UNIVERSITY OF LJUBLJANA

MASTER'S THESIS

Development and Analysis of new Activation Based Load Profiles

Author:
Jakob JENKO, BSC

Mentors:
doc. dr. Marko MEŽA and
Dr. Carolina FORTUNA

A thesis submitted in fulfillment of the requirements for the degree of Masters of Electrical Engineering

April 8, 2023

“In science, great oaks grow from little acorns.”

D. Everett

Abstract

This work explores the potential of electrical energy data and how load profiles can be used to address issues such as the optimization of electrical energy consumption patterns and the aging population. The efficient presentation of energy data through load profiles is a constant narrative throughout the thesis. Optimizing consumption has the potential to significantly reduce the human footprint since a third of electrical energy in the EU is consumed in the residential sector. Furthermore, we utilize load profiles to address issues such as the aging population. We developed an elderly care assisted living system to detect anomalies in the usage patterns of the elderly. The system identifies accidents such as falls, strokes, or dementia-induced altered behavior.

We performed a comprehensive review of existing publications and use-cases. These publications were mapped into a table, which revealed gaps in the load profiles that were not yet researched or used. Next, we analyzed the load profiles and using t-SNE presented how profiles are related in high dimensional space.

With the successful implementation of the elderly care system, we confirmed that unused load profiles are applicable. The findings of this thesis showcase the untapped potential of energy data where the table of profiles provides a foundation for further research in this area.

Keywords: load profiling, energy data, energy saving, dimensionality reduction, elderly care, anomaly detection

Povzetek

V tem delu raziščemo možnost uporabe profilov porabe električne energije za naslavljanje ovir samostojnega bivanja starejšega prebivalstva . Osrednja tema magistrske naloge je učinkovita predstavitev podatkov s pomočjo profilov porabe. Optimizacija porabe energije lahko bistveno zmanjša ogljični odtis človeka, saj se v Evropski uniji tretjina električne energije porabi v gospod sektorju.

Opravili smo obsežen pregled obstoječih publikacij in primerov uporabe. Publikacije smo prikazali v tabeli, ki je razkrila vrzeli profilov, ki še niso bili raziskani ali uporabljeni. Nato smo analizirali profile obremenitve in s pomočjo t-SNE predstavili, kako so profili povezani v visokodimenzionalnem prostoru. Z novo prodobljenim znanjem smo razvili sistem za oskrbo starejših oseb, ki jim lahko pomaga podaljšati samosotjno bivanje. Sistem preko analize profilov zazna anomalije v porabi električne energije, ki so pri starejših lahko posledica padcev, kapi ali spremenjenega vedenja zaradi demence.

Z uspešno implementacijo sistema za oskrbo starejših smo potrdili, da so do sedaj neuporabljeni profili lahko uporabni. Ugotovitve te magistrske naloge prikazujejo neizkoriščen potencial podatkov o energiji, kjer tabela profilov predstavlja osnovo za nadaljnje raziskave na tem področju.

Ključne besede: profiliranje porabe, energetski podatki, učinkovita poraba, zmanjšanje dimenzionalnosti, oskrba starejših, zaznavanje anomalij

Acknowledgements

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

Contents

List of Figures

List of Tables

List of Abbreviations

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

Chapter 1

Introduction

Climate change calls for a shift to renewable energy and restructuring of the electric power industry. Source [21] shows that 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 carbon dioxide emitter. The same source [21] also states that a third of energy is consumed by the residential sector. It is estimated, that the human population will reach 10 billion inhabitants in the next 10 years, and ever-increasing ownership of electrical appliances such as smartphones, HVACs, and EVs will further elevate this issue. Acknowledging this, reducing consumption in the residential sector could leave a significant impact on the human footprint.

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

This could be achieved through various applications and methods that use load profiling as their core technology. Authors in paper [17] proposed a method to reduce peak loads by studying consumer appliance usage patterns. Paper [20] 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 [45]. All described methods would reduce and alleviate the load off the power grid.

Load profiling in building energy consumption is not a novelty and had been in research since the 1980s. While it was thought that aggregated LPs of households are relatively predictable, recent data obtained using smart meter data showed large deviance from user to user due to different lifestyles, as the author states in paper [52]. In recent years LPs have changed due to renewable energy accelerated development of distributed energy resources such as residential photovoltaic power plants, home wind energy, and using EVs with home batteries. Socioeconomic changes such as work-from-home, also drastically reshaped the LP curve.

The thesis aims to propose and develop new, previously unused LPs, that will contribute to mitigating the raised issues. Presenting consumption with the right LPs, will help dwellers be more aware of their consumption and in terms increase their energy efficiency. Energy efficiency is the basis of our research, throughout the thesis we will explore LPs that were not yet utilized. We will perform an EDA to make sense of what they are, and what information they contain. This obtained knowledge will be used in a practical use-case to showcase that these profiles can be effectively utilized. We will design an elderly care assisted living monitoring system to detect anomalies in consumption patterns to detect strokes and falls.

Before we fully disclose our contributions, let us first have an overview of what LPs are and in which other use cases they can be utilized, besides the ones just mentioned.

1.1 Definition and Types of LPs

Author Proedrou [52] defines terms as following:

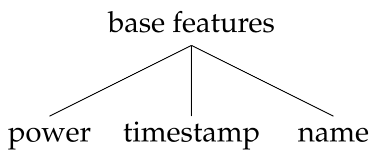
- Load: the electricity that all the electricity-powered devices in the household consume in unit time.
- Profile: a graph representing the significant features of the electricity load over time.

In other words, LPs are a graphical presentation of the consumption features of a building over time. Here, features could be anything that presents consumption. In most cases that is power. The time range used to present the consumption could be anything from daily, weekly, monthly to yearly.

One thing to mention here is, that although the buildings were mostly consuming energy in the past, nowadays, they also produce it. While this may slightly alter the definition of LPs, it also makes them more useful, as they can now be used to depict both energy consumption and production. Throughout this thesis, we will primarily focus on the use of load profiles to represent electricity consumption, but it's important to consider their potential for presenting energy production as well.

1.1.1 Feature Set

To identify the fundamental features of energy consumption in buildings, we need to examine the way that consumption is typically measured There are three main features that allow us to determine the amount of energy being used by a user:



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

We can also extract features such as the number of activations or time of operation for each activation. This can be done using sensors to detect activity or even extract this from power consumption data. In cases where we are observing individual appliances, this can be done using simple signal processing techniques. In cases where we are observing buildings, this could be achieved using more complex disaggregation algorithms also known as NILM (non-intrusive load monitoring) algorithms. NILM algorithms enable us to detect consumption patterns of multiple appliances from a single power meter.

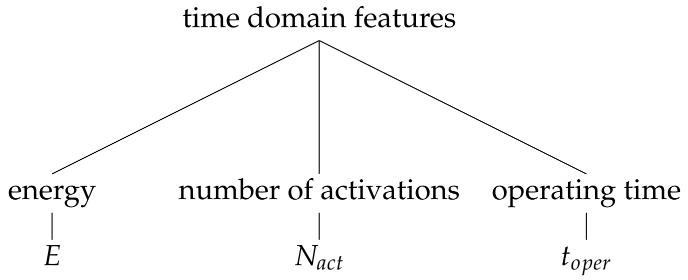
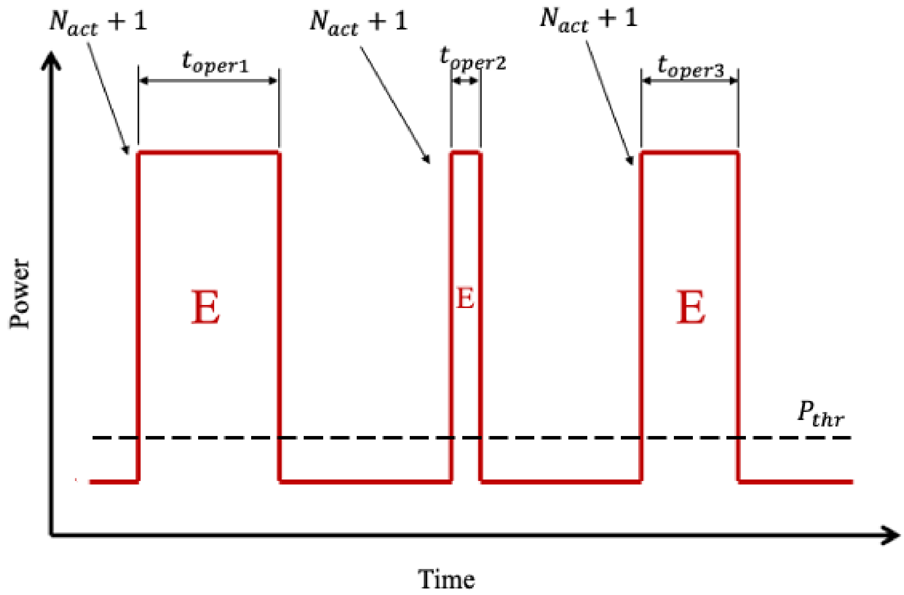


FIGURE 1.1: Simple signal processing of power consumption for a single appliance



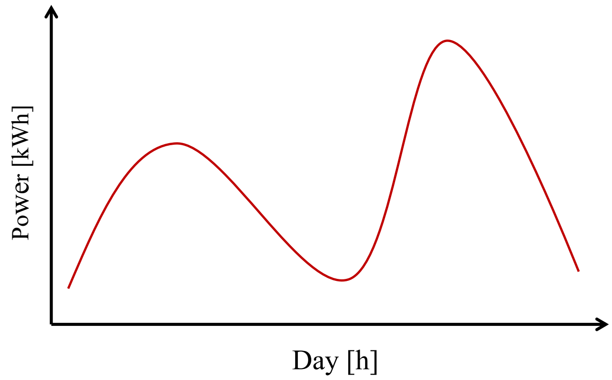
As we can see in Figure 1.1 all three-time domain features can be extracted from the graphical presentation. Energy E is equal to the area under the graphical presentation or in other words integral of power over time. N_{act} can be measured based on the number of times the power value exceeded some pre-defined threshold P_{thr}. The t_{oper} is the time between on and off events, where we use the same threshold as with N_{act}. While there are other features, such as time between activations, or total operational time that could be extracted, these were not commonly used in related work.

1.1.2 Types of LPs

Power LP

Combinations of the features result in many possible types of LPs that enable us to present the data. The most commonly used type of LP is average power consumption over some time. One such example can be seen in Figure 1.2. Here, we used daily timescale, since it is so commonly used it is also known as the standard daily LP. This LP can be used to portray per-building as well as per-appliance data. Its use is one of the most versatile, and it is used in fields such as demand response, anomaly detection and zero-energy buildings. While the LP in Figure 1.2 is a sketch, it still presents consumption trends in morning and evening peaks.

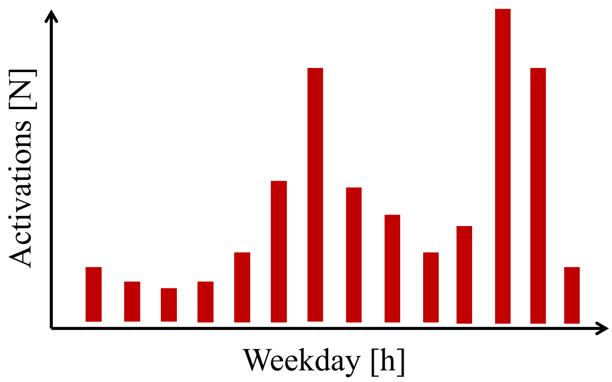
FIGURE 1.2: Average daily usage profile for an appliance or a building



Activation LP

Alternatively, we can use a histogram-based presentation to present a number of activations feature such as can be seen in Figure 1.3. Here, we split the given timescale into discrete intervals also known as buckets. These buckets are then filled with activations that had taken place in a given interval. In the case of Figure, 1.3 timescale is a day and it was split into 12 intervals. While this is not real-world data we can again observe consumption patterns throughout the day, with morning and evening peaks. Activation LP is usually used to portray per-appliance data. In order to portray per-building data, we would need to install a power meter for every appliance in the building. This LP has the very same use-cases as the power type and can be used in the same fields, but as mentioned, it is less practical for per-building LPs. While Figure 1.3 presents the same data as Figure 1.2, due to data processing, it could potentially reveal more relevant consumption patterns. The downside is that we have to invest additional time to process power data into activations.

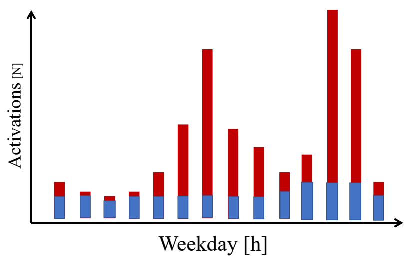
FIGURE 1.3: Histogram of daily activations profile for an appliance or a building



Per-Building Per-Appliance LP

The next two types of LP are known as per-building per-appliance LPs. They can be used to present whole building usage as well as per-appliance consumption. Two such examples can be seen in Figure 1.4a and 1.4b. In the case of activation LP, we concatenate activations of each appliance and label them accordingly. This enables us to see the contribution of each appliance to the total number of activations in each bucket. The per-building per-appliance LPs have the potential to be used in the same fields as the first two types, but practically they are only used in demand response. Additional information enables us to understand the data and potentially discover patterns that we otherwise would not be able to. The reason behind slow adoption could be that this additional information is not needed.

(A) Daily per-building per-appliance activations LP for appliances A and B



(B) Daily per-building per-appliance power LP for appliances A, B and C

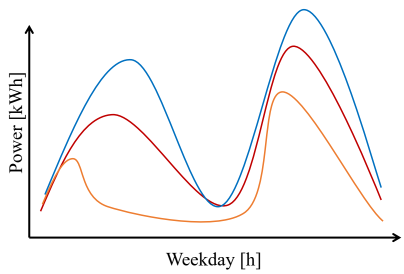


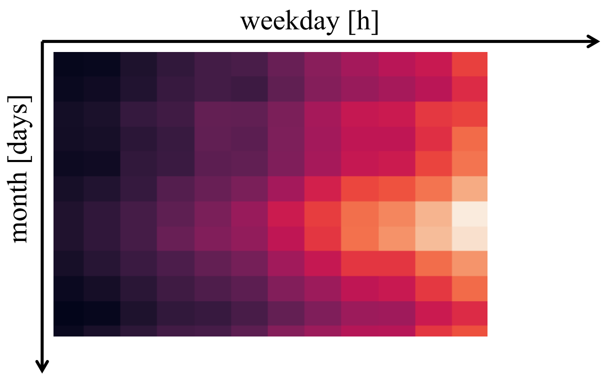
FIGURE 1.4: Per-building Per-appliance LP

Heatmap LPs

The last type is a heatmap LP. Heatmap LPs can be further divided into two sub-types.

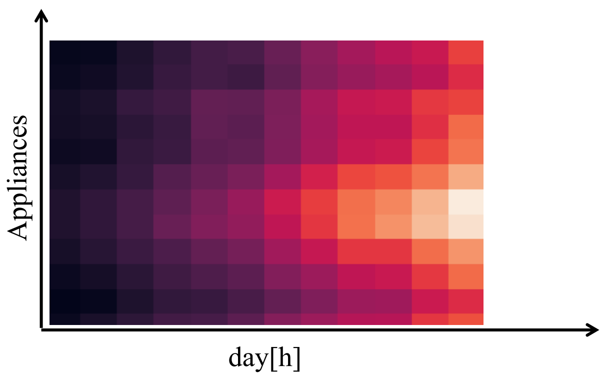
The first type is LPs which consist of two-time dimensions and use color to display consumption. This LP can be used to portray activation as well as power consumption features and could be used to present per-building consumption as well as per-appliance data. All this makes them very versatile. One such example can be seen in Figure 1.5. It is possible to see the consumption pattern throughout each day in a month. The brightness presents the activity of the household or an appliance. The brighter the plot, the more activity for that hour of that day of the month. One other thing to keep in mind when reading such a profile is that the origin is placed in the upper left corner. This originates from image processing standards.

FIGURE 1.5: Number of daily activations/power consumption of one appliance/house in one-month period



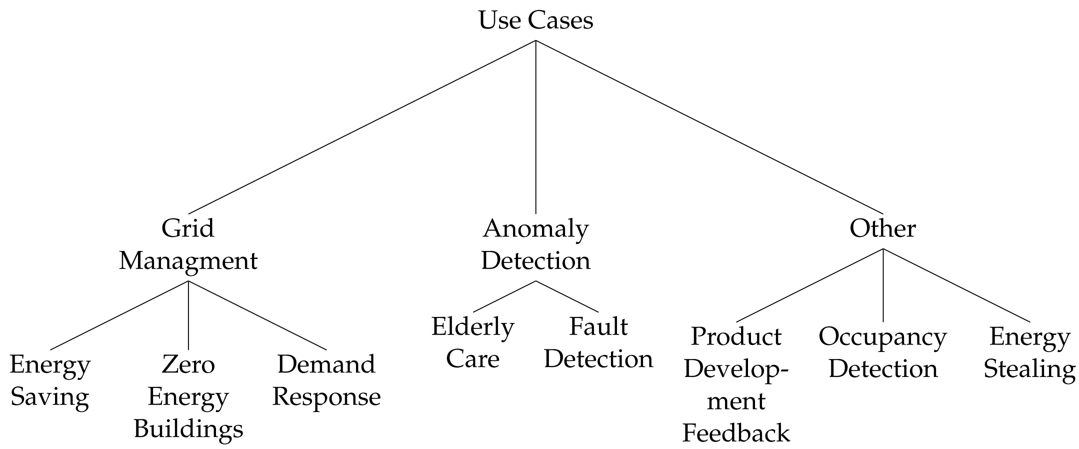
The second subtype is essentially Per-building Per-appliance LP but portrayed differently. Instead of plotting consumption data as the sum of contributions of each appliance, we plot their consumption by side. These LPs have the same uses as Per-building Per-appliance LPs since they are essentially the same. One such sketched example can be seen in Figure 1.6.

FIGURE 1.6: Consumption for each appliance in a day



While there are many features and many more types of LPs out there, we have selected the ones that are most commonly used. There are also many versions of the LPs above with different timescales, where each has a different use case. A more detailed presentation of use cases will also follow in the coming chapters, with a classification in the next section.

