UNIVERSITY OF LJUBLJANA

MASTERS THESIS

Load Profiling of Home Appliances Using Load Classification

Author:
Jakob JENKO

Supervisor:
Dr. Marko MEŽA and Dr. Carolina FORTUNA

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

in the

-
ICT

June 21, 2022

Declaration of Authorship

I, Jakob JENKO, declare that this thesis titled, "Load Profiling of Home Appliances Using Load Classification" 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:

Chapter 1

Introduction

Climate change calls for a shift to renewable energy and restructuring of the electricity sector. 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 increase energy consumption. (elevate this issue) Acknowledging that, reducing consumption in that 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 and load monitoring 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 load profiles 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 load profiles have changed due to renewable energy accelerated development of distributed energy resources such as residential photovoltaic power plants, home wind energy, and using EVs and home batteries. Socioeconomic changes such as work-from-home, also drastically reshaped the load profile curve.

Technology advancements in non-intrusive load monitoring and increased adoption of smart energy meters offer a new way of load profiling, that is NILM load profiling.

Chapter 2

Definitions

Author Proedrou, 2021 defines terms as following

- Residential: private residences, with no commercial usage, occupied by one or more persons either full-time or part-time during a calendar year.
- 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.
- Model: "a formal system that represents the combined processes" Kavousian, Rajagopal, and M. Fischer, 2013 of electricity consumption by all the electricity-powered devices in a private residence/number of residences.

Commonly load profile is a term defined as aggregated power usage of all appliances in a house. Sometimes load profile is used to describe appliance-level load profiles.

The load profile is most commonly presented with a curve, that shows daily power usage.

FIGURE 2.1: "Clustered load profiles. The graph was published by Gerbec et al., 2005"

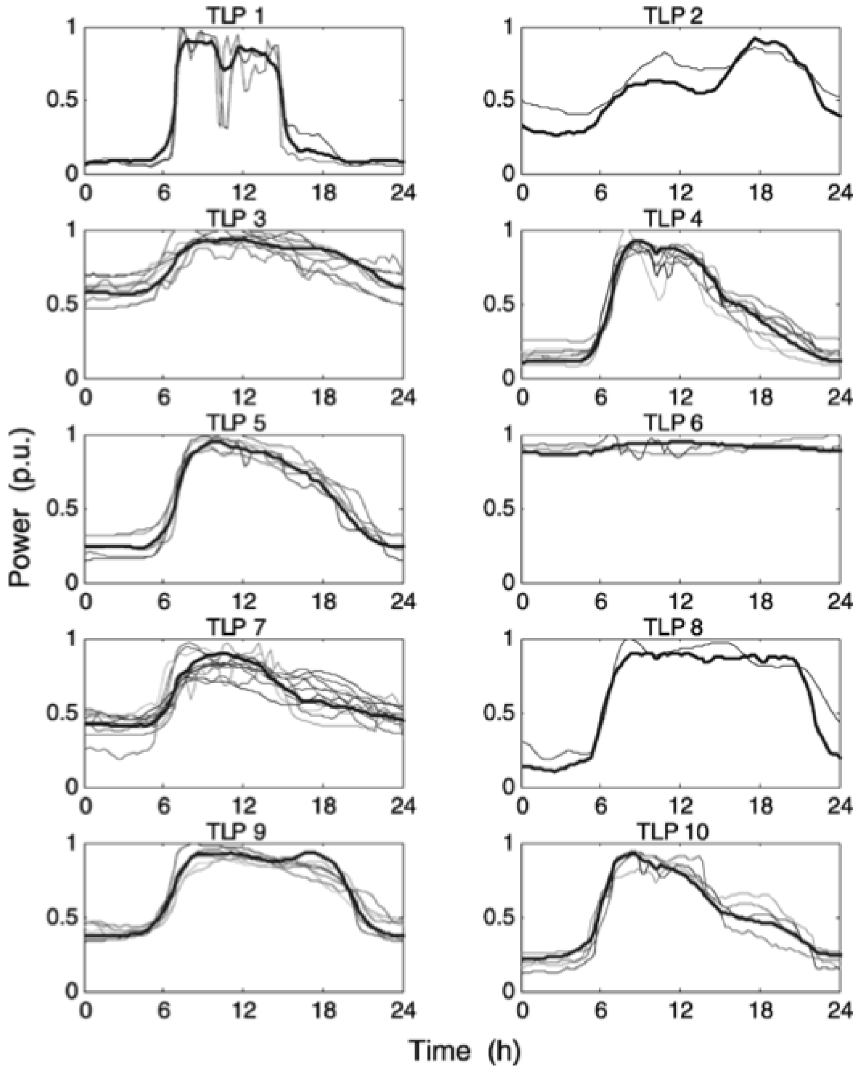


Figure 2.1 depicts 10 clusters of daily load profiles. This is not the only way to present it, for example, author Park and Son, 2019 used an image-based presentation.

Chapter 3

Related work

Work relating to load profiling can be found in two research verticals or topics. The first one is load profiling and load profile models, which in most cases study the load profile curve of the building. Few exceptions study load profiles on appliance-level. The second vertical is anomaly detection in building energy consumption data. While the first topic is closer, there are quite a few connections with the latter. If one wants to do anomaly detection, in some cases, one must first build some kind of "normal consumption profile"

3.1 Load profiling

Load profiling has been researched since 1980. Load-profiling can be performed in two ways: bottom-up and top-down.

A bottom-up approach as Swan and Ugursal, 2009 state "calculates the individual dwelling energy or electricity consumption and extrapolate these results over a target area or region" Whereas with Top-down approach as Swan and Ugursal, 2009 state "uses the total energy or electricity consumption estimates to assign them to the characteristics of the building stock" In other words, Bottom-up sub-meter data, Top-down uses aggregated data. In our case, we take a deeper dive into the bottom-up approach, since it is more relatable.

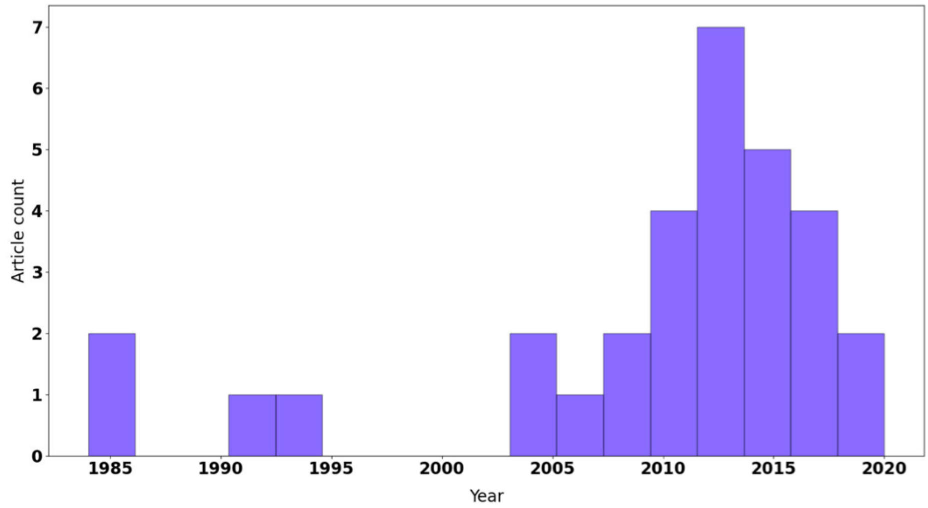
Proedrou, 2021 did a comprehensive review on load profiling. The author defined various load-profile application subgroups such as demand-side management, planning and control design of energy systems, and residential load profiles. The author also grouped modeling techniques as probabilistic models, 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. The review was structured by different methods. 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.

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 3.1 shows.

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

FIGURE 3.1: "Distribution of publications on load profiling from 1985 to 2020. The graph was published by Proedrou, 2021"



networks as a way of classification. Their methodology was tested in real use scenario.

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-side management subgroup. 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 behavioral patterns of dwellers. This method falls into the probabilistic method subgroup. 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 sum-meter data. Using this data, he made daily usage analyses on a one-hour basis. Using this information he 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 had 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 if 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. They profiled 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.

3.2 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-pint deviations for which they implemented isolation Forest algorithm and anomalous trends for which to detect, they implemented Change Point Detection.

3.3 Table of profiles

While in related work I examined load profiling in general, this chapter focuses on how data in load profiles is presented. 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. Using this information 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.

3.3.1 Feature set

Using related work and use case references, we can extract the most commonly used features to portray load profiles.

- power
- time
- operating time (how long appliance or appliances is turned on)
- appliances (a set)
- Number of activations (How many times appliance or appliances were turned on)

3.3.2 General table

Using these features we can form a table with all possible combinations. Some combinations do not make logical sense and the others repeat themselves. Combinations marked with X are such examples. Table 3.1 is then populated with references from previous chapters.

TABLE 3.1: General table of load profiles

	frequency	appliances	number of activations	power (avg)	operating time
appliances		X	X	X	X
number of activations	X	[9] [28]	X	X	X
power (avg)	X	[48]		X	X
power (array)	[28]	X	X	X	X
power (histogram)			X	X	X
operating time	X	[26]	[45] [44] [3]	[3]	X
time array	X	X	[9] [28]	[13] [15] [6] [27] [55] [19] [18] [25] [1] [29] [45] [44] [24] [3] [10] [30] [13] [8] [39] [28] [18]	[17]

Based on table 3.1 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 a bit, and it seems quite promising, we are focusing on activation-based histogram representation. Based on table 3.1 it is possible to see that not much attention was given to it.

3.3.3 Detailed table

This section will focus on exploring possible load profiles using activation-based histogram representation, while using the power feature as a baseload. Features from 3.1 will be explored in higher detail. They will be split and arranged in a way that all 21 publications and 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

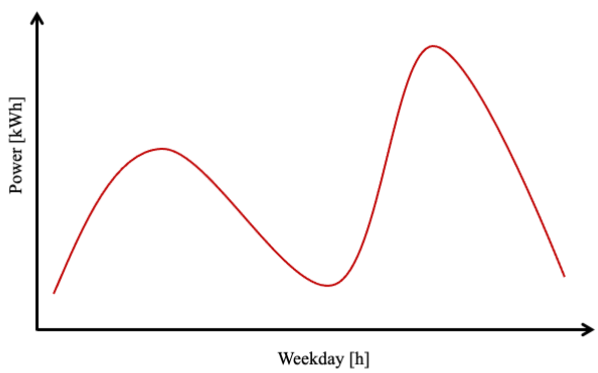
Main features were already described in section 3.3.1. It is possible to further divide them into smaller so-called sub-features. These are reshaped and grouped as follows:

- Way of presenting a profile
  - Per-house load profile
  - Per-appliance load profiles
  - Per-house and per appliance load profile
- By time range of profile
  - Daily
  - Weekly
  - Monthly
  - Yearly
- Way of measuring usage
  - Average power use
  - Activation or frequency of activation

Sketches of load profiles

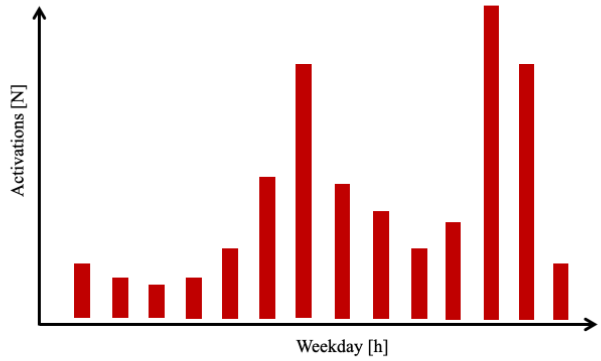
A most common way load profiles are presented is a daily power consumption profile such as shown in figure 3.2. The graph is a sketch, but it represents a standard load profile with morning and evening peaks.

FIGURE 3.2: "Average daily usage profile for an appliance or a building"



Some references include daily usage profiles as a histogram of activation at a point in a day, such as a figure 3.3.

FIGURE 3.3: "Histogram of daily activations profile for an appliance or a building"



All figures can present whole-house usage or per-device usage. Each presentation has its pros and cons. To present more information sub-meter data can be used to represent whole-house usage with per-appliance contributions. Such as on figure 3.4 and 3.5.

FIGURE 3.4: "Histogram of daily activations profile for an appliance A and B"

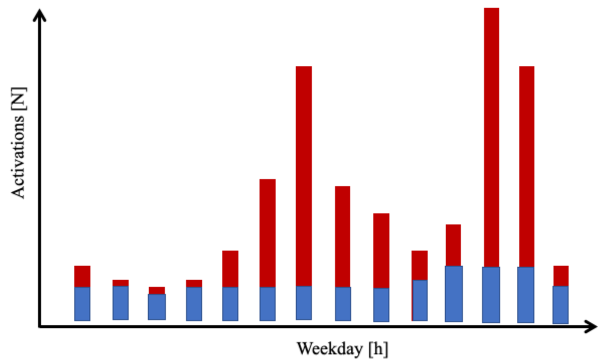
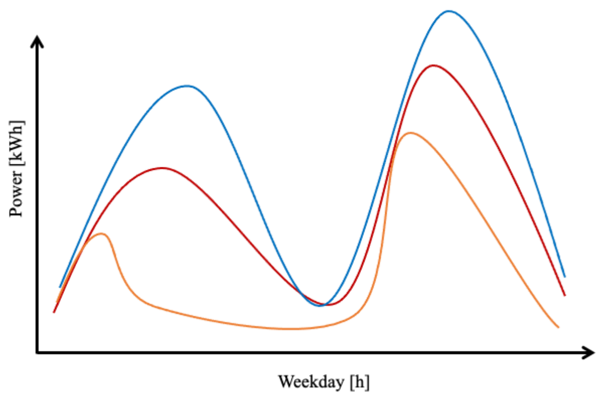


FIGURE 3.5: "Average weekday power consumption for appliances A, B and C"



To present as much information as possible all above-mentioned attributes can be presented in a multidimensional way such as heatmap in a way shown in figure 3.6 and 3.7.

FIGURE 3.6: "Number of daily activations / power consumption of one appliance / house in one month period"

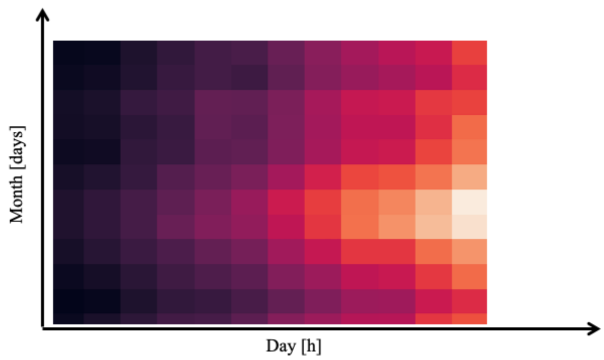
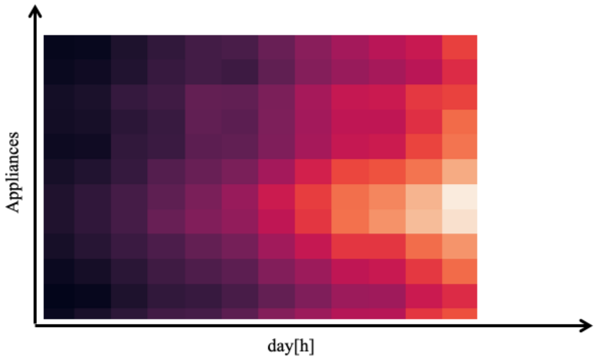


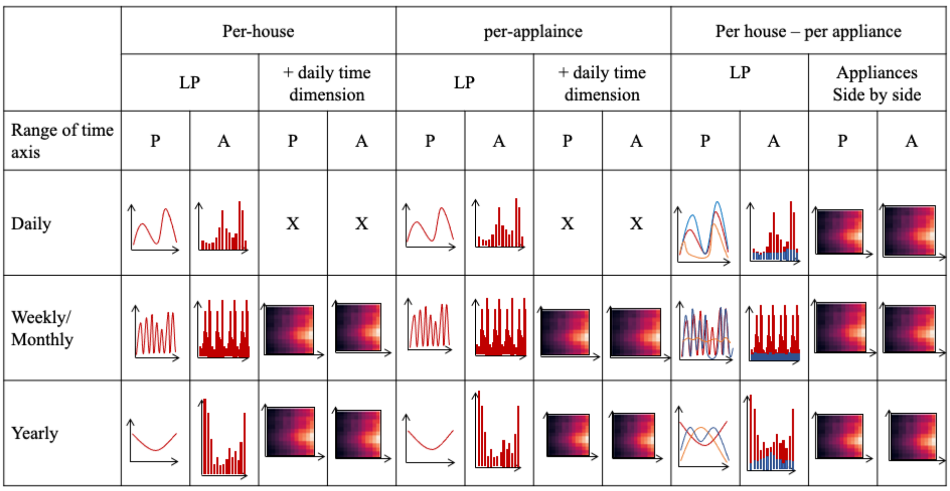
FIGURE 3.7: "Number of activations / power consumption for each appliance in one month period"



3.3.4 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 3.8 to generate and show possible combinations. Some combinations that had similar output were grouped, and some that could not be sketched were discarded.

FIGURE 3.8: "Table of combinations"



	Per-house				per-applaince				Per house – 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				
Weekly/ Monthly												
Yearly												

Figure above 3.8, uses sub-features from previous subsection 3.3.3. In general, the table is formatted in a way that features from columns are used in the x-axis of a plot, and rows are used in the y or z-axis of a plot.

The column of the table presents the time domain. Ďailym̈eans that the load profile presents average usage for one day, 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-house 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 profile group which presents the given usage unit on the y-axis and time on the x-axis. Next is a profile with a time group. 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-house subgroup includes appliances instead of time. Example for this is figure 3.7. The last columns present the usage unit, that is power (P) or a number of activations (A).

3.3.5 Mapping references to the table of profiles

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

TABLE 3.2: Table presents previously mentioned load profiles

Description	Per-house				Per-appliance				Per-house 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	[28] [13] [15] [6] [27] [8] [55] [19] [18] [25] [1] [29]		X	X	[45] [44] [24] [3] [10] [30]	[28]	X	X	[13] [8] [18]	[28]		
Weekly/ Monthly	[15] [6] [27]		[49] [39] [29]			[9]			[7]	[9]		
Yearly	[15] [6] [27]											

As can be seen from table 3.2, most of the work (14 publications) has been done with standard daily load profiles with per-house power usage such as figure 3.2. 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 3.3. During the research I focused on publications from minority classes, meaning not all existing publications for standard load profiles are included. The purpose of table 3.2 is to present missing scientific contributions and patterns of publications.

