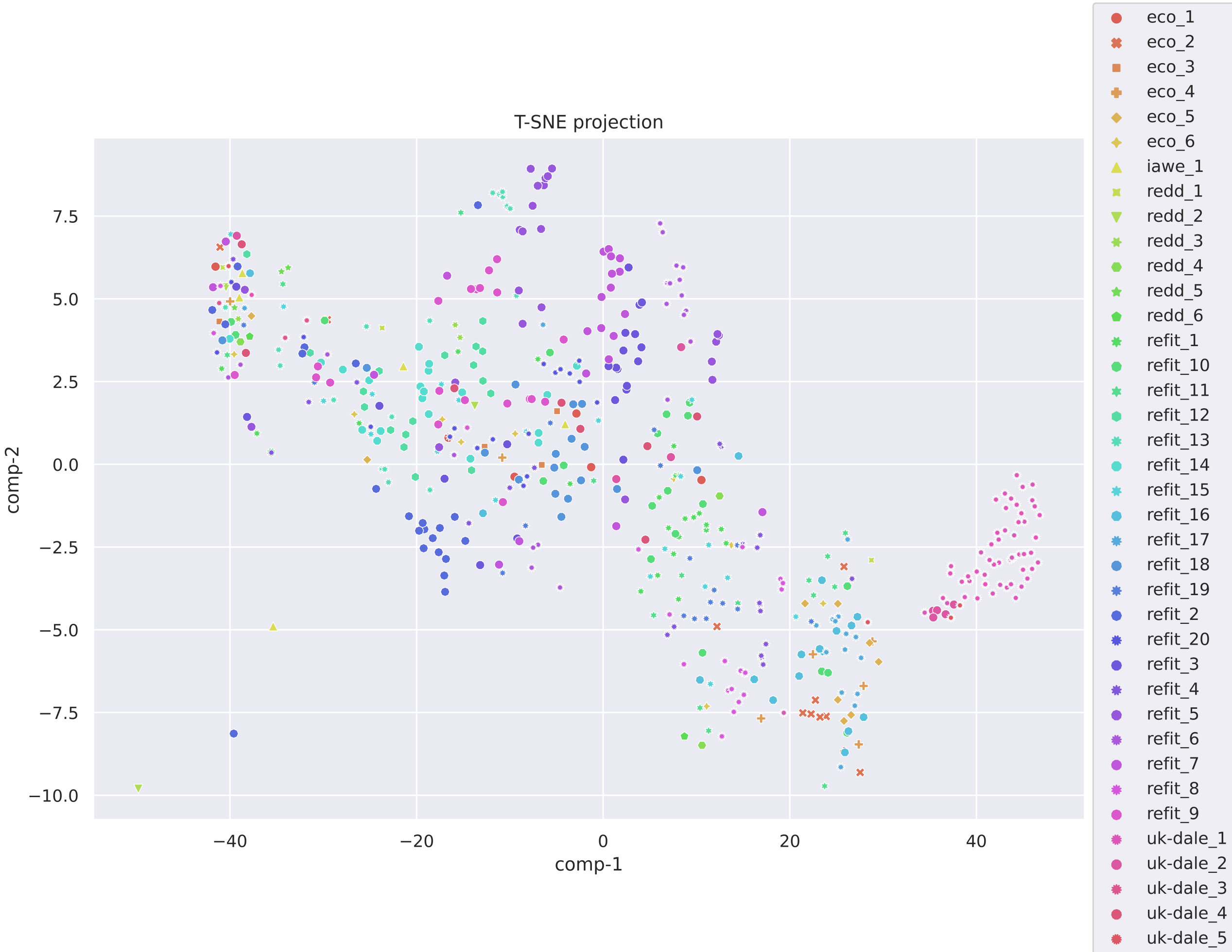
Most of the focus will be done on the per-appliance load profile since it is the most universal.

5.4.1 Results for per-building load profiles

This load profile is useful when it comes to comparing how activation patterns change over buildings and datasets. Per-building data uses combined activations of all appliances to present the aggregated usage pattern.

Figure 5.3 is using non-normalized data, meaning the number of appliances in a building will affect the end load profile. The algorithm could pick up on how many appliances are being used. In some cases, such as energy poverty detection, this information is useful, again in others we would like to find a more complex usage pattern.

FIGURE 5.3: Projection of per-building load profiles



The Figure 5.4 below presents the actual load profile for each sample. It is possible to see that on the left there are mostly samples with very little activity, and on the right, we see samples with more activity. Since the two plotted components are of a higher dimension, it is hard to determine what they present. As said t-SNE gives us the intuition of how load profiles are connected in higher-dimensional space.

The following figures are best viewed in color and a digital format. Readers reading the digital version should have the ability to zoom into each cluster, and see the actual samples. Readers reading a paper version can still explore the high-resolution figures online via the provided link at the end of the document.

FIGURE 5.4: Projection of per-building load profiles with actual samples

Normalized load profiles

To solve the issue mentioned in subsection 5.4.1 have to normalize the data between 0 and 1. The Figure 5.5 shows how normalizing samples affect the algorithm.

When comparing figures 5.3 and 5.5, it is possible to see that the samples on the latter are much closer to each other, while it is still possible to see the individual clusters. This could imply that the normalized usage pattern of users is more similar to the activation pattern of users. A normalized activation pattern tells us at what part of the day the appliances will most likely be used, and the activation pattern tells us how much will the appliance be used in each part of the day. Based on that, we can conclude the time when the appliance is used is more consistent than how much it will be used.

FIGURE 5.5: Projection of normalised per-building load profiles

The Figure 5.6 presents only the main cluster of samples. Since the smaller cluster presents mostly low entropy data, it was cut out. If the reader wants to see the samples in the cluster, the very same cluster can be found on the far left in Figure 5.4.

FIGURE 5.6: Projection of normalised per-building load profiles with actual samples

On Figure 5.6 it is possible to find various usage patterns. But the general pattern is that there is less activity during the night with one peak in the morning and evening hours. Some buildings are more active during the week and again some more during the weekend. A lot of the data is from UK-DALE building 1 (pink box). It is possible to see that the building has one big cluster where activations are generally similar, with few outliers, where the pattern completely changed. Albeit less obvious, this pattern is the same for all buildings. This happens due to events such as vacations, holidays or weather-induced behavioral changes.

5.4.2 Per-appliance

We can use per-appliance load profiles to examine how different appliances are used in a single building, how a single appliance is being used across other buildings or how many appliances are being used in many buildings.

Per appliance load profiles are built using sub-meter data, meaning each load profile should present each appliance.

Single appliance over many buildings

Using one appliance and the building as a label, allows us to examine how the same type of appliance is being used across different buildings.

Fridges are generally a bad indicator when it comes to user behavior since the user does not affect its operation. The only case when the user interacts with it is when opening the door and turning on the light inside. Usually, this event is dwarfed by the activations of a compressor. This also means that the usage pattern should be the same across all buildings. This can be seen on Figure 5.7, where apart from REFIT building 1 and 11, there are no clusters.

FIGURE 5.7: Projection of fridge load profiles for various buildings

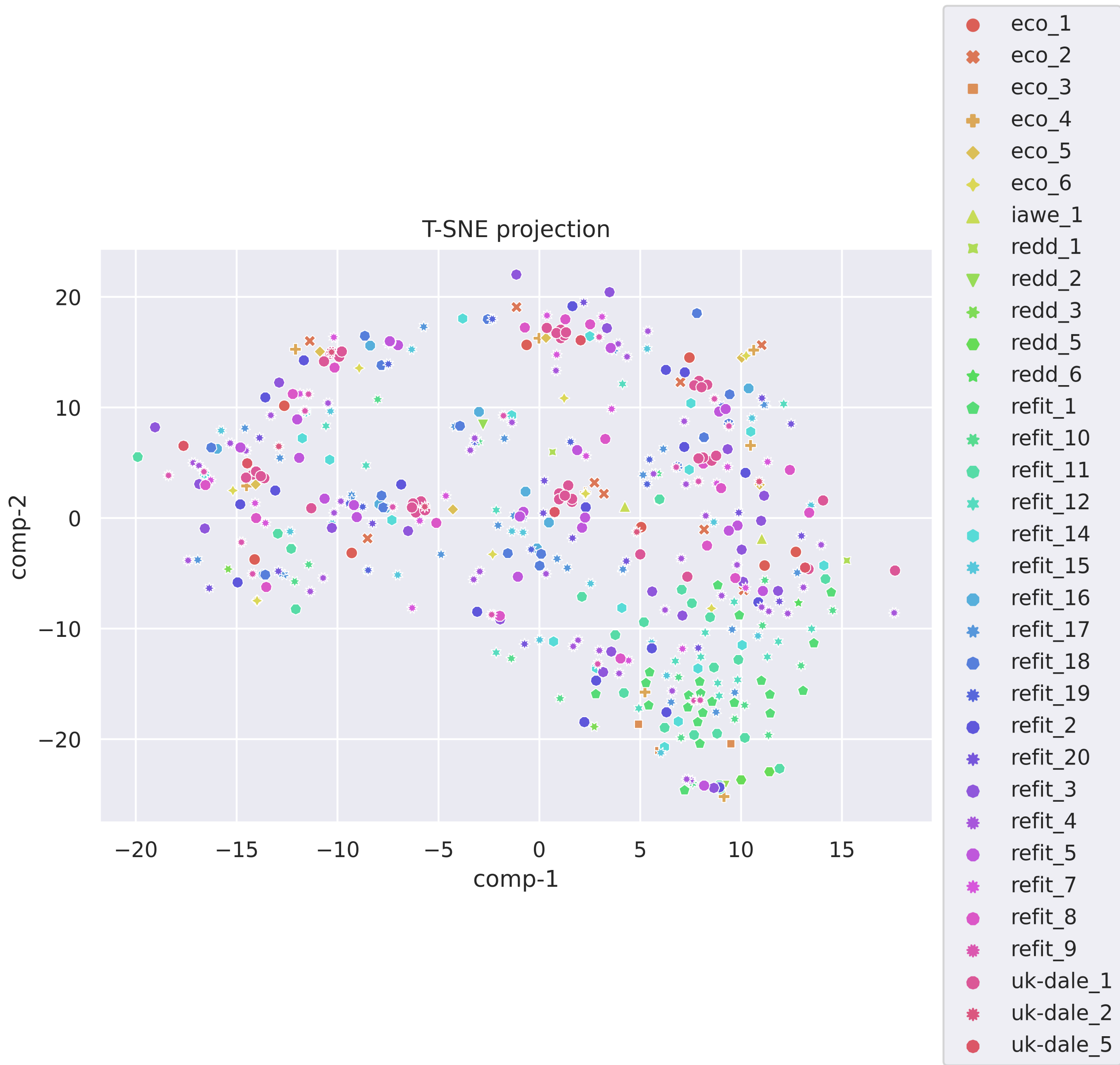


Figure 5.8 Shows mostly bright images, apart from few outliers. Load profiles scattered in a circle are generally less dynamic than the ones at the bottom. Figure 5.8 is a good example of how load profiles with little to no human interaction, can look a lot different. This could be due to different makes of the appliances, malfunctions of the appliance or the meter measuring it.

FIGURE 5.8: Projection of fridge load profiles for various buildings with actual samples

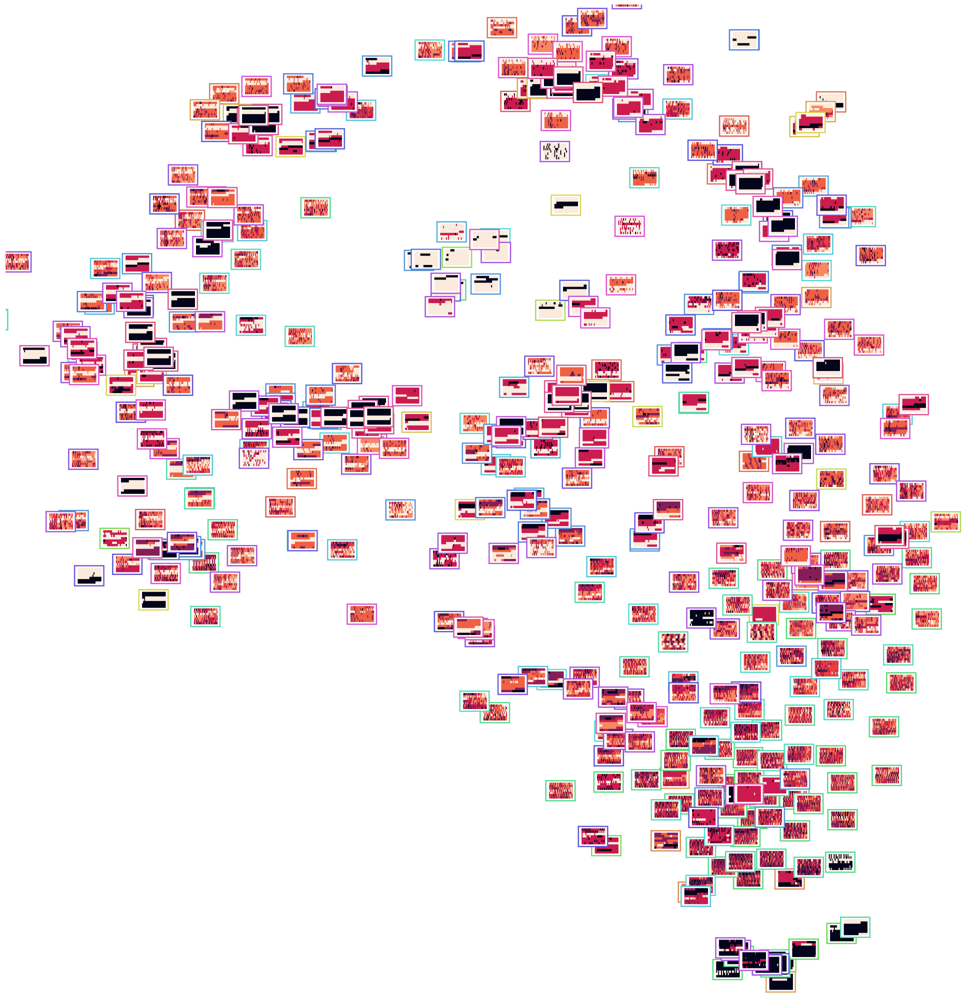
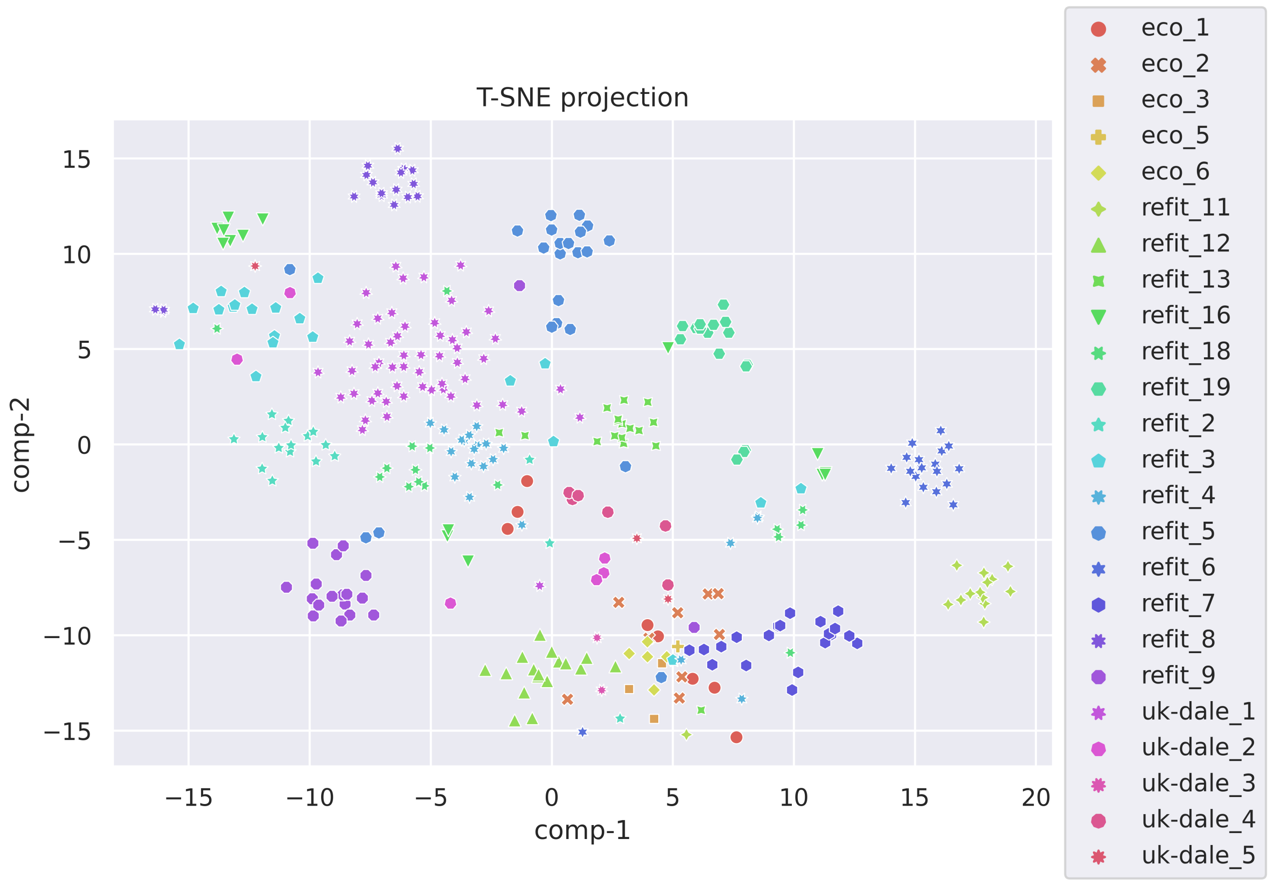


Figure 5.9 shows how, compared to fridges, kettles have many clear clusters that are spaced out between each other. This could mean that every household uses a kettle a bit differently. This cluster is a good example where we can see how strong is a routine of a user. The closer together the clusters, the higher the routine since samples are more similar to each other.

FIGURE 5.9: Projection of kettle load profiles for various buildings



The Figure 5.10 shows us that images on the lower part of the plot contain less activity than the others. Load profiles that are closer together have more similar activation patterns. Similar activation patterns are caused by similar behavior, which is essentially a routine. This means that this projection could be used to calculate how much a behavior variates in time for each building. This could be calculated by measuring the scattering of samples (variance) for each building.