1.2 LP Use-cases



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

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

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

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

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

1.3 Data

To construct the LP one needs time-series data that contains information about the consumption of the energent. While LPs are generally used to depict usage of electical energy they could be used in many areas. First, we do not have to limit to the electical energy, for an example we could use the LPs to analyse consumpton of gas, oil or even tap-water. Secondly, while in this thesis we focused on analysing the consumption of electical energy, LPs can be used to analyse the production as well. Thirdly, while we focused on optimising residential energy consumption patterns same appoach can be used in industrial and office settings.

In the thesis we used the following five datasets: UK-DALE [35], REFIT [53], ECO [7], REDD [39], and iAWE [5]. All datasets measured electrical energy consumption in residential buildings. They include main smart meter data, as well as sub-meter data for each appliance in a dwelling. While some datasets offered versions with high frequency with sampling rates up to 40 kHz, we focused on the low-frequency variations with sampling rates at around 1 Hz.

The datasets used, had frequencies ranging from 1 Hz for the ECO dataset, down to 1/8 Hz for the REFIT dataset. For datasets to be compatible, we resampled all datasets to 1/6 Hz. The missing samples were forward filled with a limit of 5, meaning if up 30 s of data was missing, its value was set to the last known value, otherwise, it was left missing. For easier handling datasets will be sliced into 1-hour intervals. The exact methodology will be presented in the methodology Chapter 3.

1.4 Contributions

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

1. Surveying the state-of-the-art LPs (Chapter 2)

The first contribution is provided by taking a look at existing research and use-cases. Using the publications, we constructed a table of LPs. We are the first to analyze LPs from this aspect. The analysis provides an overview of related work by mapping it to a table. The table reveals LPs that were not yet utilized. Using use cases we try to determine in what field each LP could be used.

2. Development of multidimensional activation LPs (Chapter 4)

Empty gaps in research motivated us to pursue the next contribution, the development of multidimensional activation LPs. Here we offer an in-depth look into the LPs, by presenting the profiles and showing how they present the consumption patterns. Each LP presents a different pattern and therefore has a different use case.

3. Visual analysis of activation LP's (Chapter 5)

The third contribution refers to exploratory data analysis (EDA) through visualizations. Here we leverage proposed and analyzed LPs and t-SNE dimensionality reduction algorithm to understand how data is related.

4. Propose a new anomaly detection method for elderly care (Chapter **??**)

This newly obtained knowledge should help us provide the last contribution. In this Chapter, we utilize LPs that haven't been considered before. We design and construct elderly care assisted living system by utilizing one of the proposed LPs. The system can detect anomalies in the daily routine of an elder. In case the anomaly is detected, the caregiver is notified to check on the caretaker. It is simple, efficient and ready for real-world use.

Chapter 2

Related Work and Table of Profiles

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

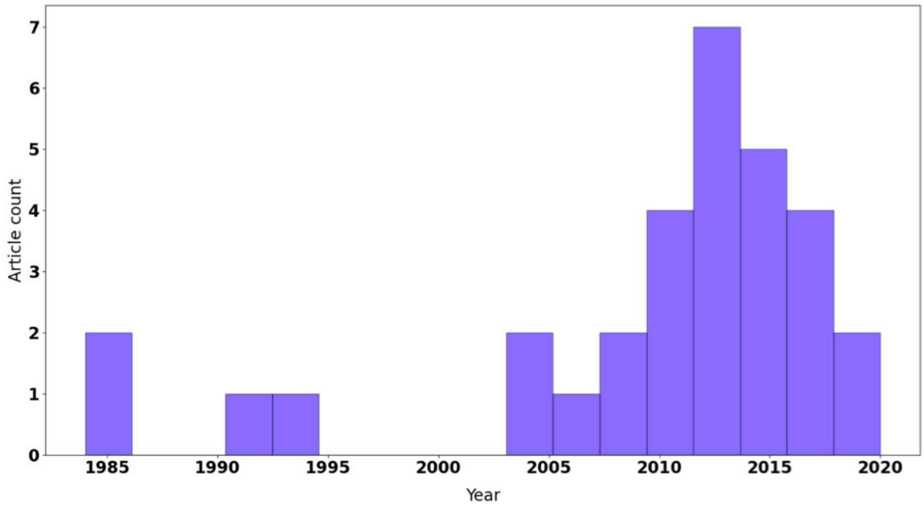
2.1 Related Work

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

2.1.1 Load Profiling

One of the first publications on load profiling was published by Train et al.[62]. They used a bottom-up approach using sub-meter data and other socioeconomic and demographic characteristics to create an LP 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 et al.[67] published a paper where they used a bottom-up approach with psychological factors to create probability models of when will an individual use an appliance. Since then there were two more in 1995. Research picked up the pace in 2005 with 7 publications in 2013 as Figure 2.1 shows.

FIGURE 2.1: Distribution of publications on load profiling from 1985 to 2020. The graph was published by [52].



Load profiling can be performed in two ways: bottom-up and top-down. A bottom-up approach as authors in [61] state "calculates the individual dwelling energy or electricity consumption and extrapolates these results over a target area or region" Whereas with top-down approach as authors in [61] state "uses the total energy or electricity consumption estimates to assign them to the characteristics of the building stock" In other more general words, bottom-up uses sub-meter data, Top-down uses aggregated data. In our case, we take a deeper dive into the bottom-up approach.

The author in [52] 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 LPs. The author also grouped modeling techniques as probabilistic models, Markov chains, and Monte Carlo. The author first disclosed the current state of load profiling and issues with past work. They made a review of existing load profiling models and asses the-state-of-the art. Next, they pointed out future research directions and applications of load profiling models. Finally, the author exposes issues that researchers face and addresses possible solutions with conclusions.

Gerbec et al.[24] tried to assign typical LPs to a particular group of consumers based on their activity. To achieve that, they used probabilistic neural networks as a way of classification. Their methodology was tested in real-use scenarios.

Gao et al.[23] makes use of the bottom-up method to build a forecasting framework for household load profiling, which takes into account the consumption patterns of residents. A model falls into the demand response use case. They have developed a "single-day extraction model", designed to select the same days by comparing environmental and household factors, which influence energy consumption. By using this approach, they have improved the accuracy of predicting the behavioral patterns of dwellers. Results show that their method successfully modeled daily usage.

Chuan et al.[17] uses load profiling to optimize energy consumption distribution during the day. This reduces peak usage and alleviates load off the grid. The author used the bottom-up method, that is, using sub-meter data. Using this data, they made daily usage analyses on a one-hour basis. Using this information they optimized the daily activation of appliances so that peak usage was not as high. Results show that peak shedding was successful.

Csoknyai et al.[20] 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.[32] used extreme points in the appliance usage curve to cluster usage profiles. Usually, the first usage peak is in the morning, and the second one is in the evening. Additionally, they used demographic characteristics that are: region, area, age, salary, etc. to improve the results. Using collected data, they clustered profiles. They discovered 6 different usage profiles, where every cluster had a physical meaning such as energy-saving, morning heavy, evening heavy, etc.

Another clustering methodology was proposed by Park et al.[48], using load image profiles and image processing. They represented time series data as an image. The image is a grid of squares where the y-axis contains monthly data with a resolution of one day, x-axis contains daily data with a resolution of one hour. Grid is color filled with an algorithm that authors developed, where red means more activity and blue less. Using digital image filters they transformed the type-1 image to type-2 and from there used a threshold to obtain type-3. Using that information they clustered data based on images similarly. They used three different clustering methods: k-means, FCM, and EM algorithm. Using the Davies-Bouldin index, they were able to prove that image-based clustering performs better than non-image.

Abreu et al.[1] clustered different LPs using electricity consumption data and surveys using data from residential homes. They used PCA and k-means resulting in 5 clusters. Similar to other load profiling papers.

Whereas most of the above-mentioned papers focused on aggregated consumption of building to build an LP, authors [31] focused on appliance-level load profiling. Their main contribution was to create a realistic per-appliance LP. They developed a wireless measurement system with smart plugs that enabled them to obtain power signatures for each appliance. They evaluated the data and based on observations they determined working cycles for each appliance. Furthermore, they concluded that 15 % of consumed power can be shifted, where they took tariffs into account.

2.1.2 Anomaly Detection in Building Energy Consumption Data

A review on anomaly detection in building energy consumption data was written by authors [27]. 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 et al.[55] 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.

The same author published another paper [54] 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.[14] 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. The dataset included data from 2018 to 2020 meaning it included covid-induced anomalies. The authors first pre-processed the data by aggregating input load in hourly energy consumption, the second derived additional features, which are the time of use and duration of the activation. They use that data to detect single-point deviations for which they implemented the isolation Forest algorithm and anomalous trends for which to detect, they implemented Change Point Detection.

2.2 Use-cases

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

2.2.1 Grid Management

Zero Energy Buildings and Energy Saving

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

On the other side, more and more photovoltaic systems are being installed, which is slowly shifting energy production towards end-users. Slowly energy grid is starting to shift towards so-called distributed energy resources or "DER" [45]. DERs include all kinds of micro-energy sources such as PV, wind power, water power, and all kinds of energy accumulators that can store and release energy when needed such as heat pumps with hot water storage, home batteries, and EVs that can be used as a battery.

With smart management, these appliances could be used in a way that would reduce the net flow of energy and alleviate the load off the power grid. A way to achieve this is via load profiling and load modeling. To manage the appliances, a control system would have to be put in place [29]. 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" [47]. They will be an essential part of the European Union's plan to reach zero-energy buildings and near-zero-energy buildings [49]. The directive was accepted in 2010 and was recast in 2021. The plan is set to be realized in the next decade.

An actual use-case would be an EV owner with an installed PV system and heat pump, who works from home on occasion. In this case, two profiles would be developed. Normal workday and work-from-home day. Additional information would be obtained from the user's calendar. On a normal workday, the system would use PV energy to heat the water and store it, based on the user profile. On work-from-home days, the system would start charging the car with the morning sun, using only the PV energy. In the evening hours, when consumption rises and production falls, EVs could inject the power back into the house. Again using appliance LPs 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 [56] [44].

One other way to use user LPs is to optimally distribute the load by studying user's usage patterns as [17] [40] proposed in their papers. This could be further extended to neighborhoods connected into peer 2 peer energy distribution networks. As mentioned earlier, the way to save energy consumption is to distribute it as locally as possible. Knowing the usage patterns of all peers, the system could optimally distribute the energy using DERs across all homes without dwellers even noticing.

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

Many papers have been published, where authors explored ways to reduce the energy consumption of users by studying user consumption patterns [60] [15] [64] [60]. 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 paper [20]. Source [19] states that as much as 20 % of energy could be saved by managing consumption.

Demand Response

An increasing percentage of renewable resources is troubling energy distributors, due to the nature of renewable resources. In the prior Chapter, it was mentioned how energy-saving measures would benefit users and their peers. One other use-case would be cooperation between end-user and energy distribution companies. Joint actions between them would benefit both as authors show in papers [2] [46].

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 [57]. One way to achieve this is to control the voltage of loads [70] the other way is to shift the loads in time [40]. This process is called direct load control [29], and it is part of demand response program [16].

"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." [16]

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

The user would have three options. The first one would be that users can use the appliances as freely as they desire, this would result in a normal tariff. The second option would be to use the appliances as regularly as possible, this would lead to lower tariffs. The third option would be to leave the management of main appliances to the electricity provider via direct load control. The provider would combine the user appliance LP and the real-time market price of energy to optimize the cost [25]. 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 [41]. Commonly whole neighborhoods are being disconnected, affecting their daily lives. Using user LPs, distribution companies could disconnect the load in a way that would minimally affect the end user. When they would need to load the grid due to low demand, they could charge EVs free of charge or even pay to do so. This benefits the company as well since they do not need to lower energy production, which can be expensive.

2.2.2 Anomaly Detection

One use-case of anomaly detection was already mentioned in the Elderly care Chapter. One more thing that could be detected, using load profiling, would be the altered operation of appliances. In the case of a fridge, the system would detect that duty cycles are too long. The increased duty cycle can be caused by cooling liquid leakage, the fridge being open or compressor motor malfunction. Heat pumps work on the same basis as fridges, meaning the same anomalies could be detected. The malfunction could also be detected in heating element appliances such as toasters or boilers. Since mentioned appliances are one of the largest consumers in a household, early enough detection could lead to large energy-saving benefits [55].

Elderly Care

The aging population is an increasing socioeconomic issue. The elderly are facing many issues when staying at home alone for extended periods. Accidents such as falls or the inability to do choirs due to health-related issues or even dementia-induced issues such as leaving appliances on for long periods could all be detected, using sub-meter data such as authors in publications [66] [50] explore in their papers.

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

An example would be: the coffee machine not turning on in the morning or the stove and kitchen vent not being used at the noon. Another issue could be detected if the appliance would be used more frequently or for extended periods of time.

This could indicate that the user forgot to turn off the stove, oven, or even a light. The same system could detect that a fridge or a freezer was left open since the duty cycles would be longer and more frequent. As soon as the issue would be detected it would notify the caregiver to check on the patient.

2.2.3 Other

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

Yip et al.[69] 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 paper [37]. Information about occupancy could be used as part of elderly care monitoring or in the case of building automation, to run certain tasks when a user enters or leaves the room or a building.

2.3 Table of Profiles

In the first part of this Chapter, we focused on the general concept of load profiling and reviewed the existing literature on the topic. In this second part, we will delve into the various ways in which load profiling data can be presented using LPs. We will begin by constructing a general LP table from previously defined features in Section 1.1.2 Next, we will map the references and use cases from the related work reviewed in previous chapters to this table and select the main features to use. Using this reduced set of features, we will create a more detailed LP table and populate it with information from the same references. Finally, we will use this information to identify potential directions for future research in this field.

2.3.1 General Table

Using these features defined in Section 1.1.2 we can form a Table with all possible combinations. Table 2.1 is then populated with references from previous chapters. To understand the table more clearly, let's imagine that each feature is used as an axis label when plotting.

TABLE 2.1: General table of LPs

	power	number of activations
time	[17] [20] [10] [34] [67] [24] [23] [32] [1] [36] [55] [54] [31] [4] [14] [37] [17] [12] [48] [35] [23]	[13] [35]
operation time	[4]	[55] [54] [4]

Table 2.1 shows a combination of base features of power and time with 21 publications. One example of such a profile can be seen in Figure 1.1 or 1.2 and is also known as standard LP (SLP).

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

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

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

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

2.3.2 Detailed Table

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

Sub-features

General features were already described in Section 1.1.2. It is possible to further divide them into smaller so-called sub-features. These are reshaped and grouped as follows:

- Way of presenting a profile

    - Per-building
    - Per-appliance
    - Per-building and per appliance

- By time range of profile
    - Daily
    - Weekly
    - Monthly
    - Yearly

- Way of measuring usage
    - Average power use
    - Number of activations

2.3.3 Table of Combinations or Detailed Table

The above-shown profiles can be combined, yielding a new way of displaying the data. Bellow, a Table 2.2 with combinations of the above-mentioned profiles is presented. The purpose of Table 2.2 is to show possible LP combinations. Some combinations that had similar output were grouped, and some that could not be sketched were discarded.

The LPs and figure graphics used in Table 2.2 were sourced from Section 1.1.

Table 2.2, uses features from the previous Subsection 2.3.2. In general, Table 2.2 is formatted in a way that features from columns (time range) are used in the x-axis of a plot, and rows (consumption data) are used in the y or z-axis of a plot.

The column of Table 2.2 presents the time domain. "Daily" means that the LP presents average usage for one day and "Weekly" means it presents usage for a week. To be clear, for one to construct a decent daily profile, one needs a few weeks of data. The same goes for yearly profiles, in that case, one needs many years' worth of data.

The top row of Table 2.2 is composed of 3 main groups. The first group focuses on per-building energy consumption. The second group examines the energy consumption of each appliance in a house separately. The third group analyses all appliances in a building.

The next row of Table 2.2 is further divided into two groups. First is the LP group which presents the given usage unit on the y-axis and time on the x-axis. Next is an LP with an additional time axis. In this case, we present the given usage unit on the z-axis and then time on the x and y-axis. Here, the second-time dimension can be anything from a week to a year. In the case of the per-building, the subgroup includes appliances instead of time. An example of this is Figure 1.6.

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

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

2.3.4 Mapping References to the Table of Profiles

To find useful LPs, references from the related work Section 2.1 must be mapped.