3.3.6 Mapping use-cases to the table of combinations

Table 3.3 includes arranged publications from chapter 5. Similar pattern emerged as in table 3.2.

TABLE 3.3: Table presents references mentioned in use cases chapter

Description	Per-house				Per-appliance				Per-house 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	[11] [50] [35] [56] [32] [20] [57]		X	X	[35] [54] [41] [30]		X	X	[13]			
Weekly/ Monthly	[50] [27]											
Yearly	[50]											

3.3.7 Table of use case groups

The table 3.4 presents same publications as 3.3, but only group names are shown. The table indicates how groups are arranged. Where anomaly detection and elderly care are dominating in the per-appliance part of the table, energy-saving and grid management are dominating in a per-house part of the table.

TABLE 3.4: Table presents references mentioned in use cases chapter

ES - energy saving GM - grid management AD - anomaly detection EC - elderly care X - unfeasible	Per-house				Per-appliance				Per-house 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	ES, GM		X	X	AD, EC, ES		X	X	GM			
Weekly/ Monthly	ES											
Yearly	ES											

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 3.5 empty spaces are filled with possible use cases for given load profiles.

TABLE 3.5: "Proposed use cases for profiles"

ES - energy saving GM - grid management AD - anomaly detection EC - elderly care X - unfeasible	Per-house				Per-appliance				Per-house 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, ES, GM,	AD, ES, GM,	X	X	AD, EC, ES, GM	AD, EC, ES, GM	X	X	AD, EC, ES, GM	AD, EC, ES, GM	AD, EC, ES, GM	AD, EC, ES, GM
Weekly/ Monthly	AD, ES, GM	AD, ES, GM,	ES, GM	ES, GM	AD, ES, GM	AD, ES, GM	AD, ES, GM	AD, ES, GM	AD, ES, GM	AD, ES, GM	AD, ES, GM	AD, ES, GM
Yearly	ES, GM	ES, GM	ES, GM	ES, GM	AD, ES, GM	AD, ES, GM	ES, GM	ES, GM	ES,GM	AD, ES, GM	AD, ES, GM	AD, ES, GM

3.3.8 Table of load profile potentials

Some combinations are indeed illogical and again others are less useful in a practical sense. The next table will try to rate the scientific 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 at 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 is the effectiveness 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 high research potential.
- 2 - The load profile does not satisfy one of the above-mentioned assumptions and has mid-research potential.
- 3 - The load profile does not suffice any of the above-mentioned assumptions and has low research potential
- X - The load profile is inexplicable.

TABLE 3.6: Proposed classification of profiles

	Per-house				Per-appliance				Per-house 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	2	X	X	1	1	X	X	1	1	1	1
Weekly/ Monthly	1	2	1	1	1	1	1	1	2	2	2	2
Yearly	1	3	3	3	2	2	3	3	3	3	2	2

3.3.9 Table of possible future research directions

Using all the above-mentioned tables we can use superposition to generate a universal table, that will present possible research directions. The load profile has to satisfy the following rules. The first is that the load profile should have no publications or yet discovered use cases. The second one is that the profile should be at least in the second class of potential defined in subsection 3.3.8.

TABLE 3.7: Possible future research contributions

	Per-house				Per-appliance				Per-house 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 3.7 presents load profiles that we will pursue in this paper. We will focus on profiles from the first class and activation frequency type of usage. When the aforementioned parameters are applied, the result is table 3.8

TABLE 3.8: Load profiles to be pursued

	Per-house				Per-appliance				Per-house 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												

Based on the table 3.8 we propose the following profiles for activation frequency: per-house daily-monthly profile, per-house weekly-yearly profile, per-appliance weekly profile, per-appliance daily-monthly, per-appliance weekly-yearly, per-house and per-appliance daily profile (stacked), per-house and per-appliance daily profile (appliance side by side)

Chapter 4

Possible use cases

The load profiling method has a lot of different use cases across different fields. It can be used to save energy by studying users usage patterns and returning feedback. Electrical energy providers could use that same data to optimize the management of their grid, with minimal impact on users daily lives. This method could also be used to help the elderly in case of an accident and help prevent one. It could be used to detect all kinds of early malfunctions in the operation of appliances and help save energy. Occupancy detection, research, and development are all areas where profiling could be used.

4.1 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, 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 includes 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 occasions. In this case, two profiles would be developed. Normal workday and work from home day. Additional information would be obtained from the users 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, EV 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 users 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 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 users 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 the consumption.

4.2 Grid management

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 lower the 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 to the user.

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. 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.

4.3 Elderly care

Demographic changes i.e. 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.

4.4 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, 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.

4.5 Other

Load profiling could also be used as feedback to the engineers and designers, of how a certain 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.

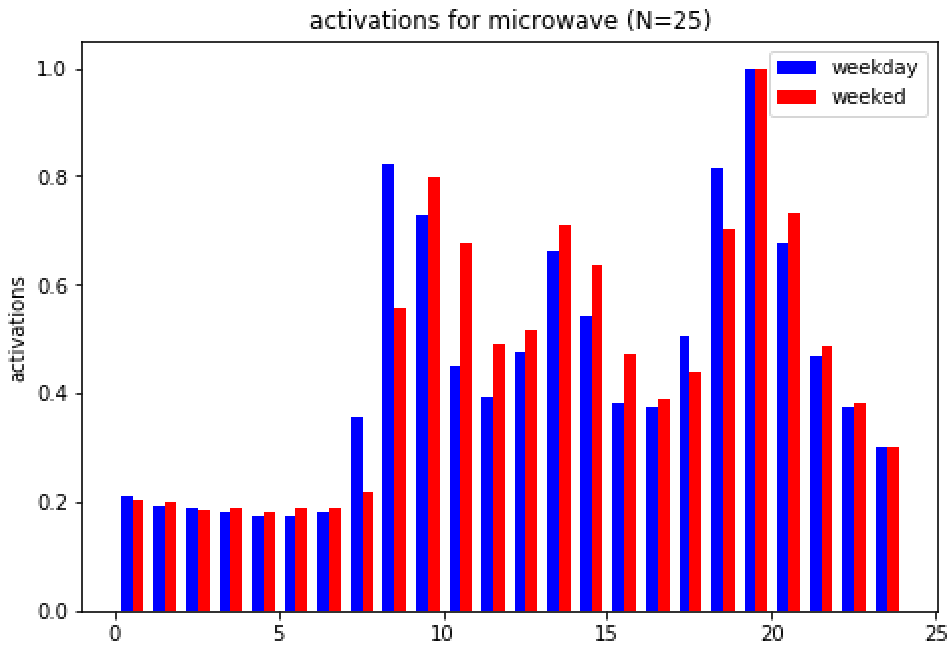
Chapter 5

Contributions

The goal of the master's thesis is to propose suitable consumption profiles for supporting residential building consumption optimization and elderly care management. To achieve this goal, we propose the following steps.

The first step is to obtain a set of datasets. In our case, this will be 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. Data will be then used to generate different daily per-appliance usage profiles.

FIGURE 5.1: "Universal normalized daily usage profile for weekend and weekday for a microwave. Superposition of data from 25 homes."



One such example can be seen in figure 5.1. The histogram shows normalized daily activation for microwaves. It consists of data from 25 homes from 4 different datasets.

The thesis is constructed from three parts, and each answers our following questions accordingly:

- How do we efficiently present big data to humans and machines?
- How is this presented data connected in a higher dimension?

- Can we use one of the profiles to build something useful?

The proposed goal will be achieved by answering these three questions.

Chapter 6

Presenting proposed profiles

Previously defined profiles will be presented in-depth. In general, each profile has its use case already assigned in table 3.3. Here, we will focus on exposing the main features, issues, and use cases.

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

6.1 Time ranges

One important thing to mention is to use cases for different time ranges of load profiles. Based on table 3.2 it is possible to see that most publications and 3.3 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.

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.

6.2 Per-house

The section will be focused on per-house 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-house 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-house 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 phase of practical use yet.

The daily per-house load profile is also known as the standard load profile. According to table 3.2 this is the most commonly used power profile. Figures 6.1a and 6.1b present usage patterns on different time ranges. The two profiles, therefore present different contextual causes.

(A) "Daily" (B) "Yearly"

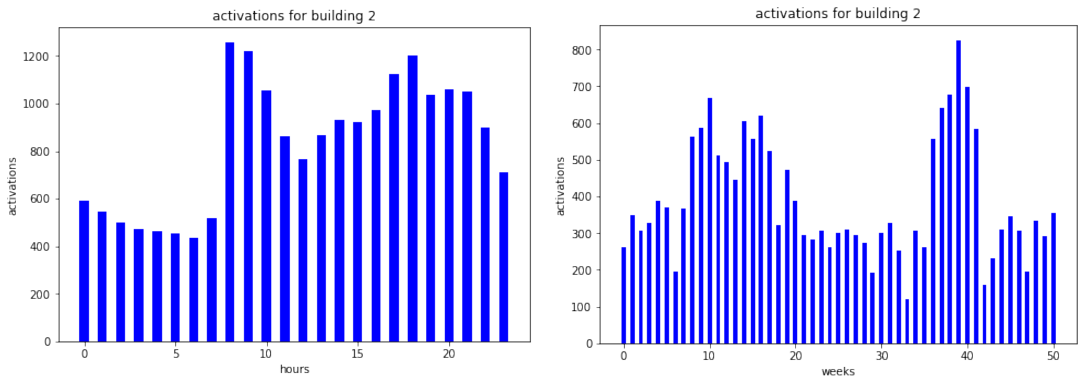
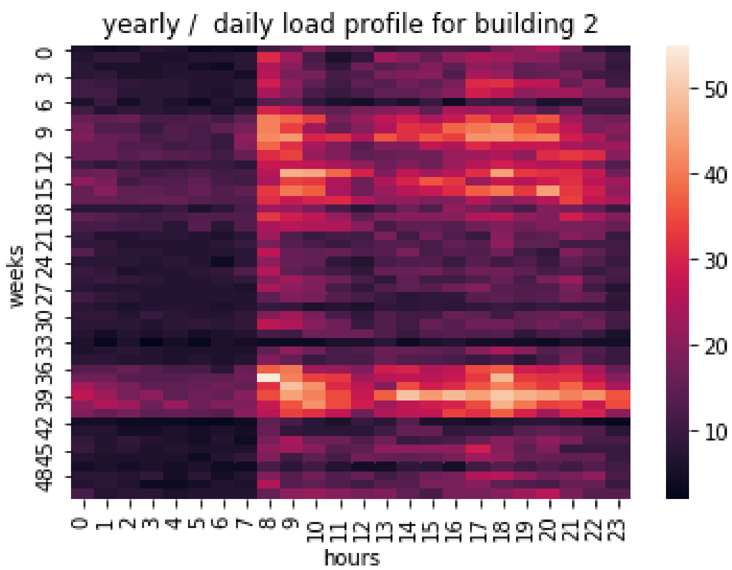


FIGURE 6.1: "per-house load profiles"

6.2.1 Per-house two-dimensional time

Alternatively, it is possible to combine figures 6.1a and 6.1b and present activations as a heat map. The result is a load profile showing more complex activation patterns.

FIGURE 6.2: "Two-dimensional time per-house load profile"

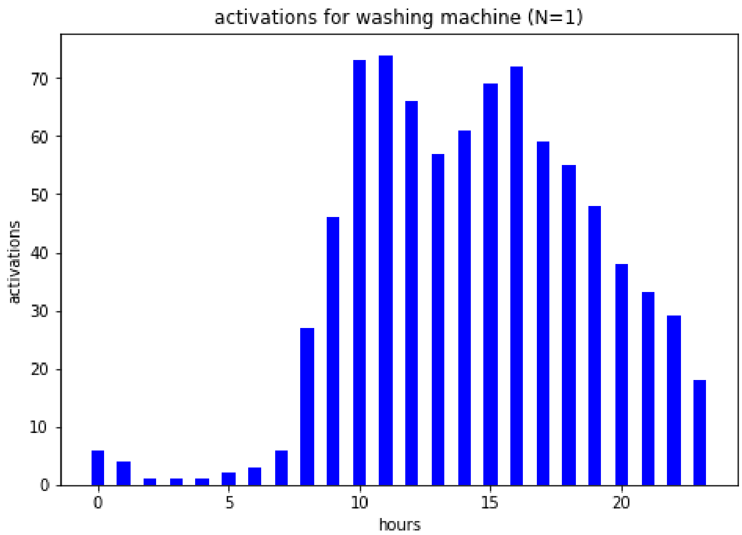


Previously it was mentioned that these kinds of profiles are the most applicable in grid management and energy-saving fields. One such example could be load shedding. Using the Load profile above, electrical energy providers could find buildings with the least activity at that time of day. Combining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

6.3 Per-appliance

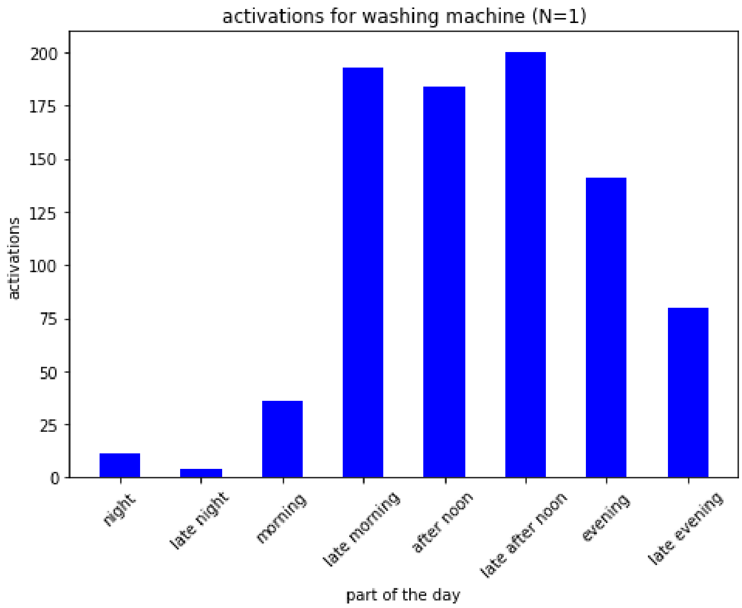
Per appliance load profiles offer a look into the consumption of each appliance. In this case activation load profiles, this is an elemental load profile, since all other profiles are built on top of it. This also means that it is one of the most universal profiles since it can be used in all previously defined use cases. Comparing power and activation profiles in the per-house chapter, it was possible to see that activation-based load profile does not bring significant advantages over power-based load profiles. In the case of per-appliance load profiles, it is possible to analyze the usage of the single appliance in greater detail

FIGURE 6.3: "Daily per-appliance load profile"



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 on figure 6.4 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. Parts of the day are:

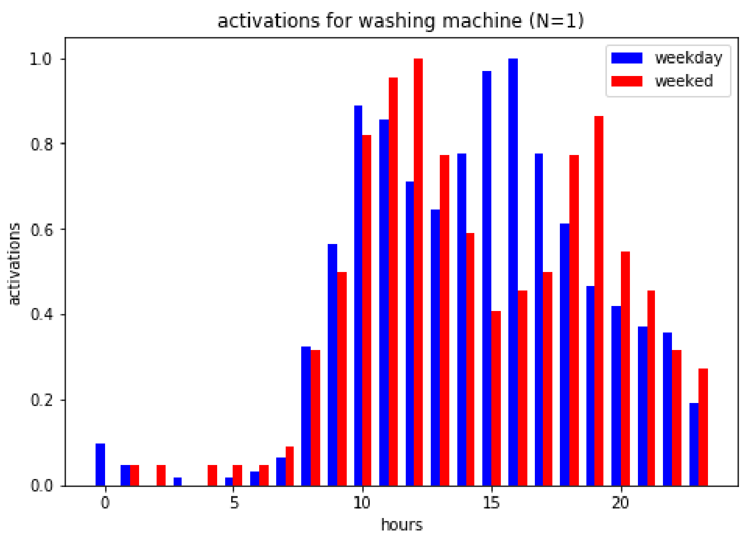
FIGURE 6.4: "Daily per-appliance load profile with larger buckets sizes"



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.

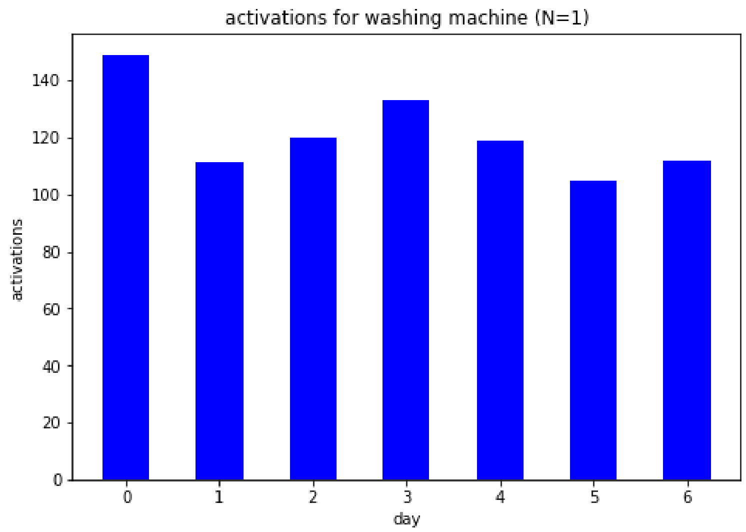
In case, dwellers have different usage patterns during the weekends, two profiles would have to be developed. It is possible to present them both at once such as in figure 6.5. This is essentially a variation of the weekly Load profile that maintains high resolution. Since there are more weekdays than weekend days, activations had to be normalized accordingly.