If we find samples that always activate in the same morning buckets, we would see that they form a straight line on the y-axis. This is the daily routine. One such example can be seen in Figure 5.9 in cluster refit 5 and refit 9, where we can see the lines and the pattern throughout the day. Since the routine is present, the samples look more similar and are therefore closer together. This does not necessarily mean that closer the samples higher the routine. They could also be together in case of "ordered chaos" such as can be seen in Figure 5.9 for building refit 16 and refit 8 where there is no pattern through the day. So the scattering is not a precise metric when it comes to the routine, but it gives us a rough idea of its presence. The strength of a routine is an important feature that will be used in Chapter 6 to build an elderly care anomaly system.

FIGURE 5.10: Projection of kettle load profiles for various buildings with actual samples

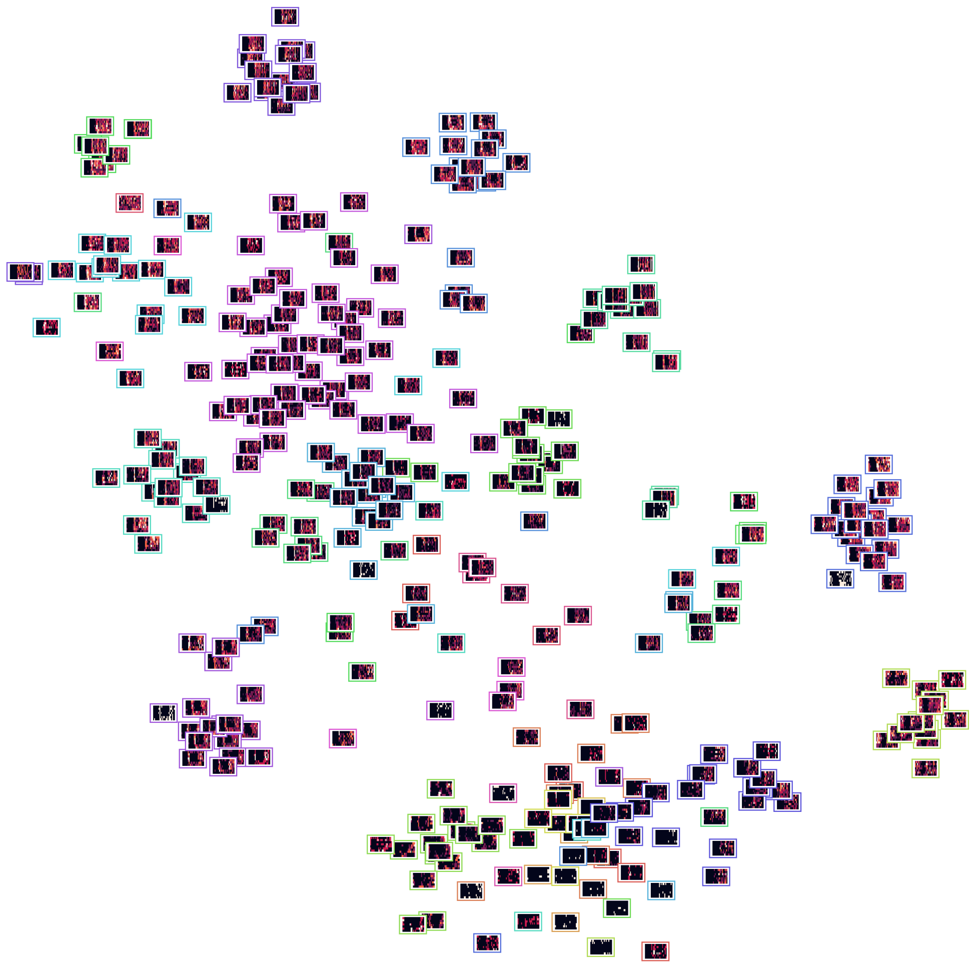


Figure 5.11 shows that microwaves are again a bit different from the kettle. They are more clustered than the fridges, and less than the kettles, even though they are used similarly. This could be due to additional electronics such as a clock that are built into the appliance. This could lead to some samples being registered as turned on due to a "dark" current. One other difference between the two is that microwave has more than one mode of operation.

FIGURE 5.11: Projection of microwave load profiles for various buildings

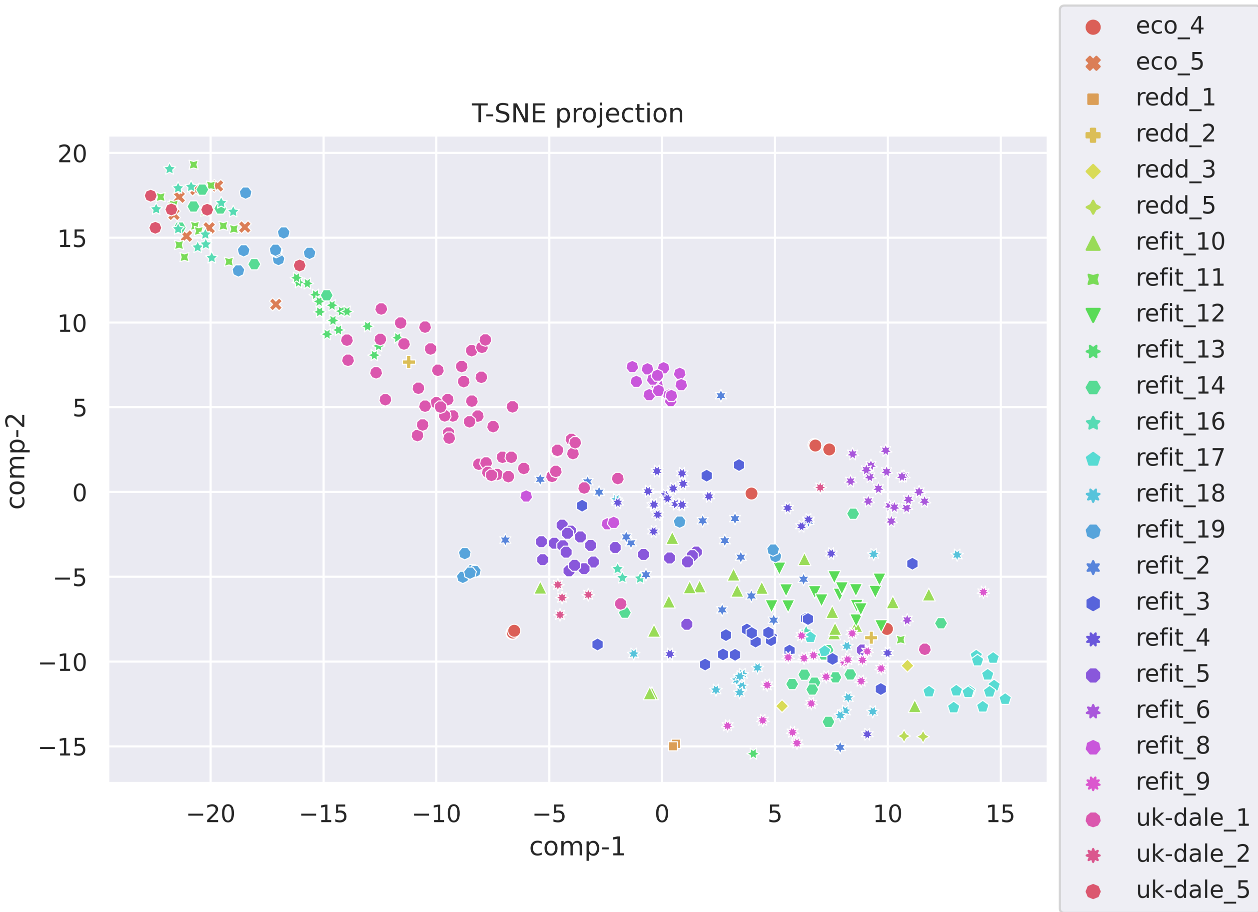
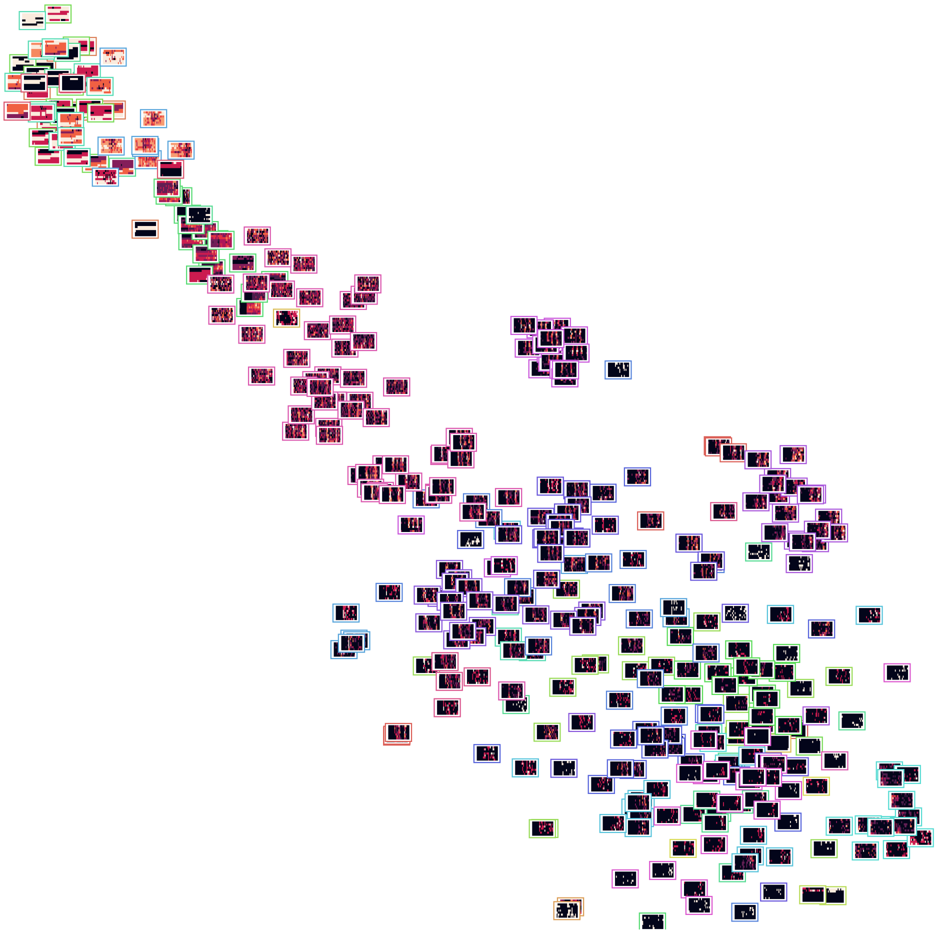


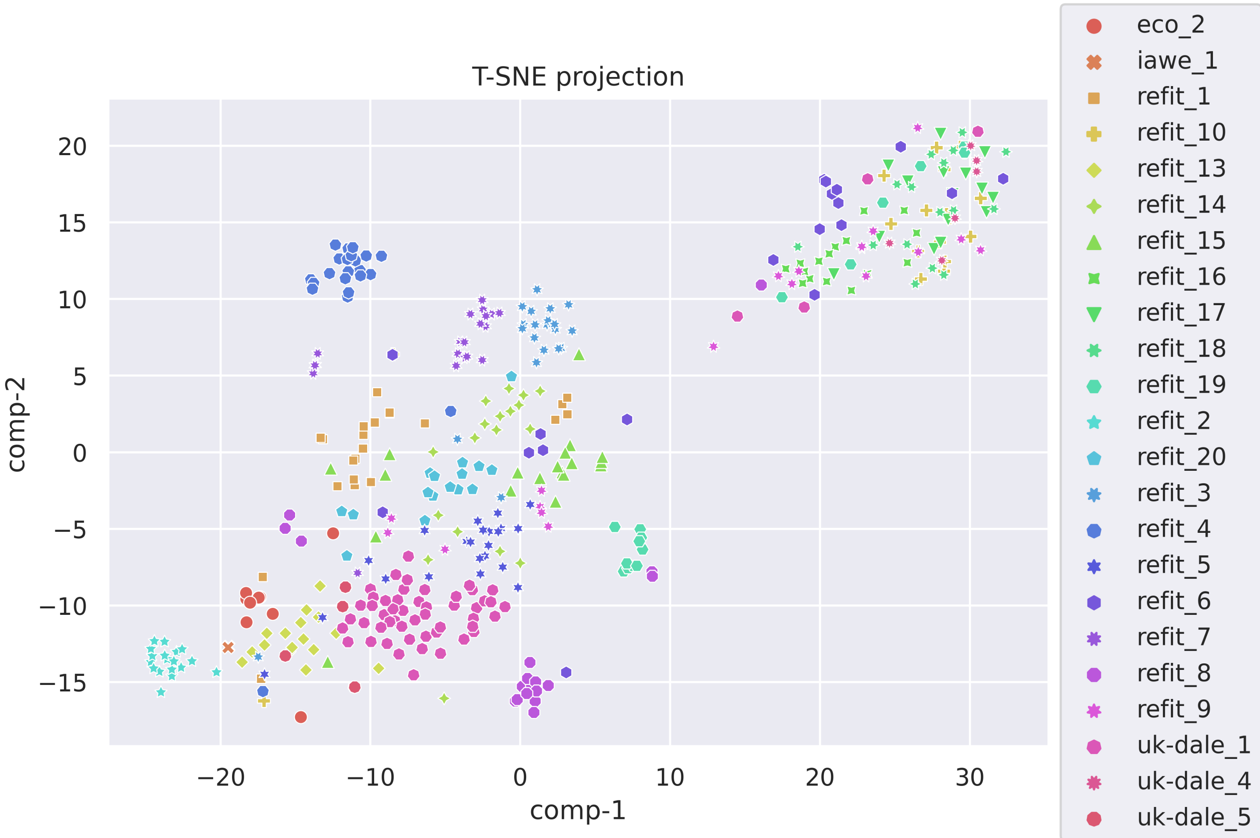
Figure 5.12 shows the faulty samples could be the ones in the upper left part of the plot since they are too bright. They do present a pattern, but it is questionable what it presents since it seems like it's turned on during the nighttime. Images at the other end show less lot less activity, which could indicate that the household does not use microwave as much. The most interesting load profiles are in the middle of the plot, where it is possible to observe clear activation patterns.

FIGURE 5.12: Projection of microwave load profiles for various buildings with actual samples



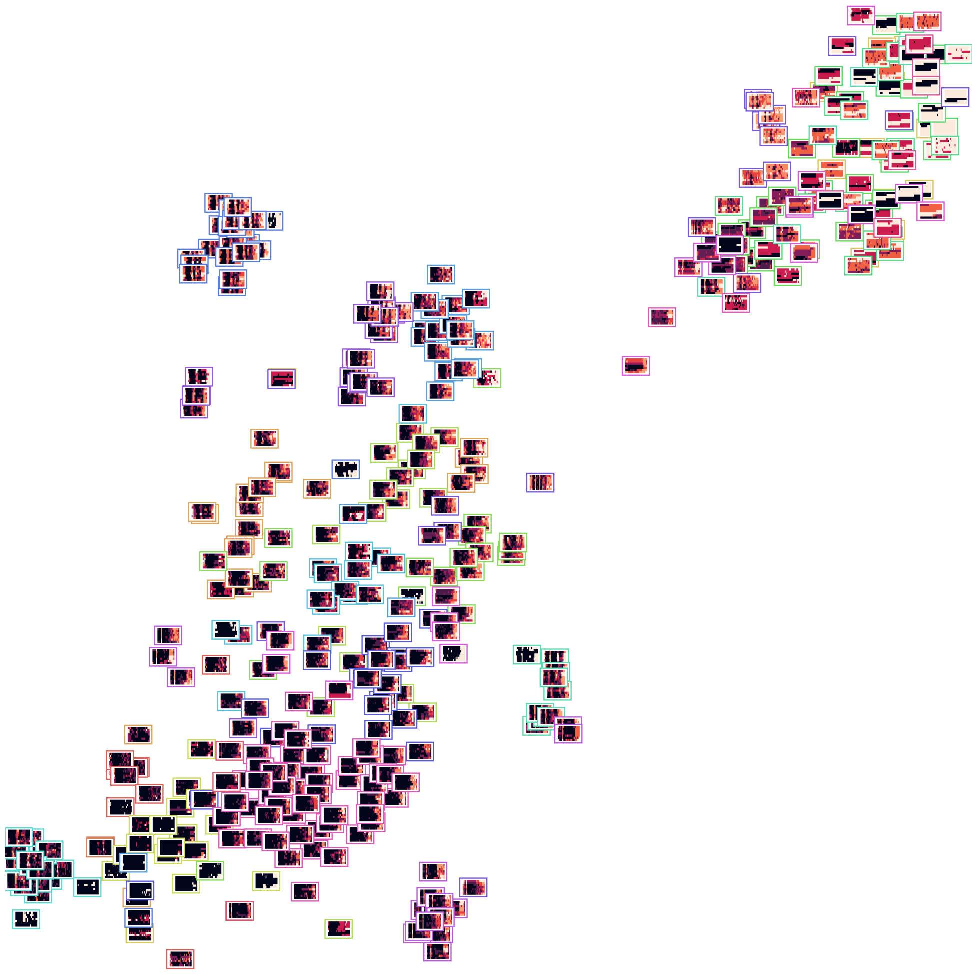
The last per-appliance example is television presented on Figure 5.13. Television was chosen since it is the most commonly occurring appliance. Interestingly enough, televisions form nice clusters with a few outliers. Clusters are separated but close together, this could mean that usage patterns across buildings are unique but not that different from one another. The load profiles in some clusters are also close to each other, which could also indicate a higher routine.

FIGURE 5.13: Projection of TV load profiles for various buildings



The images on the Figure 5.14 prove the fact that outliers' consumption is a lot different. Again the bright images could be the results of faulty appliances, faulty meters or simply odd behavior. The Figure 5.14 also enables us to see that TVs are primarily used in the evening hours. Outliers from the main cluster show slightly different behavior. One such example is the blue cluster (building REFIT 4), where appliances are mostly used in the morning hours. One other interesting observation can be made when looking at the purple cluster. This is the far low cluster for building REFIT 8. Here, the TV is being consistently used every day in the early morning hours. This is portrayed as a straight line. There could be two possible explanations for this. First is simply a high routine of a user, who turns on the TV every morning to listen to the news. The other is that the TV updates itself every morning. This is probably not the case since updates do not occur on regular basis. What is also interesting, is that the very same pattern can be observed in a few other buildings, one example being building REFIT 19.

FIGURE 5.14: Projection of TV load profiles for various buildings with actual samples.