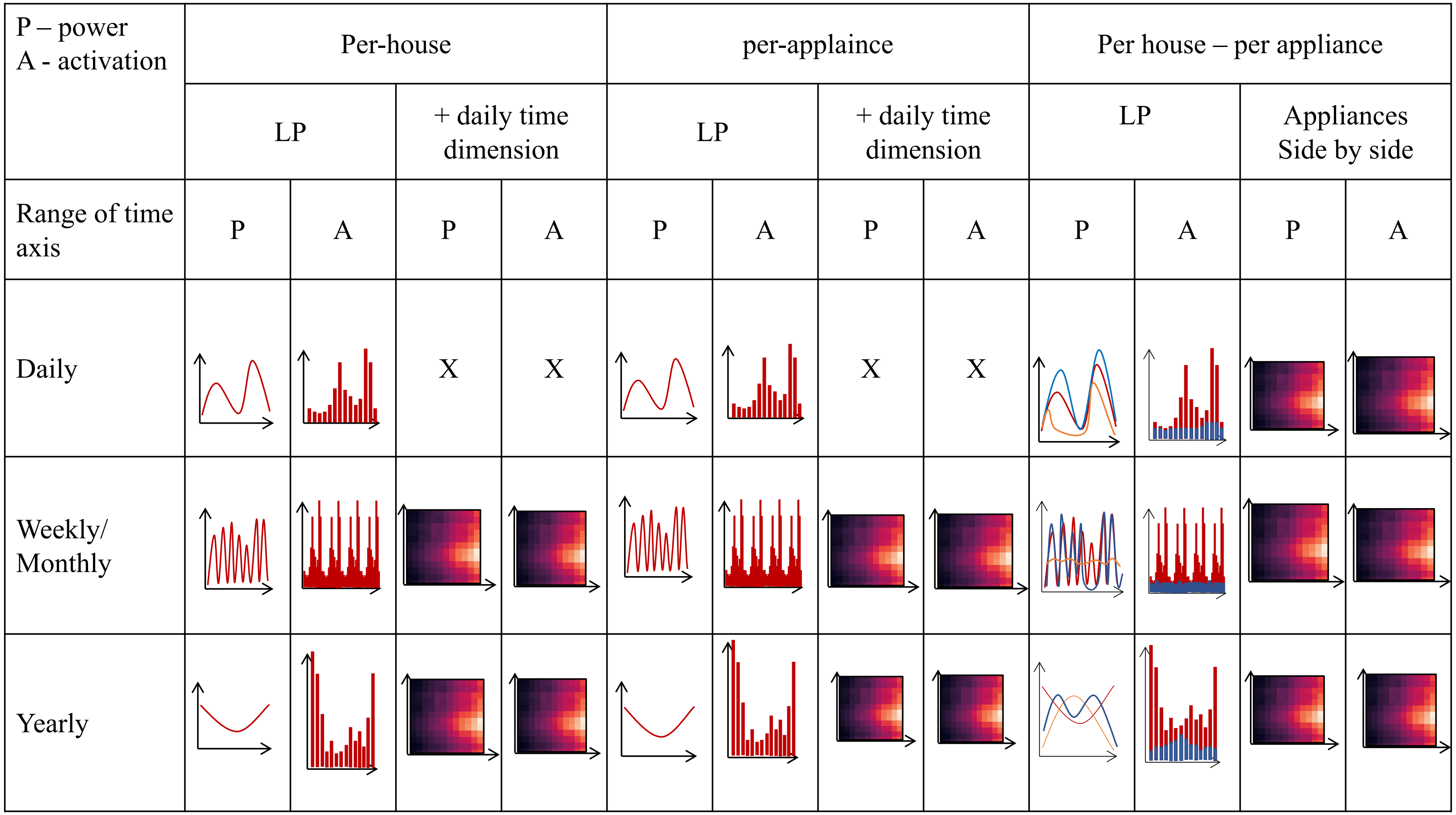


FIGURE 2.2: Table of combinations

TABLE 2.2: Table presents previously mentioned LPs

P - power A - activation	**Per-building**				**Per-appliance**				**Per-building per-appliance**			
	LP		**+ daily dim.**		**LP**		**+ daily dim.**		**LP**		**Appl. by side**	
Interval	**P**	**A**	**P**	**A**	**P**	**A**	**P**	**A**	**P**	**A**	**P**	**A**
Daily	[35] [17] [20] [10] [34] [12] [67] [24] [23] [32] [1] [36]		X	X	[55] [54] [31] [4] [14] [37]	[35]	X	X	[17] [12] [23]	[35]		
Weekly/ Monthly/	[20] [10] [34]		[59] [48] [36]			[13]			[11]	[13]		
Yearly	[20] [10] [34]											

As can be seen from Table 2.2, most of the work (14 publications) has been done with standard daily LPs with per-building power usage such as Figure 1.2. 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 LPs and a few used two-dimensional time and power presentations. Only one publication found used activation and time-based histogram such as shown in Figure 1.3. During the research we focused on publications from minority classes, meaning not all existing publications for standard LPs are included. The purpose of Table 2.2 is to present missing scientific contributions and patterns of publications.

2.3.5 Mapping Use-Cases to the Table of Profiles

Table 2.3 includes arranged publications from the use-cases Section 2.2. A similar pattern emerged as in Table 2.2.

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

P - power A - activation	**Per-building**				**Per-appliance**				**Per-building per-appliance**			
	LP		**+ daily dim**		**LP**		**+ daily dim**		**LP**		**Appl by side**	
Interval	**P**	**A**	**P**	**A**	**P**	**A**	**P**	**A**	**P**	**A**	**P**	**A**
Daily	[15] [60] [44] [69] [40] [25] [70]		X	X	[44] [66] [50] [37]		X	X	[17]			
Weekly/ Monthly/	[60] [34]											
Yearly	[60]											

2.3.6 Table of Use-Case Groups

The Table 2.4 presents same publications as Table 2.3, but only group names are shown. The groups are the main use cases from Section 2.2 and use-case tree in Chapter 1.2.

- ZEB - zero energy buildings
- DR - demand response
- AD - anomaly detection
- EC - elderly care
- X - unfeasible

The Table 2.4 indicates how groups are arranged. Where anomaly detection and elderly care are dominating in the per-appliance part of the table, zero energy buildings and demand response are dominating in a per-building part of the table.

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

P - power A - activation	**Per-building**				**Per-appliance**				**Per-building per-appliance**			
	LP		**+ daily dim**		**LP**		**+ daily dim**		**LP**		**Appl by side**	
Interval	**P**	**A**	**P**	**A**	**P**	**A**	**P**	**A**	**P**	**A**	**P**	**A**
Daily	ZEB, EC		X	X	AD, EC, ZEB		X	X	DR			
Weekly/ Monthly/	ZEB											
Yearly	ZEB											

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

TABLE 2.5: Proposed use-cases for profiles

P - power A - activation	**Per-building**				**Per-appliance**				**Per-building per-appliance**			
	LP		**+ daily dim**		**LP**		**+ daily dim**		**LP**		**Appl by side**	
Interval	**P**	**A**	**P**	**A**	**P**	**A**	**P**	**A**	**P**	**A**	**P**	**A**
Daily	AD, ZEB, DR,	AD, ZEB, DR,	X	X	AD, EC, ZEB, DR	AD, EC, ZEB, DR	X	X	AD, EC, ZEB, DR	AD, EC, ZEB, DR	AD, EC, ZEB, DR	AD, EC, ZEB, DR
Weekly/ Monthly/	AD, ZEB, DR	AD, ZEB, DR,	ZEB, DR	ZEB, DR	AD, ZEB, DR	AD, ZEB, DR	AD, ZEB, DR	AD, ZEB, DR	AD, ZEB, DR	AD, ZEB, DR	AD, ZEB, DR	AD, ZEB, DR
Yearly	ZEB, DR	ZEB, DR	ZEB, DR	ZEB, DR	AD, ZEB, DR	AD, ZEB, DR	ZEB, DR	ZEB, DR	ZEB, DR	AD, ZEB, DR	AD, ZEB, DR	AD, ZEB, DR

2.3.7 Table of LP Potentials

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

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

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

TABLE 2.6: Proposed classification of profiles

P - power A - activation	Per-building				Per-appliance				Per-building per-appliance			
	LP		+ daily dim		LP		+ daily dim		LP		Appl by side	
Interval	**P**	**A**	**P**	**A**	**P**	**A**	**P**	**A**	**P**	**A**	**P**	**A**
Daily	1	3	X	X	1	2	X	X	1	2	3	3
Weekly/ Monthly/	1	3	2	3	3	2	3	3	2	2	3	3
Yearly	1	3	3	3	3	3	3	3	3	3	3	3

2.3.8 Table of Possible Future Research Directions

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

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

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

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

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

Following these rules, Table 2.7 was constructed.

TABLE 2.7: Possible future research contributions

P - power A - activation	Per-building				Per-appliance				Per-building per-appliance			
	LP		+ daily dim		LP		+ daily dim		LP		Appl by side	
Interval	**P**	**A**	**P**	**A**	**P**	**A**	**P**	**A**	**P**	**A**	**P**	**A**
Daily		(2)	X	X			X	X			(1)	(1)
Weekly/ Monthly/		(2)		(1)	(1)		(1)	(1)			(2)	(2)
Yearly					(2)	(2)					(2)	(2)

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

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

TABLE 2.8: LPs to be pursued

P - power A - activation	**Per-building**				**Per-appliance**				**Per-building per-appliance**			
	LP		**+ daily dim**		**LP**		**+ daily dim**		**LP**		**Appl by side**	
Interval	**P**	**A**	**P**	**A**	**P**	**A**	**P**	**A**	**P**	**A**	**P**	**A**
Daily			X	X			X	X				(1)
Weekly/ Monthly/				(1)				(1)				
Yearly												

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

- Per-building daily-weekly LP
- Per-appliance daily-weekly LP

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

- Per-building Per-appliance daily LPs with appliances side by side

To build assisted living system for the elderly.

Chapter 3

Methodology

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

3.1 Data

We already briefly presented the datasets in the first chapter in Section 1.3. Here, we will do an in-depth presentation of the datasets and present how we processed and cleaned the data.

3.1.1 Non-Intrusive Load Monitoring (NILM)

While the NILM method was described in Chapter 2, we will reintroduce it as part of the methodology. NILM is a method that enables us to disaggregate consumption patterns of the whole building from a single meter. The conventional method would be to meter installed for each consumer. An alternative would be NILM. With this method, one meter can be used to find out which appliances consume the most energy.

The data used here was collected for means of research we did on Non-Intrusive load Monitoring (NILM) in publication [8]. Where we used this data to perform a classification of appliances using deep learning. Using Gramian angular fields (GAF) we transform time series into images. We used multiple images in series to form a stack, a video. Using deep learning architecture utilizing LSTM that used a stack of images as input, we were able to classify appliances with an F1 score of 80 %. Such techniques could be used to recognize appliances in unlabeled datasets, or help find possible mislabeling in existing datasets.

3.1.2 Dataset Selection

The Table 3.1 was published on the NILMTK [6] wiki page. NILMTK is a tool developed by authors in paper [6]. It intends to make the development of NILM algorithms easier by standardizing a format in which building energy consumption datasets are stored. They also developed converters to convert existing datasets into a universal format.

This enables engineers to simply load and process multiple datasets. NILMTK includes a dataset converter from most of the datasets from Table 3.1.

TABLE 3.1: List of NILM datasets.

Dataset	Sampling rate	Duration	Buildings	Subject	Country	Availability
Dataport	1 Hz to 1 minute	4+ years	1200	multiple	US	Licensed
BLOND-50	50 kHz/6.4kHz	213 days	1	office	Germany	Public
FIRED	12 kHz to 1 Hz	101 days	1	residential	Germany	Public
REDD	16500 Hz / 1 Hz	100 days	5	Residential	US	Request access
BLUED	12000 Hz	7 days	1	Residential	US	Request access
UK-DALE	16000 Hz / 1 Hz	2 years	6	Residential	UK	Public
PLAID	30000 Hz	5 seconds	55	Appliances	US	Public
WHITED	44000 Hz	5 seconds	9	Appliances	Multiple	Public
Tracebase	1 Hz	1 day	158	Appliances	Germany	Request access
DRED	1 Hz / 1 min	150 days	1	Residential	Netherlands	Public
AMPds	1 minute	2 years	1	Residential	Canada	Public
RAE	1 Hz	72 days	1	Residential	Canada	Public
iAWE	1 Hz	73 days	1	Residential	India	Public
HES	2 minutes	1 year	251	Residential	UK	Request access
REFIT	8 seconds	2 years	20	Residential	UK	Public
ECO	1 second	200 days	6	Residential	Switzerland	Public
COMBED	30 seconds	30 days		Office	India	
IHEPCDS	1 minute	4 years	1		France	
SMART	1 Hz	60 days	3		USA	
LIT-Dataset	15 kHz	30 seconds	26	Residential	Brazil	Public

Pruned version of the Table published by authors on NILMTK[6] wiki page. Full table available here `https://web.archive.org/web/20190607094329/http://wiki.nilm.eu/datasets.html`.

The reason why more datasets were not selected from the Table 3.1, was because we followed the criteria:

1. Sampling rate between 1 Hz and 1/10 Hz
2. Duration more than 30 days
3. Subject had to be a residential area building
4. Include main meter as well as sub-meter measurements
5. Has to be accessible

After applying these criteria we were left with the following datasets:

- UK-DALE [35]
- REFIT [53]
- ECO [7]
- REDD [39]
- iAWE [5].

While the Dataport dataset seems to be the best fit and of the best quality, it did sadly not meet our criteria as it is a closed dataset available only via license.

3.1.3 Processing

After datasets were obtained and converted they were ready to be processed. We decided to slice the data into hourly slices so that it will be easier to find missing data and build LPs.

Firstly we resampled the time series data 1/6 Hz. This had to be done since datasets were sampled at different frequencies. A frequency of 1/6 Hz is commonly used since it has a good ratio between resource usage and NILM algorithm performance. Resampling was done using Pandas resample. We used a forward fill parameter with a limit of 5. This means that in case of missing data, we will fill in no more than 5 samples with the last known value. Secondly, we sliced the time series data into hourly slices. One sample every 6 seconds means, there were roughly 60 samples in every slice. Thirdly, we removed slices with missing data. This was done for all slices where there was more than 20 % of data missing. In cases where less than 20 % of data were missing, we forward-filled it with the last known value. In the worst case, we forward filled 12 samples. Finally, resampled and cleaned data was stored in the .hdf file.

3.1.4 Splitting and Evaluation

In some cases, datasets had to be split, so that the algorithm we developed in Chapter **??**, may be empirically evaluated. The data was split into train and test sets, where 80 % of the data was used for training and 20 % percent of the data for testing. The data was split based on the number of samples, so in some cases where there is a lot of missing data, the time window of test data might be longer, although it contains only 20 % of the samples.

3.1.5 Dataset analysis

Table 3.2 presents all 5 datasets that we have made use of. The second column shows the number of samples for each dataset. Numbers can be a bit deceptive when it comes to the actual amount since data was resampled. To confront this, a third column was created. Here we present the actual number of samples, that is number that was obtained after resampling. We also divided the number of samples by 10^6, to make the column comprehensible.

TABLE 3.2: Summary of datasets and their characteristics

Dataset	**Samples (M)**	**Buildings**
REFIT	192.46	20
UK-DALE	55	5
ECO	21.7	6
iAWE	6.43	1
REDD	5.89	6

Note: Samples are abbreviated as M for millions.

The figures below show where time splits between train and test data were done. At the same time, we can also observe the health of each dataset. By healthy dataset, we presume a dataset that has a long uninterrupted timeline with many buildings.

REFIT

The REFIT [53] dataset included data for more than 15 buildings, as can be seen in the Figure below. The dataset in general is of the highest quality since it is the longest with the least missing data. This means this dataset should give the most relevant results.

FIGURE 3.1: Timeline for REFIT

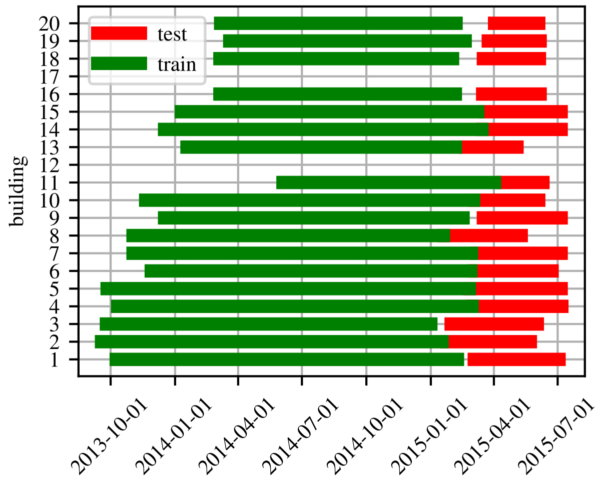


Figure 3.3 presents appliances sorted by a number of samples, where the top 10 were selected. Together, REFIT contains data for 23 different appliances from 20 homes.

TABLE 3.3: Appliances sorted by number of samples for REFIT

Appliances	**Instances**	**Samples (M)**
fridge freezer	15	47.19
television	20	40.14
freezer	13	37.91
computer	12	18.04
fridge	7	12.18
dishwasher	15	5.70
washing machine	20	5.59
microwave	17	5.49
pond pump	1	4.38
broadband router	2	2.85

Note: Samples are abbreviated in millions as M.

UK-DALE

Through the UK-DALE [35] dataset is of similar size, most of the data is from building 1. In general, it includes 5 years of data, but only for some appliances, where many appliances are rarely used. When taking all of this into account, there were too many issues with building 1, and it was simply ignored. Another issue that can be seen in Figure 3.2 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 3.2: Timeline for UK-DALE

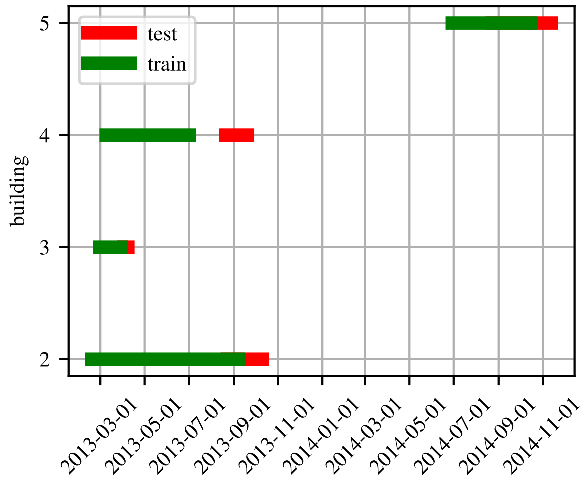


Figure 3.4 presents appliances sorted by a number of samples, where top 10 were selected. Together, UKDALE contains data for 53 different appliances.

TABLE 3.4: Summary of datasets and their characteristics

Appliances	Instances	Samples (M)
light	15	9.86
fridge freezer	2	8.83
HTPC	1	4.87
solar thermal pumping station	1	4.25
audio amplifier	2	4
boiler	2	3.75
computer monitor	4	3.2
television	3	2.59
desktop computer	3	2.55
laptop computer	4	2.1
microwave	3	1.8

ECO

ECO [7] 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 3.3: Timeline for ECO

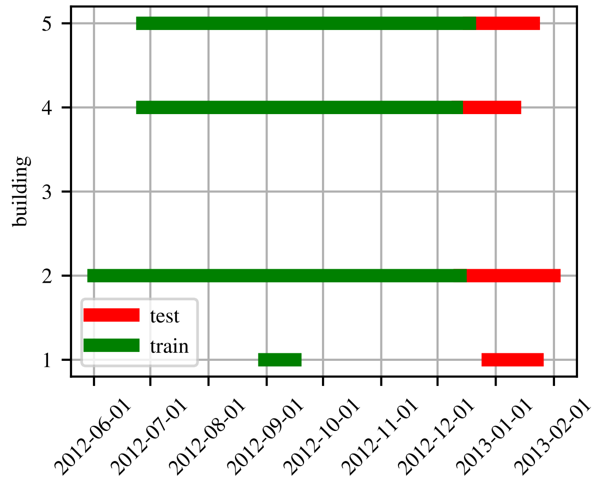


Table 3.5 shows that there were a lot of samples for the ECO dataset. This number was reduced by a factor of 6 after resampling was done.