FIGURE 6.5: "Normalized daily per-appliance with weekday and weekend load profiles"



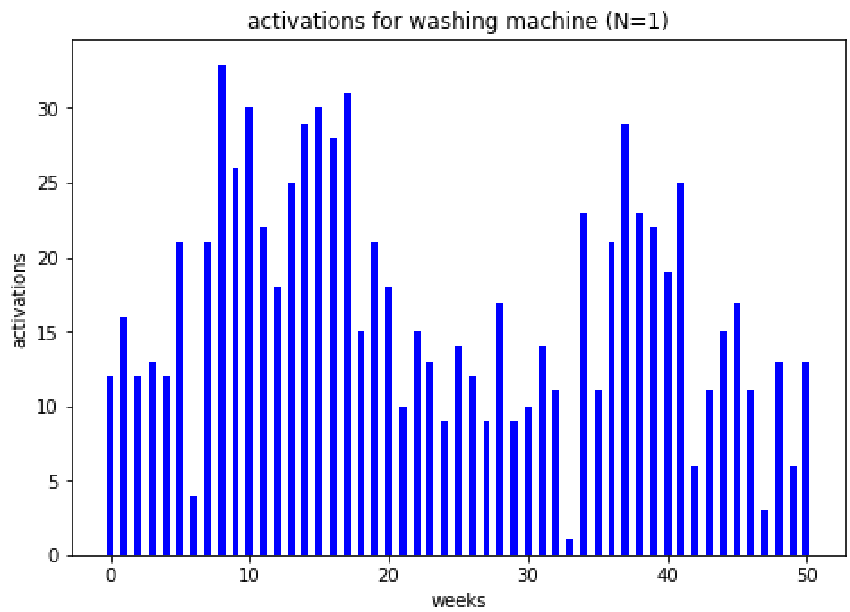
Another way to present weekly data is figure 6.6 This resolution offers a look into how consumption pattern changes over the week. This is useful for applications such as grid management and energy saving. In this particular case, it is possible to see that the user most commonly uses the washing machine on Mondays and Wednesdays. Using a weekly weather report that would indicate high energy production on Wednesday, 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 6.6: "Weekly per-appliance load profile"



As mentioned earlier, the monthly presentation does not show any significant usage pattern, where yearly presentation again shows the more broad usage pattern. This is useful for grid management and energy-saving, where one could detect seasonal changes in usage of an appliance.

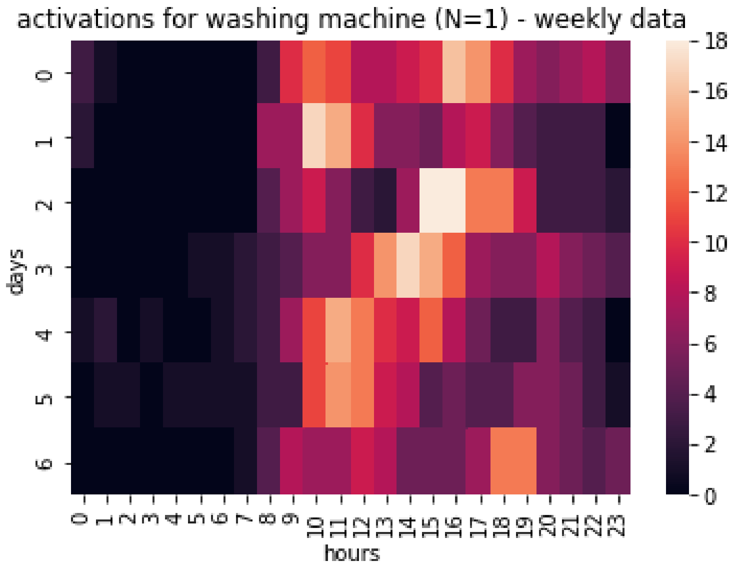
FIGURE 6.7: "Yearly per-appliance load profile"



6.3.1 Two-dimensional time per-appliance load profiles

Using a combination of figures 6.3 and 6.6, it is possible to generate a heatmap 6.8.

FIGURE 6.8: "Two-dimensional time per-appliance load profile"



In this case, a similar use case could be fitted as in the first example. The first example used load shedding for when the demand is too high. On the contrary, it can also occur 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 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 such as wind and cloud coverage. By combining weekly wind forecast, weekly cloud coverage, and users consumption profiles energy providers could notify users to turn on the appliances at peak usage times.

By analyzing figure 6.8 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 should 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 operator or electrical distribution company 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

The figures below 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.

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

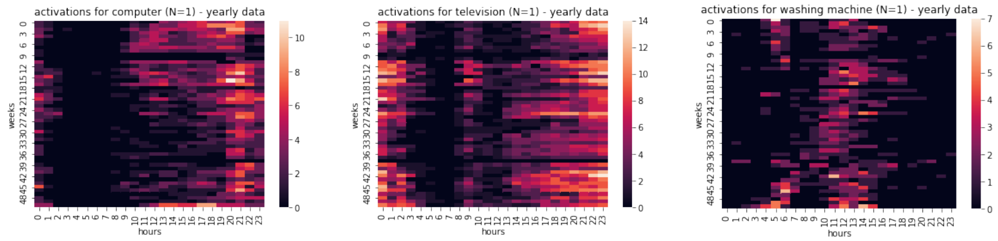


FIGURE 6.9: Various yearly two-dimensional load profile

Another example worth mentioning is one 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"

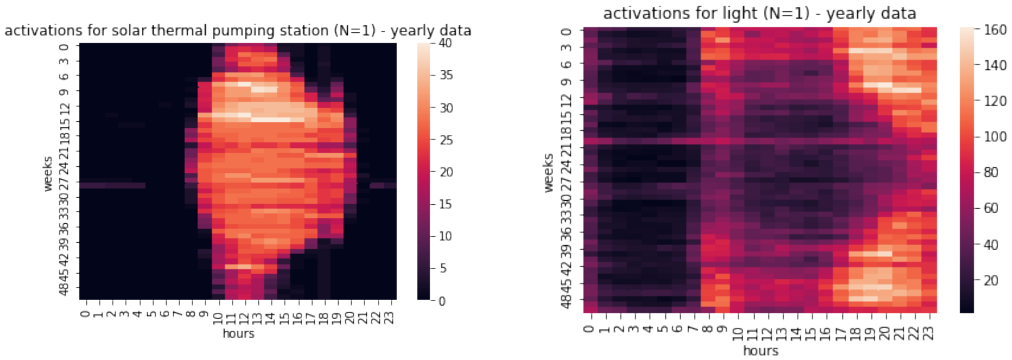


FIGURE 6.10: "Effect of seasonal changes on load profiles"

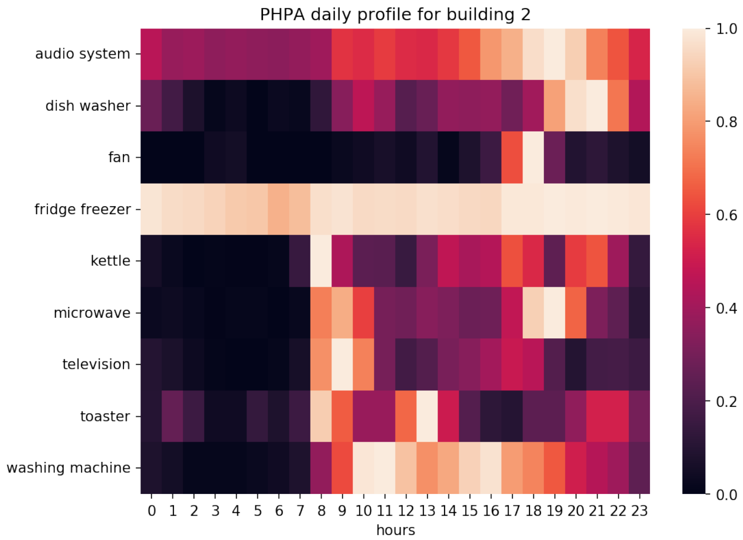
Figures 6.10a and 6.10b show how weather and season affect the usage pattern of appliances.

6.4 Per-house per-appliance

The last group of profiles is a combination of per-house 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 stove and kettle are commonly used together each morning this use could translate to an event such as breakfast. In order to achieve this, one needs to observe all appliances at once such as shown on figure 6.11.

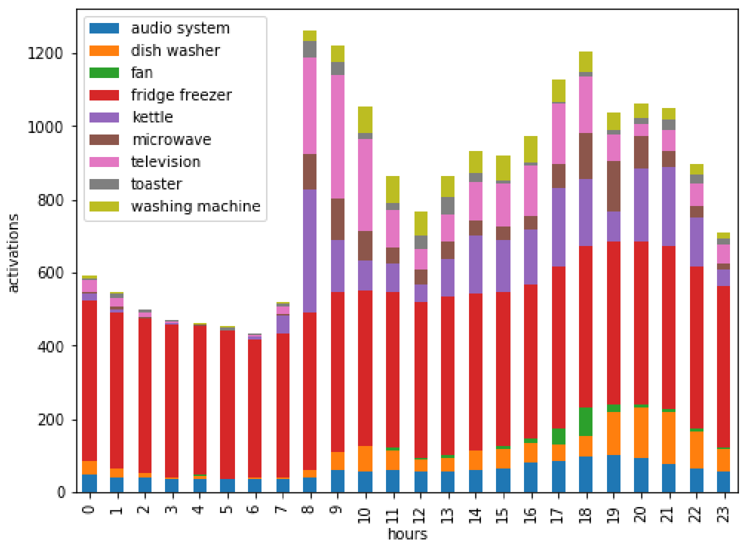
The figure is also a good example of elderly care system, that would detect an anomaly such as fall, or person unable to get up from the bed in the morning. This profile shows that first thing in the morning used are kettle and toaster, and with delay of one hour, microwave and TV. This enables us to 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 appliance. In case elder falls during cooking, toasting bread or opening 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 6.11: "Daily per-appliance per-house building load profile"



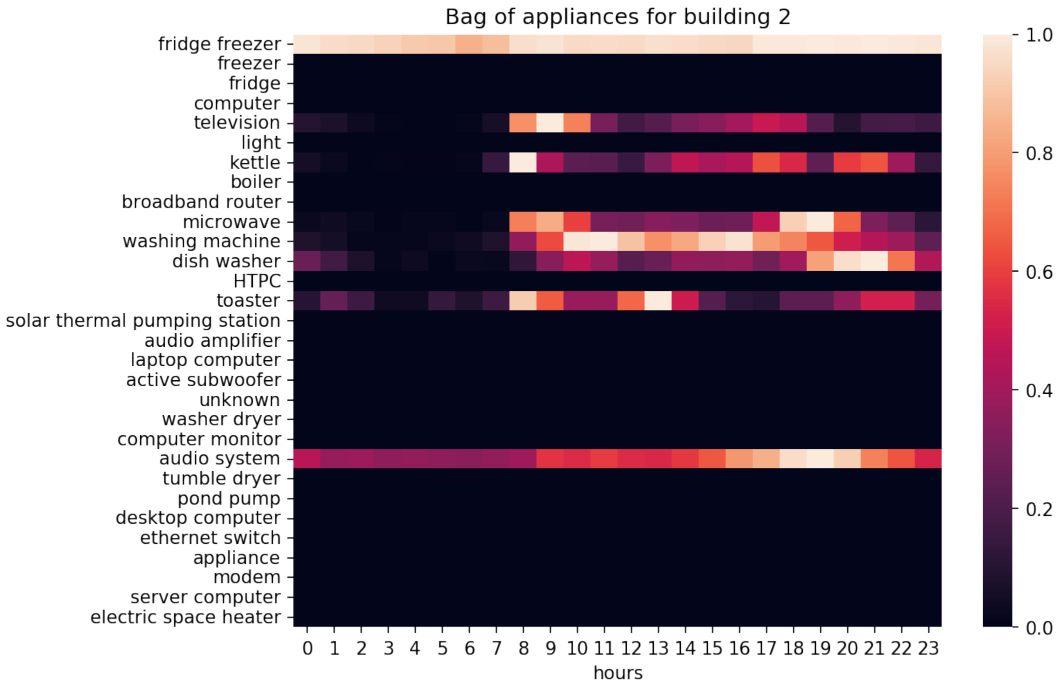
The very same data can be presented in an alternative way, such as shown in figure 6.12. The usage pattern is the same as on 6.1a, except that it is possible to see the contribution of each appliance.

FIGURE 6.12: "Stacked daily per-appliance per-house building load profile"



These load profiles are useful when it comes to analyzing the usage pattern in one building. In order to be able to process the load-profiles a cross many buildings a new profile must be introduced. The idea is derived from the bag-of-words method used in text processing, where a list of 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 number of activations and then only top 30 appliances are selected. Using this list it is possible to present the usage of each building universally.

FIGURE 6.13: "Universal presentation of per-house per-appliance load profile"



Chapter 7

t-SNE

7.1 Introduction

The load profiles can present data in a new, previously unseen 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 how their consumption differs.

To measure the similarity of activation profiles, the t-SNE algorithm will be used.

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

7.2 Goals

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

7.3 Methodology

7.3.1 Load profiles

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. Days 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 in a week, this would be a Monday from midnight to one o'clock. The lower-right corner presents the last hour of a week. Since there are roughly 4 weeks in each month, each pixel will present 4 samples.

7.3.2 Datasets

To present the results REFIT, UK-DALE, ECO, REDD and iAWE 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.

7.3.3 t-SNE

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 on 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. Geoffrey Hinton and Roweis, 2003 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 in order to calculate the similarity.

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

7.3.4 Tools used

To process the data and to obtain results Google Colab was used. For handling the data python libraries such as h5py, pandas and NumPy were used. To present the data Seaborn and Matplotlib libraries were used. To implement the t-SNE a SciPy library was used.

7.4 Results

The results will be presented in three subsections, these are

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

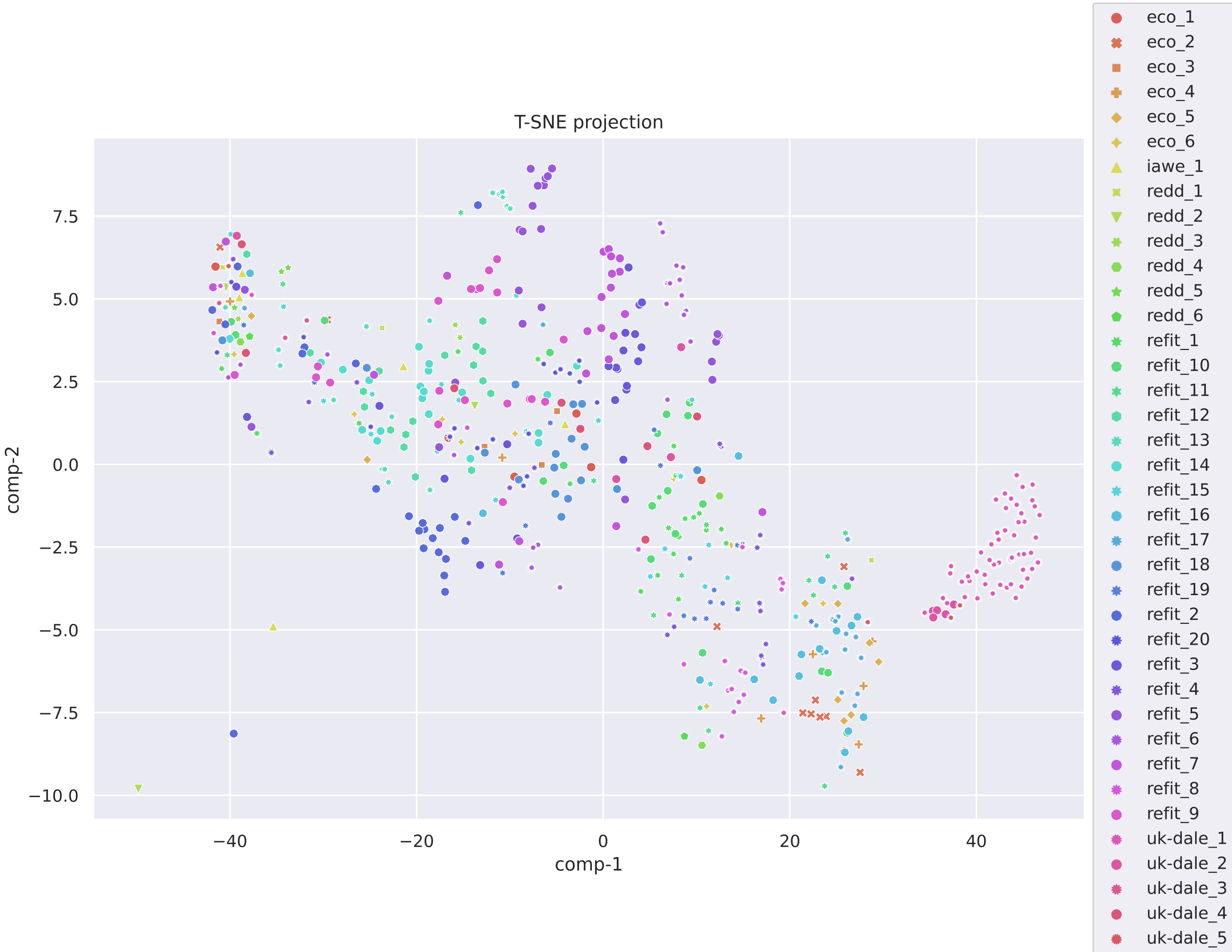
Most of the focus will be done on the per-appliance load profile since it is the most universal.

7.4.1 Results for per-building load profiles

This load profile is useful when it comes to comparing how activation patterns change over buildings and datasets.

Figure 7.1 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 much a certain appliance is used. In some cases this information is useful, again in others we would like to find a more complex usage pattern.

FIGURE 7.1: "Projection of per-building load profiles"



The figure 7.2 presents the actual load profile behind each sample. It is possible to see that on the left there are mostly samples with very little activity, and on the left, we see samples with more activity. Since the two plotted components are of a higher dimension, it is hard to guess what they present. As said t-SNE gives us the intuition of how load profiles are connected in higher-dimensional space.

FIGURE 7.2: "Projection of per-building load profiles with actual samples"

Normalized load profiles

To solve the issue mentioned we have to normalize the data between 0 and 1. The figure 7.3 shows how normalizing samples affect the algorithm. The samples are much closer to each other, while it is still possible to see the individual clusters. This could imply that the usage pattern of users is more similar than the amount of it they use.

FIGURE 7.3: "Projection of normalised per-building load profiles"

The figure 7.4 presents only the main cluster of samples, since the smaller cluster presents mostly low entropy data. Same cluster can be seen on far left on figure 7.2.

FIGURE 7.4: "Projection of normalised per-building load profiles with actual samples"

On figure 7.4 it is possible to find all kinds of usage patterns. But the general one is that there is less activation during the night, Then there is one peak during the morning and again once in the evening. Some buildings are more active during the week and again others during the weekend. Most 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, then there are some outliers, where the pattern completely changed. Albeit less obvious, the same happens overall buildings. This happens due to events such as vacations, holidays or weather-induced behavioral changes.

7.4.2 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.