Per-appliance load profiles - comparing appliances

Figure 5.15 presents the general picture of where each appliance lays in comparison to the other. One obvious issue here is that there are too many appliances, and it is impossible to comprehend the plot.

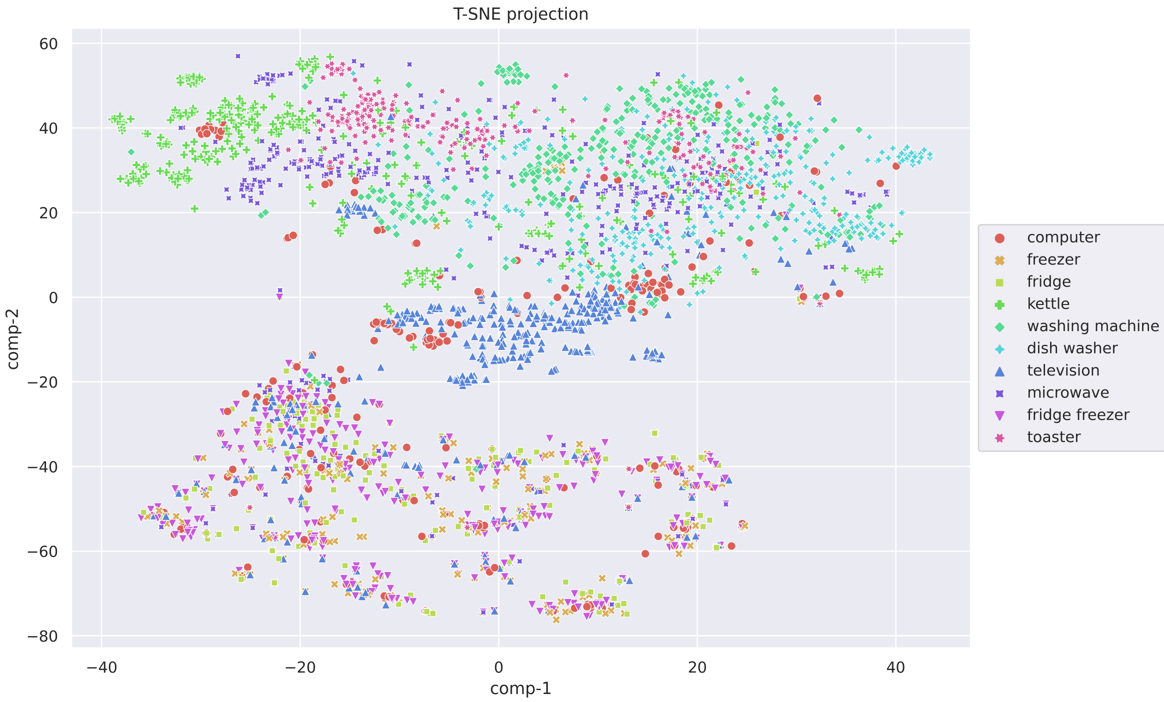
FIGURE 5.15: Projection of per-appliance load profiles

The same goes for image presentation on Figure 5.16. We can see, that most active appliances are the ones in the bottom left, by moving to the upper right part of the corner, we can see less activity. Less activity does not necessarily mean that load profiles contain less information about user behavior.

FIGURE 5.16: Projection of per-appliance load profiles with actual samples

To get a general idea of where each appliance group lies, let's filter out all appliances that have less than 150 samples. Applying this filter yields Figure 5.17.

FIGURE 5.17: Projection of filtered per-appliance load profiles



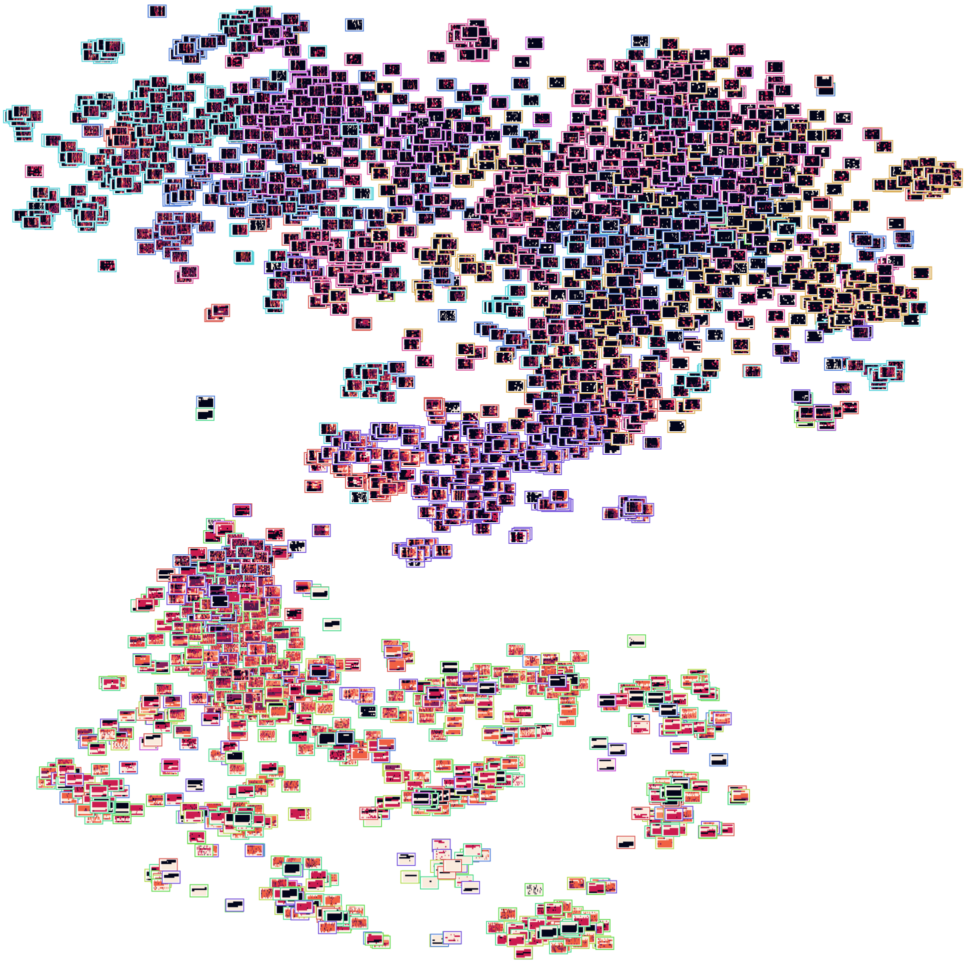
The Figure 5.17 shows how these 10 appliances ares connected in high dimensional space. Kettles, microwaves and toasters are quite similar when it comes to usage patterns. They are operated for a short amount of time and are usually used in users' routines in the morning or evening. These appliances are located in the upper left part of the plot.

The second group of appliances that are quite near each other is white goods (without fridges) such as washing machines, dishwashers, dryers etc. Let's say that they are white goods with a program. This group of appliances is located in the upper right part of the plot.

The third group of appliances is white goods with a compressor. They are usually not affected by human interaction and are therefore harder to cluster. They are located in the lower part of the plot.

The final group of appliances is televisions and computers. They lie on a bridge between the fridges and other groups.

FIGURE 5.18: Projection of filtered per-appliance load profiles with actual samples



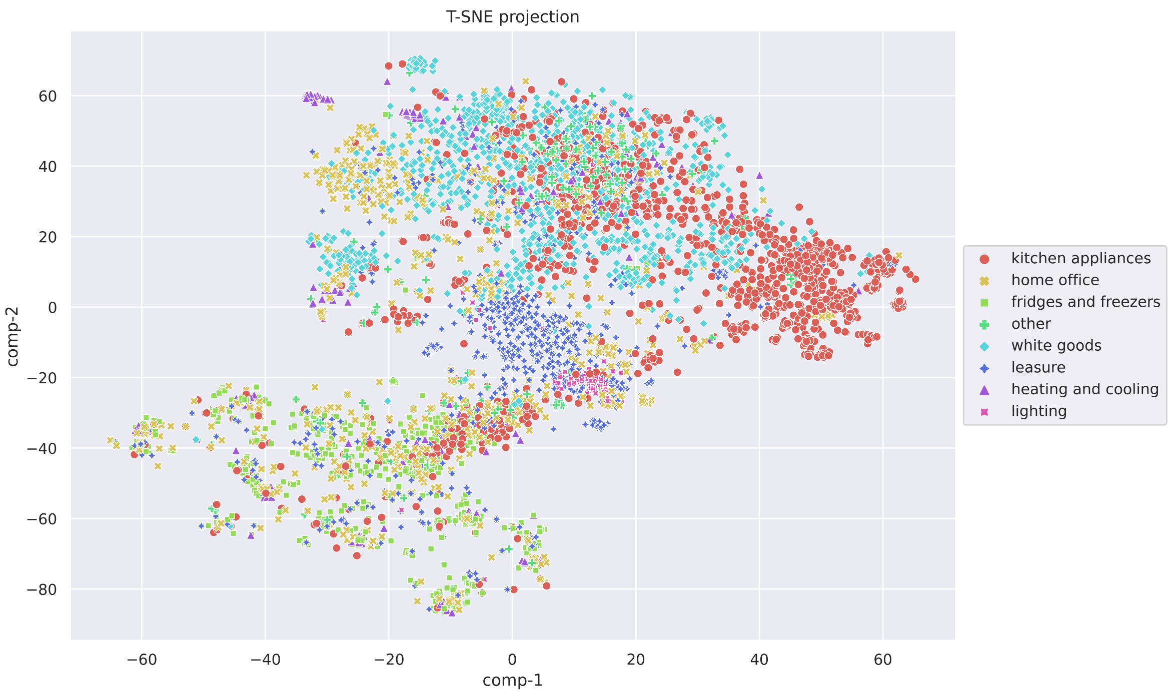
Even though the set of load profiles is different due to filtering, the Figure 5.18, retains a similar structure to the previous Figure 5.17.

Knowing that a pattern exists, we can use the newly found group to define new appliance groups. The following 8 groups will be defined

- Kitchen appliances - toasters, ovens, microwaves, etc.
- Fridges and freezers - contains fridges, freezers and fridge freezers or white goods with a compressor
- White goods - washers, dryers, dishwashers i.e. white goods with a program
- heating and cooling - Electric radiators, dehumidifiers and HVACs
- leisure - Living room appliances such as TVs, games consoles, audio amps, HTPCs, etc.
- home office - Computer, laptops, printers, network equipment, chargers, etc.
- lightning - lights and lamps
- Others - unknown and unlabeled appliances

Applying these groups yields Figure 5.19. The new plot shows how, although appliances could be used by a different user, maybe even by users in a different part of the EU or world, they can be grouped in a high-dimensional space.

FIGURE 5.19: Projection of grouped per-appliance load profiles



The Figure 5.20 below is the same as the first Figure 5.16 in the subsection, except it is easier to use color to see the appliance they present

FIGURE 5.20: Projection of grouped per-appliance load profiles with actual samples

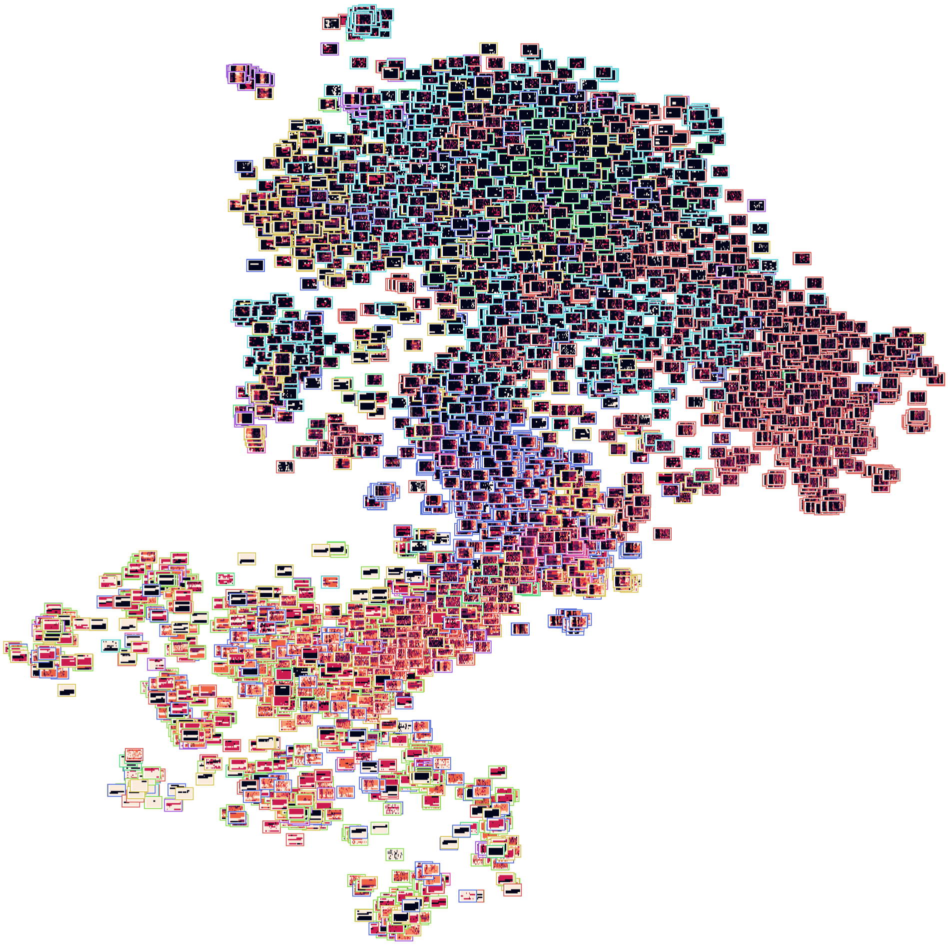
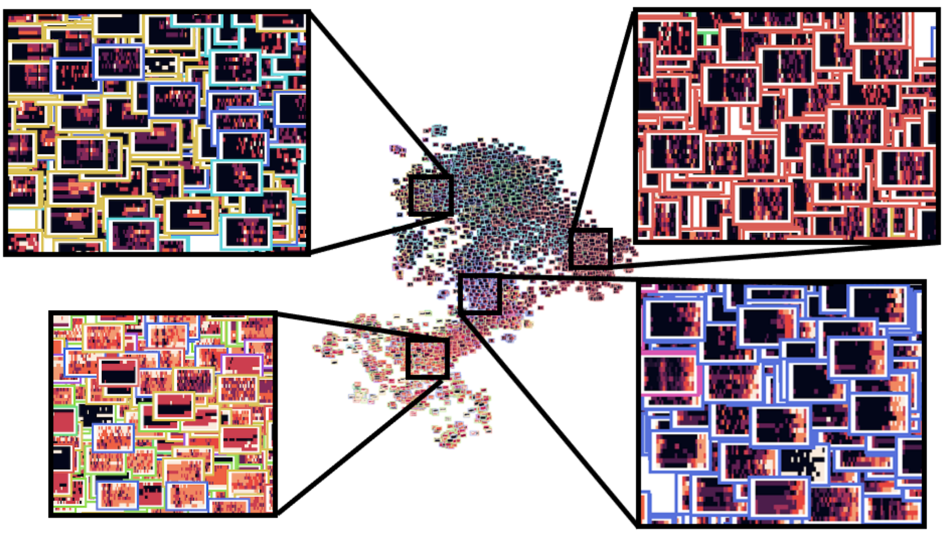


Figure 5.21 shows the four main types of profiles for readers that cannot zoom in.

FIGURE 5.21: Projection of grouped per-appliance load profiles with actual samples



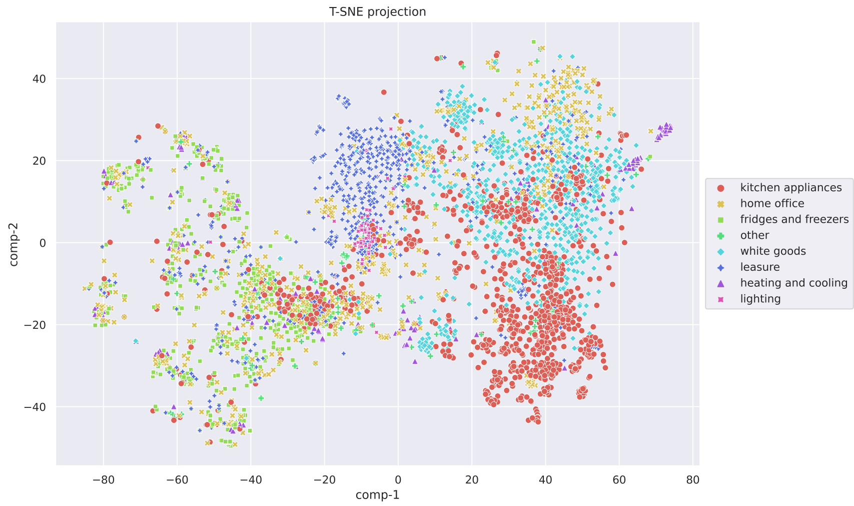
One issue that causes the t-SNE algorithm an issue is low entropy data or in other words, images that are almost completely dark or white, due to various faults in appliances or measurements.

If we calculate the entropy for each image and set a threshold, it is possible to filter out these samples. By setting an entropy threshold of 0.5, we filter out around 5 % of all samples.

FIGURE 5.22: Projection of entropy filtered per-appliance load profiles

Again, we can apply appliance grouping and get nicely formed clusters, such as can be seen in Figure 5.23.