Appliances	Instances	Samples (M)
freezer	4	5.58
fridge	6	4.26
computer	3	2.69
HTPC	5	2.61
audio system	1	0.98
laptop computer	5	0.85
television	1	0.70
lamp	3	0.56
broadband router	1	0.16
washing machine	1	0.12

TABLE 3.5: Summary of appliances in the ECO dataset

3.2 Activation Detection

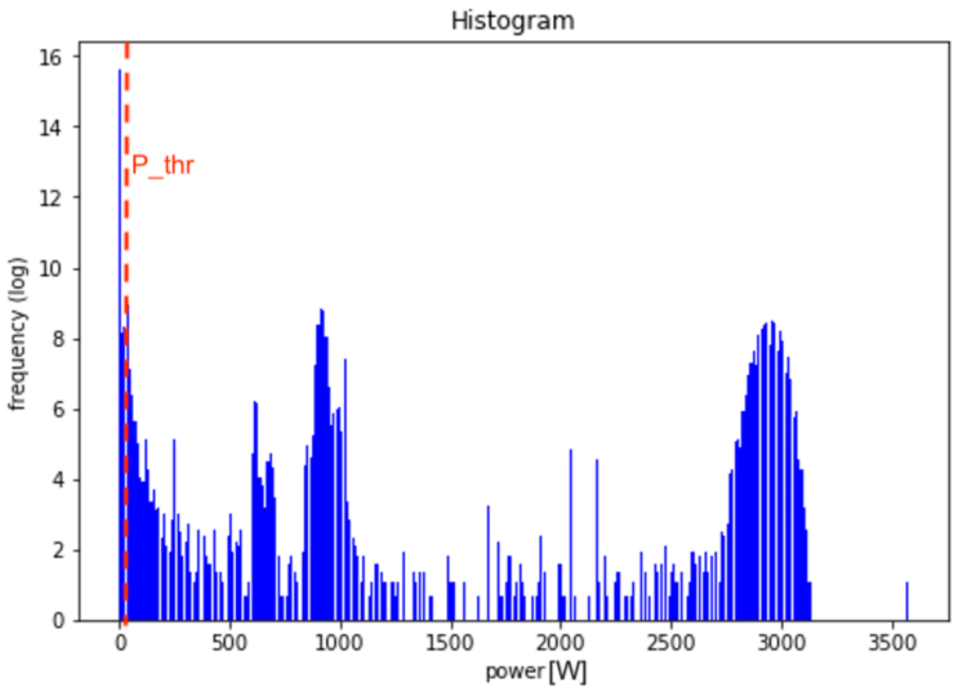
How appliance activations are extracted was already mentioned in Subsection 1.1.2. There we said, the activation occurs when consumption exceeds P_{thr}. This is portrayed in Figure 1.1 in the same subsection.

This threshold was selected as the standard value of 10 W. This value is used as a standard threshold in NILMTK [6].

This hard-set value is an issue for appliances that consume small amounts of energy, but still, show interesting usage patterns. One such example would be a mobile phone charger or broadband router. Issues could occur even with larger consumers such as smart TVs, that could consume more than 10W even when they are not operating.

In order to check that this was not an issue we came up with a test. We created a histogram of power values, with a resolution of 10 W, where one such example can be seen in Figure 3.4. For appliances that are mostly off, the first bucket should be the most populated. This was true for the majority of appliances and with that, they passed the test. For the ones that this was not true, we manually checked them and were either discarded or given a new threshold. The new threshold was manually set between the first and the second frequency peak.

FIGURE 3.4: Histogram of power values for Toaster



When we are observing Figure 3.4, we have to keep in mind that the frequency scale is logarithmic. Another thing to note is, that Figure 3.4 is also an LP, that we mention expanded table of LPs in Appendix B. This LP is useful for the detection operation modes. In this case, we can see that this Toaster has three operating modes, where each peak is a unique mode. One at 3 kW, the second at 1 kW and the third at 0.7 kW. Setting thresholds around these peaks could enable us to build 3 different LPs, where each one would present a different usage pattern.

3.3 Infrastructure and Software Used

To process the data and to obtain the results the environment and virtual machines from Google Colab [9] were used. They offer access to Google GPU-accelerated compute machines with 12 GB of RAM. Colab also offers access to Drive cloud storage, where the dataset and results were stored. While running the experiments, we made use of Drives 100 TB pooled cloud storage, which is available to students of the University of Ljubljana. For development and version control, GitHub was used.

Within the Colab which uses a Jupyter [38] environment at its core, various python libraries were used. To store and read the datasets in hdf5 format we used h5py [18] and Pickle [63]. To load datasets into RAM and then handle them, the pandas [43] library was used. For handling the large matrices and calculating we used NumPy [26]. To present the data with graphs we have used Matplotlib [30] and to present data with heatmap Seaborn [68]. For easier implementation, such as of the t-SNE, a Scikit [51] and SciPy [65] libraries were used.

Chapter 4

Presenting Proposed LPs

The Chapter will provide an analysis and in-depth presentation of LPs from Section 2.3.3. In general, each profile has its use-case already assigned in Table 2.3. Here, we will focus on exposing the main features, issues, and use cases of the aforementioned LPs.

Using the same pattern as in Table 2.2, we will first present per-building LPs with different time ranges. We will start with simple LPs and then move to more advanced LPs with two-time dimensions. Secondly, using the same pattern, we will present per-appliance profiles. Finally, we will present per-building per-appliance LPs. Data for profiles in this chapter was used from building 2 from the REFIT dataset discussed in Section 3.1.

4.1 Time Ranges

Time ranges are an important part of the LP since each reveals a different usage pattern. Throughout the thesis, we used four different time ranges: daily, weekly, monthly and yearly.

The daily profiles are the most commonly used LPs, as can be seen in Table 2.2 Generally, they are the easiest 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 LPs need a few months of data, monthly few months, and yearly few years. And this is the main issue, there is rarely enough data to build such profiles. Even then, usage patterns could change over a long period such as a decade. Combining that with a smaller number of use cases for such profiles, reveals why such profiles were not looked into as much in Table 2.2.

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

4.2 Per-Building LPs

The section will be focused on per-building profiles. Per-building profiles refer to representations where whole building usage is presented as a single LP. This kind of presentation is useful for observing general activation trends in a building. Possible use cases for per-building LPs are grid management and energy saving.

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

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

(A) Daily (B) Yearly

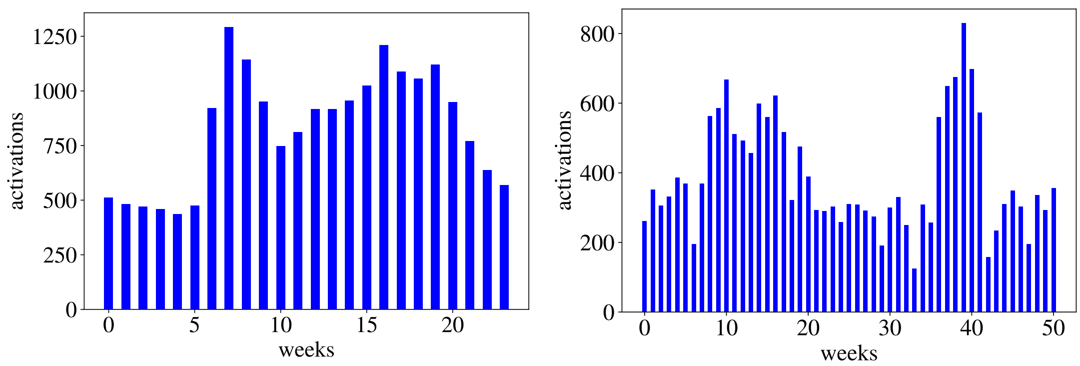


FIGURE 4.1: Per-building LPs

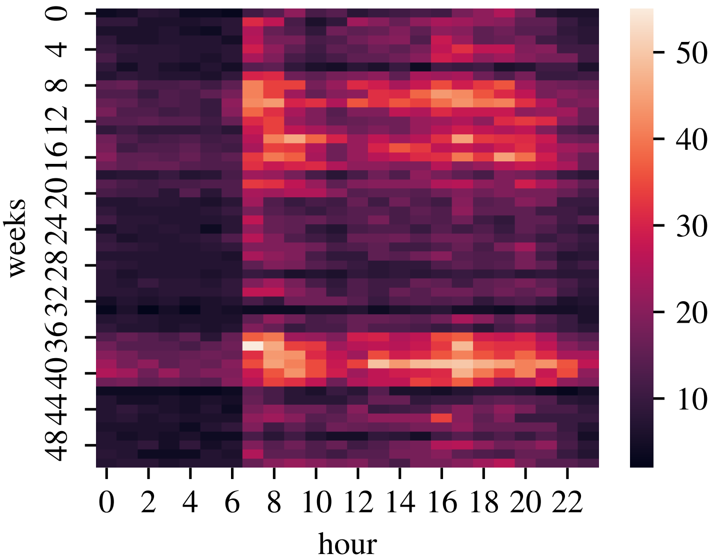
The first Figure 4.1a shows how activations change throughout the day. It is possible to see that there is some activity even throughout the night and early morning. These can mostly be related to fridges or other appliances that are not directly activated by users. At around 8 in the morning, it is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

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

4.2.1 Per-Building Two-Dimensional Time LPs

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

FIGURE 4.2: Two-time-dimensional per-building LP



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

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

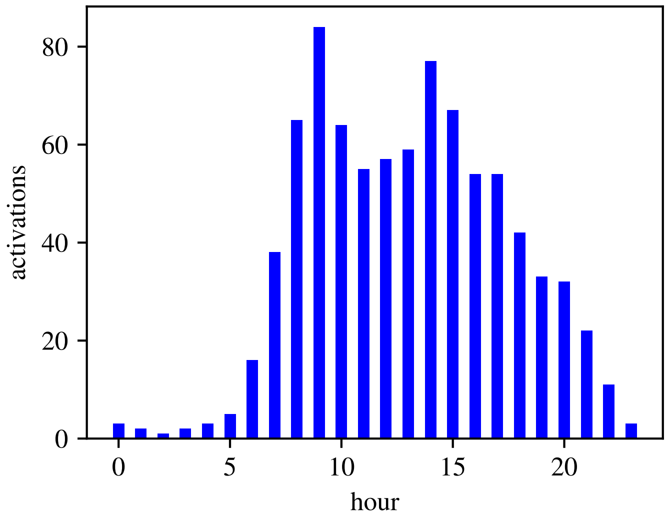
One more thing to mention is that the increased activity at the start of the fall increased activity throughout the night and day. This could point to that some new appliance was installed, which increased the number of activations.

Previously, in Section 2.2.1, it was mentioned that these kinds of profiles are the most applicable in grid management. One such example could be load shedding. Using the LP above, electrical energy providers could find buildings with the least activity at that time of day. Combining that with power data, it could disconnect the buildings with the least activity and most power consumption.

4.3 Per-appliance

Per-appliance LPs offer a look into the consumption of each appliance. In the case of activation LP, this is an elemental LP, since all other activation LPs are built on top of it. This also means that it is one of the most universal profiles since it can be used in use-cases defined in Section 2.2

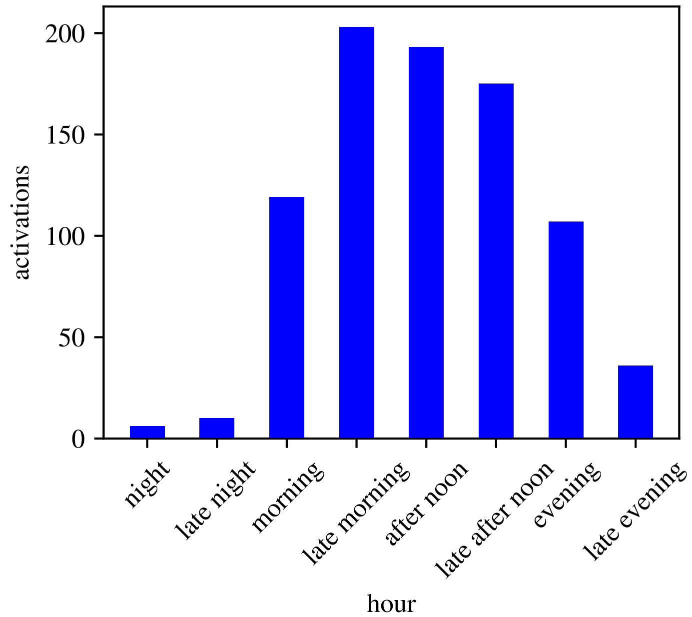
FIGURE 4.3: Daily per-appliance LP



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

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

FIGURE 4.4: Daily per-appliance LP with larger buckets sizes

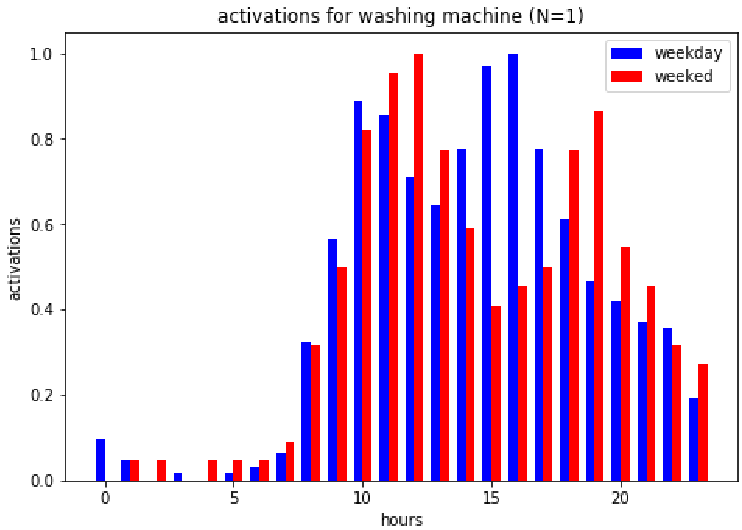


While the low resolution is useful for contextual presentation, high resolution is needed for time-sensitive applications such as elderly care, where we have to detect an accident as soon as possible. The hourly resolution would mean that in case of an accident, system would need at least an hour to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

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

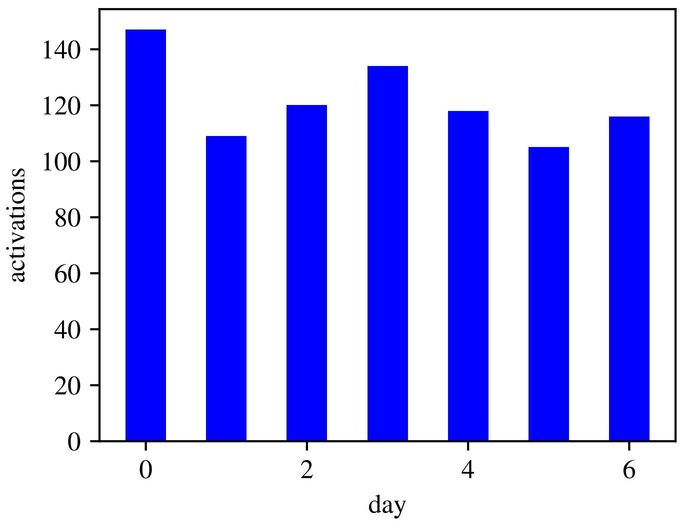
Figure 4.5 again shows the same pattern as in Figure 4.3. What can we observe here is how these two patterns are the same but are shifted in time. On weekdays the first peak occurs at around 10 AM and the second at around 3 PM. On weekends the first peak does not occur until 12 AM and the second at around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

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



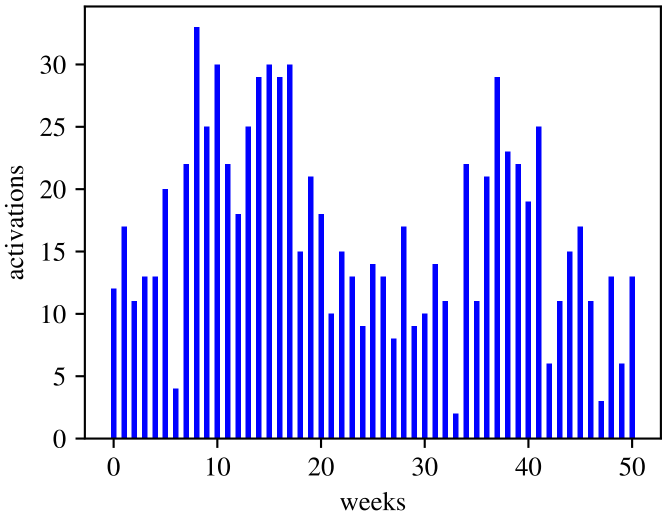
Another way to present weekly data is shown in Figure 4.6. In this case, weekdays are numbered, where 0 stands for Monday and 6 for Sunday. This resolution offers a look into how consumption pattern changes over the week. This is useful for applications such as grid management or energy saving. In this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

FIGURE 4.6: Weekly per-appliance LP



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

FIGURE 4.7: Yearly per-appliance LP

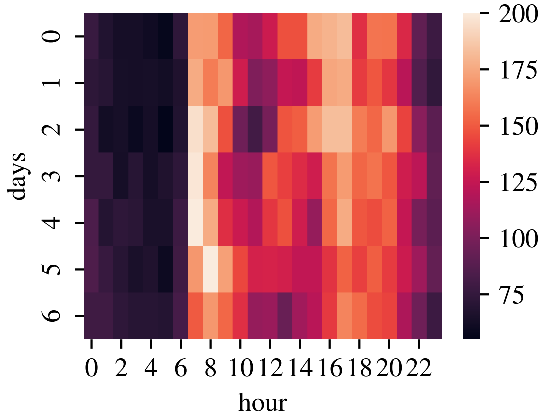


When comparing the pattern from Figure 4.7 to pattern from Figure 4.1b it is possible to see the very same pattern. When making a quick comparison, they seem like the same image, only when taking a closer look it is possible to see that differences do exists. We can make a similar conclusion here, as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

4.3.1 Two-Dimensional Time Per-Appliance LPs

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

FIGURE 4.8: Two-dimensional time per-appliance LP



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

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

Other Two-Dimensional Presentations

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

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

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

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

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

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

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

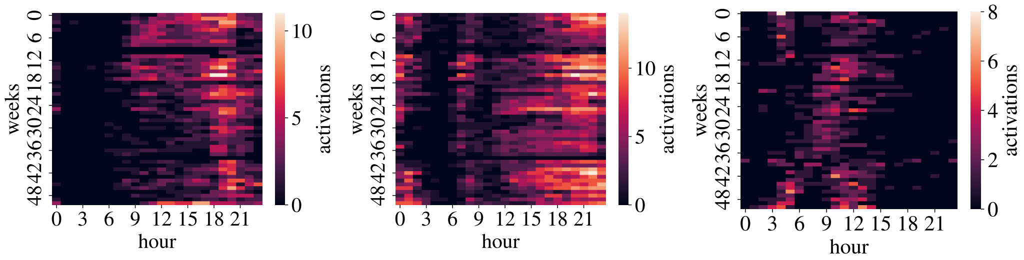


FIGURE 4.9: Various yearly two-dimensional LPs for building 4 from REFIT.