Single appliance over many buildings

Using only one appliance and using 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, apart from opening the door and turning on the light inside. Usually, this event is dwarfed by the number of activations of a compressor. This also means that the usage pattern should be the same across all buildings. This can be seen on figure 7.5, where apart from REFIT building 1 and 11, there are no clusters.

FIGURE 7.5: "Projection of fridge load profiles for various buildings"

Figure 7.6 Shows mostly bright images, apart from few outliers. Load profiles scattered in a circle are generally less dynamic than the ones at the bottom. The figure 7.6 is a good example how even little to none human interaction, load profiles 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 7.6: "Projection of fridge load profiles for various buildings with actual samples"

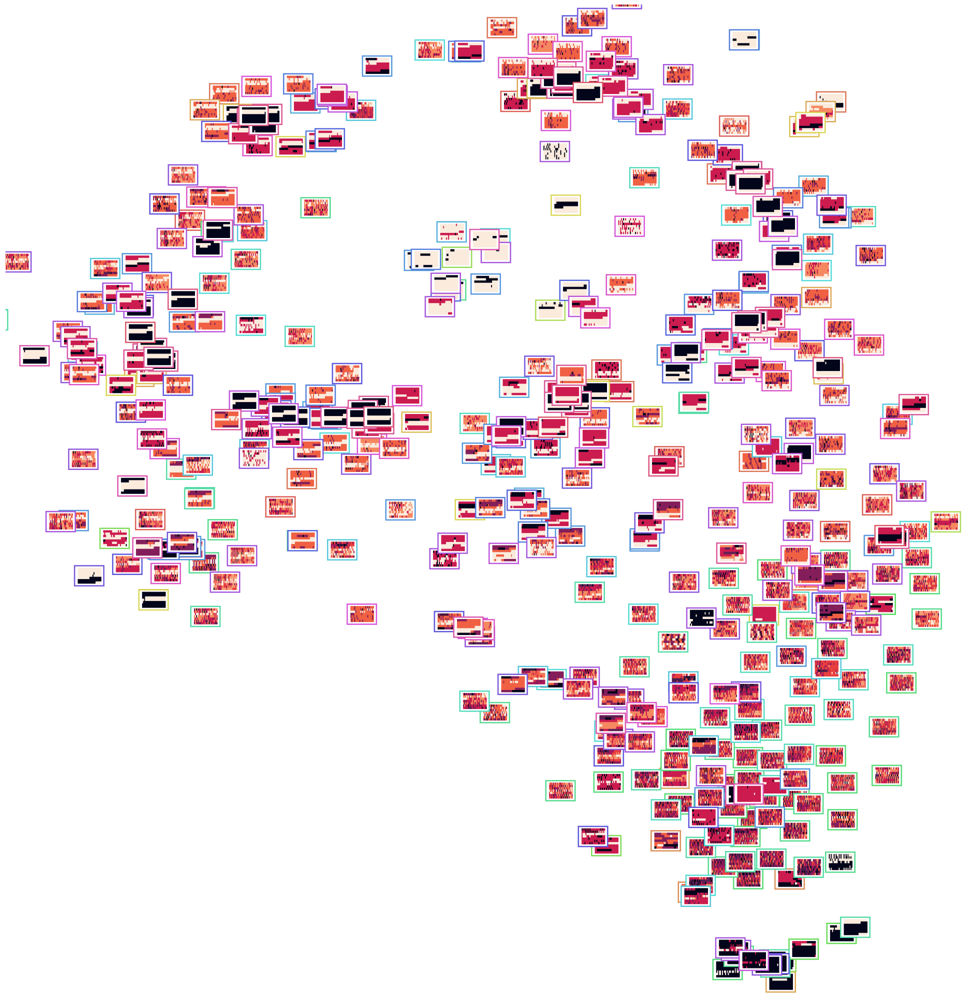
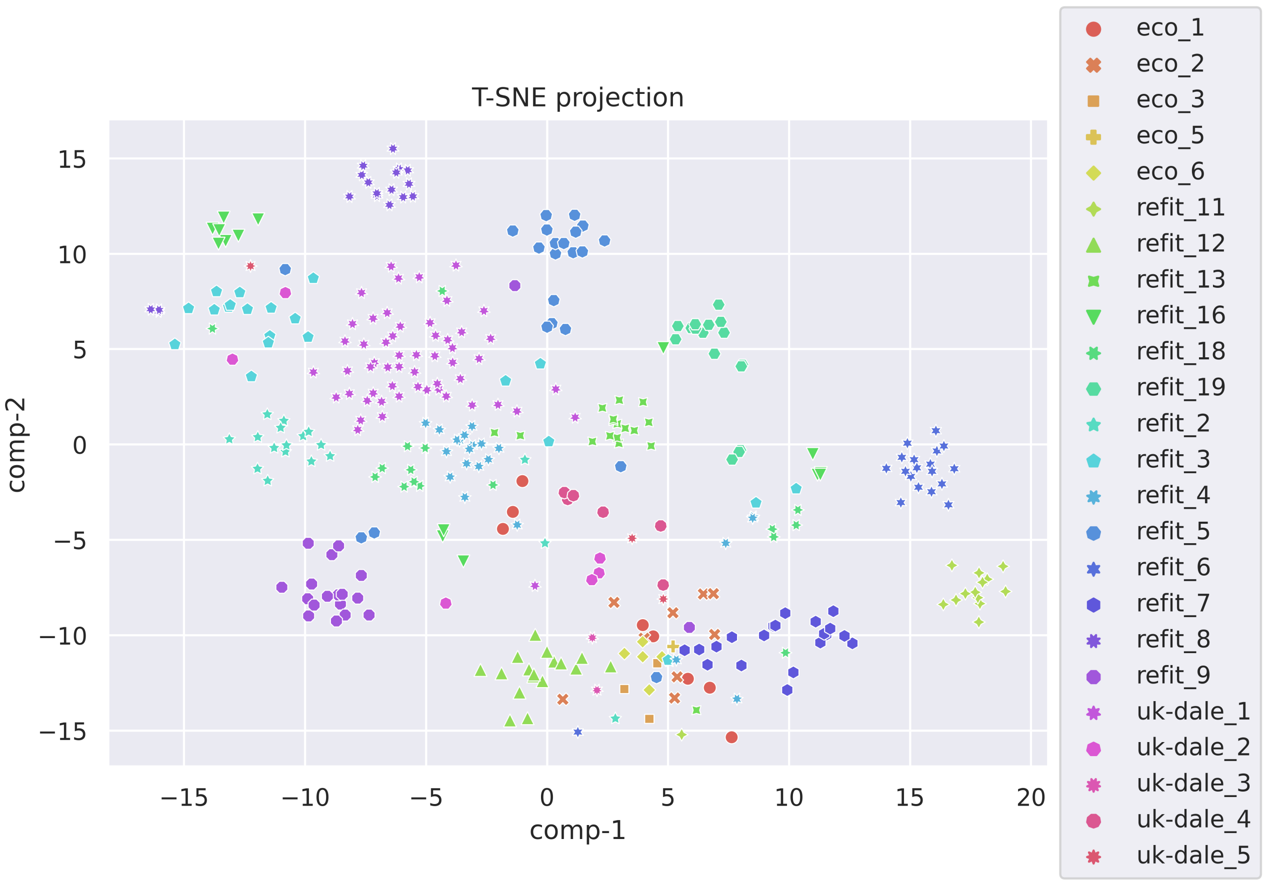


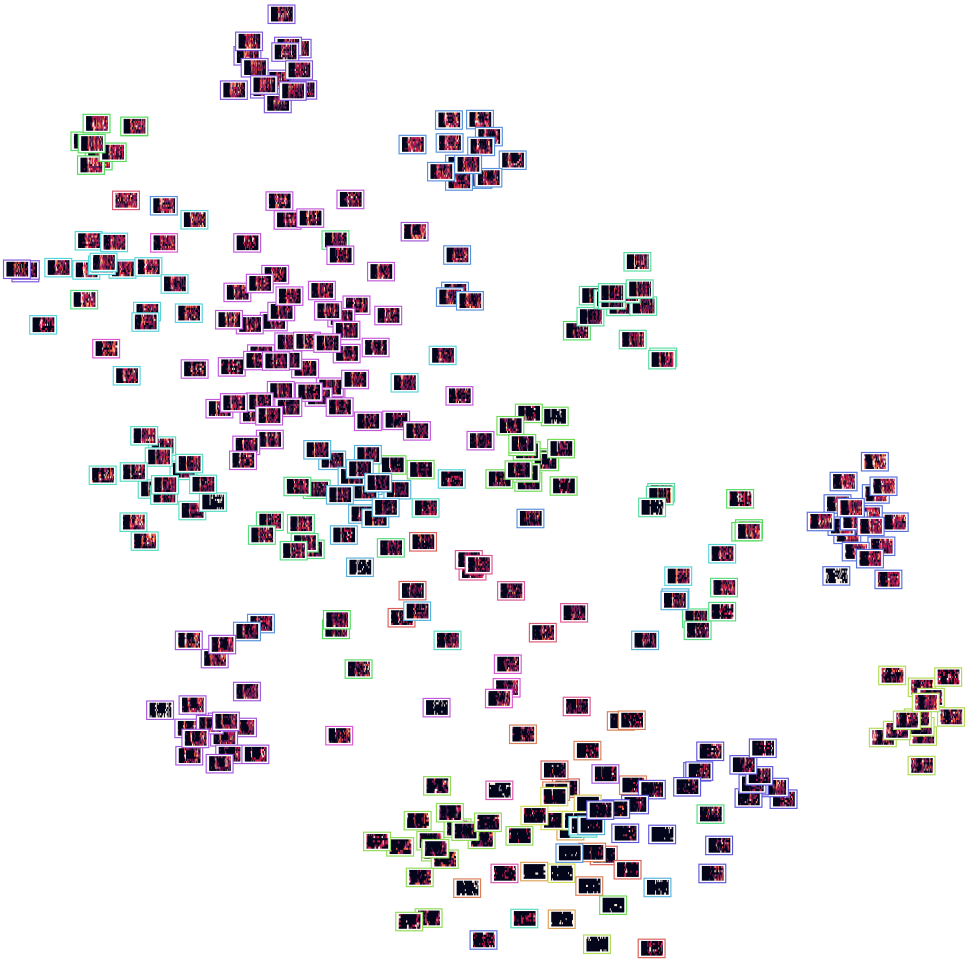
Figure 7.7 shows how, compared to fridges, kettles have much more clear clusters that are spaced out between each other. This could mean that every household uses a kettle a bit differently. This cluster is a good example where we can see how strong is a routine of a user, The closer together the clusters, the higher the routine since samples have to be close to each other.

FIGURE 7.7: "Projection of kettle load profiles for various buildings"



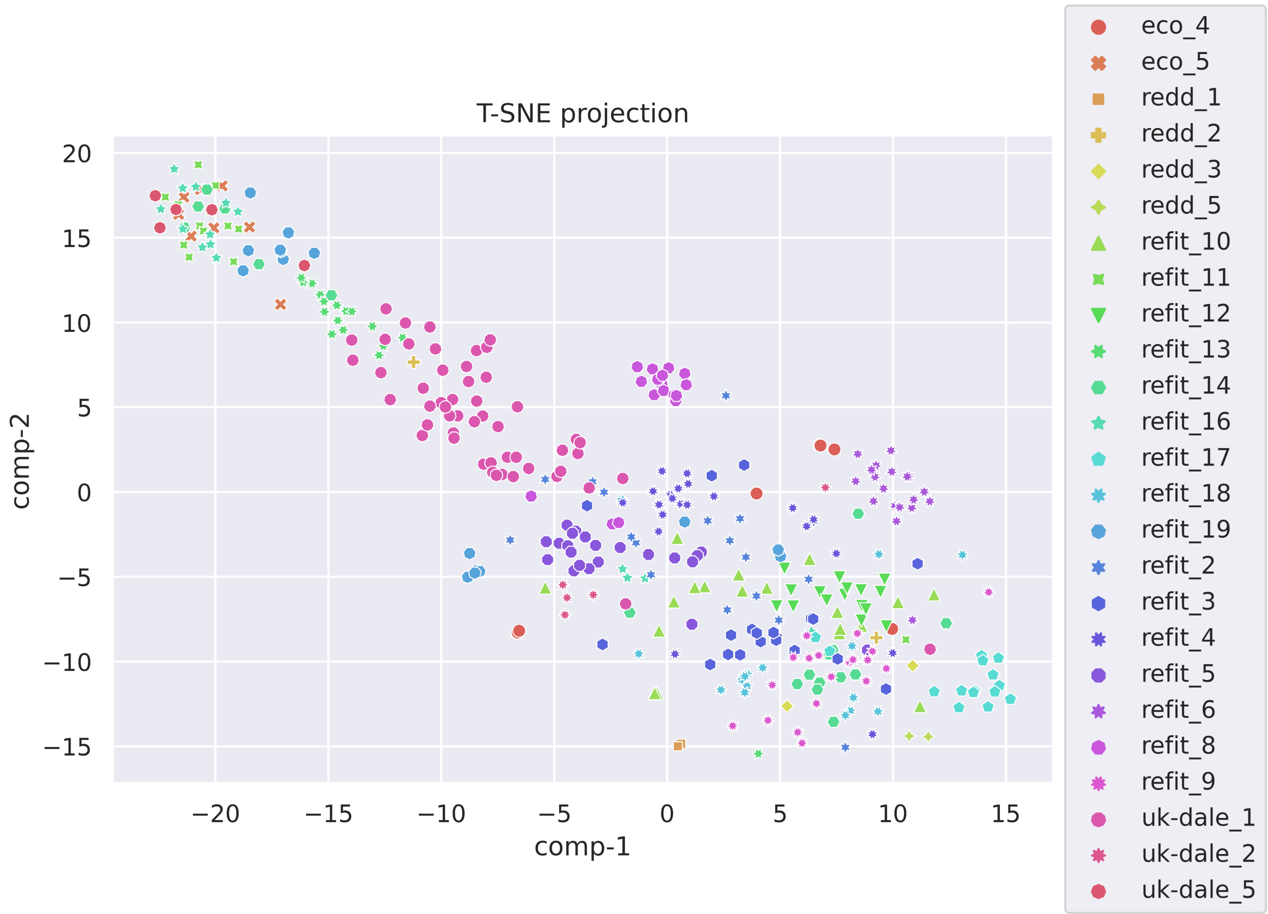
The figure 7.8 shows us that images on the lower part of the plot contain less activity than the others. It is also possible to observe that load profiles that are closer together have more similar activation patterns. The strength of a routine is an important feature that will be used in the next chapter to build an elderly care anomaly system.

FIGURE 7.8: "Projection of kettle load profiles for various buildings with actual samples"



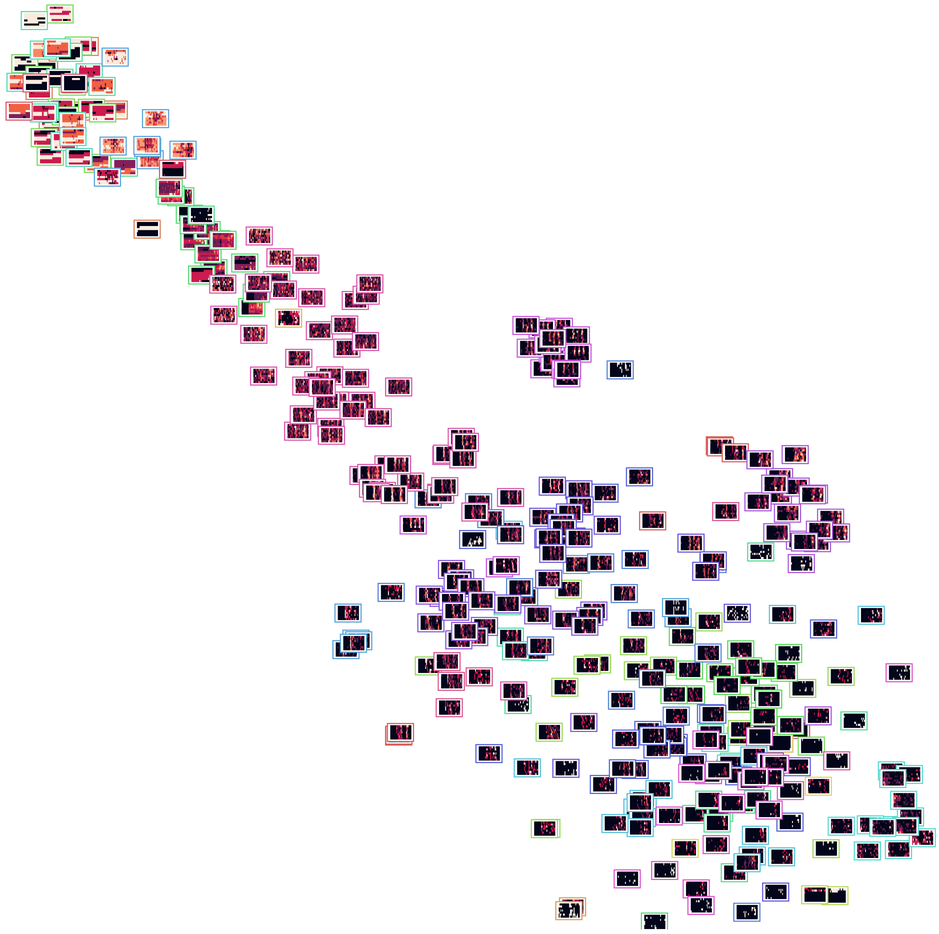
The figure 7.9 shows that microwaves are again a bit different from the kettle. They are more clustered than the fridges, and less than the kettles, even though they are used similarly. This could be due to additional electronics such as a clock that are built into, the appliance, this could lead to some samples being required as turned on due to "dark" current. One other difference between the two is that microwave has more than one mode of operation.