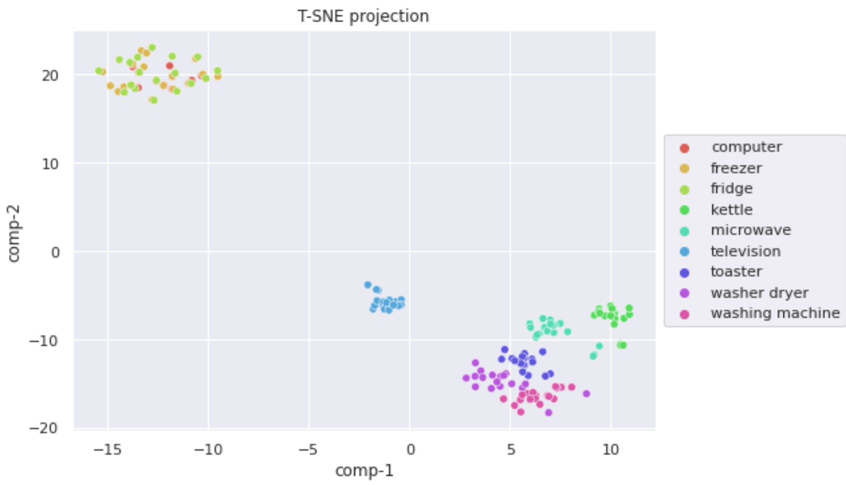
FIGURE 5.23: Projection of entropy filtered per-appliance load profiles with actual samples



Comparing appliances in a building

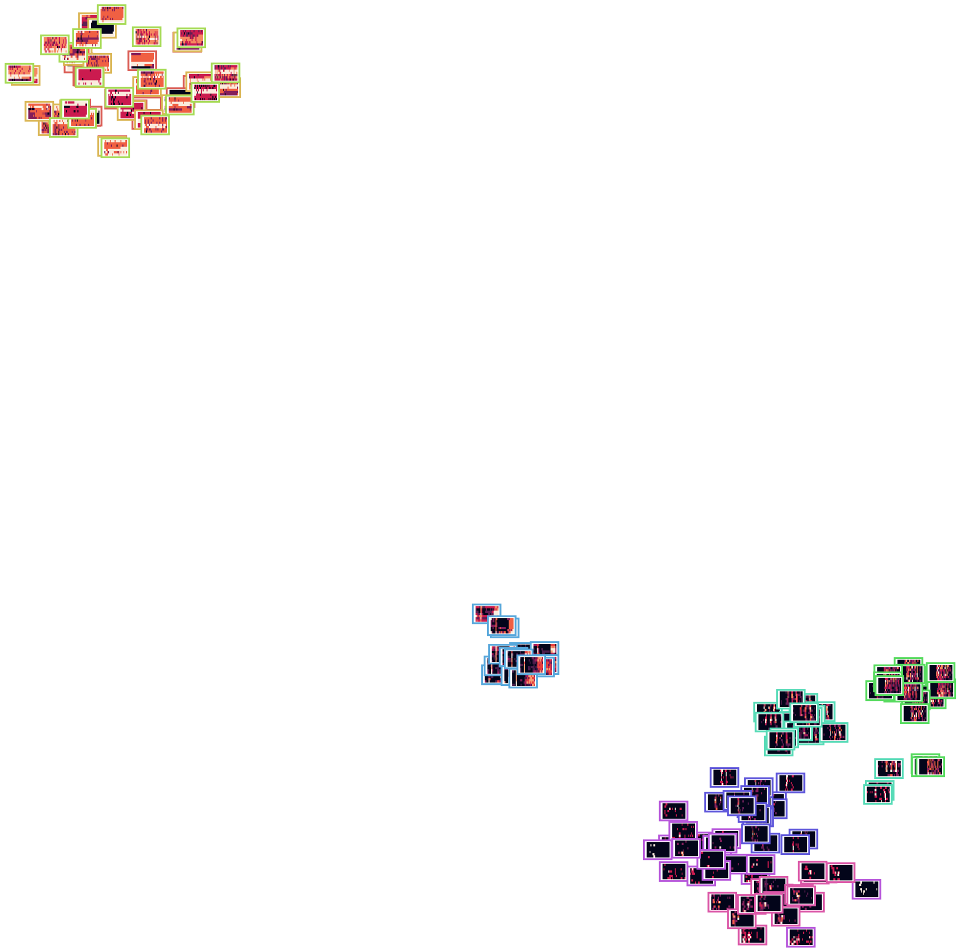
It is also possible to use per-appliance data to study individual buildings, and how each appliance is used. In this case, we have used building 8 from REFIT as an example.

FIGURE 5.24: Projection of per-appliance load profiles in a single building



In general, the scattering is similar to before. Fridges and freezers are located opposite of white goods and kitchen appliances. The television lies somewhere in between.

FIGURE 5.25: Projection of per-appliance load profiles in a single building with actual samples



Similar as before Figure 5.25 shows fridge cluster as the most active, with the least interesting information. In the middle, we can again observe the TV that is mostly being used in the evening hours. The samples also point to the possibility of a high user routine in the early morning hours. The high routine is portrayed as a straight line on the figures.

In this case, when observing white goods in pink and purple boxes, it is possible to see that this user primarily uses them during the night hours. This could point out that the user is making use of cheaper tariffs.

One other interesting observation that can be made here is comparing kettles, microwaves and toasters. Usually, these appliances are used similarly and in similar parts of the day. Here the toaster and microwave are being used periodically, but the kettle is being used throughout the day with no general pattern. It is self-obvious that some users have higher routines than others, but this observation would add that some users can have a higher routine for some appliances and lower for others, of the same type.

5.4.3 Per-appliance per-building

To study the usage by comparing all appliances between buildings, we have to use one of the proposed load profiles and in this case, this is a Bag of appliances.

Bag of appliances

This load profile is a combination of the load profiles above, except it offers a larger detail when observing groups of appliances. Since we are using one dimension for appliances, we will use only the daily dimension.

To construct such a profile we need a universal way of constructing it. This is done by measuring how many times each appliance occurs in the datasets, then this list is sorted from most common to least common, and finally, the top 30 are selected.

The problem with such a comparison is, that it is best if all buildings would use the same appliances. Since that is not the case, missing appliances are portrayed as always off.

This is the main reason why we can see in Figure 5.26 the clusters are separated quite a bit. We can still see that some clusters are closer than others, meaning they are more similar.

FIGURE 5.26: Projection of a bag of appliances load profiles for various buildings

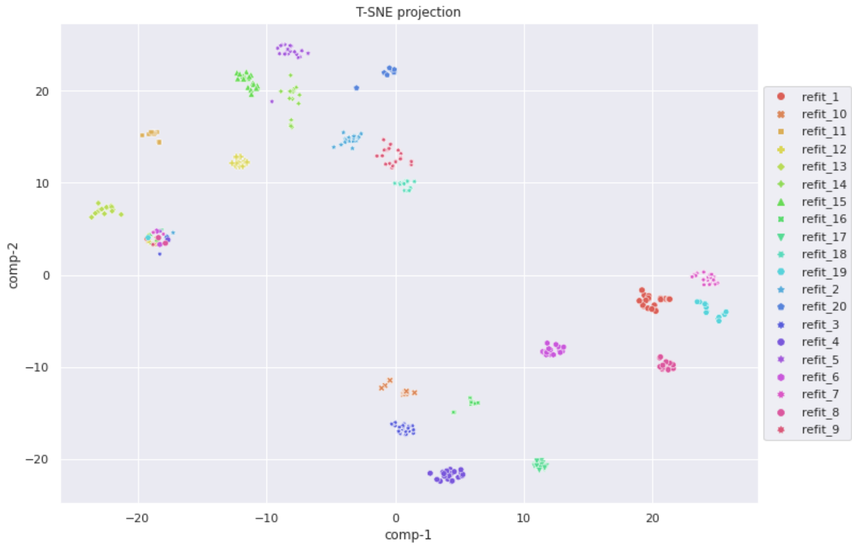
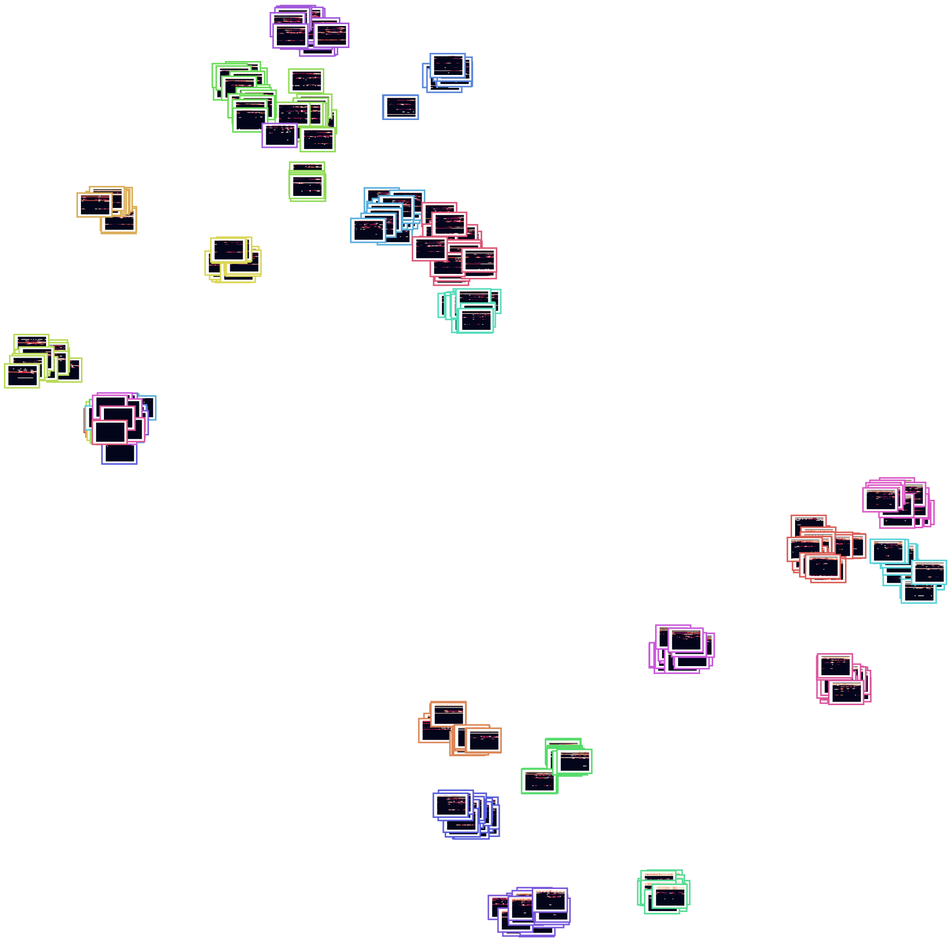


Figure 5.27 shows that load profiles are split between two poles. By observing the Figure it is possible to see that all the bottom clusters have more than one active white good with a compressor (fridges and freezers), while the top ones have only one. In general, the bottom buildings have more appliances, with more activity than the top ones.

FIGURE 5.27: Projection of a bag of appliances load profiles for various buildings with actual samples



5.5 Discussion

The main goal of the chapter was to show how data is related in high-dimensional space. For this to be achieved, three different types of load profiles were utilized.

Per-building load profiles offered a look into how activations patterns differ across different buildings and datasets. Per-building per-appliance bag of appliance load profiles offered the same thing, but in greater detail. Per-appliance load profiles were the most versatile and were utilized in the most various ways. First, we have shown how the same type of appliance is being used across various buildings. Next, we compared appliances with each other. Since the plot was hard to comprehend, we have defined appliance groups. These new groups formed clusters, which cleared up the plot and revealed the connection between types of profiles. Finally, we compared how appliance load profiles are connected in a single building.

5.6 Conclusion

We have seen how samples are connected and related in higher dimensions, this should enable us to be more efficient when using these load profiles in the next chapter. The comprehensive analysis of the plots pointed out that we can roughly estimate the strength of a routine based on the scattering of load profiles. This information will be useful in the next chapter where we will utilize it to build the elderly care system.

While there are no empirical results to present, seeing the actual samples placed on a map contributes to the understanding of the data we are working with. We could go into greater detail and use other dimensionality-reduction algorithms for comparison, but that was not the goal. The goal was to show and present the patterns in the data and other features such as routine strength. While doing so, we also proved that the load profiles that we proposed, can be efficiently used to do that.

Chapter 6

Elderly care demo

6.1 Introduction

Elderly care has been addressed by many EU-funded research projects since the aging population is one of the main issues the EU is facing. There are many solutions to this problem. One approach is invasive such as wearables, sound sensors, IR occupancy detectors, etc. This approach has been addressed by thousands of publications, such as reviews Blackman et al., 2016, Stavropoulos et al., 2020 and Baig et al., 2019 show and present.

Authors Belley et al., 2014 and Dai, Wang, and Meng, 2021 tried to solve this issue using a non-invasive approach with NILM algorithms. In the case of a non-invasive approach, no additional meters need to be installed, since per-appliance usage can be disaggregated. While this is practical from the "no additional equipment needed" side, it is a bit less practical from the efficiency and accuracy side, especially for larger buildings.

There is a middle way between invasive and non-invasive approaches, such as the authors explored in Visconti et al., 2019 and Patrono, Rametta, and Meis, 2018. It is possible to use sub-meters for each appliance and indirectly observe the usage pattern. The advantage of this approach is that the elder does not need to wear the device. The disadvantage is, that new meters need to be installed for the most commonly used appliances.

Our approach will use the latter.

6.2 Goal

The chapter will focus on building an elderly care system that will use users' periodic usage patterns to detect an anomaly. The anomaly could be anything from a fall, stroke or altered usage pattern due to dementia. The algorithm will be designed based on the load profile 4.11, which we discussed in chapter 4. Figure shows, that the first thing in the morning used are a kettle and toaster, and with a delay of one hour, microwave and TV. If none of these appliances are used within that hour, then that hour is considered anomalous. This means that the algorithm will be able to detect the anomaly within 1 hour of the accident.

6.3 Methodology

6.3.1 Defining an anomaly

Since the elderly care system is based on anomaly detection, we have to define it first. In our case, the anomaly occurs when something that should operate, does not.

Based on this definition we will develop an anomaly detection algorithm.

6.3.2 Building anomaly detection algorithm

The next section will present the steps taken while designing this algorithm.

Step one

To detect the anomalies one first needs to build a daily activation profile for each appliance, such as the one previously shown in Figure 4.11. In this specific case, we will be using 2h buckets, yielding a total of 12 buckets.

Step two

The second step is to ignore appliances that are always on by calculating the standard deviation of activations for each bucket. The activations are normalized between 0 and 1. This step is important so that appliances that are always on, such as fridges or freezers get ignored. These appliances are detected based on the width of their activation normal distribution. Periodic (on an hourly basis) appliances should have narrow distributions and the more dynamic should have wider distributions. This can be seen in examples from building 2.

866	842	810	772	868	854	859	854	864	887	889	883

FIGURE 6.1: Daily activations for fridge $\sigma = 0.036$

86	77	75	74	100	118	113	127	168	202	171	121

FIGURE 6.2: Daily activations for audio system $\sigma = 0.2$

9	4	4	3	212	122	85	80	102	260	134	48

FIGURE 6.3: Daily activations for microwave $\sigma = 0.3$

Based on results from all appliances a threshold of $\sigma = 0.1$ was set. This method will also get rid of appliances that are always on due to their specific nature such as server computers or fridges.

Third step

Next, appliances that trigger together must be grouped. This means we must find part of the day that they are operating together. Due to the filter in the previous step, we are left with appliances whose usage variate throughout the day. Some appliances are on even when the user is not necessarily using them, this can be seen in figure 6.2. One of many ways to do this is to normalize the activations, this yields a metric that tells us the probability of that appliance being turned on compared to the rest of the day. If we do this for the same appliances as above the result is the following:

0.43	0.38	0.37	0.37	0.5	0.58	0.56	0.63	0.83	1.	0.85	0.6

FIGURE 6.4: Daily activations for audio system $\sigma = 0.2$

0.03	0.02	0.02	0.01	0.82	0.47	0.33	0.31	0.39	1.	0.52	0.18

FIGURE 6.5: Daily activations for microwave $\sigma = 0.3$

Finally, a suitable threshold must be selected. The threshold of 0.5 was selected, which yields the following vectors:

0	0	0	0	0	1	1	1	1	1	1	1

FIGURE 6.6: Daily activations for audio system

0	0	0	0	1	0	0	0	0	1	1	0

FIGURE 6.7: Daily activations for microwave with one usage peak in the morning and the other in the evening

The vectors show us that the microwave has two usage peaks, where the audio system can be used anytime throughout the day. It is possible to do this for all appliances, which results in a 2D matrix. Using this matrix we can build rules for which appliances are being used together. The Figure 6.8 uses rows for appliances and columns for buckets. If we use terminology from image processing the matrix 6.8 is essentially a highly saturated load profile 4.11, which can be easily processed by computer algorithms due to binary encoding.

0	0	0	0	0	1	1	1	1	1	1	1
0	0	0	0	0	0	0	0	0	1	1	1
0	0	0	0	0	0	0	0	1	1	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	1	0	0	1	1	1	1	0
0	0	0	0	1	1	0	0	1	0	0	0
0	0	0	0	1	1	0	1	0	0	0	0
0	0	0	0	1	0	1	0	0	0	1	0
0	0	0	0	0	1	1	1	1	1	0	0

FIGURE 6.8: Activation matrix

It is possible to display the matrix 6.8 as an image. The Figure below shows how the load profile is transformed.

(A) Input data

(B) Figure 6.8 as an image

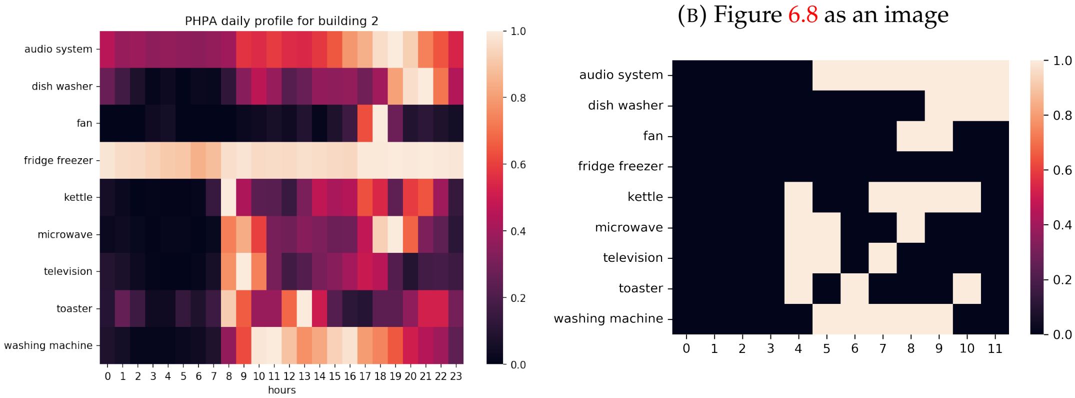


FIGURE 6.9: Transformation of source load profile to black and white

Step four

Previously, we have defined that an anomaly occurs when something that should activate does not. Using the matrix 6.8 we can compile an algorithm that will detect the anomaly using current activations being tested and comparing it to the adjacent column in matrix 6.8. Let us use the fifth bucket as an example. That is data from 8 to 10 o'clock.