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

(A) Solar thermal pumping station (B) Light

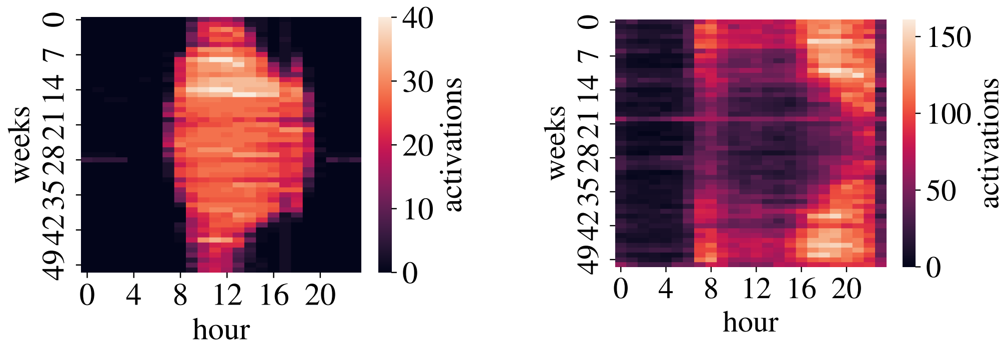


FIGURE 4.10: Effect of seasonal changes on LPs

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

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

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

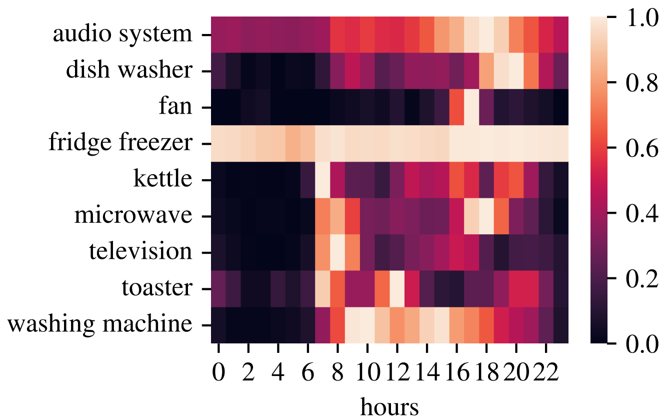
4.4 Per-Building Per-Appliance

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

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

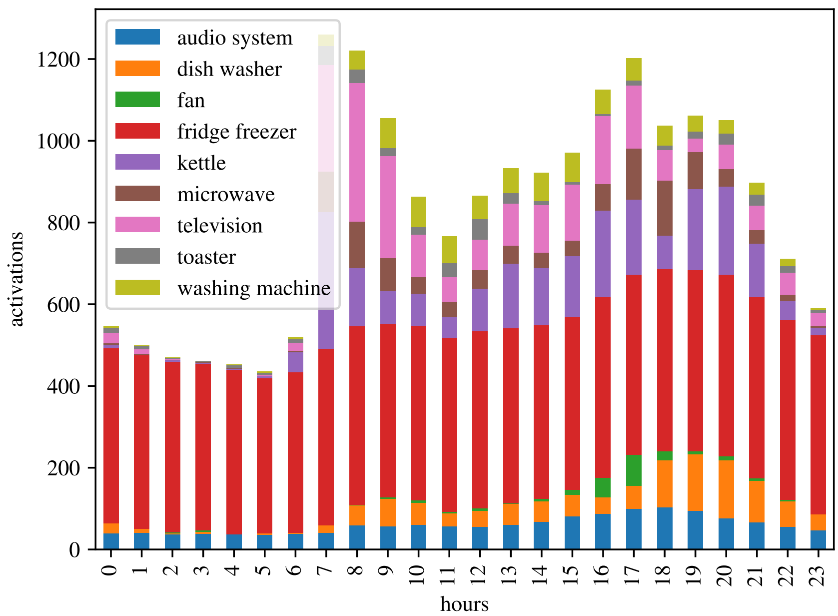
Figure 4.11 is also a good example of the elderly care system, that would detect an anomaly such as a fall, or a person unable to get up from the bed in the morning. This profile shows that the first thing in the morning used are a kettle and toaster, and with a delay of one hour, microwave and TV. This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

FIGURE 4.11: Daily per-appliance per-building building LP



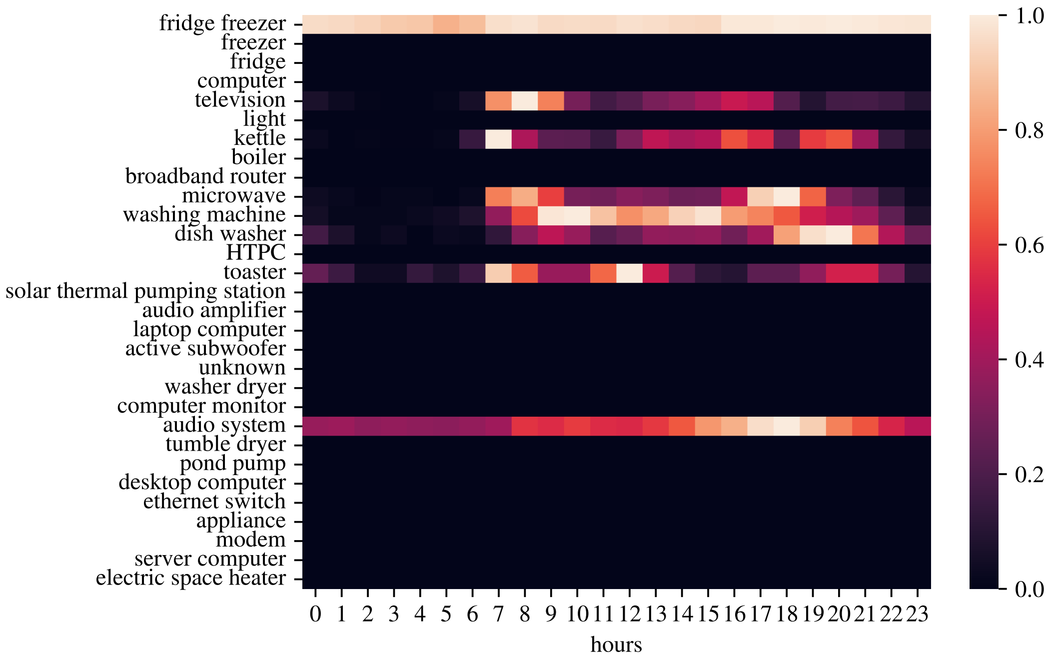
The very same data can be presented in an alternative way, such as shown in Figure 4.12. The usage pattern is the same as on 4.1a, except that it is possible to see the contribution of each appliance.

FIGURE 4.12: Stacked daily per-appliance per-building building LP



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

FIGURE 4.13: Universal presentation of per-building per-appliance LP



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

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

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

4.5 Summary

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

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

Chapter 5

Exploratory data analysis of LPs using t-SNE

5.1 Introduction

LPs can be used to understand the consumption patterns of appliances or buildings. The one thing they do not offer is a comparison between activation patterns. To achieve this we can utilize various dimensionality reduction algorithms. In the process of dimensionality reduction, these algorithms map similar LPs closer together compared to dissimilar LPs. This enables us to have an insight into similar activation patterns across various entities. It enables us to visualize and compare LPs of buildings and appliances, to find the differences and similarities in their activation patterns.

In this chapter, we will explore the use of t-distributed stochastic neighbor embedding (t-SNE) for Exploratory Data Analysis (EDA) on LPs. The t-SNE is a non-linear dimensionality reduction algorithm, used to visualize high dimensional data in usually two or three dimensions. We will delve into the details of what t-SNE is and how can it be applied to the LPs.

To achieve this goal, we will first provide a brief overview of t-SNE and its application to LPs. Next, we will describe our methodology for using t-SNE to analyze LPs and compare activation patterns. Finally, we will present the results of our analysis and discuss their implications for understanding energy consumption patterns.

The clustering of similar LPs was researched many times before, as it was described in related work Chapter 2. We will be working with dimensionality reduction, where clusters are usually formed as a side product. The following clustering publications are worth mentioning. We have seen that authors [24], [32] and [1] have clustered regular one-dimensional LPs, as well as with 2D image-based load profiling in publications published by authors [48].

The publication by authors [3] compared various dimensionality reduction techniques for clustering and visualization of LPs. Their goal was to compare Principal Component Analysis, Isometric Feature Mapping, Sammon Mapping, Locally Linear Embedding and Stochastic Neighbor Embedding. They used daily power LPs from residential and industrial areas. This publication was of the closest resemblance to our goals, that we were able to find.

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

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

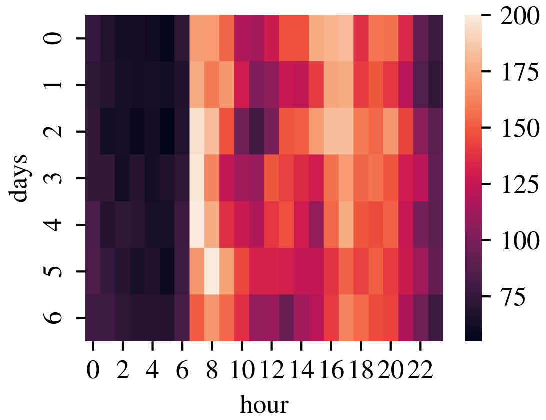
5.2 Methodology

5.2.1 LPs

Weekly-Daily LP

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

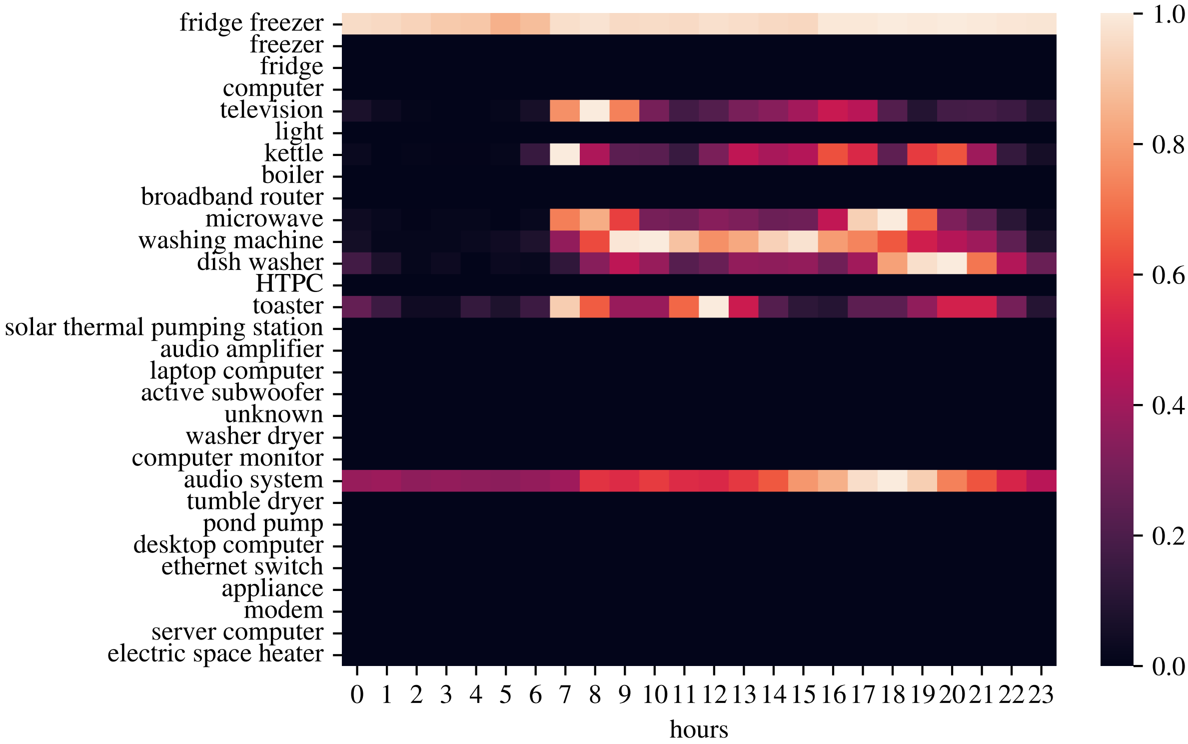
FIGURE 5.1: Weekly per-appliance LP



Bag of Appliances LP

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

FIGURE 5.2: Universal presentation of per-building per-appliance LP



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

5.2.2 Data

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

More detailed methodological approaches were discussed in Section 3.1.

5.2.3 T-SNE Algorithm

The t-SNE [42] or t-distribution stochastic neighboring embedding is a method for portraying high dimensional data in low dimensional space. This process is also known as dimensionality reduction.

One of the well-known dimensionality reduction algorithms is PCA. The key difference between the two is that one is linear, and the other is non-linear. PCA, linear, projects data in new space and finds the one with the least variance between data points. SNE [28], 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 [28]. The pairs with high similarity have a high probability, and pairs with lower a low probability. Second, it uses Kullback-Leibler divergence to minimize it with respect to a location on a map. To achieve this it uses gradient descent to minimize the cost function. Over many iterations, similar data points should be close together and far away from dissimilar objects. Similar data points usually form clusters. t-SNE uses SNE as a basis, except that it uses t-student distribution instead of normal to calculate the similarity.

A good example that showcases the non-linearity of t-SNE can be seen in Figure 5.3. In this simple task, projecting all data points to the y-axis would leave us with a different solution than one we can see in Figure 5.3.

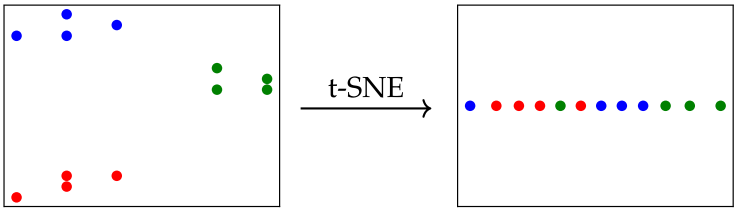


FIGURE 5.3: 2D data point transformed into 1D data point using t-SNE

In order to calculate the t-SNE for a set of data points, we first need to calculate the conditional probability. This is calculated based on the Equation 5.1 below. The author of t-SNE Van der Mateen [42] states: "The similarity of datapoint x_j to datapoint x_i is the conditional probability, p_{ij}, that x_i would pick x_j as its neighbor if neighbors were picked in proportion to their probability density under a Gaussian centered at x_i."

$$p_{ij} = \frac{\exp(-\|x_i - x_j\|^2/2\sigma_i^2)}{\sum_{k\neq l}\exp(-\|x_k - x_l\|^2/2\sigma_i^2)} \tag{5.1}$$

In Equation 5.1 x_i and x_j are two data points and $|x_i - x_j|$ is the Euclidean distance between the two. The nominator in Equation 5.1 is equal to the similarity between two points normalized by the variance $2\sigma_i^2$. The whole expression is run through $exp()$ function to ensure the value stays positive and within boundaries. The denominator in Equation 5.1 serves as a normalisation factor, to ensure that the sum of probabilities for data point x_i will sum to 1.

The σ_i is also known as Gaussian bandwidth, Gaussian kernel or just variance is picked for each data point based on the number of neighbors in its vicinity. In areas where data points are more crowded, σ_i is usually smaller than in less crowded areas. It is pre-calculated for every point using binary search. A search is complete when σ_i outputs probability distribution P_i that matches user-defined perplexity $Perp(P_i)$.

$$Perp(P_i) = 2^{H(P_i)}$$

Here, $H(P_i)$ is the entropy of the conditional probability distribution P_i. The entropy of conditional probability distribution is a measure of perplexity. Perplexity is one of the parameters defined by the user, and it's used as a measure of the number of effective neighbors, between which we will compute similarities. High perplexity means that the distribution of the Gaussian kernel will be wide and contain more data points between which similarity will be computed. Low perplexity means that the kernel will be narrow, so fewer data points will fit into it and therefore fewer data points will be compared.

The output of the algorithm is a map of every data point y_i. These points are low dimensional counterparts of x_i. Usually, these data points contain a comprehensible number of dimensions where $y_i \in \mathbb{R}^2$ or $\mathbb{R}^3$. Similarly, as in equation 5.1 we can now use low dimensional data points y_i and y_j to calculate probability q_{ij} in equation 5.2. Here, t-student distribution with one degree of freedom is utilized to calculate the similarities.

$$q_{ij} = \frac{(1 + \|y_i - y_j\|^2)^{-1}}{\sum_{k \neq l}(1 + \|y_k - y_l\|^2)^{-1}} \tag{5.2}$$

q_{ij} is again a conditional probability of finding y_i and y_j near each other but for fewer dimensions.