FIGURE 7.9: "Projection of microwave load profiles for various buildings"



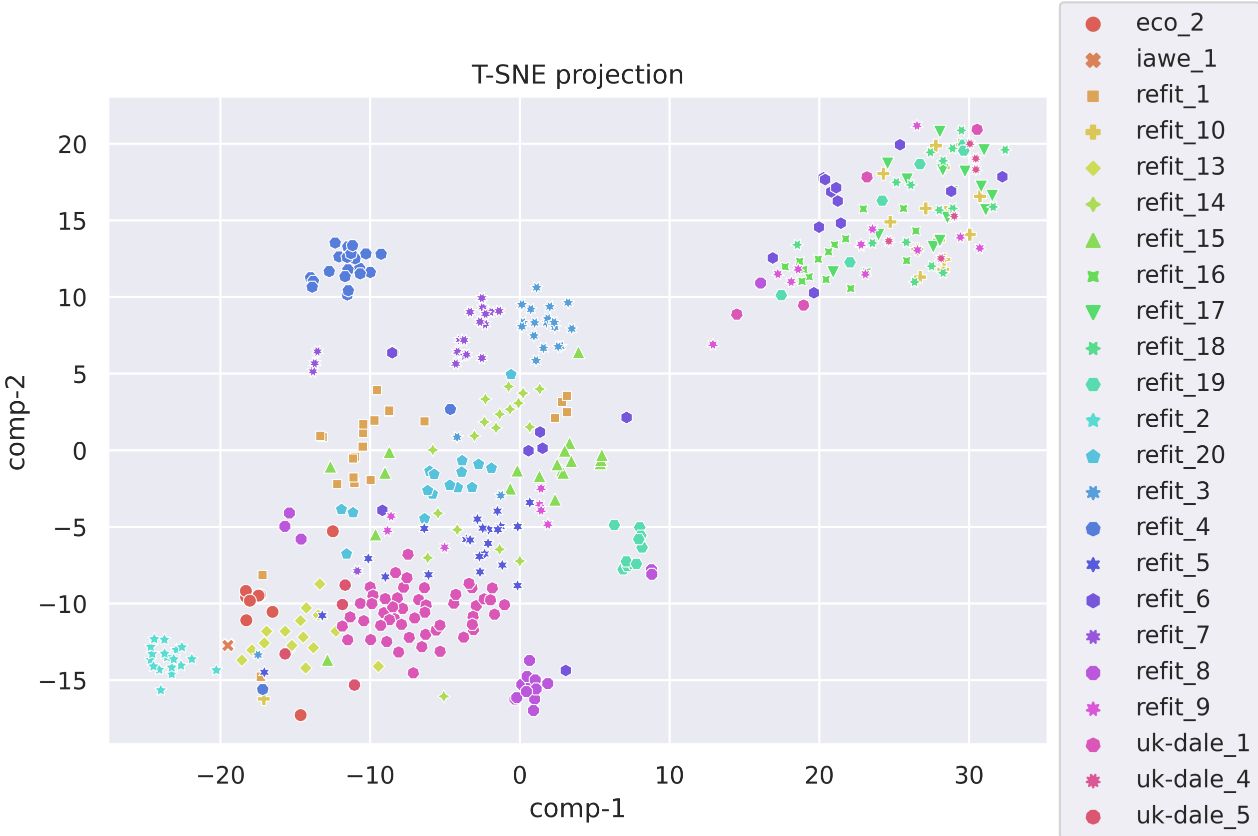
The faulty samples could be the ones in the upper left part of the plot since they are too bright. They do present a pattern, but it is questionable what it presents since it seems like it's turned on during the nighttime. Images at the other end show less lot less activity, which could indicate that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

FIGURE 7.10: "Projection of microwave load profiles for various buildings with actual samples"



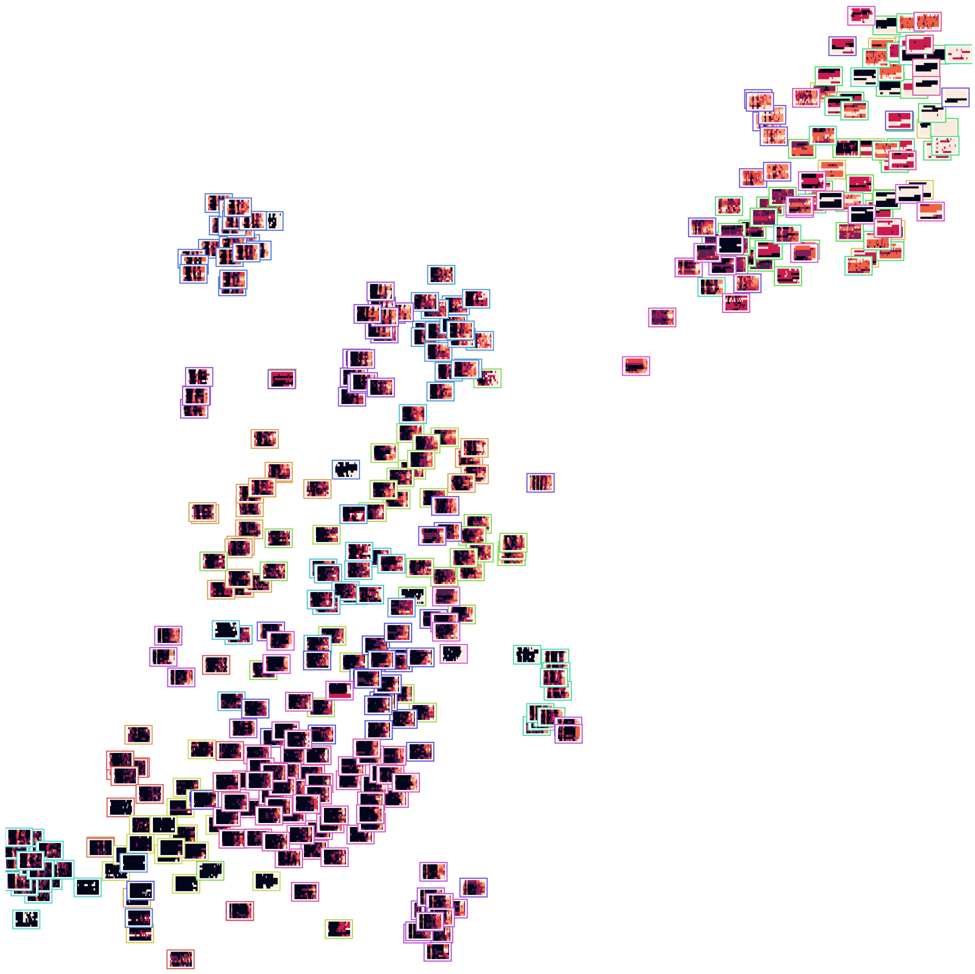
The last per-appliance example is television presented on figure7.11. Television was chosen since they were common across all datasets. Interestingly enough, televisions form nice clusters with a few outliers. Clusters are separated but close together, this could mean that usage patterns are unique but 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 7.11: "Projection of TV load profiles for various buildings"



The images on the figure behind clusters on proving 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.

FIGURE 7.12: "Projection of TV load profiles for various buildings with actual samples"



Per-appliance load profiles - comparing appliances

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

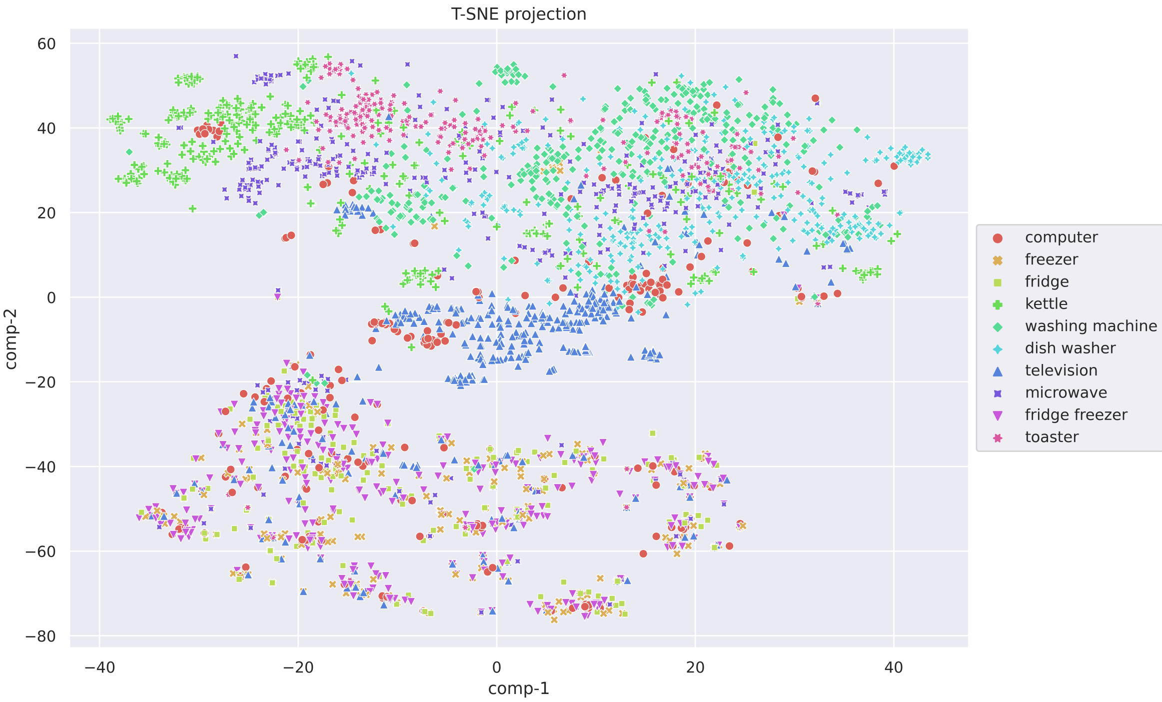
FIGURE 7.13: "Projection of per-appliance load profiles"

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

FIGURE 7.14: "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 7.15.

FIGURE 7.15: "Projection of filtered per-appliance load profiles"



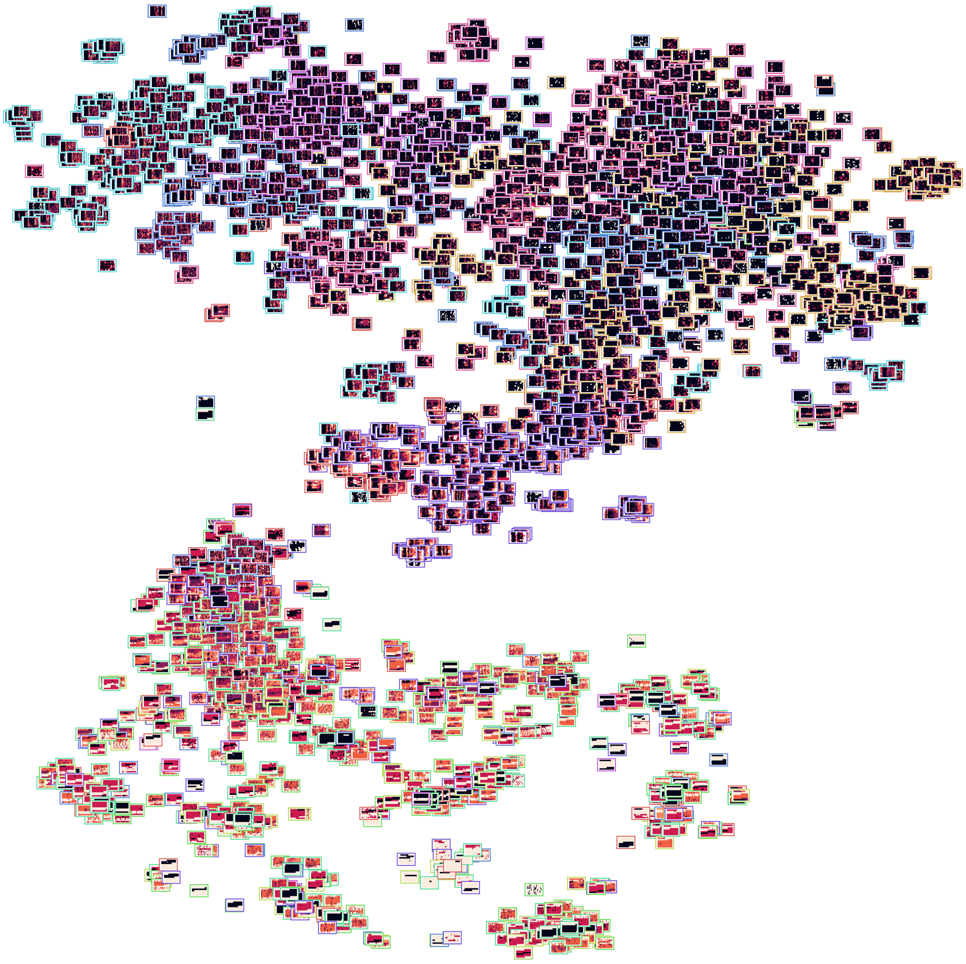
The figure 7.15 shows how these 10 appliances ares connected in high dimensional space. Kettle, microwave and toaster are quite similar when it comes to usage patterns. They are operated for a short amount of time and are usually used in users' routines in the morning or evening. These appliances are 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 7.16: "Projection of filtered per-appliance load profiles with actual samples"



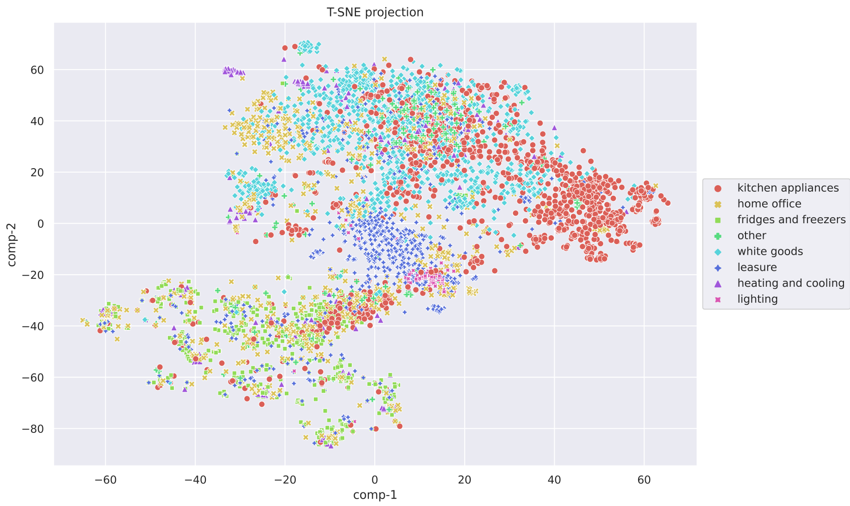
All through the set of load profiles is different due to filtering, the figure 7.16, retains a similar structure to the previous figure 7.15.

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

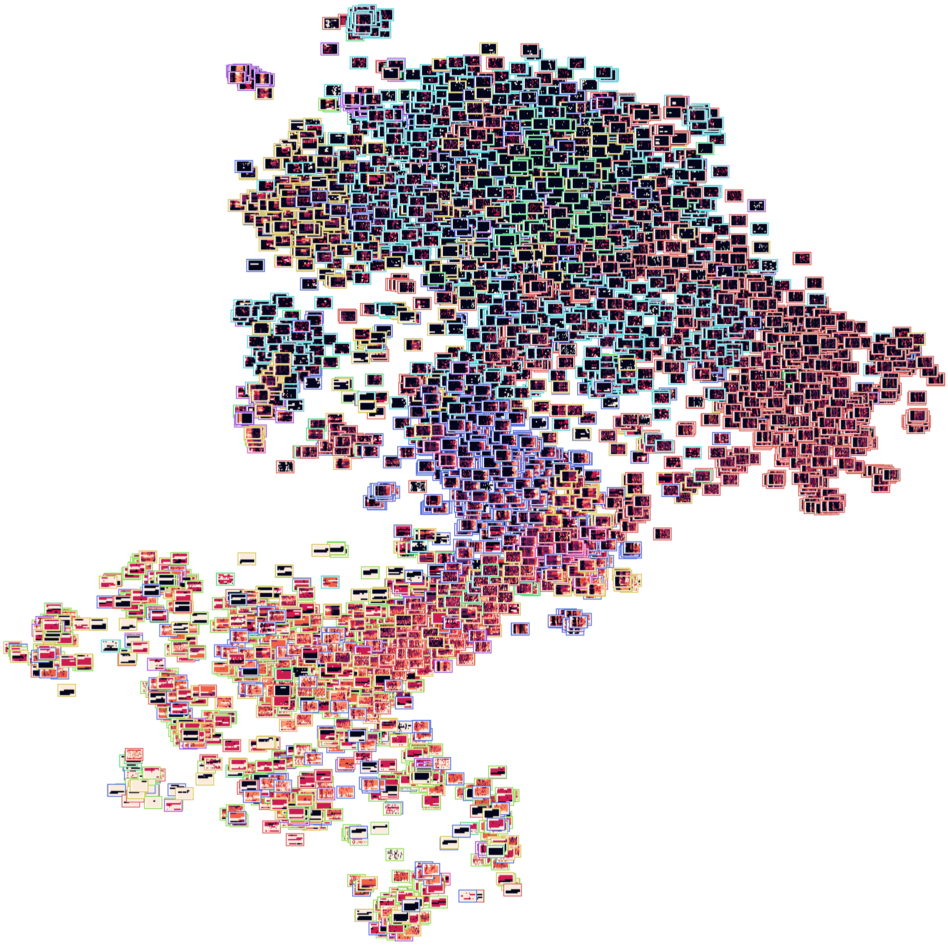
Applying these groups yields figure 7.17. The new plot shows how, although appliances could be used by a different user, maybe even by users in a different part of the EU or world, they can be grouped in a high-dimensional space.

FIGURE 7.17: "Projection of grouped per-appliance load profiles"



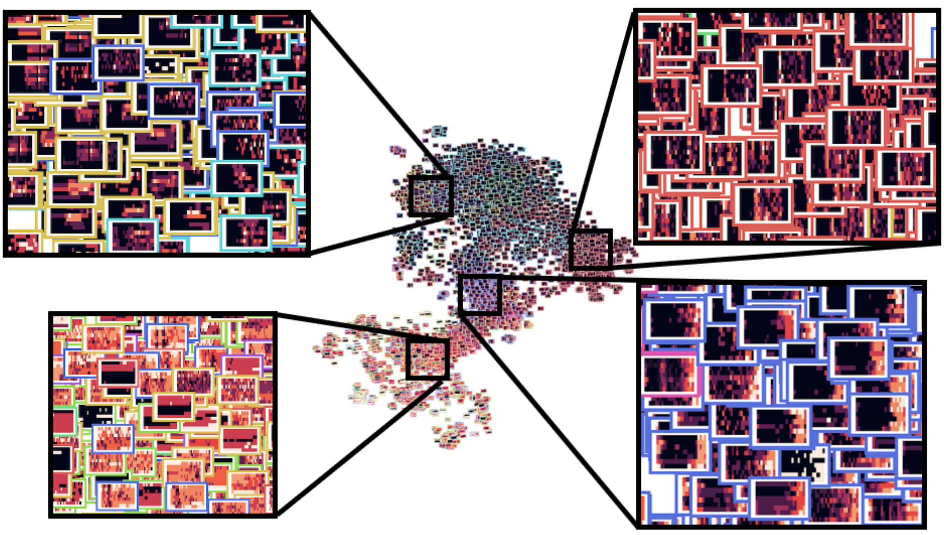
The image below is the same as the first one in the subsection, except it is easier to use color to see the appliance they present

FIGURE 7.18: "Projection of grouped per-appliance load profiles with actual samples"



The figure 7.19 shows the four main types of profiles for readers that do not have the ability to zoom in.