The tested sample is considered normal if at least two appliances that are normally being used are activated. Otherwise, the tested sample is considered anomalous. Our implementation multiplies the adjacent matrix column to the tested sample. We sum the elements of the resulting array and check if it is larger or equal to 2. If cases where this rule is false, samples are considered anomalous.

	profile		sample		result	
audio system	0		1		0	
dish washer	0		0		0	
fan	0		0		0	
fridge freezer	0		0		0	
kettle	1	X	1	=	1	IF SUM >= 2 not an anomaly
microwave	1		0		0	
television	1		0		0	
toaster	1		1		1	
washing machine	0		0		0	

FIGURE 6.10: The evaluation of the test sample compared to the adjacent column from the matrix. An example is for a fifth bucket or fifth row from the matrix.

This process is done for all samples, where we count normal and anomalous samples for each bucket The important thing to note here is that we are evaluating the samples from train data, from which the profile was built.

472	469	468	466	57	153	288	187	123	84	75	281

FIGURE 6.11: Aggregated anomalies for each bucket

0	0	0	0	409	312	181	280	342	384	394	188

FIGURE 6.12: Aggregated normal samples for each bucket

Step five

The next step is to combine these two arrays so that we calculate the percentage of anomalous samples for each bucket with an equation.

$$\frac{N_{anom}}{N_{anom} + N_{norm}} \tag{6.1}$$

Where N_{anom} is a number of anomalous samples and N_{norm} is a number of normal samples.

We can alter the equation 6.1 so that it will measure a number of normal samples out of all. The result is the equation 6.2 In other words, we are measuring the strength of a routine that user has in each bucket.

$$R_{outine} = \frac{N_{norm}}{N_{anom} + N_{norm}} \tag{6.2}$$

Using the equation 6.2 we can populate the array 6.13.

0.0	0.0	0.0	0.0	0.88	0.7	0.39	0.6	0.74	0.82	0.84	0.4

FIGURE 6.13: Aggregated anomalies for each bucket

In other words, the array 6.13 tells us how persistent is the user's routine in each bucket or part of the day. The higher the metric the higher the routine. Since routine is detected based on the usage of appliances it cannot be picked up during the night.

It is possible to see that the routine is quite high during the morning and evening hours. The anomaly detection algorithm will work best when the metric above is high. A good trait of the elderly is that their routine is quite high even during the day.

One more thing to do is to ignore the parts of the day when the user has no routine. This is done by using the array 6.13 and setting a threshold of 0.7.

A threshold of 0.5 would mean that we could detect false positive anomalies every other day. Setting the rate to 0.7 reduces this to every third day. Here, compromises must be made, the lower the threshold the more accurate the algorithm will be. This also means that it will be less sensitive. In our case, there is not much harm in false positive detections, since the caregiver can call the elder to check if it is okay.

0	0	0	0	1	1	0	0	1	1	1	0

FIGURE 6.14: Using the above-mentioned threshold a new mask is made, to check only buckets with high routine.

Step six

The last step is to repeat steps 4 and 5 with test data. When using test data, we skip the buckets with low routine rates by using the mask on Figure 6.14. Since the profile has never seen the data being used, this should give us a good presentation of actual performance in a real-world scenario.

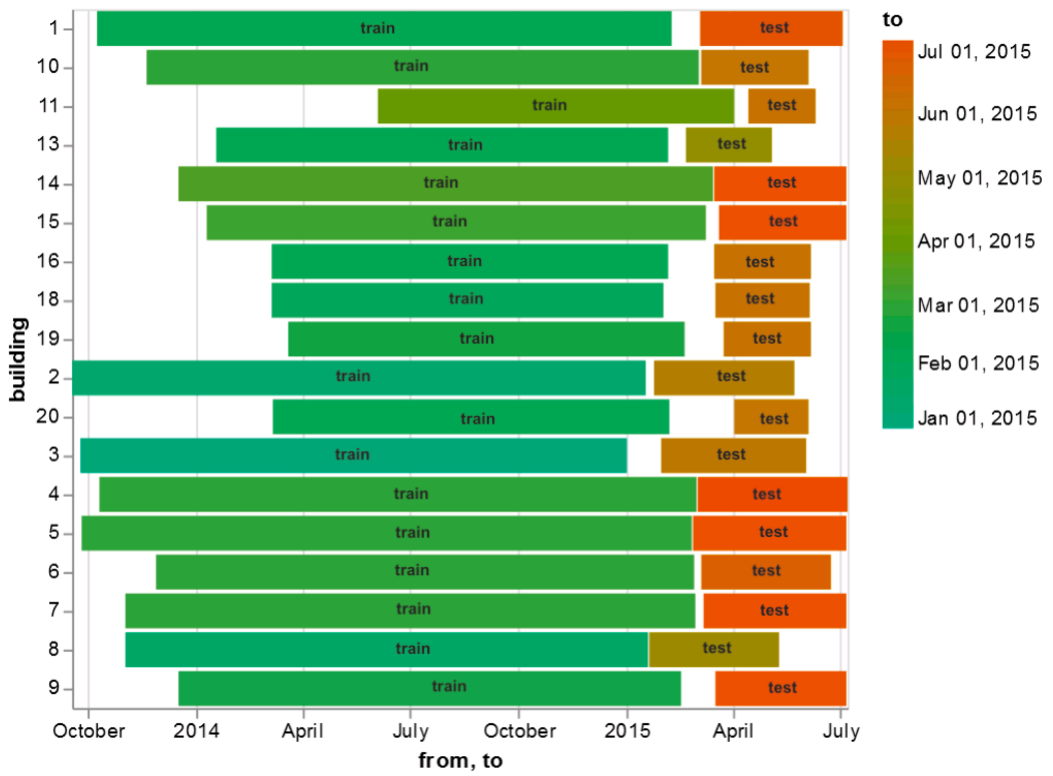
6.3.3 Datasets and evaluation

The data was split into train and test sets, where 80 % of the data was used for training and 20 % percent of the data for testing. The data was split based on the number of samples, so in some cases where there is a lot of missing data, the time window of test data might be longer, although it contains only 20 % of the samples.

REFIT

The REFIT Rashid, 2019 dataset included data for more than 15 buildings, as can be seen in the Figure below. The dataset in general is of the highest quality since it is the longest with the least missing data. This means this dataset should give the most relevant results.

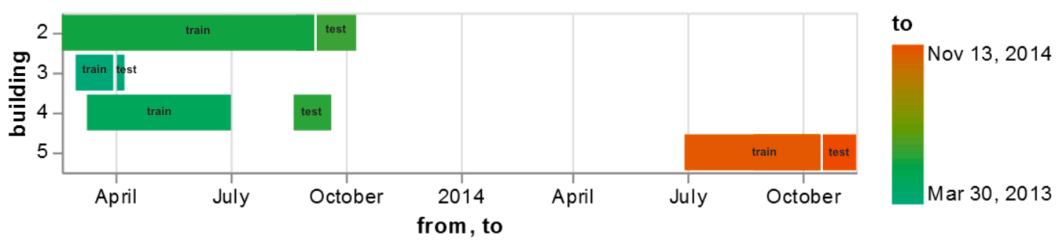
FIGURE 6.15: Timeline for REFIT



UK-DALE

Through the UK-DALE Kelly and Knottenbelt, 2015 dataset is of similar size, most of the data is from building 1. In general, it includes 5 years of data, but only for some appliances, where many appliances are rarely used. When taking all of this into account, there were too many issues with building 1, and it was simply ignored. Another issue that can be seen in Figure 6.16 is that there is not enough data for building 3. The test includes only a week of data, which is not enough for representative results, therefore it was ignored. The rest of the buildings seem healthy.

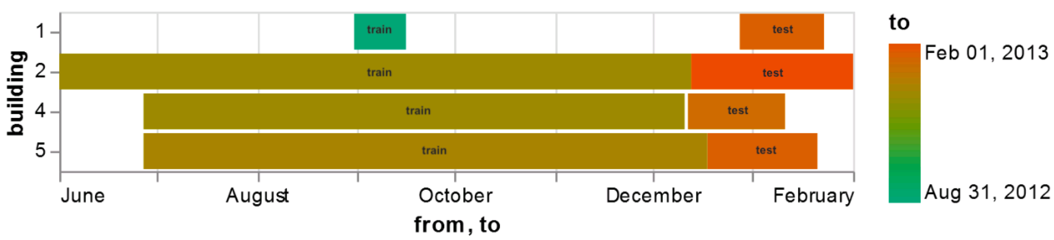
FIGURE 6.16: Timeline for UK-DALE



ECO

ECO Beckel et al., 2014 dataset has a length of data similar to UK-DALE. The only issue is building 1, where there is a lot of missing data. This is a good example of how data is split, it is split based on several samples, meaning that there is 80 % in the train bar, due to missing data the second bar is longer.

FIGURE 6.17: Timeline for ECO



6.3.4 The metric - routine rate

Due to the lack of ground truth data of actual accidents, it is hard to determine the exact accuracy of this algorithm. Every anomaly detected is not necessarily an actual accident, it could be that the user decided to lie in bed a bit longer, or decided to go to bed early in the evening. One metric that we can use to determine how well the algorithm functions is the routine rate metric 6.13. The reason behind that is, that if the routine rate is high it means that it will be easier to detect the actual anomaly.

- Routine rate of 0 would mean that for that bucket household has no routine at all.
- Routine rate of 0.5 would mean that the routine is broken every second day.
- Routine rate of 0.8 would mean that routine is broken on average every fifth day.
- Routine rate of 1 would mean that this household has a routine that is never broken.

An example of when a user routine rate is close to 1. When a true anomaly occurs such as a fall, the dweller, though he had the same strong routine for the past year, would not be able to practice it, and the algorithm will be quite sure that this is an actual anomaly. Therefore, the lower the routine rate the less sure we are that an actual anomaly such as a fall occurred. This is a good alternative measurement, that tells us how well this algorithm will perform. Since sometimes it is easier to read when results are presented with percentages, we will sometimes use this way of presenting it.

6.4 Results

Results were obtained for 3 datasets. REDD and iAWE datasets were not used, since they were too small. They contained less than a month of data.

6.4.1 The routine rate over a period of time

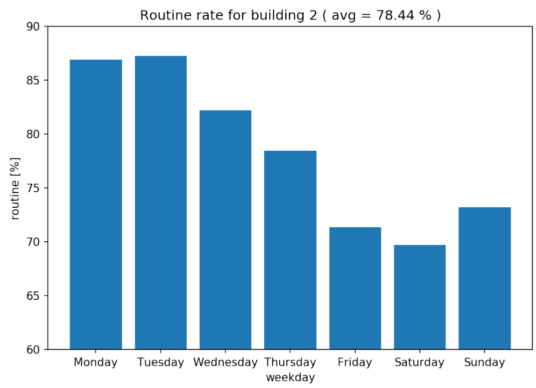
In the following sections, we will present how the metric changes over given periods of time. This will enable us to see that there are patterns that this metric helps reveal.

Since we have more than a year of training data, this will enable us to see how the metric changes over years. This enables us to see how routine changes over the year. We cannot use testing data in this case, since there is not enough of it.

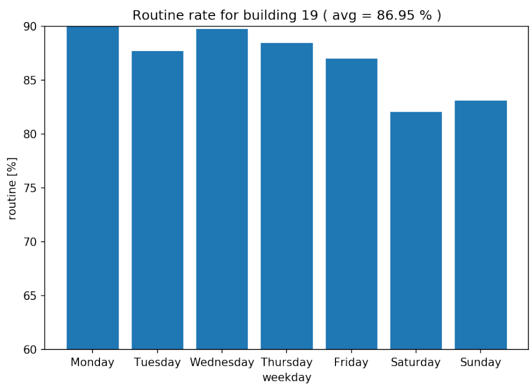
The routine rate through the week

As the behavior of the dweller changes, so does the accuracy of the algorithm. One observation that was made, was that the routine was higher during the week than during the weekends, as can be seen in the Figure 6.18 below. The only exception is Figure 6.18d, which shows that the observation does not hold for all houses.

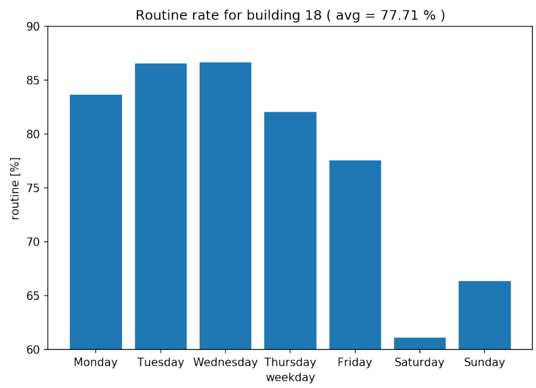
(A) Building 2



(B) Building 19



(C) Building 18



(D) Building 5

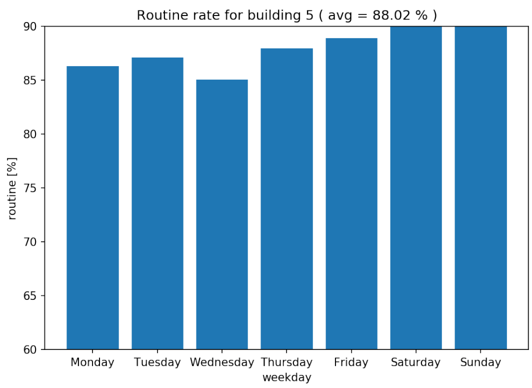


FIGURE 6.18: Routine rate through the week (train data)

Since we are dealing with the elderly, they have a higher routine, and it does not change that much during the weekends. Usually, assisted living systems are put in place since elders are alone in the dwelling. Taking all of this into account, we could assume that the routine of the elderly is the same trough the week and simply ignore the weekends. This should yield more relevant results.

Routine rate through a year

The rate at which the routine is being practiced also changes over a year. While on average the routine rate is higher during the winter, spring and fall, it is lower during the summer, due to vacation. This can be seen in Figure 6.19 below. It is possible to observe dips in routine. In some cases, these dips occur in summer and others in springtime. Without metadata, we cannot know for sure, what was the event behind these dips. There is a high chance most of them are vacations or other events where one or more dwellers are away from home for extended periods of time.

(A) Building 2 (B) Building 19

(C) Building 18 (D) Building 5

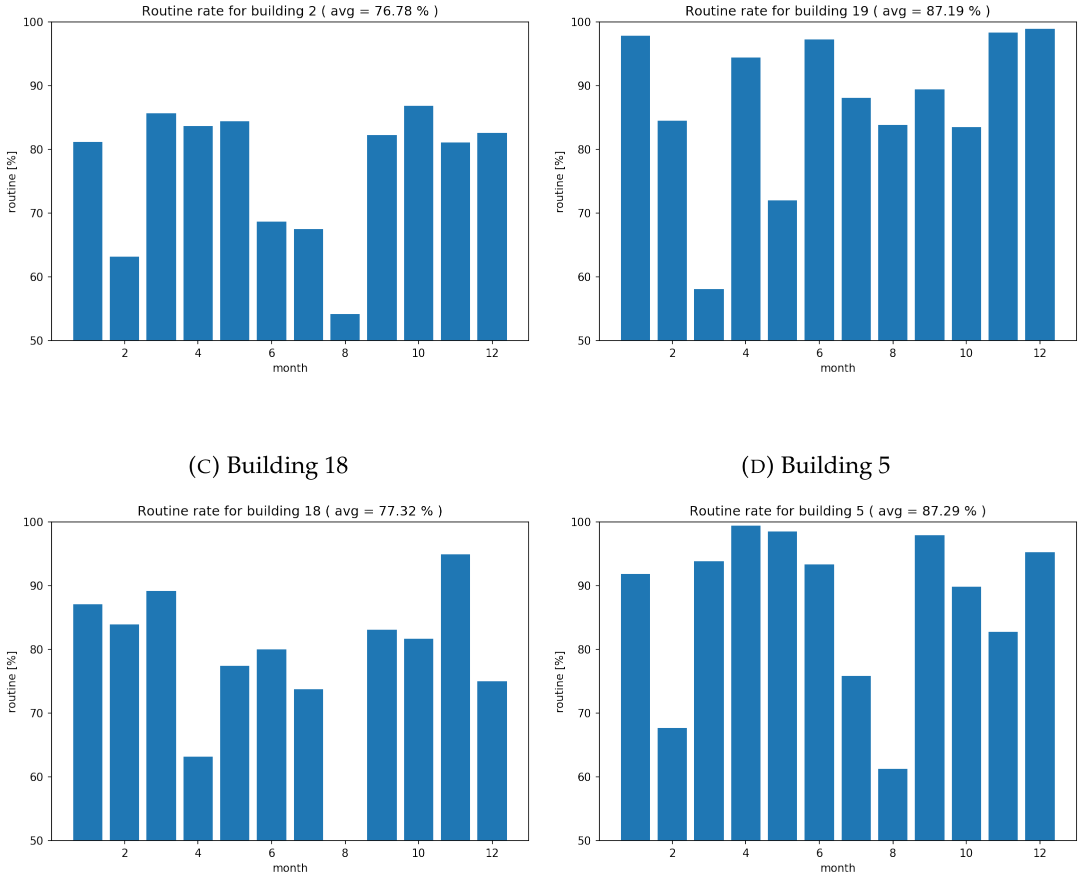


FIGURE 6.19: Routine through the year (train data)

Effectiveness of anomaly detection through the day

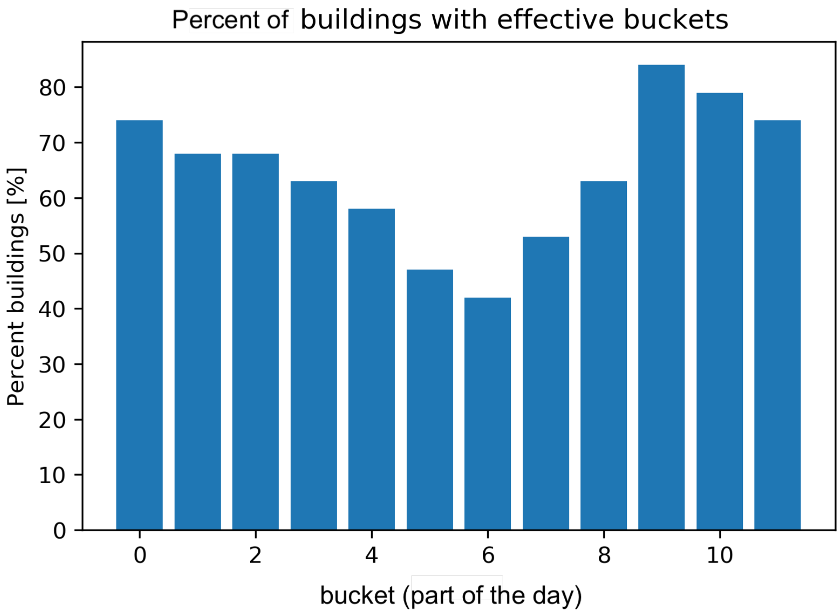
The following subsection will show how the effectiveness of anomaly detection changes throughout the day.