Setting up a cost function, which tries to minimize the difference between q_{ij} and p_{ij} should result in a low dimensional map where similar points should be near eachother. The cost function is also known as Kullback-Leibler divergence seen in Equation 5.3. The equation is the sum of all pairwise similarities between low and high-dimensional data points. The smaller the C the closer the similar data points are in low dimensional space.

$$C = \sum_{i \neq j}^{n} p_{ij} \log \frac{p_{ij}}{q_{ij}} \tag{5.3}$$

The similarity is achieved over many iterations where we use gradient descent to minimize the Kullback-Leibler divergence seen in Equation 5.3. The process can be seen in Figure 5.4.

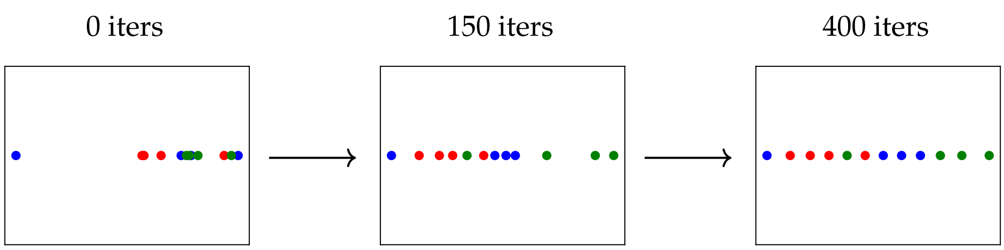


FIGURE 5.4: Iterations of t-SNE
The input data can be seen in 5.3

In our case, two dimensions will be used. Since this is a non-linear dimensionality reduction, the axis usually presents dimensions that are hard to comprehend by the brain. It is important to keep in mind that the resulting low-dimensional representation is not necessarily interpretable in the same way as the original high-dimensional data. This also means that the axes labels on the graphical presentations are meaningless. In our case, we labeled the two axes as $dimension - 1$ and $dimension - 2$.

5.3 Results

The results will be presented in three subsections

- Per-building LP
- Per-appliance LP

- Per-building per-appliance LP

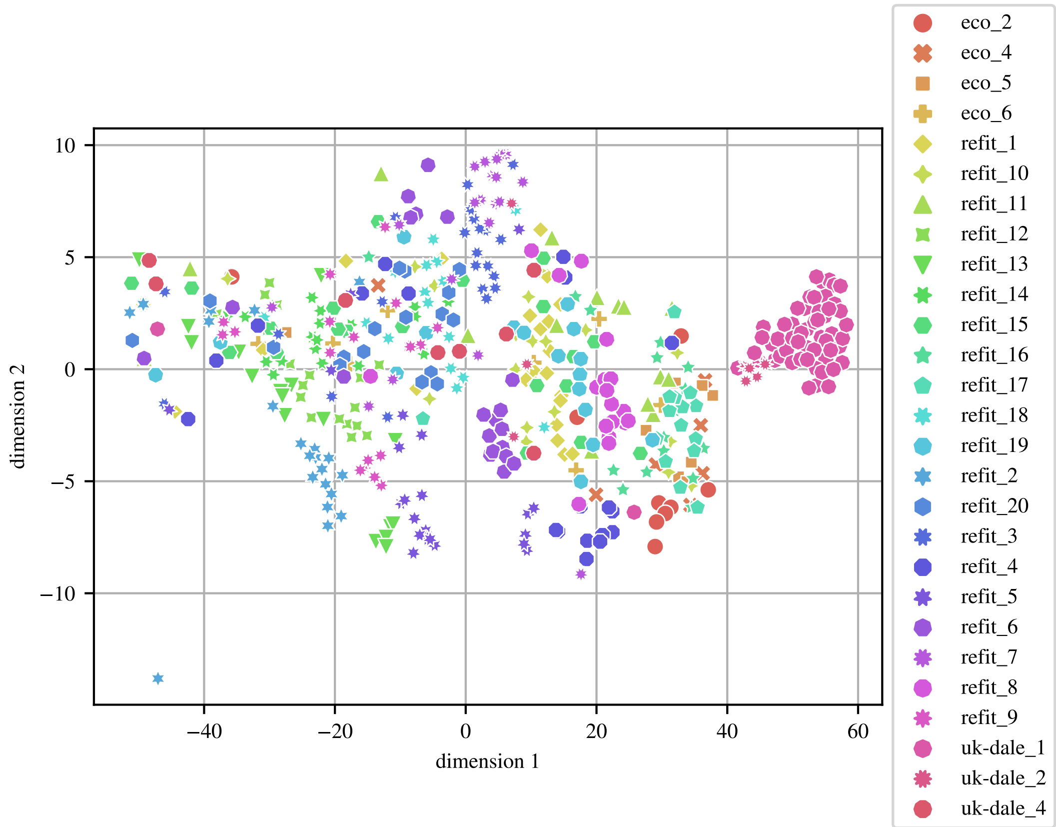
Most of the focus will be done on the per-appliance LP since it is the most universal.

5.3.1 Results for Per-Building LPs

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

Figure 5.5 is using non-normalized data, meaning the number of appliances in a building will affect the end LP. The algorithm could pick up on how many appliances are being used. In some cases, such as energy poverty detection, this information is useful, again in othersm, where would like to find more complex usage patterns, we are better off using normalized profiles.

FIGURE 5.5: Projection of per-building LPs



Full resolution figure: https://github.com/jenkoj/msc/tree/main/Figures/TSNE/TSNE_per_building/scatter_per_building.pdf

Figure 5.6 below presents the actual LP for each sample. It is possible to see that on the left there are mostly samples with very little activity, and on the right, we see samples with more activity. Since the two plotted components are of a higher dimension, it is hard to determine what they present. As said t-SNE gives us the intuition of how LPs are connected in higher-dimensional space.

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

FIGURE 5.6: Projection of per-building LPs with actual samples

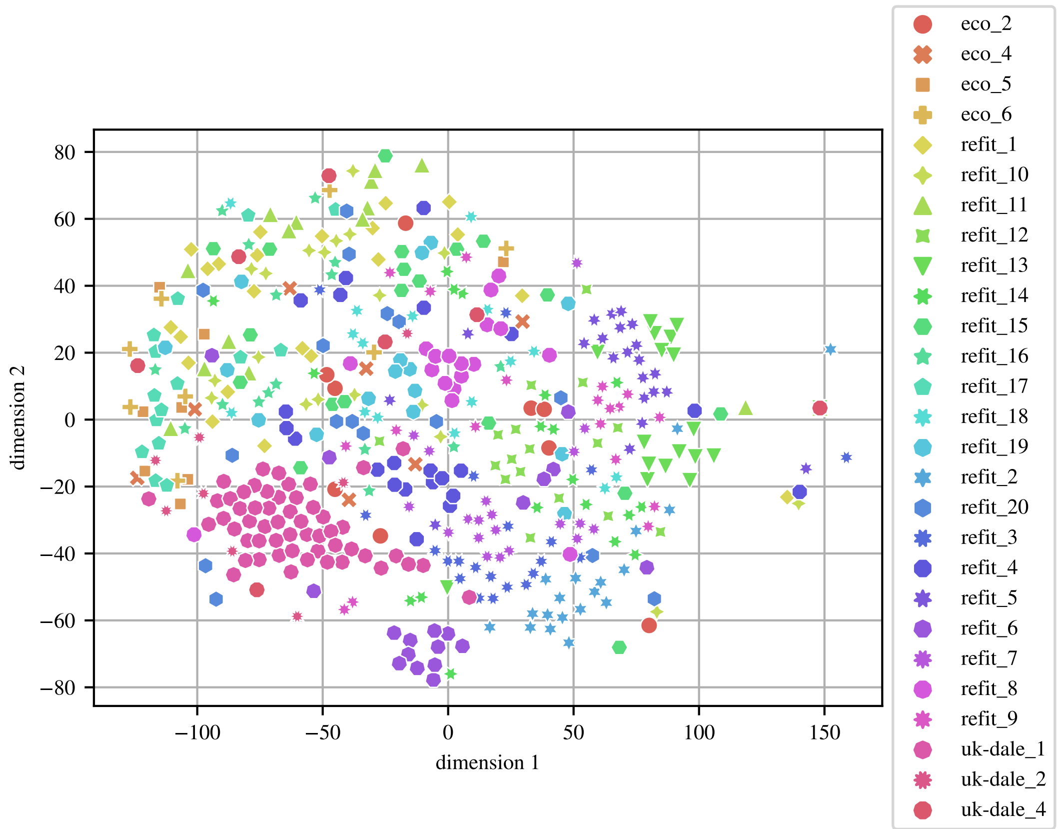
Full resolution figure: https://github.com/jenkoj/msc/tree/main/Figures/TSNE/TSNE_per_building/img_scatter_per_building.png

Normalized LPs

To solve the issue mentioned in Subsection 5.3.1 have to normalize the data between 0 and 1. Figure 5.7 shows how normalizing samples affect the algorithm.

When comparing figures 5.5 and 5.7, it is possible to see that the samples on the latter are much closer to each other, while it is still possible to see the individual clusters. This could imply that the normalized usage pattern of users is more similar to the activation pattern of users. A normalized activation pattern tells us at what part of the day the appliances will most likely be used, and the activation pattern tells us how much will the appliance be used in each part of the day. Based on that, we can conclude the time when the appliance is used is more consistent than how much it will be used.

FIGURE 5.7: Projection of normalised per-building LPs



Full resolution figure: `https://github.com/jenkoj/msc/tree/main/Figures/TSNE/TSNE_per_building/scatter_per_building_norm.pdf`

Figure 5.8 presents only the main cluster of samples. Since the smaller cluster presents mostly low entropy data, it was cut out. If the reader wants to see the samples in the cluster, the very same cluster can be found on the far left in Figure 5.6.

FIGURE 5.8: Projection of normalised per-building LPs with actual samples

Full resolution figure: https://github.com/jenkoj/msc/tree/main/Figures/TSNE/TSNE_per_building/img_scatter_per_building_norm.png

In Figure 5.8 it is possible to find various usage patterns. But the general pattern is that there is less activity during the night with one peak in the morning and evening hours. Some buildings are more active during the week and again some more during the weekend. A lot of the data is from UK-DALE building 1 (pink box). It is possible to see that the building has one big cluster where activations are generally similar, with few outliers, where the pattern completely changed. Albeit less obvious, this pattern is the same for all buildings. This happens due to events such as vacations, holidays or weather-induced behavioral changes.

5.3.2 Per-Appliance

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

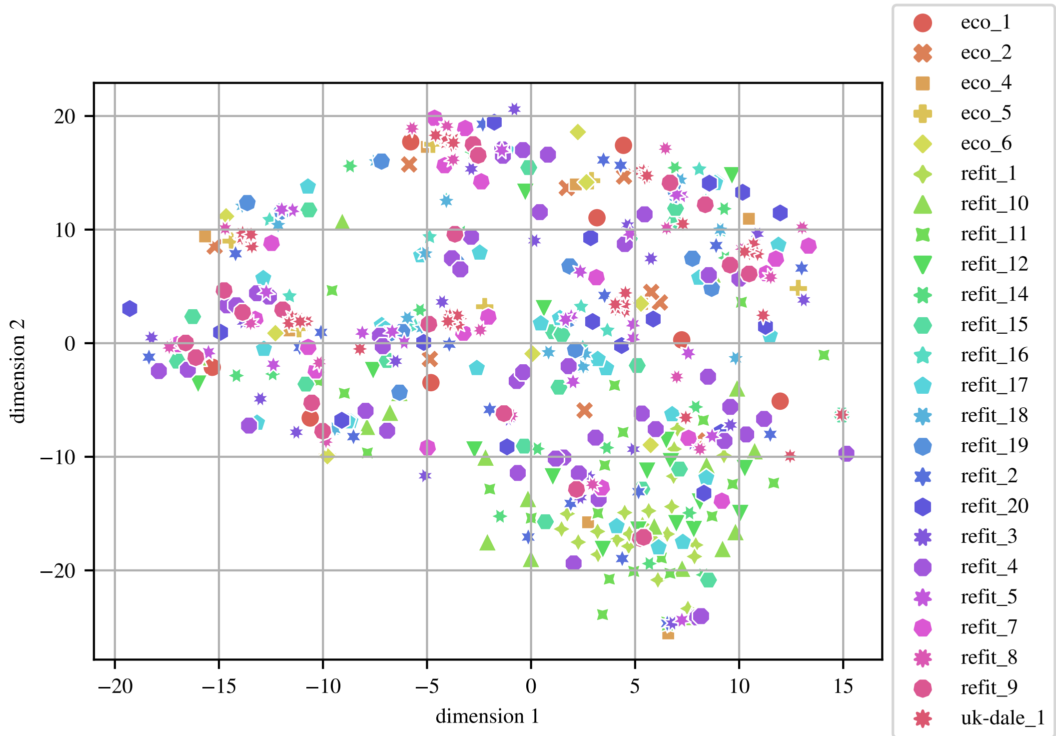
Per appliance LPs are built using sub-meter data, meaning each LP should present each appliance.

Single Appliance Over Many Buildings

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

Fridges are generally a bad indicator when it comes to user behavior since the user does not affect its operation. The only case when the user interacts with it is when opening the door and turning on the light inside. Usually, this event is dwarfed by the activations of a compressor. This also means that the usage pattern should be the same across all buildings. This can be seen in Figure 5.9, where apart from REFIT buildings 1 and 11, there are no clusters.

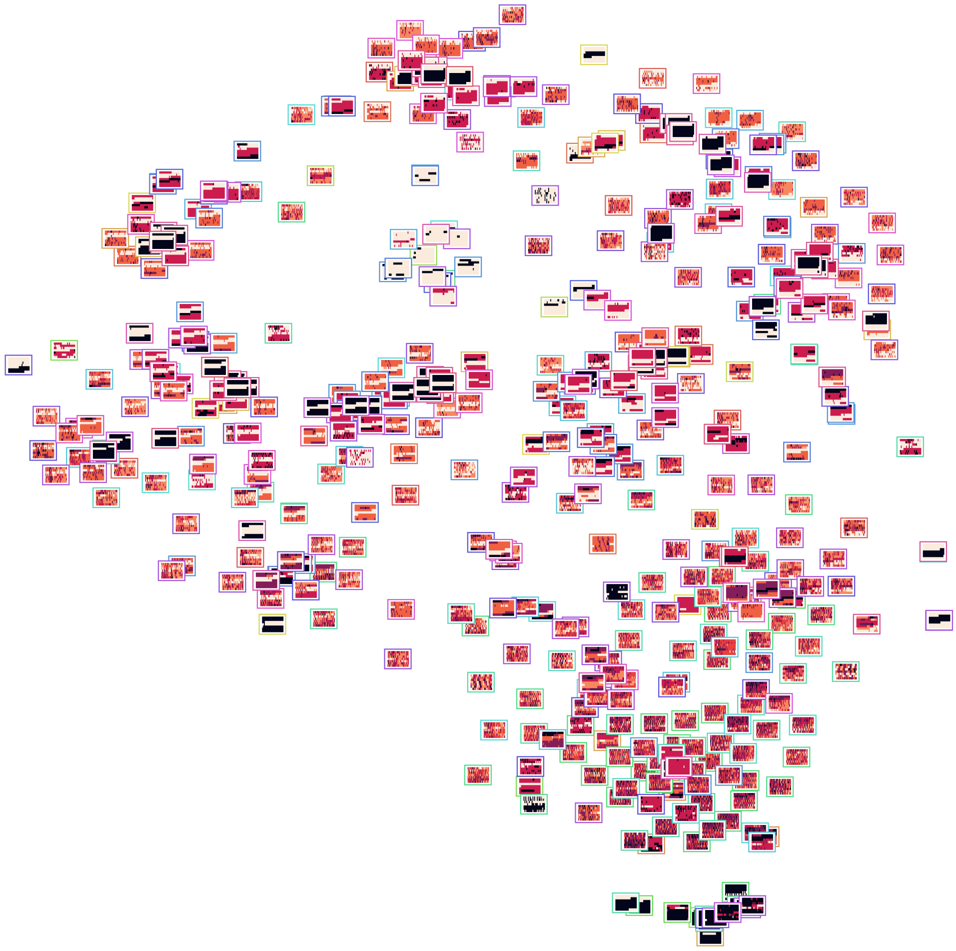
FIGURE 5.9: Projection of fridge LPs for various buildings



Full resolution figure: `https://github.com/jenkoj/msc/tree/main/Figures/TSNE/TSNE_per_appliance/scatter_refit_fridge_freeezer_fridge_freezer.pdf`

Figure 5.10 Shows mostly bright images, apart from a few outliers. LPs scattered in a circle are generally less dynamic than the ones at the bottom. Figure 5.10 is a good example of how LPs with little to no human interaction, can look a lot different. This could be due to different makes of the appliances, malfunctions of the appliance or the meter measuring it.

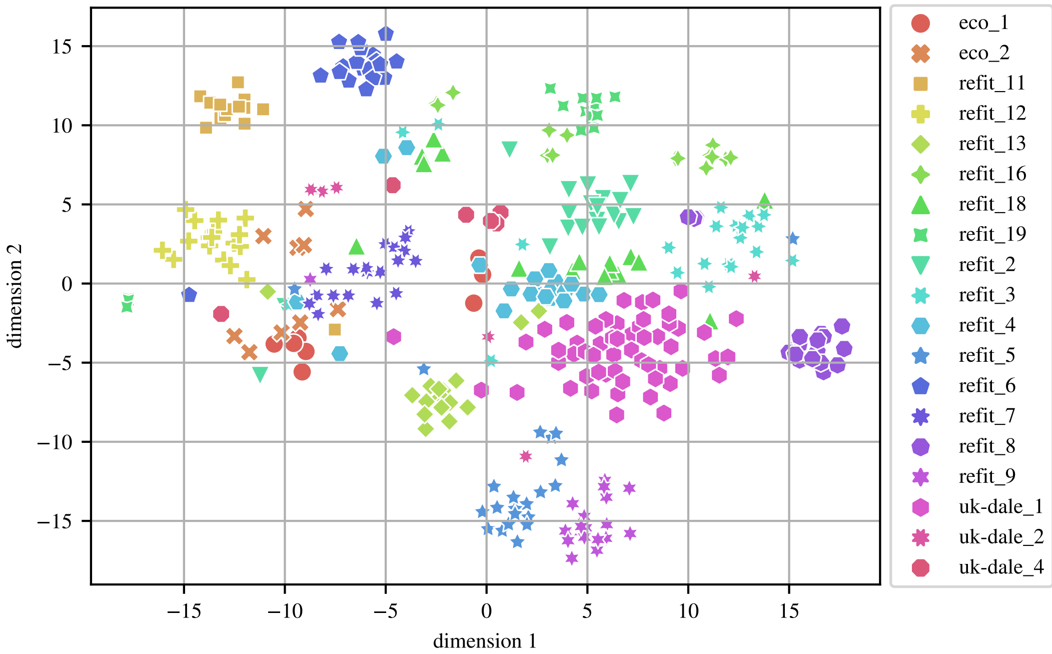
FIGURE 5.10: Projection of fridge LPs for various buildings with actual samples



Full resolution figure: https://github.com/jenkoj/msc/tree/main/Figures/TSNE/TSNE_per_appliance/img_scatter_refit_fridge_freeezer_fridge_freezer.png

Figure 5.11 shows how, compared to fridges, kettles have many clear clusters that are spaced out between each other. This could mean that every household uses a kettle a bit differently. This cluster is a good example where we can see how strong is a routine of a user. The closer together the clusters, the higher the routine since samples are more similar to each other.