FIGURE 7.19: "Projection of grouped per-appliance load profiles with actual samples"



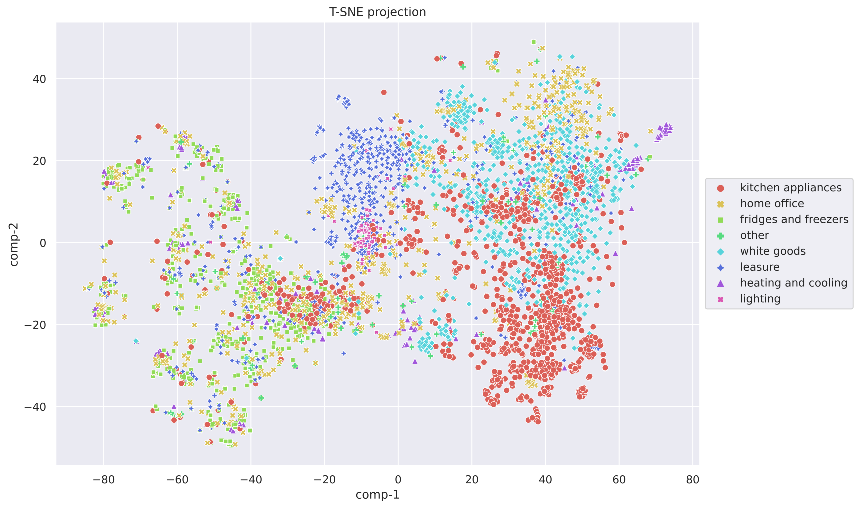
One issue that causes the t-SNE algorithm an issue is low entropy data or in other words, images that are almost completely dark or white, due to various faults in appliances or measurements.

If we calculate 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 7.20: "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 7.21.

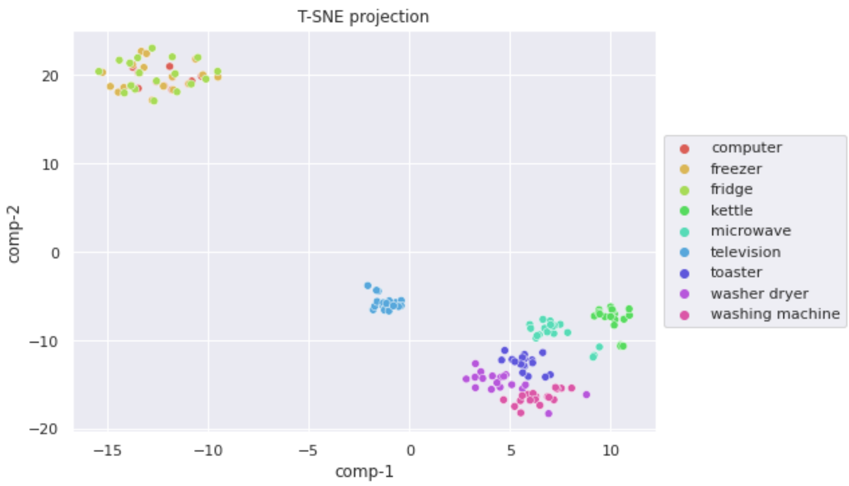
FIGURE 7.21: "Projection of entropy filtered per-appliance load profiles with actual samples"



Comparing appliances in a building

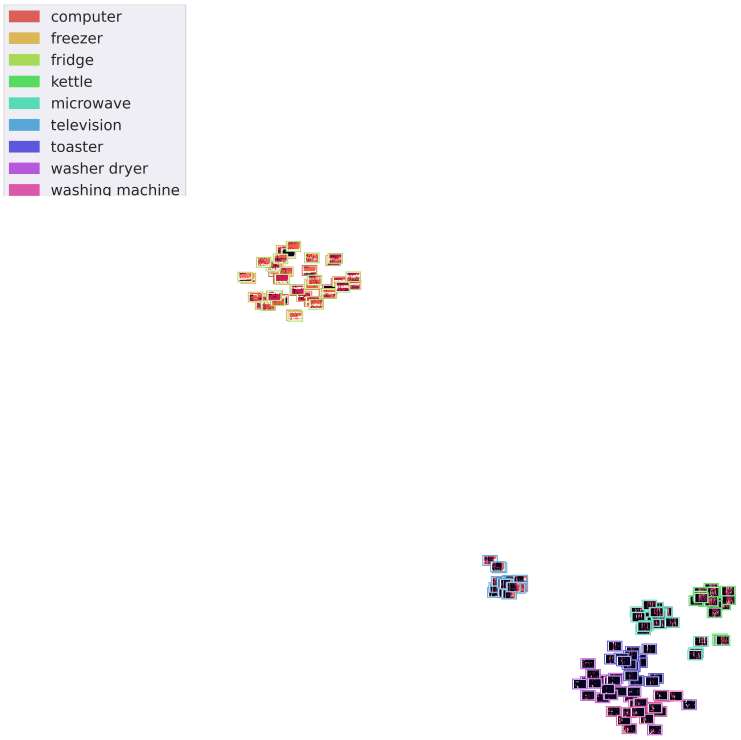
It is also possible to use per-appliance data to study individual buildings, and how each appliance is used.

FIGURE 7.22: "Projection of per-appliance load profiles in a single building"



In general, the scattering is similar to before fridge and freezers are placed opposite of white goods and kitchen appliances and televisions lie somewhere in between.

FIGURE 7.23: "Projection of per-appliance load profiles in a single building with actual samples"



7.4.3 Per-appliance per-building

To study the usage by comparing all appliances between buildings, we have to use one of the proposed load profiles.

Bag off 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, it is cut at 30 Appliances.

The results are appliances in the following order:

- fridge freezer
- freezer
- fridge

- computer
- television
- light
- kettle
- boiler
- broadband router
- microwave
- washing machine
- dish washer
- HTPC
- toaster
- solar thermal pumping station
- audio amplifier
- laptop computer
- active subwoofer
- unknown
- washer dryer
- computer monitor
- audio system
- tumble dryer
- pond pump
- desktop computer
- ethernet switch
- appliance
- modem
- server computer
- electric space heater

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

This is the main reason why we can see in figure 7.24 the clusters are nicely formed but are separated quite a lot. We can still see that some clusters are closer than others, meaning they are more similar.

FIGURE 7.24: "Projection of bag of appliances load profiles for various buildings"

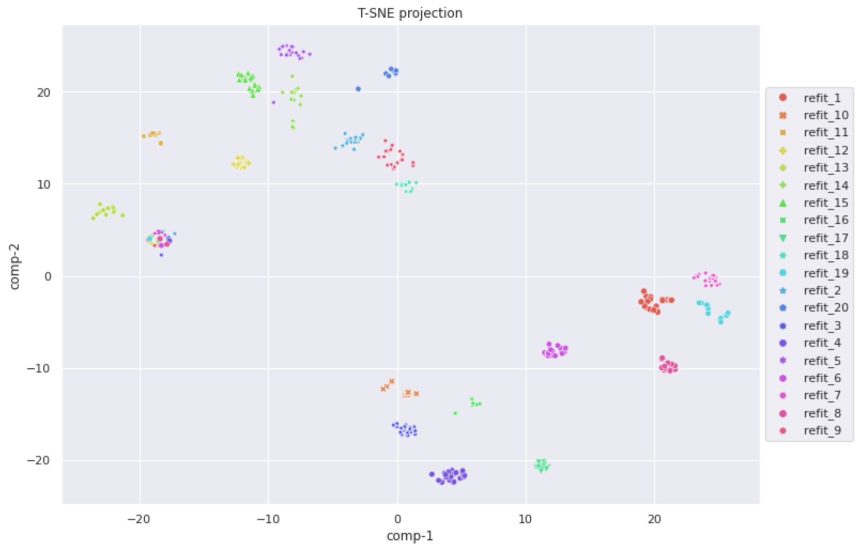
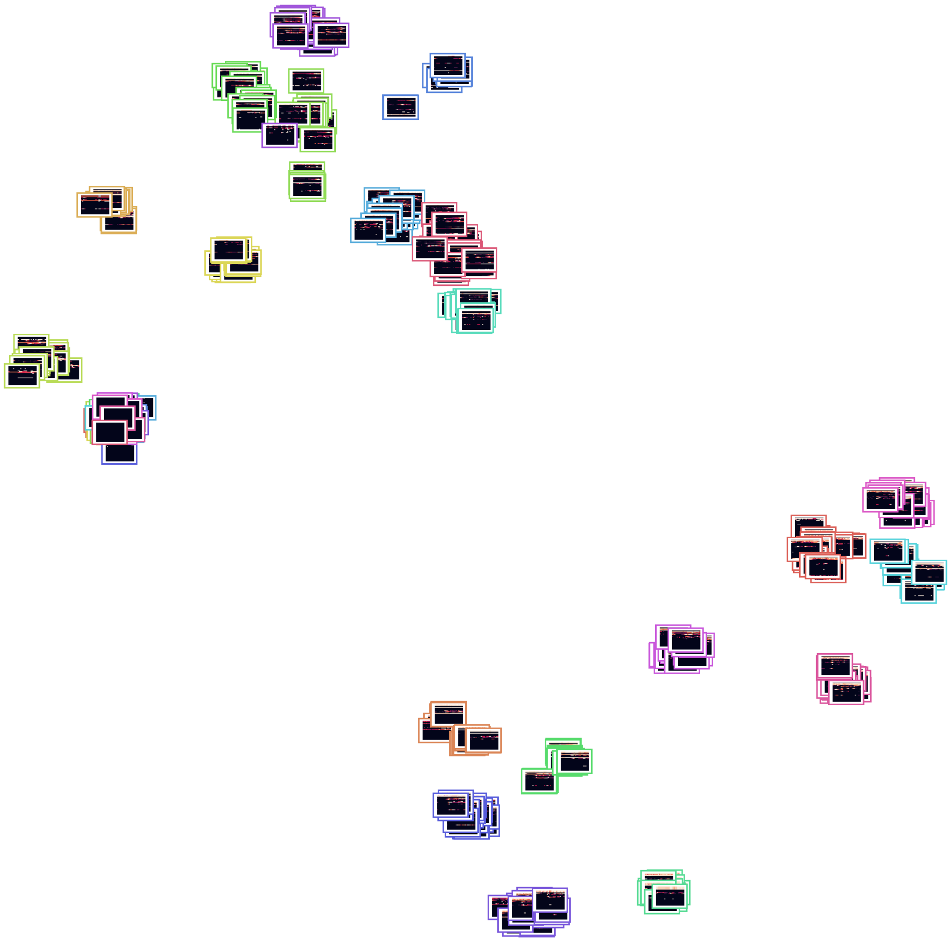


Figure 7.25 shows that load profiles are split between two poles. By observing the figure it is possible to see that the bottom Clusters all have more than one active white good with a compressor (fridges and freezers), while the top ones have only one. In general, the bottom buildings have more appliances, with more activity than the top ones.

FIGURE 7.25: "Projection of bag of appliances load profiles for various buildings with actual samples"



Chapter 8

Elderly care demo

The chapter will first address the issue being solved, then the methodology will be presented, and finally, the results will be disclosed.

Elderly care has been addressed by many EU-funded research projects since the aging population is one of the main issues the union is facing.

There are many solutions to this problem. One approach is invasive such as wearables, sound sensors, IR occupancy detectors, etc. Few authors 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 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 upside of this approach is that the elder does not need to wear the device. The downside is, that new meters need to be installed for most commonly used appliances.

8.1 Goal

The chapter will focus on building the 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 6.11 from the previous chapter 5. Figure shows, that first thing in the morning used are 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.

8.2 Methodology

8.2.1 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 build an anomaly detection.

8.2.2 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 mentioned 6.11. 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 fridge or freezer get ignored. These appliances are detected based on the width of their activation normal distribution. Where periodic (hourly basis) appliances should have narrow distributions, 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 8.1: Daily activations for fridge $\sigma = 0.036$

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

FIGURE 8.2: Daily activations for audio system $\sigma = 0.2$

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

FIGURE 8.3: 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 are rarely used.

Third step

Next, appliances that trigger together must be grouped. This means we must find part of the day that they are operating. 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 8.2. One out 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 same appliances as above the result is 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 8.4: 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 8.5: 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 8.6: Daily activations for audio system

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

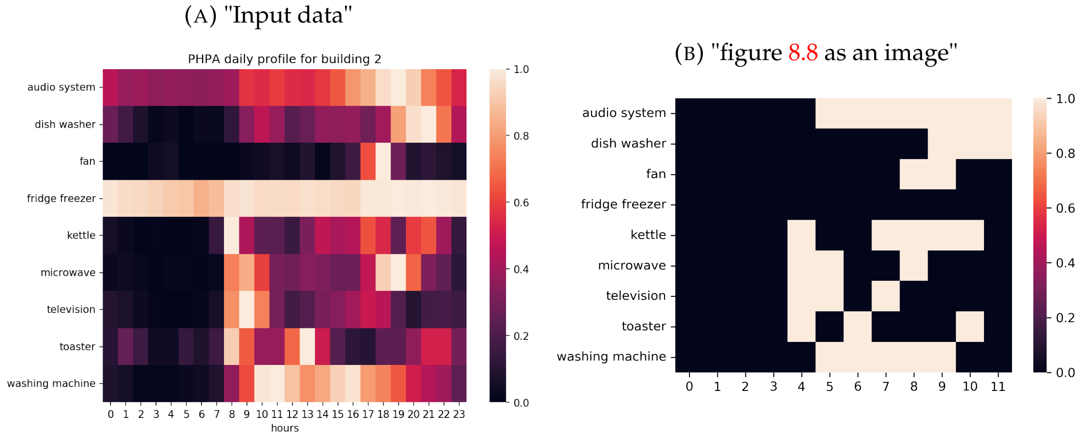
FIGURE 8.7: Daily activations for microwave with one usage peak in the morning and the other in the evening

The vectors show us that microwave has two usage peaks, where the audio system can be used anytime through 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 8.8 uses rows for appliances and columns for buckets. If we use terminology from image processing the matrix 8.8 is essentially a highly saturated load profile 6.11, 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	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	0	1	0	0	0	1	0
0	0	0	0	0	1	1	1	1	1	0	0

FIGURE 8.8: Activation matrix

It is possible to display the matrix 8.8 as an image. The figure below shows how the load profile is transformed.



(A) "Input data"

(B) "figure 8.8 as an image"

FIGURE 8.9: "Trasformation of source load profile to black and white"

Step four

Using the matrix 8.8 we can write an algorithm that will use current activations and compare it to the matrix. For example, if we have the test sample for the fifth bucket, that is data from 8 to 10 o clock. What needs to be done is to use the fifth column and multiply the current activations. Then anomaly is detected based on a rule: that at least two appliances must be activated, for the bucket to be labeled as normal, or else the bucket is labeled as anomalous. In this case, we sum the elements of an array and check if it is larger or equal to 2.

FIGURE 8.10: "Process of evaluating an anomaly"

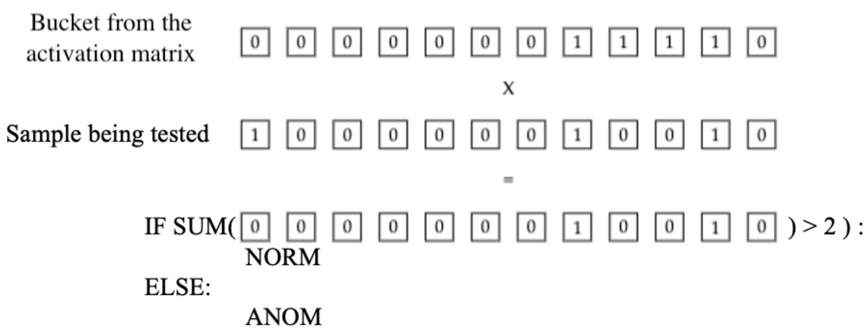


FIGURE 8.11: The evaluation of new bucket compared to matrix. Example is for fifth bucket or fifth row from the matrix

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 which the profile was built.

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

FIGURE 8.12: Agregated anomalies for each bucket

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

FIGURE 8.13: Agregated normal samples for each bucket

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

$$\frac{N_{anom}}{N_{anom} + N_{norm}} \tag{8.1}$$

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

We can alter the equation 8.1 so that it will measure a number of normal samples out of all. The result is the equation 8.2 In other words, we are measuring the strength of a routine that user has in each bucket.

$$R_{outine} = \frac{N_{norm}}{N_{anom} + N_{norm}} \tag{8.2}$$

Using the equation 8.2 we can populate the array 8.14.

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 8.14: Agregated anomalies for each bucket

In other words, the array 8.14 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, which means that the anomaly detection will function better throughout the day.

One more thing to do is to ignore the parts of the day when the user has no routine, by using the array 8.14 and setting a threshold of 0.7.

A threshold of 0.5 would mean that every other day could be a false positive anomaly. Setting the rate at 0.7 reduces that 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 positives 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 8.15: Using above-mentioned threshold a new mask is made, to check only buckets with high routine

Step five

The last step is to repeat steps 4 and 5 with data, that is not included in the built profile.

8.2.3 Datasets and evaluation

The data was split into train and test sets, where 80 % of the data was for training and 20 % percent of the data was used 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

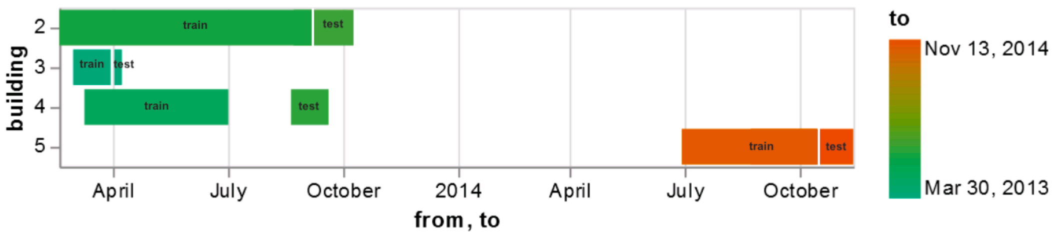
The REFIT Rashid, 2019 dataset included data for more than 15 buildings, as can be seen on the figure below. The dataset in general is of the highest quality since it is the longest with the least missing data. This means this dataset should give the most relevant results.