One thing to keep in mind is that this algorithm can detect anomalies only when the routine is high, and when more than two appliances are used in given buckets.

Figure 6.20 shows which buckets are most commonly used for the detection of an anomaly. The graph includes averaged values from all buildings and datasets. In other words, the graph presents how strong is average routine throughout the day.

This means that the higher the routine, the higher the chance that this bucket will be used for anomaly detection. During the night, it is possible to see that the average routine rate is quite high. This can be seen in Figure 6.20 this is because most users are routinely sleeping during that period. As we can see in Figure 6.20, the high routine rate does not necessarily mean the buckets are useful.

FIGURE 6.20: Effectivity of anomaly detection through the day



To find the usable buckets, an additional filter must be applied. The rule is that at least two appliances must be commonly used in that bucket. After applying this rule the following Figure emerges 6.21

FIGURE 6.21: Actual effectiveness of anomaly detection through the day

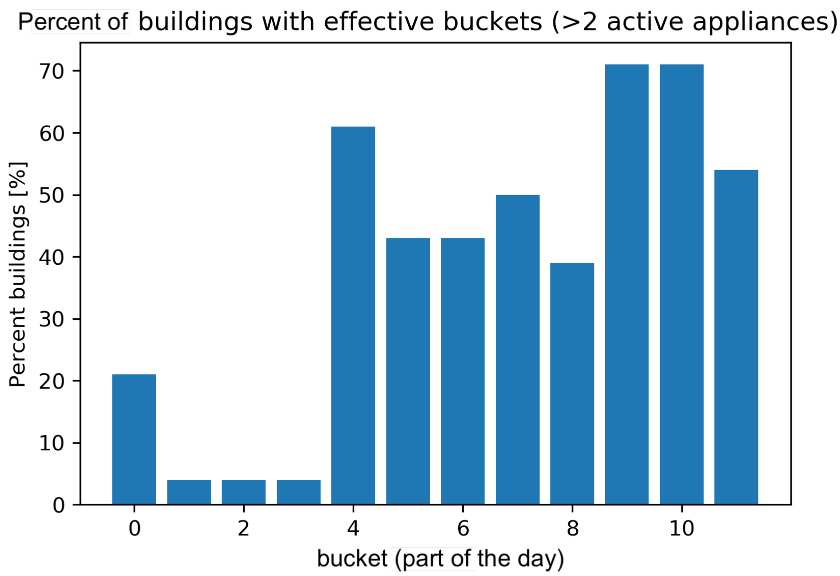


Figure 6.21 shows that there are two peaks. One in the morning and the other, a wider one, in the evening.

This means that on an average home the algorithm would perform best in the morning and evening because the average person is at school or work during noon. The elderly, are usually at home at noon, which could extend the effective detection window.

The anomaly detection during the night

We have seen that anomalies can be detected throughout the day, but are hard to detect through the night, since appliances are off.

This is because, in our current state, an anomaly occurs when something that should operate, does not. When the user is sleeping, an anomaly occurs when something that shouldn't operate, does. To implement this additional rule, we would have to build two models. One would be online during the day, and the other when the user is sleeping.

To obtain information about the user's sleep schedule, we could either have a schedule obtained from the user or we could extract it based on the usage pattern of appliances. It is possible to detect when most of the appliances are inactive and build a sleep profile based on this information.

Using the sleep schedule, it is possible to switch between the two operating modes. This new implementation would further extend the time windows within which we can detect the anomalies and thus further improve users' safety.

The main issue is not the detection itself but efficiently detecting when the user is sleeping.

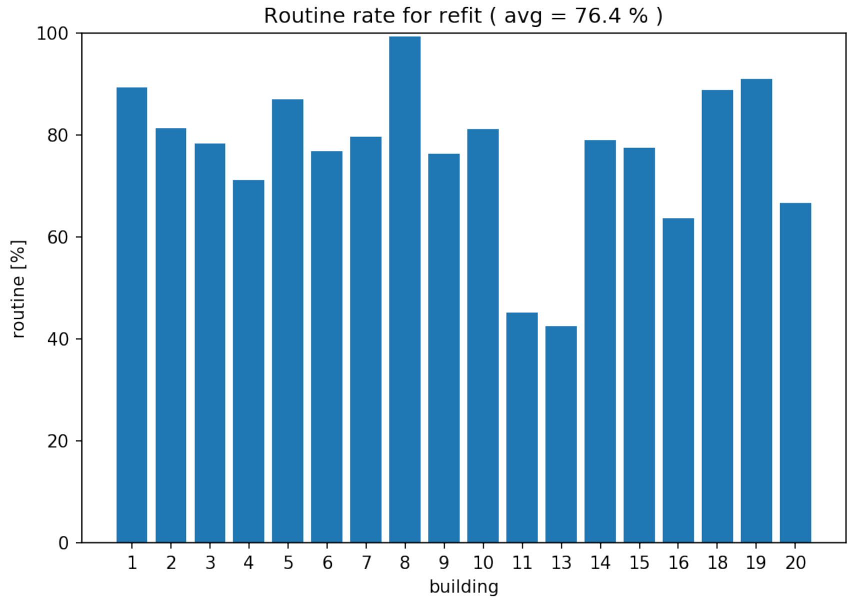
The examples above were a demonstration and a look into data and metrics. The examples shown were trained and evaluated on the same data. To show true performance, we will use test data to determine the actual performance.

6.4.2 Per-building results

REFIT

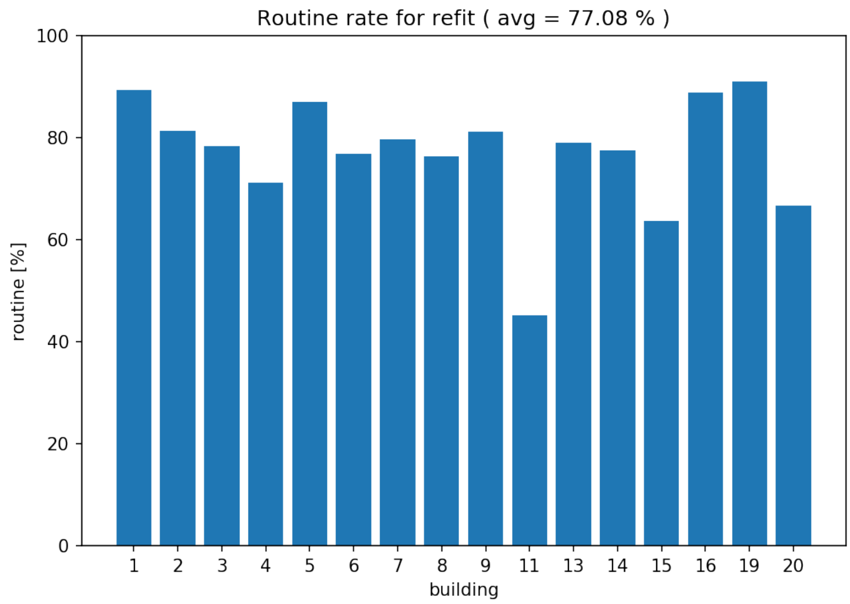
Results show, that the method is on average 76.4 % efficient for REFIT. In Figure 6.22 it is possible to see that building 8 yields much better results than the rest. Results show that the building reached a routine rate of almost 100 %, which is highly unlikely in the real world. On the contrary, buildings 11 and 13 performed much worse than the others with routine rates of around 40 %. It is hard to know the exact reasons why the buildings performed in such a way. This could be due to various dataset errors that occurred during sampling.

FIGURE 6.22: Results for REFIT



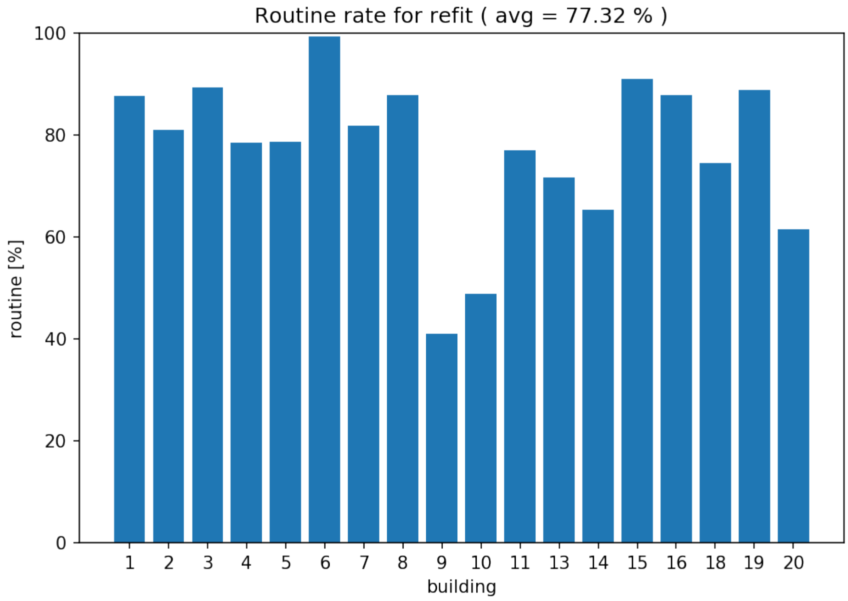
For more relevant results we can ignore the outliers by removing one maximum and minimum value, such as can be seen in Figure 6.23 This yields a result of 77.08 %. If we were to repeat this process the result would be 79.77 %. Since all outliers are removed, the result converges towards 79 %, which is the relevant value.

FIGURE 6.23: Results for REFIT with removed outliers



As mentioned in the sub-sub section 6.4.1, the average routine is different during the week and on weekends. The assumption was that the routine of elderly people does not change significantly over the week, therefore results should be more relevant if we ignore the weekends. The results in Figure 6.24 show that result improved to 77.08 %.

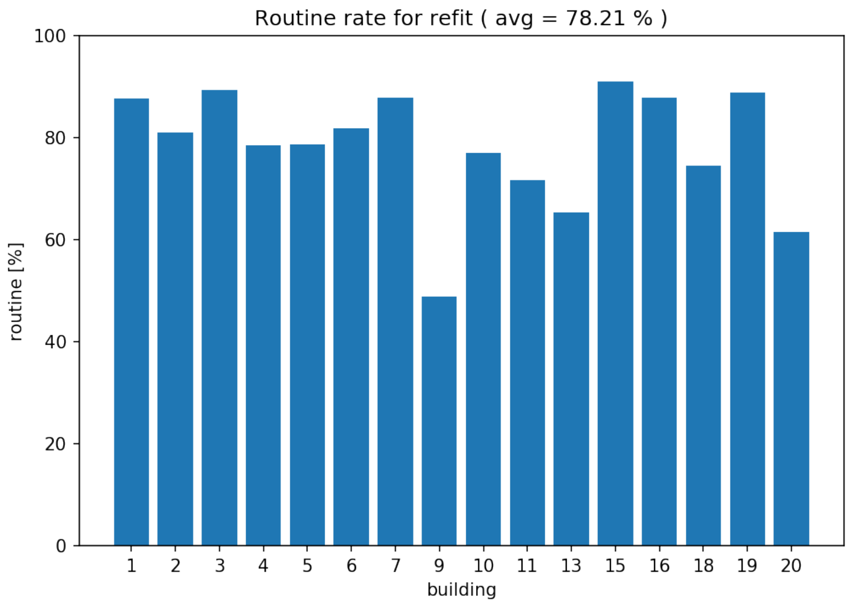
FIGURE 6.24: Results for REFIT weekday only



By ignoring the minimal and maximal outliers the results increase to 78.21 %. Repeating the process one more time the result increases to 80.20 %, since all outliers were removed, the result converges toward this value.

If we remove the weekend data, the results improved by 1.2 %.

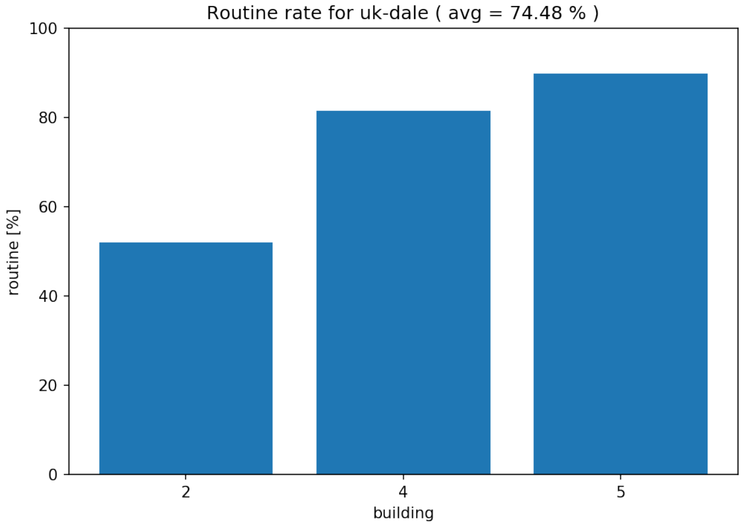
FIGURE 6.25: Results for REFIT weekday only and removed outliers



UK-DALE

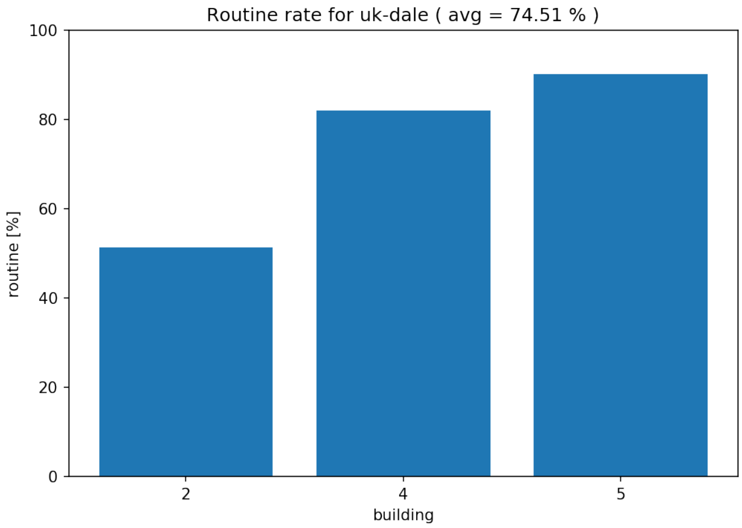
As mentioned in subsection 6.3.3, the UK-DALE is not as big and clean of errors as the previous dataset, so the results could be less relevant. The results in Figure 6.26, show that the average result is 74.48 %. Due to the low number of buildings, it is not possible to detect and ignore outliers.

FIGURE 6.26: Results for UK-DALE



The same as for REFIT, the weekend data can be ignored, In this case, this does not improve the result.

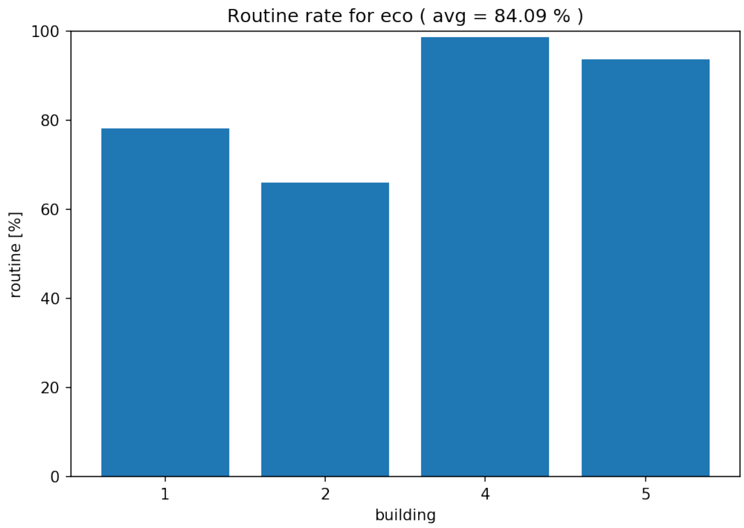
FIGURE 6.27: Results for UK-DALE omitting weekends



ECO

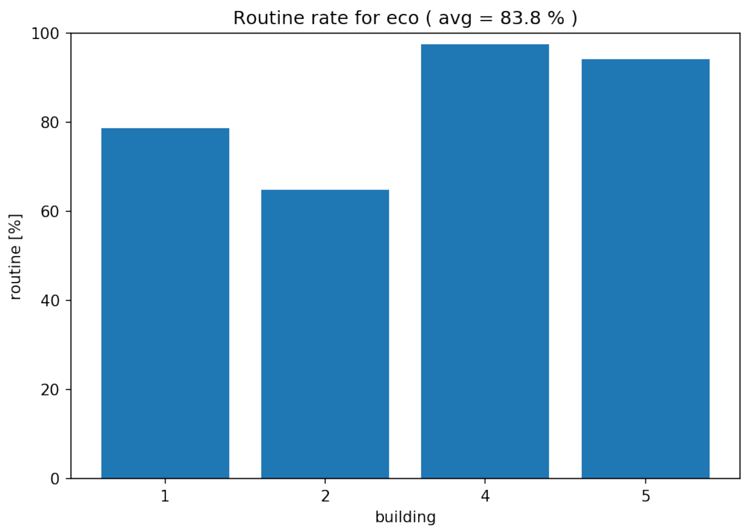
ECO is of a similar quality as UK-DALE when taking into account the number of buildings and the length of data, as can be seen in subsection 6.3.3. The results in Figure 6.28, show that this dataset performed the best, with results of 84.09 %.

FIGURE 6.28: Results for ECO



The same as before we can omit weekend data, which can be seen in Figure 6.29. This brings the result down to 83.80 %.

FIGURE 6.29: Results for ECO omitting weekends



6.4.3 Combined results

After combining results from all 25 buildings, Table 6.1 can be populated. The most relevant results can be seen in the last row.

TABLE 6.1: Combined percentage of anomalous samples [%]

N = 25	Including weekend data		Excluding weekend data	
Removed min/max outliers	train	test	train	test
0	84.73	77.35	86.20	78.07
1	84.63	77.91	86.16	78.75
2	86.53	78.53	86.13	79.23

Results show that the algorithm is 78 % efficient at detecting true anomalies. On average, the algorithm would label 22 % of samples as false positives, in other words, every fifth sample could be a false positive.