FIGURE 5.11: Projection of kettle LPs for various buildings

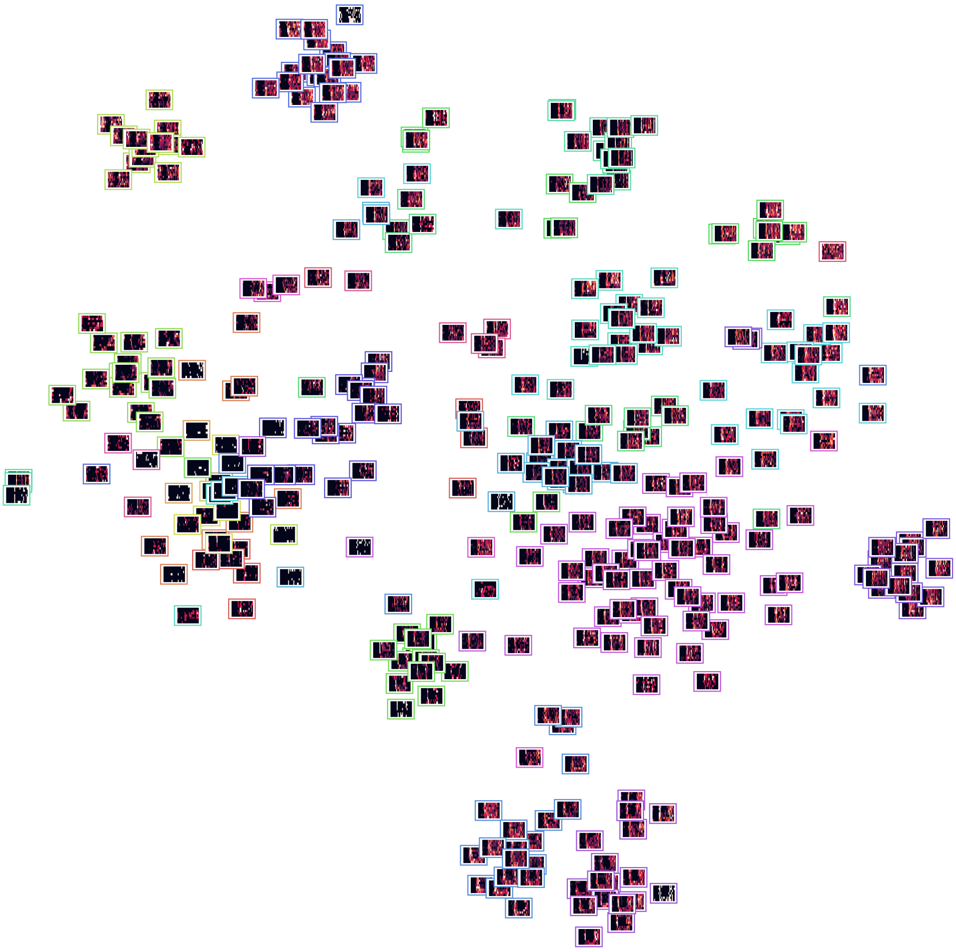


Full resolution figure: https://github.com/jenkoj/msc/tree/main/Figures/TSNE/TSNE_per_appliance/scatter_refit_kettle.pdf

Figure 5.12 shows us that images on the lower part of the plot contain less activity than the others. LPs that are closer together have more similar activation patterns. Similar activation patterns are caused by similar behavior, which is essentially a routine. This means that this projection could be used to calculate how much a behavior variates in time for each building. This could be calculated by measuring the scattering of samples (variance) for each building.

If we find samples that always activate in the same morning buckets, we would see that they form a straight line on the y-axis. This is the daily routine. One such example can be seen in Figure 5.11 in cluster refit 5 and refit 9, where we can see the lines and the pattern throughout the day. Since the routine is present, the samples look more similar and are therefore closer together. This does not necessarily mean that the closer the samples higher the routine. They could also be together in case of "ordered chaos" such as can be seen in Figure 5.11 for building refit 16 and refit 8 where there is no pattern through the day. So the scattering is not a precise metric when it comes to the routine, but it gives us a rough idea of its presence. The strength of a routine is an important feature that will be used in Chapter **??** to build an elderly care anomaly system.

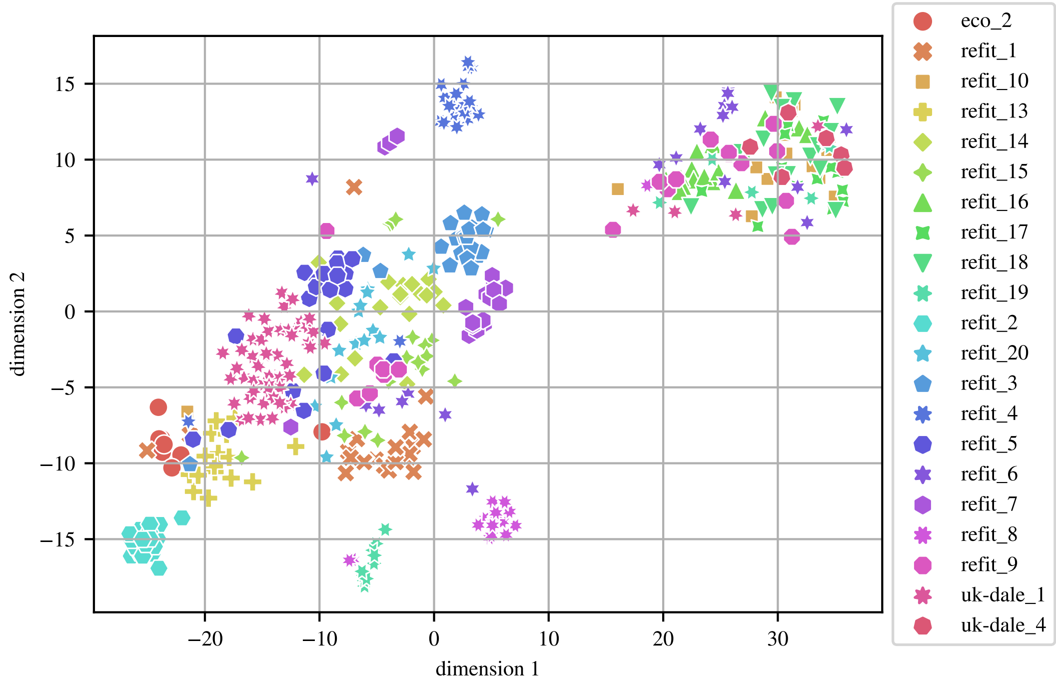
FIGURE 5.12: Projection of kettle LPs for various buildings with actual samples



Full resolution figure: https://github.com/jenkoj/msc/tree/main/Figures/TSNE/TSNE_per_appliance/img_scatter_refit_kettle.png

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

FIGURE 5.13: Projection of TV LPs for various buildings



Full resolution figure: https://github.com/jenkoj/msc/tree/main/Figures/TSNE/TSNE_per_appliance/scatter_refit_television.pdf

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

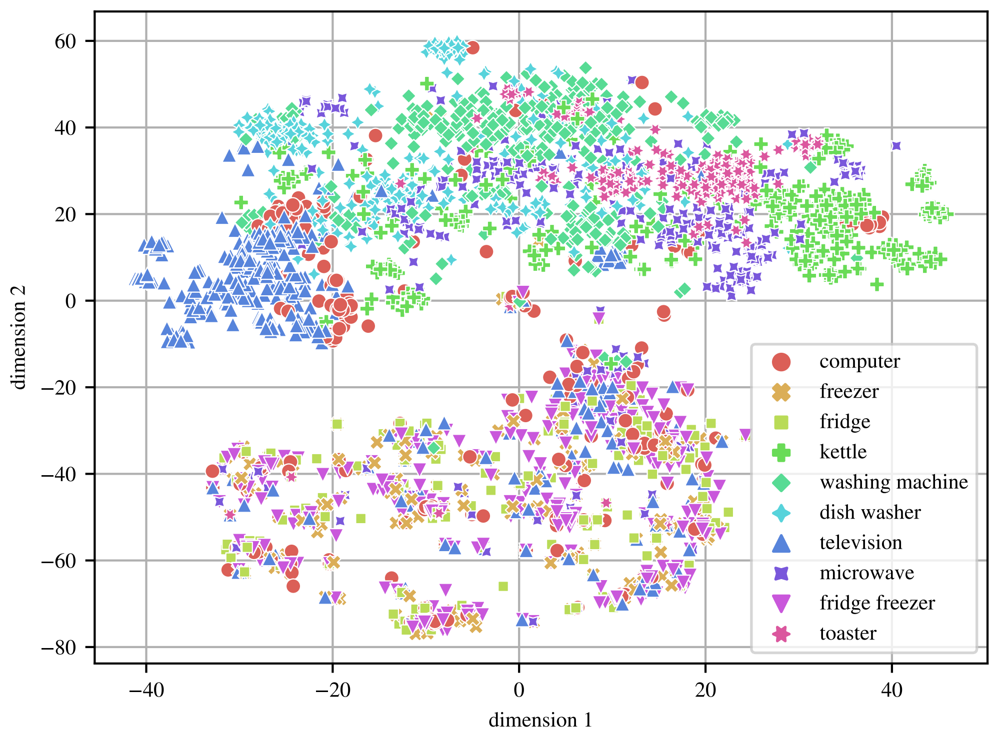
FIGURE 5.14: Projection of TV LPs for various buildings with actual samples.

Full resolution figure: https://github.com/jenkoj/msc/tree/main/Figures/TSNE/TSNE_per_appliance/img_scatter_refit_television.png

Per-Appliance LPs - Comparing Appliances

To get a general idea of where each appliance group lies, let's filter out all appliances that have less than 150 samples. Applying this filter yields Figure 5.15.

FIGURE 5.15: Projection of filtered per-appliance LPs



Full resolution figure:
https://github.com/jenkoj/msc/tree/main/Figures/TSNE/TSNE_PHPA/phpa_reduced_15.pdf

Figure 5.15 shows how these 10 appliances are connected in high dimensional space. Kettles, microwaves and toasters are quite similar when it comes to usage patterns. They are operated for a short amount of time and are usually used in users' routines in the morning or evening. These appliances are located in the upper left part of the plot.

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

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

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

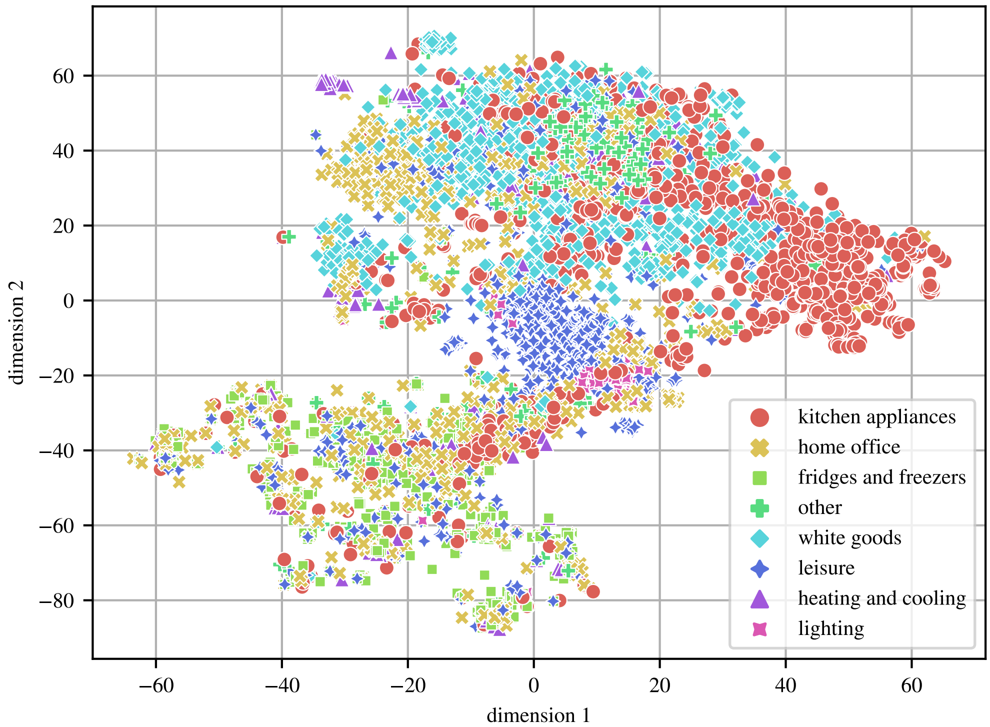
Knowing that a pattern exists, we can use the newly found group to define new appliance groups. The following 8 groups will be defined

- Kitchen appliances - toasters, ovens, microwaves, etc.
- Fridges and freezers - contains fridges, freezers and fridge freezers or white goods with a compressor
- White goods - washers, dryers, dishwashers i.e. white goods with a program
- heating and cooling - Electric radiators, dehumidifiers and HVACs
- leisure - Living room appliances such as TVs, games consoles, audio amps, HTPCs, etc.

- home office - Computer, laptops, printers, network equipment, chargers, etc.
- lightning - lights and lamps
- Others - unknown and unlabeled appliances

Applying these groups yields Figure 5.16. The new plot shows how, although appliances could be used by a different user, maybe even by users in a different part of the EU or world, they can be grouped in a high-dimensional space.

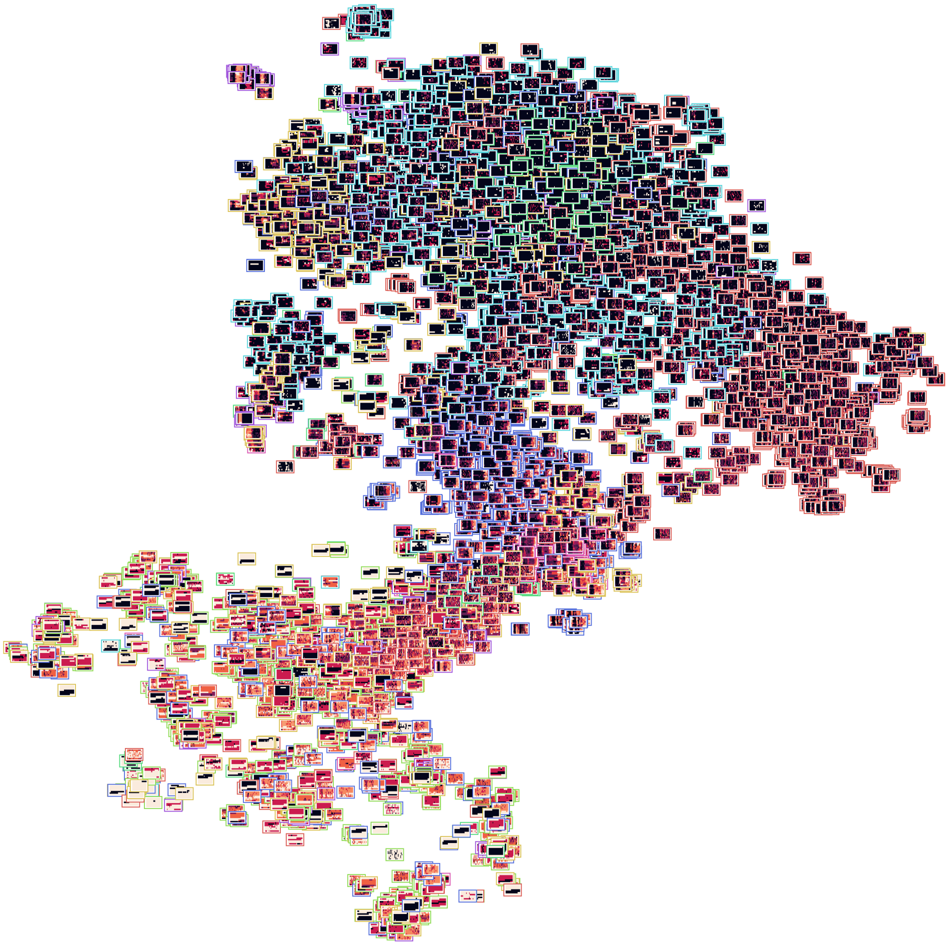
FIGURE 5.16: Projection of grouped per-appliance LPs



Full resolution figure:
https://github.com/jenkoj/msc/tree/main/Figures/TSNE/TSNE_PHPA/phpa_grouped_15.pdf

The Figure 5.17 below is the same as the first Figure **??** in the subsection, except it is easier to use color to see the appliance they present

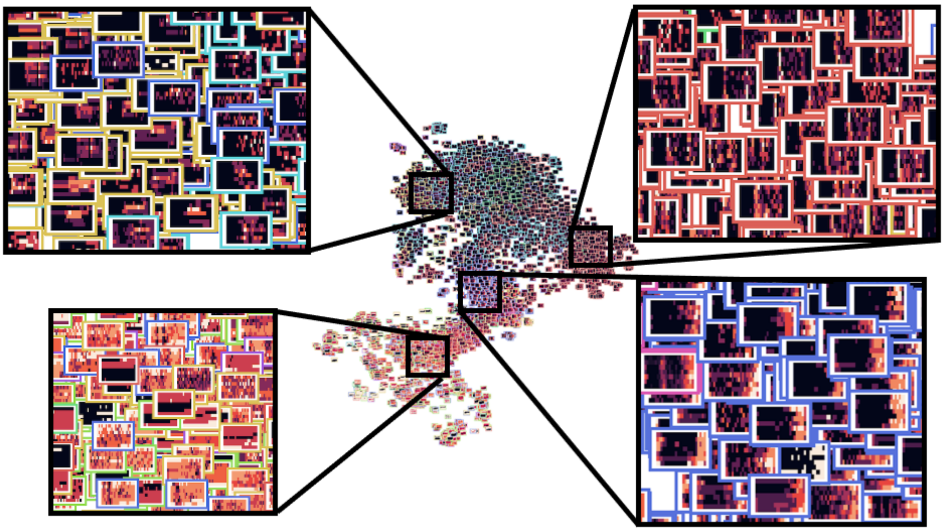
FIGURE 5.17: Projection of grouped per-appliance LPs with actual samples



Full resolution figure: `https://github.com/jenkoj/msc/tree/main/Figures/TSNE/TSNE_PHPA/img_scatter_all_all_groups.png`

To better emphasize the details from Figure 5.17 and 5.16 we present zoomed-in areas of key locations with Figure 5.18.