FIGURE 8.16: "Timeline for REFIT "

UK-DALE

All through the UK-DALE Kelly and Knottenbelt, 2015 dataset is of similar size, the 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 on 8.17 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.

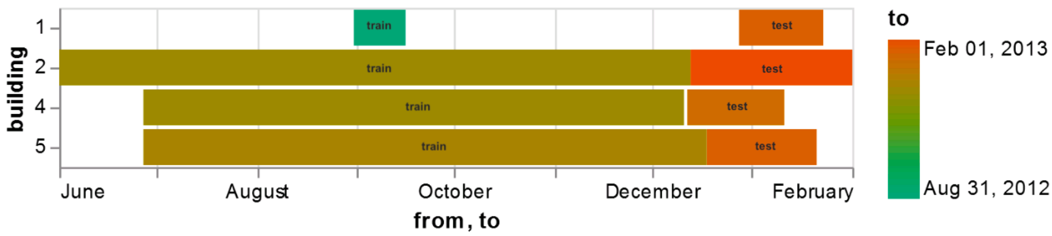
FIGURE 8.17: "Timeline for UK-DALE "



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 8.18: "Timeline for ECO"



8.2.4 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 8.14. 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 being "practiced" every second day
- routine rate of 0.8 would mean that routine is broken on average every 5 day
- routine rate of 1 would mean that this household has a routine that is never broken.

When a true anomaly occurs such as a fall, the dweller, all through he had the same strong routine for the past year (the routine rate would be close to 1), would not be able to practice the routine, 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.

8.3 Results

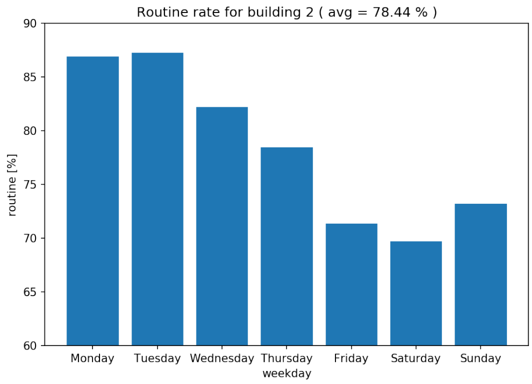
8.3.1 The routine rate over a period of time

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 emerge. To present these patterns, training data will be used. This enables us to see how routine changes over the year.

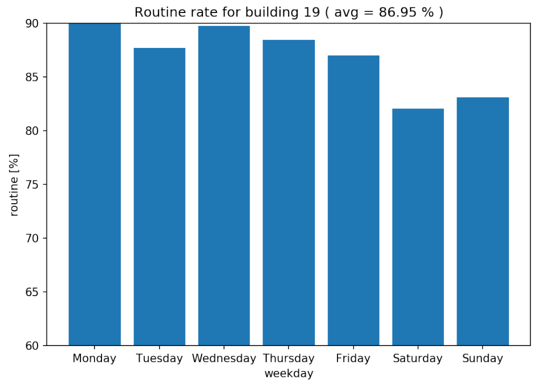
The routine rate through the week

As the behavior of the dwellers 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 figures below.

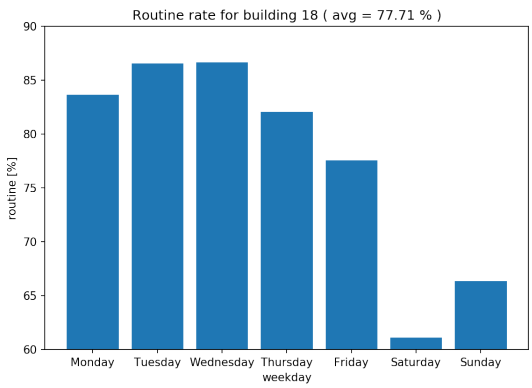
(A) "Building 2"



(B) "Building 19"



(C) "Building 18"



(D) "Building 5"

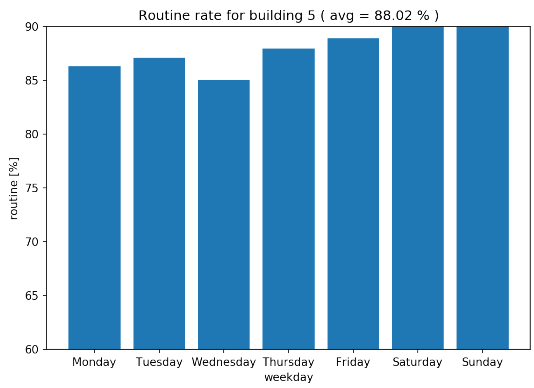
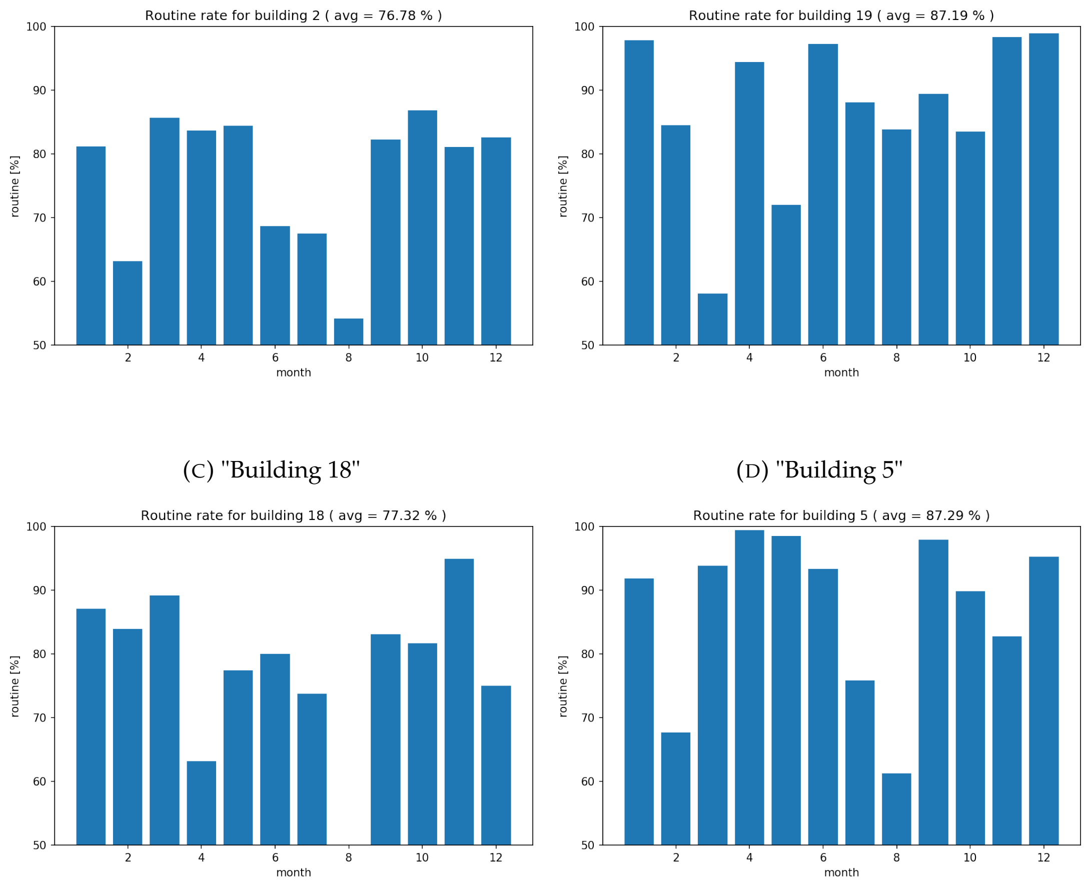


FIGURE 8.19: "Routine rate through the week (train data)"

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.

8.3.2 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 summer, it is lower during the summer due to vacation. This can be seen in the figures below.



(A) "Building 2" (B) "Building 19" (C) "Building 18" (D) "Building 5"

FIGURE 8.20: "Routine through the year (train data)"

Since we are observing how routine changes, and not the actual performance, training data was used. This is because there train we have more than a year of train data.

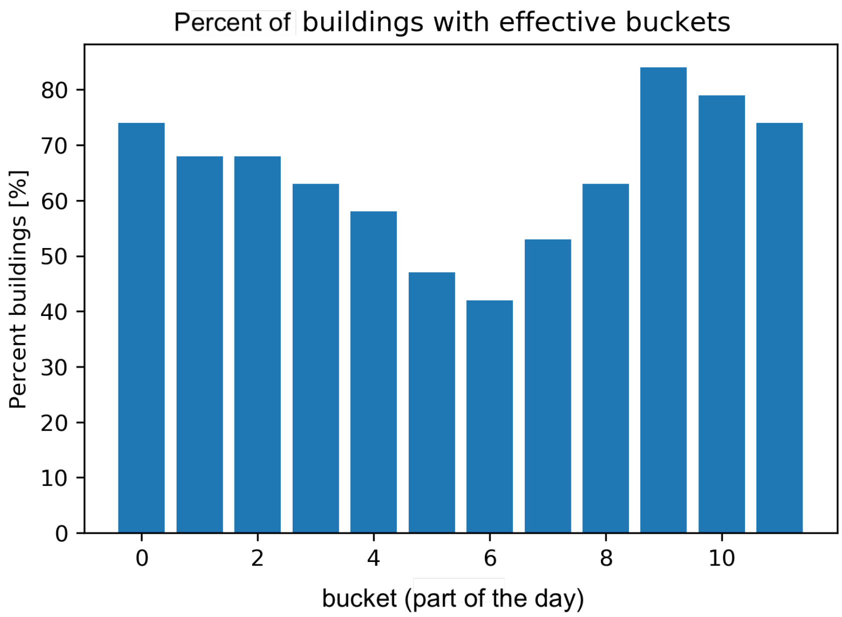
8.3.3 Effectiveness of anomaly detection through the day

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

Figure 8.21 shows which buckets were 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 through the day.

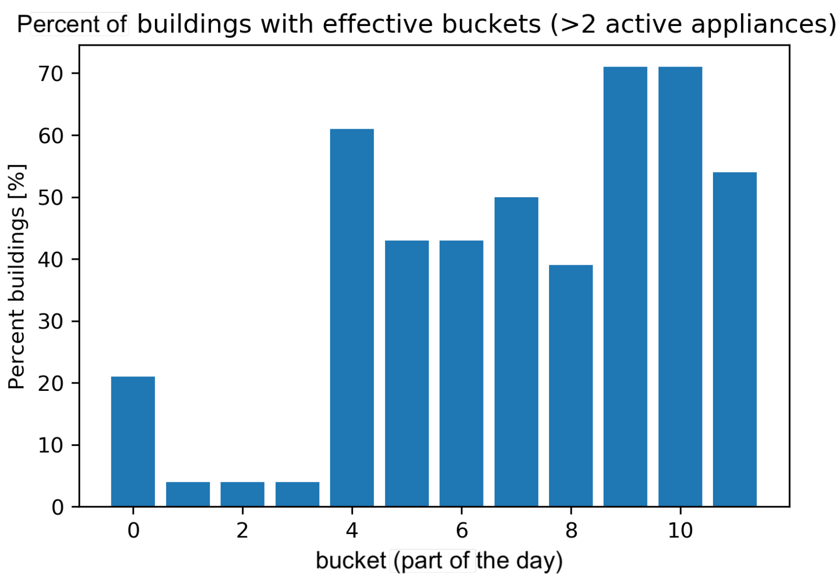
This means that the higher the routine, the higher the chance that this bucket will be used for anomaly detection. During the night, it is possible to see that the average routine rate is quite high. This can be seen in figure 8.21 this is because most users are routinely sleeping during that period. But as we can see in figure 8.21, the high routine rate does not necessarily mean the buckets are useful.

FIGURE 8.21: "Effectivnes 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 8.22

FIGURE 8.22: "Actual effectiveness of anomaly detection through the day"



The figure 8.22 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. This is because the average person is at school or work during noon. This can be confirmed on figure 8.22. 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.

The main issue here 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.

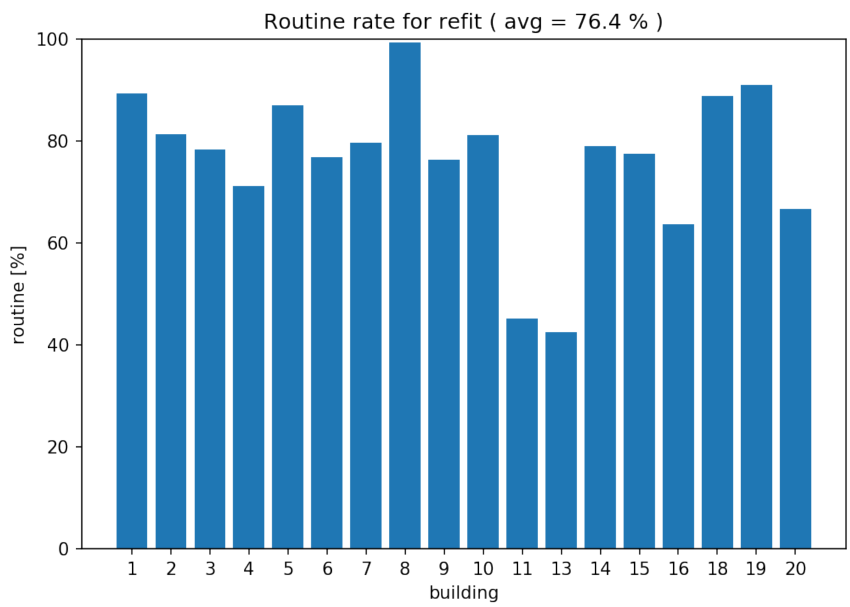
Results were obtained for 3 datasets. REDD and iAWE datasets were not used, since the dataset was too short. They contained less than a month of data.

8.3.4 Per-building results

REFIT

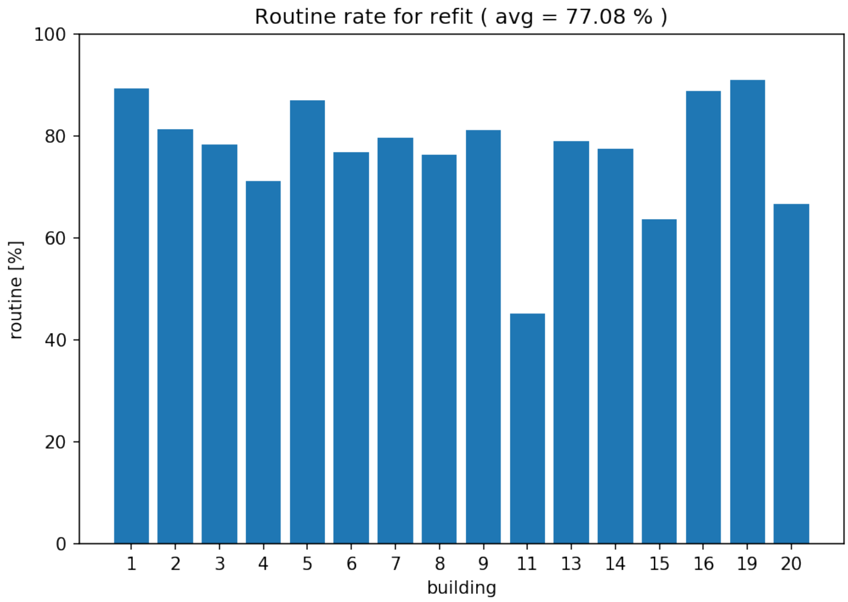
Results show, that the method is on average 76.4 % efficient for REFIT. In figure 8.23 it is possible to see that house 6 yields much better results and house 13 much worse than the rest. This could be due to various dataset errors that occurred during sampling.

FIGURE 8.23: "Results for REFIT "



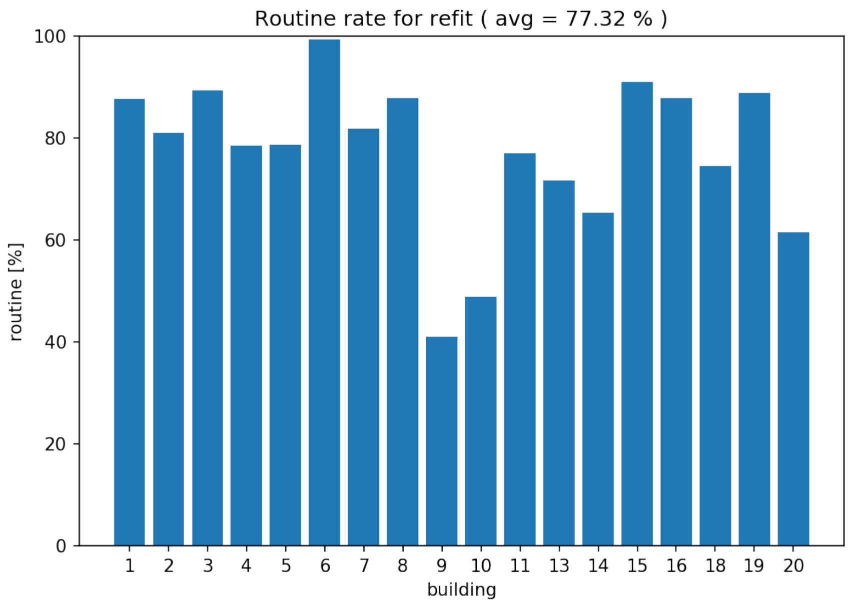
For more relevant results we can ignore the outliers by removing one maximum and minimum value, such as can be seen in figure 8.24 This yields a result of 77.08 %. If we were to repeat this process the result would be 79.77 %. Since all outliers are removed, the result converges towards 79 %, which is the relevant value.

FIGURE 8.24: "Results for REFIT with removed outliers "



As mentioned in the sub-sub section 8.3.1, the average routine is different during the week and 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 on figure 8.25 show that result improved to 77.08 %.