The nature of this system is that there is more harm done if we do not detect an anomaly than if we detect a few false positives. Therefore, we can claim that the performance of this algorithm is adequate to be used in the real world.

6.5 Discussion

When analyzing these results one important to keep in mind is, that we do not have metadata available to know what kind of socio-economic status dwellers have. Socio-economic status and other features encompass attributes such as age, income, number of children, geolocation, etc. They may also encompass the age of the building, type of insulation, number of dwellers in the buildings, etc. Since datasets do provide them, it is hard to make any other conclusions other than the algorithm works well on an average building.

We know that the reason for installing such a system is that the user is left alone. We can assume that on average there is more than one dweller living in the buildings we tested on. Since this system would usually be used by a single dweller, this would be in favor of our algorithm since it would be easier to extract the routine.

One other thing that would be in our favor is that the average person spends less time at home than an elderly person. If we take a look at the results, it is possible to see that, the average home has a low routine during the noon. This is because the average person is not at home during noon. This can be seen in Figure 6.20. Since the elderly are usually home at that time, this would increase the time windows where we can detect the accident.

We could also assume that the older the dweller, the higher the routine. The nature of the elderly is that they are more conservative when it comes to changes, and prefer to stick to their routine. Since the algorithm, works better when usage is periodic, this would also be in our favor.

Taking all of these assumptions into account, there is a possibility that this algorithm would work better on the elderly due to their nature.

Since the results on the average building are promising a test study should be performed. This would also prove our assumption that this algorithm works better on the elderly.

6.6 Iterative learning system

In the case of practical use of this algorithm, it is important the system is put online as fast as possible and that it improves over time.

This can be achieved with the implementation of iterative learning. The system will build a load profile based on the first month of data. Using this load profile, the system can be put online. At the end of the month, it can use this data to improve the load profile. This can then be repeated indefinitely.

6.6.1 Methodology

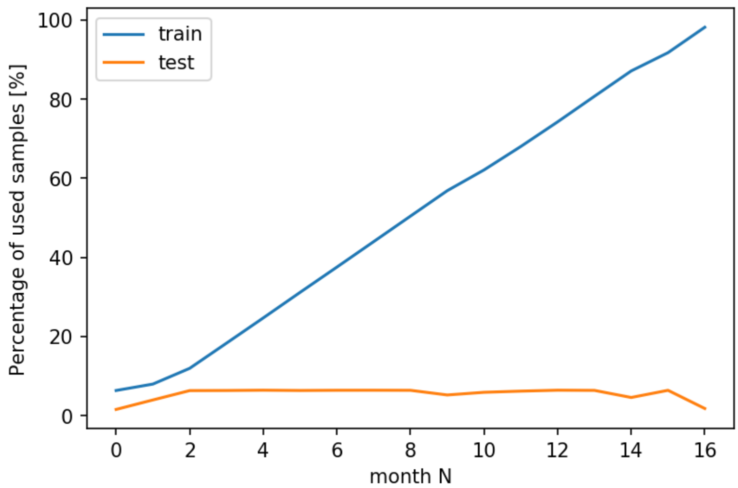
The tools, metric and other methodology is the same as in a normal learning system. The only change was made on the data preparation side.

Data preparation

For this evaluation, only REFIT (Rashid, 2019) data was used. As it can be seen in Figure 6.15, Refit buildings have long and relatively similar timelines, compared to other datasets.

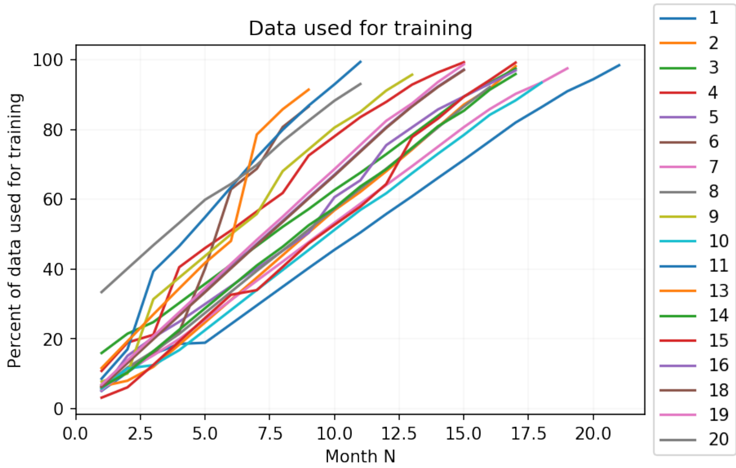
On Figure 6.30 it is possible to see, how the amount of training and testing data changes over 16 months.

FIGURE 6.30: Data for building 1 over 16 months



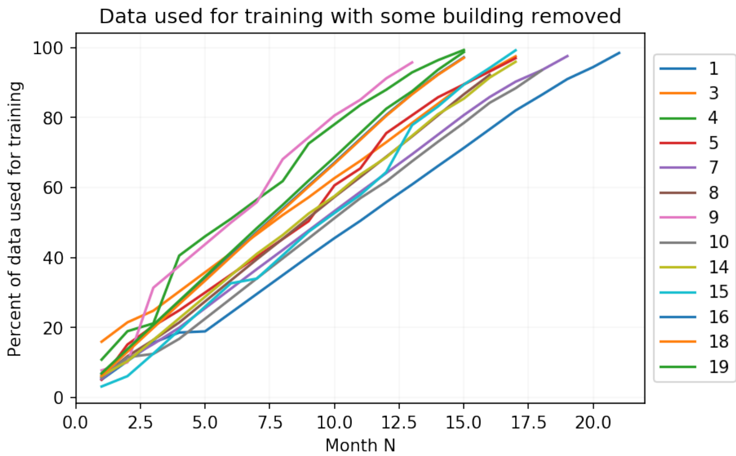
We can also plot how the amount of data changes for all buildings. This can be seen in Figure 6.31.

FIGURE 6.31: Data used for training



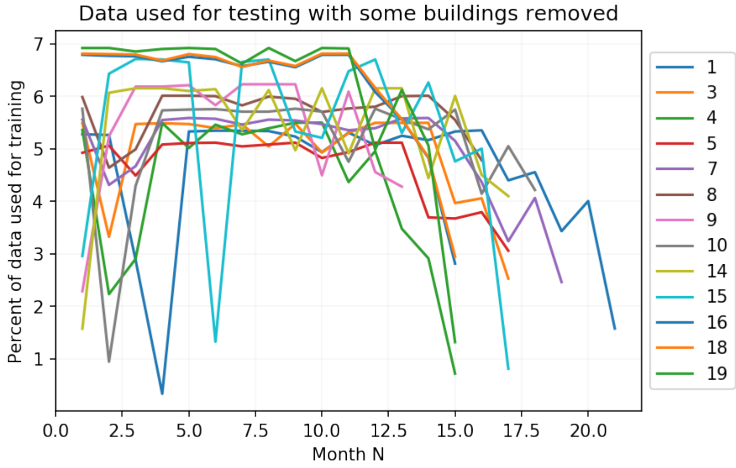
To analyze the results, at least 1 year of usable data should be available. Figure 6.32 shows only buildings containing at least one year of data.

FIGURE 6.32: Data used for training, with removed buildings



Similarly, we can check how test data changes over the months. In this case, data is not being aggregated, but only one month of it is used at a time. Figure 6.33 shows, that after one year the amount of data used for training starts to decline. To get more accurate results we will only observe the performance using one year of data.

FIGURE 6.33: Data used for training, with removed buildings



6.6.2 Results

To show the effect of training data on the metric, the Figure 6.34 is presented. The Figure 6.34 contains 12 months of data for each house.

FIGURE 6.34: Effect of new data on metric

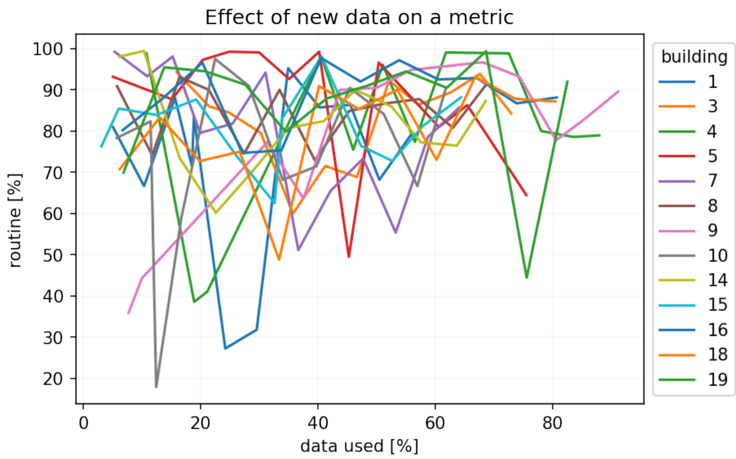


Figure 6.34 shows, that in most cases, results converge towards 80 %. In some cases, the results are good from the beginning, but sooner or later the routine rate will dip. With more data, these dips become smaller and less frequent. If the behavior in the household radically changes, it can still lead to a dip.

FIGURE 6.35: Metric over 12 months

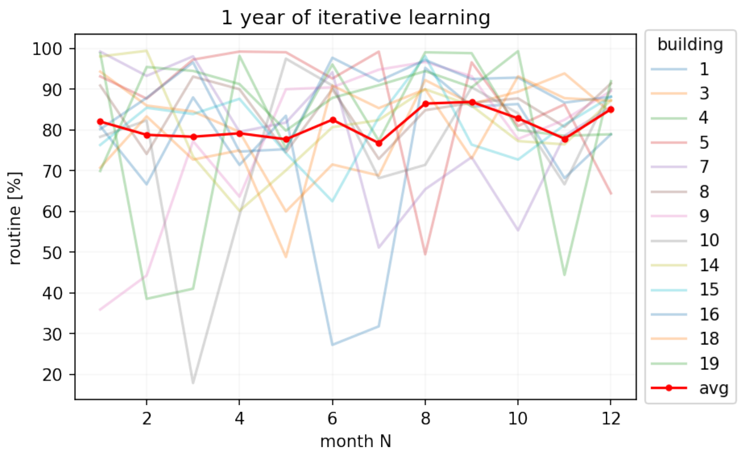


Figure 6.35 shows how the same data can also be presented so that it shows how the metric changes over a year. The same as in the previous Figure 6.34 we can observe the dips getting less frequent and smaller. Here we can also observe the average line. The average value seems to be on average at around 80 - 85 %,

6.6.3 Discussion

It is hard to compare these results to the ones from normal learning. Even though the same data was used, different sections of it were used.

Let's take the last point on the graph where the average is at the 85 % mark for an example. Here, the amount of training data is different, since we limited it to one year. The train set is also different since only last month was used, and not 20 %. There are many differences between train and test sets, therefore we can not compare them. The results do prove that the method works and that the true performance is at around the expected 80 %.

By increasing the amount of data, the algorithm becomes more stable. In some cases, where users' behavior does not change, the algorithm could work from the first month forward. In other cases, where behavior is more dynamic, the algorithm needs a month or two to stabilize.

An important thing to keep in mind is that the routine of the users does not increase with more data, our perceived one does.

6.6.4 Conclusion

The question that we tried to answer was: is this method good enough to be able to efficiently detect anomalies? The question that should be asked is, is the behavior of the users periodic enough, to be able to efficiently detect the anomaly? The answer is: yes, it is.

Chapter 7

Conclusion

In the introduction of chapter 1, it was said that the goal of the thesis will be achieved by contributing the following:

1. Surveying the state-of-the-art LPs (2)
2. Development of multidimensional activation LPs (ALP's) (4)
3. Visual analysis of ALP's (5)
4. Propose a new anomaly detection method for elderly care (6)

With the first contribution, we have found new, previously unused ways of presenting the data. This was achieved by building a detailed table of profiles such as we have seen in chapter 2. This table presented the missing gaps, and which presentations were not used by the community. We knew that not all unused profiles were useful, by using other publications we classified them based on their impact. We have selected the few with the highest impact and utilized used them in the following chapters.

Furthermore, we presented all the load profiles in high detail. This was done so that the reader was able to understand what the load profiles look like and what they present. While doing so, we pointed out how some profiles could be used, and how we will use them to prove that they are useful.

The third was contributed in chapter 5, where we have shown how data is connected in high-dimension space using t-SNE for dimensionality reduction. Here we have shown how some buildings have more similar activation patterns than others. Furthermore, we have shown which appliances are being used similarly. We have grouped the appliances into appliance groups and showed that appliances from different datasets are being used similarly, and how this method and groups can help us label unlabeled data. The formed clusters showed that a routine and persistent usage pattern does exist. This laid the groundwork for elderly care, where we have used this routine at the center of the algorithm.

The last was contributed in chapter 6, by building functioning elderly care assisted living system. The results proved that we successfully used one of the proposed load profiles in a real-world scenario. The main goal was to efficiently extract the routine, and build a working system around it. The results show that we have succeeded in doing so and that the algorithm is adequate to be used in the real world. To further prepare the algorithm for the real world, we have implemented an iterative learning system. The system could be put online a month after the installation of the system and continues to improve over time.

We could go into larger details and use other dimensionality reduction algorithms for comparison, or even have more empirical proofs. In the case of elderly care, we could use the results or algorithms of other publications to compare it to ours, or even compare it to the other intrusive methods. This could all be done in greater detail.

In the end, that was not the goal. The goal of the thesis was to prove that the proposed profiles can be efficiently utilized. By doing that, we have achieved the main goal of the thesis, that was "to propose suitable consumption profiles for supporting residential building consumption optimization and elderly care management". With that, we can conclude the thesis with the following words:

The ever-increasing amount of data is available to the scientific community. This data can be fully utilized if we find ways to efficiently extract the information that it is holding. The sole purpose of the load profiles is to reveal patterns, contextual features and information itself in the vast sea of data. With the proposed load profiles, we have hopefully contributed new tools that will help researchers to uncover the truths that the datasets hold. While we have filled in a few gaps in the table of profiles, it is up to scientists community to fill in the rest.

Appendix A

The source code, high-resolution figures and datasets

The following appendix contains mostly links that point to GitHub. GitHub should be a valid and safe repository for such projects, where links should persist indefinitely. In case the links do eventually break, you can find the repository of thesis and demos under the user name "jenkoj", under "msc" and "appliance-profiling". In case the Google Drive link breaks, send an email to jakob.jenko@gmail.com and I will try to forward the documents.

A.1 The source code

The source code used in this chapter can be found in a GitHub repository:
https://github.com/jenkoj/appliance-profiling

Individual scripts can be found in the following Jupyter Notebooks:

The source code for generating the figures can be found at:
https://github.com/jenkoj/appliance-profiling/blob/main/profilng_slices.ipynb

The source code for t-SNE can be found at:
https://github.com/jenkoj/appliance-profiling/blob/main/profilng_slices.ipynb

the source code for elderly care can be found at:
https://github.com/jenkoj/appliance-profiling/blob/main/elderly_care_demo.ipynb

A.2 High resolution figures

High-resolution figures can be found in the thesis repository https://github.com/jenkoj/msc.
More precisely in the "figures" subfolder https://github.com/jenkoj/msc/tree/main/Figures.

A.3 Data and datasets

We cannot share the data since it is not ours to share, but we can share the spreadsheet that points to each dataset. The spreadsheet includes other datasets that could be used for the very same purpose. The spreadsheet can be found here.

https://github.com/jenkoj/msc/blob/main/Appendices/datasets_and_sources.pdf

An overview of the data in the datasets was made, and it can be seen in the following spreadsheet.
https://github.com/jenkoj/msc/blob/main/Appendices/dataset_overview.pdf

The sliced hourly datasets can be found here.
https://drive.google.com/drive/folders/1tIsG-bqxoJdbU1p8xa_LCTaKNSk_LylZ?usp=sharing

Appendix B

Expanded General Table

TABLE B.1: Expanded general table of load profiles

	frequency	appliances	number of activations	power (avg)	operating time
appliances		X	X	X	X
number of activations	X	[14] [37]	X	X	X
power (avg)	X	[60]		X	X
power (array)	[37]	X	X	X	X
power (histogram)			X	X	X
operating time	X	[35]	[57] [56] [4]	[4]	X
time array	X	X	[14] [37]	[18] [21] [10] [36] [70] [26] [25] [34] [1] [38] [57] [56] [33] [4] [15] [39] [18] [13] [50] [37] [25]	[24]

Bibliography

1. Abreu, Joana M., Câmara Pereira, Francisco, and Ferrão, Paulo (2012). "Using pattern recognition to identify habitual behavior in residential electricity consumption". In: *Energy and Buildings* 49, pp. 479–487. ISSN: 0378-7788. DOI: https://doi.org/10.1016/j.enbuild.2012.02.044. URL: https://www.sciencedirect.com/science/article/pii/S0378778812001363.
2. Albadi, M.H. and El-Saadany, E.F. (2008). "A summary of demand response in electricity markets". In: *Electric Power Systems Research* 78.11, pp. 1989–1996. ISSN: 0378-7796. DOI: https://doi.org/10.1016/j.epsr.2008.04.002. URL: https://www.sciencedirect.com/science/article/pii/S0378779608001272.
3. Aréchiga, A. et al. (2016). "Comparison of Dimensionality Reduction Techniques for Clustering and Visualization of Load Profiles". In: *2016 IEEE PES Transmission and Distribution Conference and Exposition Latin America*, 1 to 6. DOI: 10.1109/TDC-LA.2016.7805661.
4. Azizi, Elnaz, Beheshti, Mohammad T. H., and Bolouki, Sadegh (Nov. 2021). "Appliance-Level Anomaly Detection in Nonintrusive Load Monitoring via Power Consumption-Based Feature Analysis". In: *IEEE Transactions on Consumer Electronics* 67.4, pp. 363–371. ISSN: 1558-4127. DOI: 10.1109/TCE.2021.3129356.
5. Baig, Mirza Mansoor et al. (June 2019). "A systematic review of wearable sensors and IoT-based monitoring applications for older adults - a focus on ageing population and independent living". en. In: *J. Med. Syst.* 43.8, p. 233.
6. Batra, N. et al. (2013). "It's Different: Insights into home energy consumption in India". In: *Proceedings of the 5th ACM Workshop on Embedded Systems For Energy-Efficient Buildings*, 1 to 8.
7. Beckel, Christian et al. (Nov. 2014). "The ECO Data Set and the Performance of Non-Intrusive Load Monitoring Algorithms". In: DOI: 10.1145/2674061.2674064.
8. Belley, Corinne et al. (2014). "An efficient and inexpensive method for activity recognition within a smart home based on load signatures of appliances". In: *Pervasive and Mobile Computing* 12, pp. 58–78. ISSN: 1574-1192. DOI: https://doi.org/10.1016/j.pmcj.2013.02.002. URL: https://www.sciencedirect.com/science/article/pii/S1574119213000370.
9. Bisong, Ekaba (2019). "Google Colaboratory". In: *Building Machine Learning and Deep Learning Models on Google Cloud Platform A Comprehensive Guide for Beginners*. Berkeley, CA: Apress, 59 to 64. ISBN: 978-1-4842-4470-8. DOI: 10.1007/978-1-4842-4470-8_7. URL: https://doi.org/10.1007/978-1-4842-4470-8_7.
10. Bitterer, R. and B. Schieferdecker, Prof. Dr. habil. (1999). *Repräsentative VDEW-Lastprofile Aktionsplan Wettbewerb, M-32 99*. VDEW.
11. Blackman, Stephanie et al. (2016). In: *Journal of Intelligent Systems* 25.1, pp. 55–69. DOI: doi:10.1515/jisys-2014-0136. URL: https://doi.org/10.1515/jisys-2014-0136.
12. Boßmann, Tobias, Schleich, Joachim, and Schurk, Robert (June 2015). "Unravelling load patterns of residential end-uses from smart meter data". In:

13. Capasso, A. et al. (May 1994). “A bottom-up approach to residential load modeling”. In: *IEEE Transactions on Power Systems* 9.2, pp. 957–964. ISSN: 1558-0679. DOI: 10.1109/59.317650.
14. Carpino, Cristina et al. (Jan. 2020). “Energy performance gap of a nearly Zero Energy Building in Denmark the influence of occupancy modelling”. In: *Building Research Information* 48.
15. Castangia, Marco et al. (2021). “Detection of Anomalies in Household Appliances from Disaggregated Load Consumption”. In: *2021 International Conference on Smart Energy Systems and Technologies (SEST)*, pp. 1–6. DOI: 10.1109/SEST50973.2021.9543232.
16. Cellura, Maurizio et al. (2013). “The role of the building sector for reducing energy consumption and greenhouse gases: An Italian case study”. In: *Renewable Energy* 60, pp. 586–597. ISSN: 0960-1481. DOI: https://doi.org/10.1016/j.renene.2013.06.019. URL: https://www.sciencedirect.com/science/article/pii/S0960148113003121.
17. Chen, Chen (2018). “4 - Demand response: An enabling technology to achieve energy efficiency in a smart grid”. In: *Application of Smart Grid Technologies*. Ed. by Lisa Ann Lamont and Ali Sayigh. Academic Press, pp. 143–171. ISBN: 978-0-12-803128-5. DOI: https://doi.org/10.1016/B978-0-12-803128-5.00004-0. URL: https://www.sciencedirect.com/science/article/pii/B9780128031285000040.
18. Chuan, Luo, Rao, D. M. K. K. Venkateswara, and Ukil, Abhisek (2014). “Load profiling of Singapore buildings for peak shaving”. In: *2014 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC)*, pp. 1–6. DOI: 10.1109/APPEEC.2014.7065998.
19. Collette, Andrew (2013). *Python and HDF5*. O’Reilly.
20. Commission, Communication from the et al. (2006). “Action Plan for energy efficiency: realising the potential”. In: *COM (2006)* 545.
21. Csoknyai, Tamás et al. (2019). “Analysis of energy consumption profiles in residential buildings and impact assessment of a serious game on occupants’ behavior”. In: *Energy and Buildings* 196, pp. 1–20. ISSN: 0378-7788. DOI: https://doi.org/10.1016/j.enbuild.2019.05.009. URL: https://www.sciencedirect.com/science/article/pii/S0378778818334790.
22. Dai, Shuang, Wang, Qian, and Meng, Fanlin (2021). “A telehealth framework for dementia care an ADLs patterns recognition model for patients based on NILM”. In: *2021 International Joint Conference on Neural Networks*, pp. 1–8. DOI: 10.1109/IJCNN52387.2021.9534058.
23. Eurostat (2022). “Gross and net production of electricity and derived heat by type of plant and operator”. In: 62. URL: https://ec.europa.eu/eurostat/databrowser/view/nrg_ind_peh/default/table?lang=en.
24. Fischer, David, Härtl, Andreas, and Wille-Haussmann, Bernhard (2015). “Model for electric load profiles with high time resolution for German households”. In: *Energy and Buildings* 92, pp. 170–179. ISSN: 0378-7788. DOI: https://doi.org/10.1016/j.enbuild.2015.01.058. URL: https://www.sciencedirect.com/science/article/pii/S0378778815000845.
25. Gao, Bingtuan, Liu, Xiaofeng, and Zhu, Zhenyu (2018). “A Bottom-Up Model for Household Load Profile Based on the Consumption Behavior of Residents”. In: *Energies* 11.8. ISSN: 1996-1073. DOI: 10.3390/en11082112. URL: https://www.mdpi.com/1996-1073/11/8/2112.

26. Gerbec, D. et al. (May 2005). "Allocation of the load profiles to consumers using probabilistic neural networks". In: *IEEE Transactions on Power Systems* 20.2, pp. 548–555. ISSN: 1558-0679. DOI: 10.1109/TPWRS.2005.846236.
27. Graditi, Giorgio et al. (Feb. 2015). "Heuristic-Based Shiftable Loads Optimal Management in Smart Micro-Grids". In: *IEEE Transactions on Industrial Informatics* 11.1, pp. 271–280. ISSN: 1941-0050. DOI: 10.1109/TII.2014.2331000.
28. Harris, Charles R. et al. (2020). "Array programming with NumPy". In: *Nature* 585.7825, 357 to 362. DOI: 10.1038/s41586-020-2649-2. URL: https://doi.org/10.1038/s41586-020-2649-2.
29. Himeur, Yassine et al. (2021). "Artificial intelligence based anomaly detection of energy consumption in buildings: A review, current trends and new perspectives". In: *Applied Energy* 287, p. 116601. ISSN: 0306-2619. DOI: https://doi.org/10.1016/j.apenergy.2021.116601. URL: https://www.sciencedirect.com/science/article/pii/S0306261921001409.
30. Hinton, Geoffrey and Roweis, Sam (2003). "Stochastic Neighbor Embedding". In: *Advances in neural information processing systems* 15. Ed. by S Thrun S Becker and KEditors Obermayer, 833 to 840.
31. Hledik, Ryan and Lee, Tony (2021). "Chapter 9 - Load flexibility: Market potential and opportunities in the United States". In: *Variable Generation, Flexible Demand*. Ed. by Fereidoon Sioshansi. Academic Press, pp. 195–210. ISBN: 978-0-12-823810-3. DOI: https://doi.org/10.1016/B978-0-12-823810-3.00001-7. URL: https://www.sciencedirect.com/science/article/pii/B9780128238103000017.
32. Hunter, J. D. (2007). "Matplotlib: A 2D graphics environment". In: *Computing in Science and Engineering* 9.3, 90 to 95. DOI: 10.1109/MCSE.2007.55.
33. Issi, Fatih and Kaplan, Orhan (2018). "The Determination of Load Profiles and Power Consumptions of Home Appliances". In: *Energies* 11.3. ISSN: 1996-1073. DOI: 10.3390/en11030607. URL: https://www.mdpi.com/1996-1073/11/3/607.
34. Jeong, Hyun Cheol et al. (2021). "Clustering of Load Profiles of Residential Customers Using Extreme Points and Demographic Characteristics". In: *Electronics* 10.3. ISSN: 2079-9292. DOI: 10.3390/electronics10030290. URL: https://www.mdpi.com/2079-9292/10/3/290.
35. Kattan, Emily, Halasah, Suleiman, and hamed, Tareq a (May 2018). "Practical Challenges of Photovoltaic Systems in the Rural Bedouin Villages in the Negev". In: *Journal of Fundamentals of Renewable Energy and Applications* 8. DOI: 10.4172/20904541.1000258.
36. Kavousian, Amir, Rajagopal, Ram, and Fischer, Martin (2013). "Determinants of residential electricity consumption: Using smart meter data to examine the effect of climate, building characteristics, appliance stock, and occupants' behavior". In: *Energy* 55, pp. 184–194. ISSN: 0360-5442. DOI: https://doi.org/10.1016/j.energy.2013.03.086. URL: https://www.sciencedirect.com/science/article/pii/S0360544213002831.
37. Kelly, Jack and Knottenbelt, William (Mar. 2015). "The UK-DALE dataset, domestic appliance-level electricity demand and whole-house demand from five UK homes". In: *Scientific Data* 2.1, p. 150007. ISSN: 2052-4463. DOI: 10.1038/sdata.2015.7. URL: https://doi.org/10.1038/sdata.2015.7.
38. Klaassen, Elke, Frunt, Jasper, and Slootweg, Han (June 2015). "Assessing the impact of distributed energy resources on LV grids using practical measurements". In:
39. Kleiminger, Wilhelm et al. (Nov. 2013). "Occupancy Detection from Electricity Consumption Data". In: pp. 1–8. DOI: 10.1145/2528282.2528295.

40. Kluyver, Thomas et al. (2016). "Jupyter Notebooks, a publishing format for reproducible computational workflows". In: *Positioning and Power in Academic Publishing: Players, Agents and Agendas*. Ed. by F. Loizides and B. Schmidt. IOS Press, 87 to 90.
41. Kolter, J. Zico and Johnson, Matthew J. (2011). "REDD: A Public Data Set for Energy Disaggregation Research". In: *IN SUSTKDD*.
42. Li, Congmiao, Srinivasan, Dipti, and Reindl, Thomas (Nov. 2015). "Real-time scheduling of time-shiftable loads in smart grid with dynamic pricing and photovoltaic power generation". In: *2015 IEEE Innovative Smart Grid Technologies - Asia (ISGT ASIA)*, pp. 1–6. DOI: 10.1109/ISGT-Asia.2015.7387165.
43. Lopes, J.A.P., Moreira, C.L., and Madureira, A.G. (May 2006). "Defining control strategies for MicroGrids islanded operation". In: *IEEE Transactions on Power Systems* 21.2, pp. 916–924. ISSN: 1558-0679. DOI: 10.1109/TPWRS.2006.873018.
44. Maaten, L.J.P. van der and Hinton, G.E. (2008). "Visualizing High-Dimensional Data Using tSNE". In:
45. McKinney, Wes et al. (2010). "Data structures for statistical computing in python". In: *Proceedings of the 9th Python in Science Conference*. Vol. 445. Austin, TX, 51 to 56.
46. Mehrjerdi, Hasan and Hemmati, Reza (2020). "Coordination of vehicle-to-home and renewable capacity resources for energy management in resilience and self-healing building". In: *Renewable Energy* 146, pp. 568–579. ISSN: 0960-1481. DOI: https://doi.org/10.1016/j.renene.2019.07.004. URL: https://www.sciencedirect.com/science/article/pii/S0960148119310249.
47. Moreno Jaramillo, Andres F. et al. (2021). "Load modelling and non-intrusive load monitoring to integrate distributed energy resources in low and medium voltage networks". In: *Renewable Energy* 179, pp. 445–466. ISSN: 0960-1481. DOI: https://doi.org/10.1016/j.renene.2021.07.056. URL: https://www.sciencedirect.com/science/article/pii/S0960148121010612.
48. Moslehi, Khosrow and Kumar, Ranjit (June 2010). "A Reliability Perspective of the Smart Grid". In: *IEEE Transactions on Smart Grid* 1.1, pp. 57–64. ISSN: 1949-3061. DOI: 10.1109/TSG.2010.2046346.
49. Parag, Yael and Sovacool, Benjamin K. (Mar. 2016). "Electricity market design for the prosumer era". In: *Nature Energy* 1.4, p. 16032. ISSN: 2058-7546. DOI: 10.1038/nenergy.2016.32. URL: https://doi.org/10.1038/nenergy.2016.32.
50. Park, Keon-Jun and Son, Sung-Yong (2019). "A Novel Load Image Profile-Based Electricity Load Clustering Methodology". In: *IEEE Access* 7, pp. 59048–59058. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2019.2914216.
51. Parliament, The European and Council of the European Union, the (2021). "Directive 2010/31/EU of the European Parliament and of the Council of 19 May 2010 on the energy performance of buildings (recast)". In:
52. Patrono, Luigi, Rametta, Piercosimo, and Meis, Jochen (June 2018). "Unobtrusive Detection of Home Appliance's Usage for Elderly Monitoring". In: *2018 3rd International Conference on Smart and Sustainable Technologies (SpliTech)*, pp. 1–6.
53. Pedregosa, F. et al. (2011). "Scikit learn: Machine Learning in Python". In: *Journal of Machine Learning Research* 12, 2825 to 2830.
54. Proedrou, Elisavet (2021). "A Comprehensive Review of Residential Electricity Load Profile Models". In: *IEEE Access* 9, pp. 12114–12133. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2021.3050074.
55. Rashid, Haroon (2019). *Annotated load anomalies from the REFIT Dataset*. DOI: 10.15129/9729A2A0-11CE-4CCE-B0D0-144C483FCB33. URL: https://pureportal.strath.ac.uk/en/datasets/9729a2a0-11ce-4cce-b0d0-144c483fcb33.

56. Rashid, Haroon, Singh, Pushpendra, et al. (2019). "Can non-intrusive load monitoring be used for identifying an appliances anomalous behaviour?" In: *Applied Energy* 238, pp. 796–805. ISSN: 0306-2619. DOI: https://doi.org/10.1016/j.apenergy.2019.01.061. URL: https://www.sciencedirect.com/science/article/pii/S0306261919300613.
57. Rashid, Haroon, Stankovic, Vladimir, et al. (May 2019). "Evaluation of Non-intrusive Load Monitoring Algorithms for Appliance-level Anomaly Detection". In: *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 8325–8329. DOI: 10.1109/ICASSP.2019.8683792.
58. Robledo, Carla B. et al. (2018). "Integrating a hydrogen fuel cell electric vehicle with vehicle-to-grid technology, photovoltaic power and a residential building". In: *Applied Energy* 215, pp. 615–629. ISSN: 0306-2619. DOI: https://doi.org/10.1016/j.apenergy.2018.02.038. URL: https://www.sciencedirect.com/science/article/pii/S0306261918301636.
59. Shen, Jingshuang, Jiang, Chuanwen, and Li, Bosong (2015). "Controllable Load Management Approaches in Smart Grids". In: *Energies* 8.10, pp. 11187–11202. ISSN: 1996-1073. DOI: 10.3390/en81011187. URL: https://www.mdpi.com/1996-1073/8/10/11187.
60. shuma-iwisi, Mercy (June 2009). "Domestic appliances standby power losses: The case of eleven suburbs in the greater Johannesburg," in:
61. Sidqi, Yousra et al. (Oct. 2020). "Flexibility quantification in households: a swiss case study". In: *Energy Informatics* 3. DOI: 10.1186/s42162-020-00126-4.
62. Spataru, Catalina and Gauthier, Stephanie (2014). "How to monitor people 'smartly' to help reducing energy consumption in buildings?" In: *Architectural Engineering and Design Management* 10.1-2, pp. 60–78. DOI: 10.1080/17452007.2013.837248. eprint: https://doi.org/10.1080/17452007.2013.837248. URL: https://doi.org/10.1080/17452007.2013.837248.
63. Stavropoulos, Thanos G et al. (May 2020). "IoT wearable sensors and devices in elderly care A literature review". en. In: *Sensors (Basel)* 20.10, p. 2826.
64. Swan, Lukas G. and Ugursal, V. Ismet (2009). "Modeling of end-use energy consumption in the residential sector: A review of modeling techniques". In: *Renewable and Sustainable Energy Reviews* 13.8, pp. 1819–1835. ISSN: 1364-0321. DOI: https://doi.org/10.1016/j.rser.2008.09.033. URL: https://www.sciencedirect.com/science/article/pii/S1364032108001949.
65. Train, Kenneth, Herriges, Joseph, and Windle, Robert (1985). "Statistically adjusted engineering (SAE) models of end-use load curves". In: *Energy* 10.10, pp. 1103–1111. ISSN: 0360-5442. DOI: https://doi.org/10.1016/0360-5442(85)90025-8. URL: https://www.sciencedirect.com/science/article/pii/0360544285900258.
66. Van Rossum, Guido (2020). *The Python Library Reference*. Python Software Foundation.
67. Verbong, Geert P.J., Beemsterboer, Sjouke, and Sengers, Frans (2013). "Smart grids or smart users? Involving users in developing a low carbon electricity economy". In: *Energy Policy* 52. Special Section: Transition Pathways to a Low Carbon Economy, pp. 117–125. ISSN: 0301-4215. DOI: https://doi.org/10.1016/j.enpol.2012.05.003. URL: https://www.sciencedirect.com/science/article/pii/S0301421512004004.
68. Virtanen, Pauli et al. (2020). "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python". In: *Nature Methods* 17, 261 to 272. DOI: 10.1038/s41592-019-0686-2.

69. Visconti, Paolo et al. (June 2019). "A sensors-based monitoring system of electrical consumptions and home parameters remotely managed by mobile app for elderly habits' control". In: *2019 IEEE 8th International Workshop on Advances in Sensors and Interfaces (IWASI)*, pp. 264–269. DOI: 10.1109/IWASI.2019.8791399.
70. Walker, C.F. and Pokoski, J.L. (1985). "Residential Load Shape Modelling Based on Customer Behavior". In: *IEEE Transactions on Power Apparatus and Systems* PAS-104.7, pp. 1703–1711. DOI: 10.1109/TPAS.1985.319202.
71. Waskom, Michael L. (2021). "seaborn: statistical data visualization". In: *Journal of Open Source Software* 6.60, p. 3021. DOI: 10.21105/joss.03021. URL: https://doi.org/10.21105/joss.03021.
72. Yip, Sook-Chin et al. (2018). "An anomaly detection framework for identifying energy theft and defective meters in smart grids". In: *International Journal of Electrical Power and Energy Systems* 101, pp. 189–203. ISSN: 0142-0615. DOI: https://doi.org/10.1016/j.ijepes.2018.03.025. URL: https://www.sciencedirect.com/science/article/pii/S0142061517318719.
73. Zakariazadeh, Alireza et al. (2014). "A new approach for real time voltage control using demand response in an automated distribution system". In: *Applied Energy* 117, pp. 157–166. ISSN: 0306-2619. DOI: https://doi.org/10.1016/j.apenergy.2013.12.004. URL: https://www.sciencedirect.com/science/article/pii/S0306261913009884.