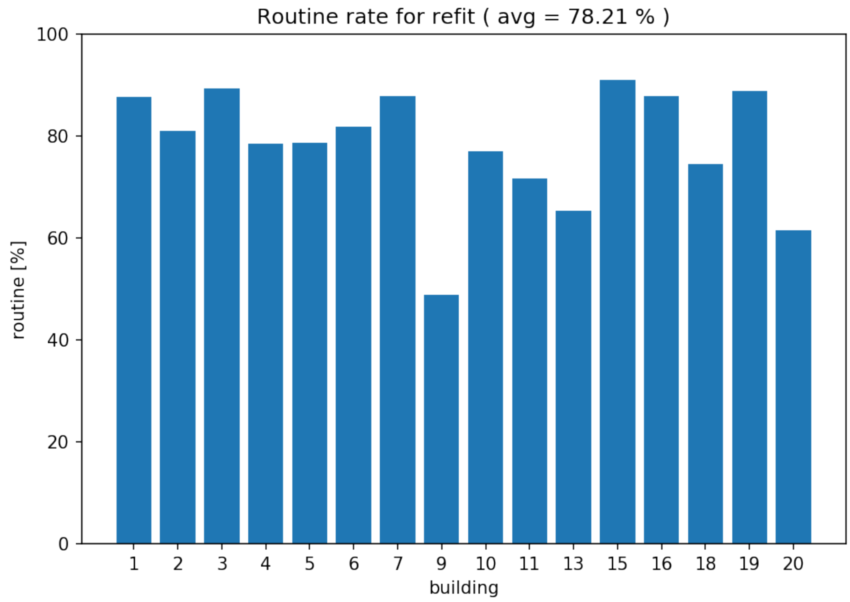
FIGURE 8.25: "Results for REFIT weekday only"



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

By removing the weekend data, the results improved by 1.2 %.

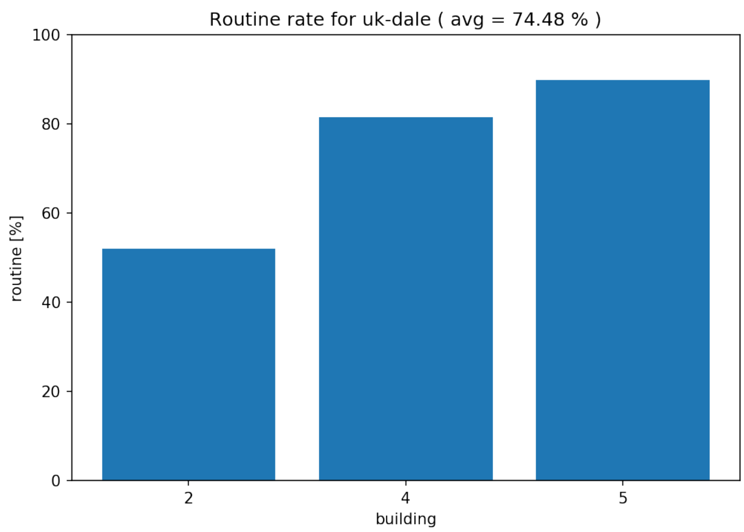
FIGURE 8.26: "Results for REFIT weekday only and removed outliers"



UK-DALE

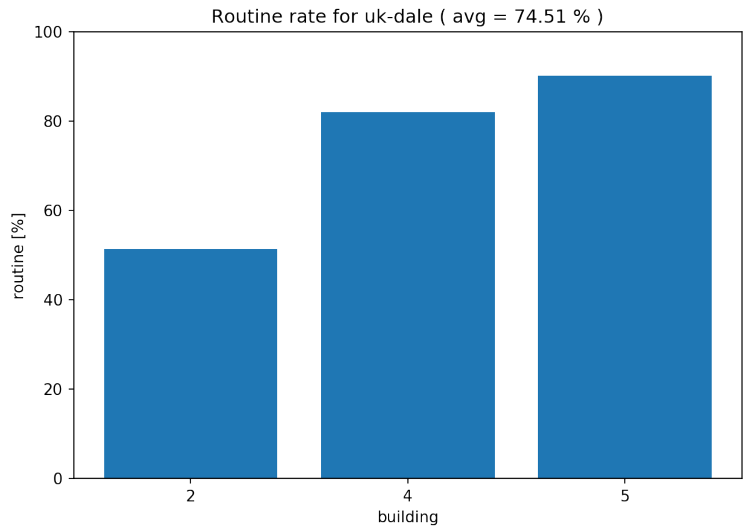
As mentioned in subsection 8.2.3, the UK-DALE is not as big and cleaned as the previous dataset, so the results could be less relevant. The results on figure 8.27, show that the average result is 74.48 %. Due to the low number of buildings, it is not possible to detect and ignore outliers.

FIGURE 8.27: "Results for UK-DALE"



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

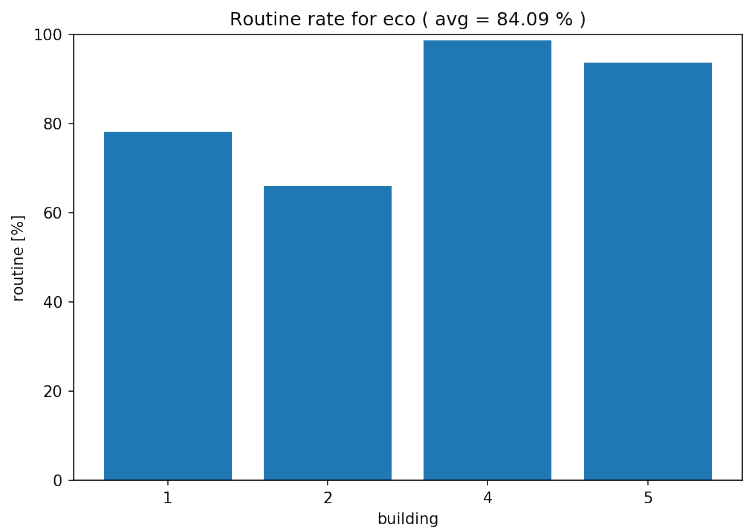
FIGURE 8.28: "Results for UK-DALE omitinng weekends "



ECO

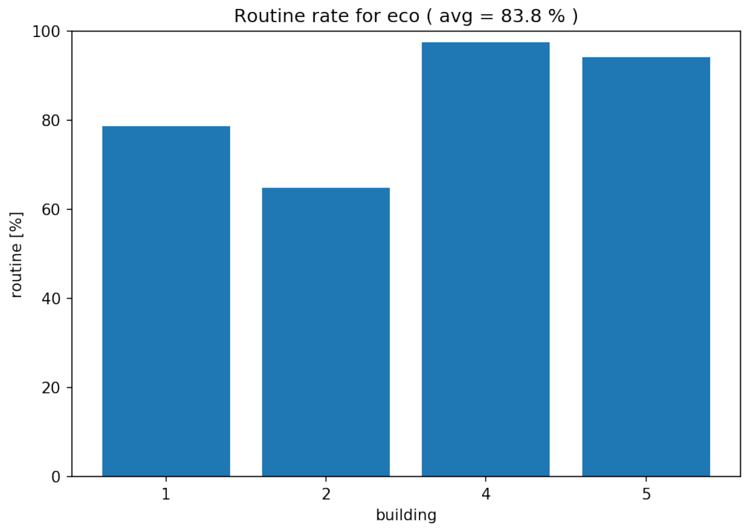
ECO is of a similar quality as UK-DALE regarding the number of buildings and the length of data. This can be seen in 8.2.3 The results 8.29, show that this dataset performed the best, with results of 84.09 %.

FIGURE 8.29: "Results for ECO "



The same as before we can omit weekend data, which can be seen on 8.30. This brings the result down to 83.80 %.

FIGURE 8.30: "Results for ECO omiting weekends "



8.3.5 Combined results

After combining results from all 25 buildings, the table 8.1 can be populated. The most relevant results can be seen in the last row.

TABLE 8.1: Combined percentage of anomolous samples [%]

N = 25	Including weekend data		Excluting 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.

8.4 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 do dwellers have. Socio-economic status encompasses attributes such as age, income, geo-location, 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 in use 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 on figure 8.21. 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.

8.5 Iterative learning system

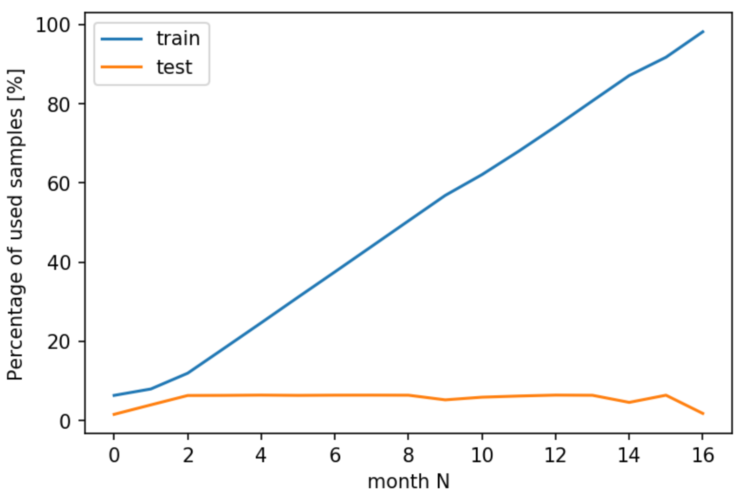
In the case of practical use of this algorithm, it is important the system is put online as fast as possible and that it improves over time. This can be achieved with the implementation of iterative learning. The system will build a load profile based on the first month of data. Using this load profile, 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.

8.5.1 Results

For this evaluation, only REFIT Rashid, 2019 data was used. As it can be seen on figure 8.16, Refit buildings have long and relatively similar timelines, compared to other datasets.

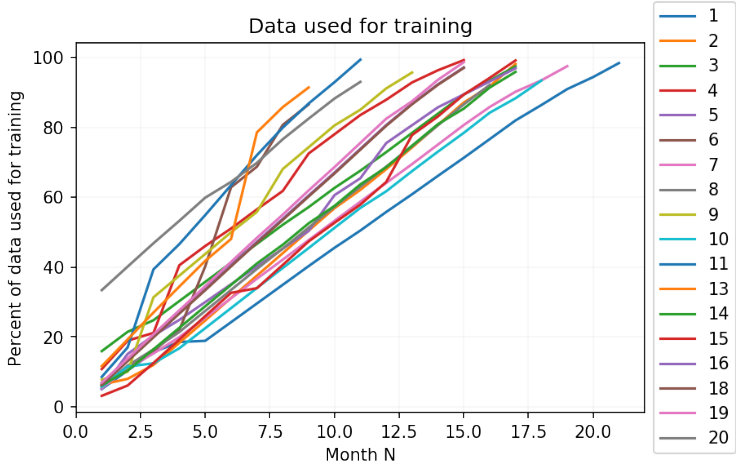
On figure 8.31 it is possible to see, how the amount of training and testing data changes over 16 months.

FIGURE 8.31: "Data for building 1 over 16 months"



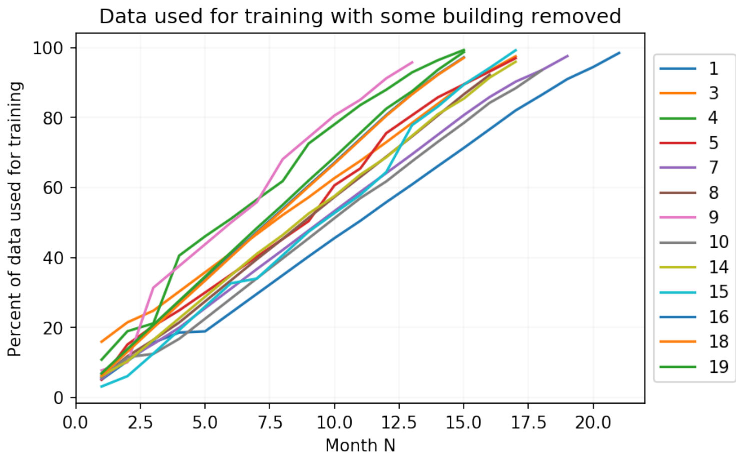
We can also plot how the amount of data changes for all buildings. This can be seen on figure 8.32.

FIGURE 8.32: "Data used for training"



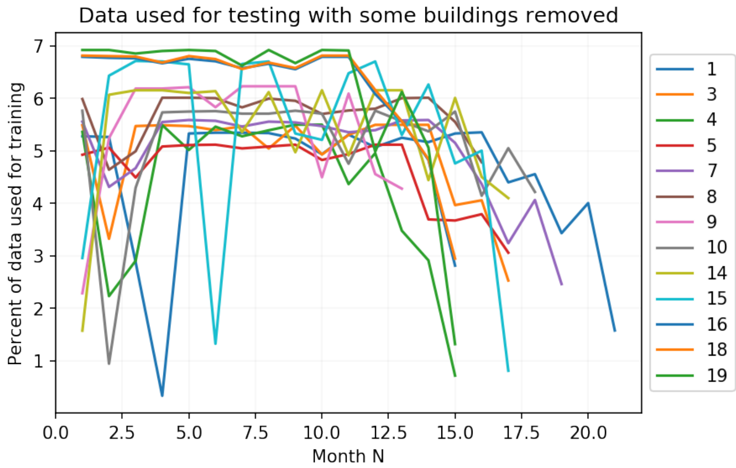
To analyze the results, at least 1 year of usable data should be available. Figure 8.33 shows only buildings contain at leas one year of data.

FIGURE 8.33: "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 8.34 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 8.34: "Data used for training, with removed buildings"



To show the effect of training data on the metric, the figure 8.35 is presented. The figure 8.35 contains 12 months of data for each house, since some have 17 months of data, they never reach the 100 % mark.

FIGURE 8.35: "Effect of new data on metric"

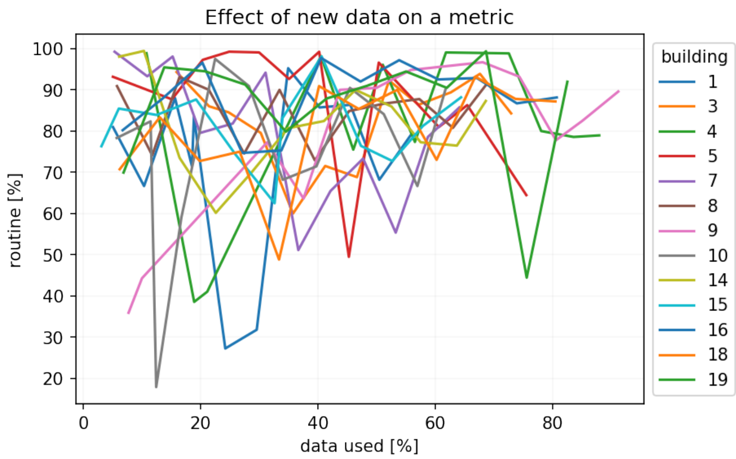


Figure 8.35 shows, that in most cases, results converge towards 80 %. In some cases, the results are good from the beginning, but sooner or later the routine rate will dip. With more data, these dips become smaller and less frequent. If the behavior in the household radically changes, it can still lead to a dip.

FIGURE 8.36: "Metric over 12 months"

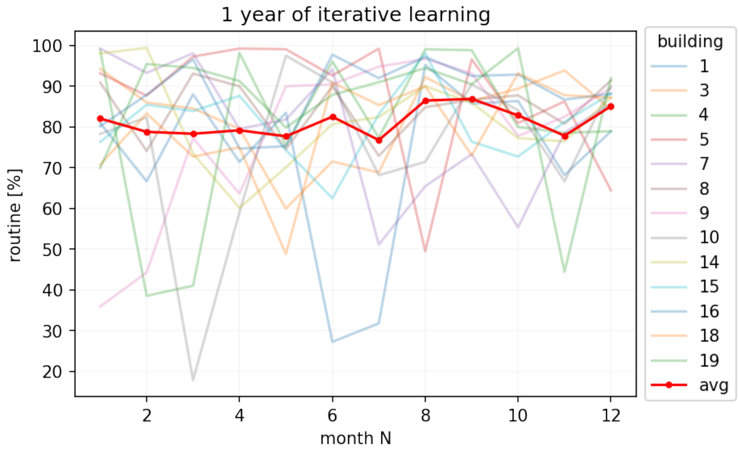


Figure 8.36 shows how the same data can also be presented so that it shows how the metric changes over a year.

8.5.2 Discussion

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 could need 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.

8.5.3 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 9

Conclusion

In contributions chapter 5, it was said that the goal will be achieved by answering the three questions:

- How do we efficiently present big data to humans and machines?
- How is this presented data connected in a higher dimension?
- Can we use one of the profiles to build something useful?

By answering the first question 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 5.

The second was answered in chapter 7, where we have shown how data is connected in high dimension space using t-SNE for dimensionality reduction.

The last was answered in chapter 8, 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.

While we have filled in a few gaps in the table of profiles, it is up to scientific commonly to fill in the rest.

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. 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.
4. 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.
5. 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.
6. Bitterer, R. and B. Schieferdecker, Prof. Dr. habil. (1999). *Repräsentative VDEW-Lastprofile Aktionsplan Wettbewerb, M-32 99*. VDEW.
7. Boßmann, Tobias, Schleich, Joachim, and Schurk, Robert (June 2015). "Unravelling load patterns of residential end-uses from smart meter data". In:
8. 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.
9. 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.
10. 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.
11. 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.
12. 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.

13. 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.
14. Commission, Communication from the et al. (2006). “Action Plan for energy efficiency: realising the potential”. In: *COM (2006)* 545.
15. 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.
16. 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.
17. 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.
18. 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.
19. 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.
20. 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.
21. 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.
22. 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.
23. 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.
24. 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.
25. 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.

26. 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.
27. 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.
28. 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.
29. Klaassen, Elke, Frunt, Jasper, and Slootweg, Han (June 2015). “Assessing the impact of distributed energy resources on LV grids using practical measurements”. In:
30. Kleiminger, Wilhelm et al. (Nov. 2013). “Occupancy Detection from Electricity Consumption Data”. In: pp. 1–8. DOI: 10.1145/2528282.2528295.
31. Kolter, J. Zico and Johnson, Matthew J. (2011). “REDD: A Public Data Set for Energy Disaggregation Research”. In: *IN SUSTKDD*.
32. 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.
33. 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.
34. Maaten, L.J.P. van der and Hinton, G.E. (2008). “Visualizing High-Dimensional Data Using tSNE”. In:
35. 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.
36. 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.
37. 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.
38. 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.
39. 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.

40. 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:
41. 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.
42. 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.
43. 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.
44. 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.
45. 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.
46. 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.
47. 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.
48. shuma-iwisi, Mercy (June 2009). "Domestic appliances standby power losses: The case of eleven suburbs in the greater Johannesburg," in:
49. 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.
50. 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.
51. 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.
52. 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.

53. 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.
54. 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.
55. 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.
56. 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.
57. 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.