FIGURE 5.18: Projection of grouped per-appliance LPs with actual samples



Full resolution figure:
https://github.com/jenkoj/msc/tree/main/Figures/TSNE/TSNE_PHPA/t-sne_zoomed.png

5.3.3 Per-Appliance Per-Building

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

Bag of Appliances

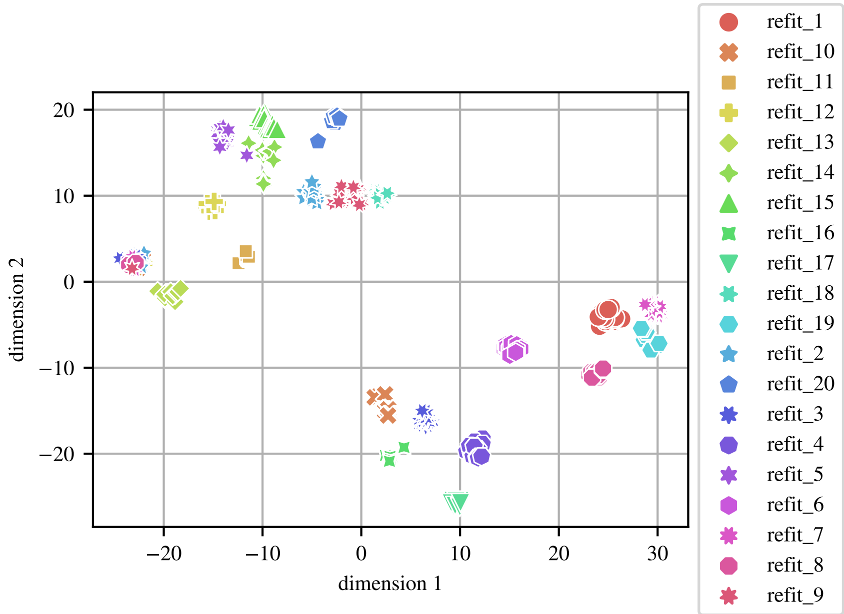
This LP is a combination of the LPs above, except it offers a larger detail when observing groups of appliances. Since we are using one dimension for appliances, we will use only the daily dimension.

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

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

This is the main reason why we can see in Figure 5.19 the clusters are separated quite a bit. We can still see that some clusters are closer than others, meaning they are more similar.

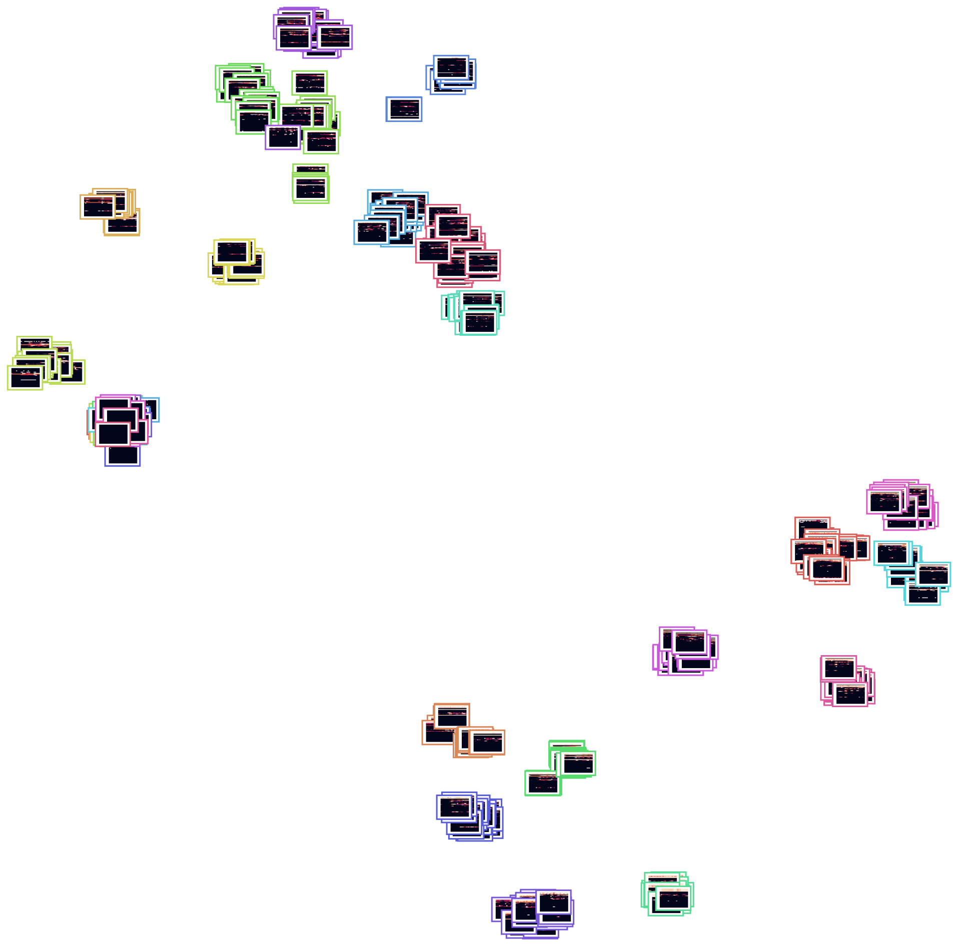
FIGURE 5.19: Projection of a bag of appliances LPs for various buildings



Full resolution figure:
https://github.com/jenkoj/msc/tree/main/Figures/TSNE/TSNE_BOA/scatter_refit_boa.pdf

Figure 5.20 shows that LPs are split between two poles. By observing the Figure it is possible to see that all the bottom clusters have more than one active white good with a compressor (fridges and freezers), while the top ones have only one. In general, the bottom buildings have more appliances, with more activity than the top ones.

FIGURE 5.20: Projection of a bag of appliances LPs for various buildings with actual samples



Full resolution figure:
https://github.com/jenkoj/msc/tree/main/Figures/TSNE/TSNE_BOA/img_scatter_boa.png

5.4 Discussion

We used t-SNE to show how LPs are related in high-dimensional space, by mapping them into two-dimensional space. We used three different types of LPs: per-building, per-building per-appliance, a bag of appliances, and per-appliance. Per-building load profiles offered a look into how activations patterns differ across different buildings and datasets. Per-building per-appliance bag of appliance load profiles offered the same thing, but in greater detail. Per-appliance load profiles were the most versatile and were utilized in the most various ways: First, we have shown how the same type of appliance is being used across various buildings. Next, we compared appliances with each other. Since the plot was hard to comprehend, we have defined appliance groups. These new groups formed clusters, which furthermore revealed the relation between LPs. Finally, we compared how appliance load profiles are connected in a single building.

One of the main findings of this chapter was the formation of appliance groups. Such groups enable us to look into the similarity of their activation profiles and enable us to understand which groups have similar usage patterns. Another important piece of information these groups contain is the strength of the user's routine. The closer the samples, the more similar their activation is, which means the user has a higher routine. Such a routine will be useful in the next chapter, where we will try to evaluate if it is strong enough to detect anomalies.

5.5 Summary

The analysis provided a look into the relationships between LPs and their consumption patterns. We were able to group appliances into categories and found a presence of routine in the LPs These findings will be valuable in the next chapter where we continue to explore the potential applications of LPs.

Appendix A

The source code, high-resolution figures and datasets

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

A.1 The source code

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

Individual scripts can be found in the following Jupyter Notebooks:

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

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

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

A.2 High resolution figures

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

A.3 Data and datasets

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

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

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

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

Appendix B

Expanded General Table

TABLE B.1: Expanded general table of load profiles

	frequency	appliances	number of activations	power (avg)	operating time
appliances		X	X	X	X
number of activations	X	[13] [35]	X	X	X
power (avg)	X	[58]		X	X
power (array)	[35]	X	X	X	X
power (histogram)			X	X	X
operating time	X	[33]	[55] [54] [4]	[4]	X
time array	X	X	[13] [35]	[17] [20] [10] [34] [67] [24] [23] [32] [1] [36] [55] [54] [31] [4] [14] [37] [17] [12] [48] [35] [23]	[22]

Bibliography

1. Joana M. Abreu, Francisco Câmara Pereira, and Paulo Ferrão. "Using pattern recognition to identify habitual behavior in residential electricity consumption". In: *Energy and Buildings* 49 (2012), 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. M.H. Albadi and E.F. El-Saadany. "A summary of demand response in electricity markets". In: *Electric Power Systems Research* 78.11 (2008), 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. A. Aréchiga et al. "Comparison of Dimensionality Reduction Techniques for Clustering and Visualization of Load Profiles". In: *2016 IEEE PES Transmission and Distribution Conference and Exposition Latin America*. 2016, 1 to 6. DOI: 10.1109/TDC-LA.2016.7805661.
4. Elnaz Azizi, Mohammad T. H. Beheshti, and Sadegh Bolouki. "Appliance-Level Anomaly Detection in Nonintrusive Load Monitoring via Power Consumption-Based Feature Analysis". In: *IEEE Transactions on Consumer Electronics* 67.4 (Nov. 2021), pp. 363–371. ISSN: 1558-4127. DOI: 10.1109/TCE.2021.3129356.
5. N. Batra et al. "It's Different: Insights into home energy consumption in India". In: *Proceedings of the 5th ACM Workshop on Embedded Systems For Energy-Efficient Buildings*. 2013, 1 to 8.
6. Nipun Batra et al. "NILMTK". In: *Proceedings of the 5th international conference on Future energy systems*. ACM, June 2014. DOI: 10.1145/2602044.2602051. URL: https://doi.org/10.1145%2F2602044.2602051.
7. Christian Beckel et al. "The ECO Data Set and the Performance of Non-Intrusive Load Monitoring Algorithms". In: Nov. 2014. DOI: 10.1145/2674061.2674064.
8. Blaž Bertalanič, Jakob Jenko, and Carolina Fortuna. *Dimensionality Expansion of Load Monitoring Time Series and Transfer Learning for EMS*. 2022. DOI: 10.48550/ARXIV.2204.02802. URL: https://arxiv.org/abs/2204.02802.
9. Ekaba Bisong. "Google Colaboratory". In: *Building Machine Learning and Deep Learning Models on Google Cloud Platform A Comprehensive Guide for Beginners*. Berkeley, CA: Apress, 2019, 59 to 64. ISBN: 978-1-4842-4470-8. DOI: 10.1007/978-1-4842-4470-8_7. URL: https://doi.org/10.1007/978-1-4842-4470-8_7.
10. R. Bitterer and Prof. Dr. habil. B. Schieferdecker. *Repräsentative VDEW-Lastprofile Aktionsplan Wettbewerb, M-32 99*. VDEW. 1999.
11. Tobias Boßmann, Joachim Schleich, and Robert Schurk. "Unravelling load patterns of residential end-uses from smart meter data". In: June 2015.
12. A. Capasso et al. "A bottom-up approach to residential load modeling". In: *IEEE Transactions on Power Systems* 9.2 (May 1994), pp. 957–964. ISSN: 1558-0679. DOI: 10.1109/59.317650.

13. Cristina Carpino et al. “Energy performance gap of a nearly Zero Energy Building in Denmark the influence of occupancy modelling”. In: *Building Research Information* 48 (Jan. 2020).

14. Marco Castangia et al. “Detection of Anomalies in Household Appliances from Disaggregated Load Consumption”. In: *2021 International Conference on Smart Energy Systems and Technologies (SEST)*. 2021, pp. 1–6. DOI: 10.1109/SEST50973.2021.9543232.

15. Maurizio Cellura et al. “The role of the building sector for reducing energy consumption and greenhouse gases: An Italian case study”. In: *Renewable Energy* 60 (2013), 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.

16. Chen Chen. “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, 2018, 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.

17. Luo Chuan, D. M. K. K. Venkateswara Rao, and Abhisek Ukil. “Load profiling of Singapore buildings for peak shaving”. In: *2014 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC)*. 2014, pp. 1–6. DOI: 10.1109/APPEEC.2014.7065998.

18. Andrew Collette. *Python and HDF5*. O’Reilly, 2013.

19. Communication from the Commission et al. “Action Plan for energy efficiency: realising the potential”. In: *COM (2006)* 545 (2006).

20. Tamás Csoknyai et al. “Analysis of energy consumption profiles in residential buildings and impact assessment of a serious game on occupants’ behavior”. In: *Energy and Buildings* 196 (2019), 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.

21. Eurostat. “Gross and net production of electricity and derived heat by type of plant and operator”. In: 62 (2022). URL: https://ec.europa.eu/eurostat/databrowser/view/nrg_ind_peh/default/table?lang=en.

22. David Fischer, Andreas Härtl, and Bernhard Wille-Haussmann. “Model for electric load profiles with high time resolution for German households”. In: *Energy and Buildings* 92 (2015), 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.

23. Bingtuan Gao, Xiaofeng Liu, and Zhenyu Zhu. “A Bottom-Up Model for Household Load Profile Based on the Consumption Behavior of Residents”. In: *Energies* 11.8 (2018). ISSN: 1996-1073. DOI: 10.3390/en11082112. URL: https://www.mdpi.com/1996-1073/11/8/2112.

24. D. Gerbec et al. “Allocation of the load profiles to consumers using probabilistic neural networks”. In: *IEEE Transactions on Power Systems* 20.2 (May 2005), pp. 548–555. ISSN: 1558-0679. DOI: 10.1109/TPWRS.2005.846236.

25. Giorgio Graditi et al. “Heuristic-Based Shiftable Loads Optimal Management in Smart Micro-Grids”. In: *IEEE Transactions on Industrial Informatics* 11.1 (Feb. 2015), pp. 271–280. ISSN: 1941-0050. DOI: 10.1109/TII.2014.2331000.

26. Charles R. Harris et al. “Array programming with NumPy”. In: *Nature* 585.7825 (2020), 357 to 362. DOI: 10.1038/s41586-020-2649-2. URL: https://doi.org/10.1038/s41586-020-2649-2.
27. Yassine Himeur et al. “Artificial intelligence based anomaly detection of energy consumption in buildings: A review, current trends and new perspectives”. In: *Applied Energy* 287 (2021), p. 116601. ISSN: 0306-2619. DOI: https://doi.org/10.1016/j.apenergy.2021.116601. URL: https://www.sciencedirect.com/science/article/pii/S0306261921001409.
28. Geoffrey Hinton and Sam Roweis. “Stochastic Neighbor Embedding”. In: *Advances in neural information processing systems* 15 (2003). Ed. by S Thrun S Becker and KEditors Obermayer, 833 to 840.
29. Ryan Hledik and Tony Lee. “Chapter 9 - Load flexibility: Market potential and opportunities in the United States”. In: *Variable Generation, Flexible Demand*. Ed. by Fereidoon Sioshansi. Academic Press, 2021, 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.
30. J. D. Hunter. “Matplotlib: A 2D graphics environment”. In: *Computing in Science and Engineering* 9.3 (2007), 90 to 95. DOI: 10.1109/MCSE.2007.55.
31. Fatih Issi and Orhan Kaplan. “The Determination of Load Profiles and Power Consumptions of Home Appliances”. In: *Energies* 11.3 (2018). ISSN: 1996-1073. DOI: 10.3390/en11030607. URL: https://www.mdpi.com/1996-1073/11/3/607.
32. Hyun Cheol Jeong et al. “Clustering of Load Profiles of Residential Customers Using Extreme Points and Demographic Characteristics”. In: *Electronics* 10.3 (2021). ISSN: 2079-9292. DOI: 10.3390/electronics10030290. URL: https://www.mdpi.com/2079-9292/10/3/290.
33. Emily Kattan, Suleiman Halasah, and Tareq a hamed. “Practical Challenges of Photovoltaic Systems in the Rural Bedouin Villages in the Negev”. In: *Journal of Fundamentals of Renewable Energy and Applications* 8 (May 2018). DOI: 10.4172/20904541.1000258.
34. Amir Kavousian, Ram Rajagopal, and Martin Fischer. “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 (2013), 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.
35. Jack Kelly and William Knottenbelt. “The UK-DALE dataset, domestic appliance-level electricity demand and whole-house demand from five UK homes”. In: *Scientific Data* 2.1 (Mar. 2015), p. 150007. ISSN: 2052-4463. DOI: 10.1038/sdata.2015.7. URL: https://doi.org/10.1038/sdata.2015.7.
36. Elke Klaassen, Jasper Frunt, and Han Slootweg. “Assessing the impact of distributed energy resources on LV grids using practical measurements”. In: June 2015.
37. Wilhelm Kleiminger et al. “Occupancy Detection from Electricity Consumption Data”. In: Nov. 2013, pp. 1–8. DOI: 10.1145/2528282.2528295.

38. Thomas Kluyver et al. "Jupyter Notebooks, a publishing format for reproducible computational workflows". In: *Positioning and Power in Academic Publishing: Players, Agents and Agendas*. Ed. by F. Loizides and B. Schmidt. IOS Press. 2016, 87 to 90.

39. J. Zico Kolter and Matthew J. Johnson. "REDD: A Public Data Set for Energy Disaggregation Research". In: *IN SUSTKDD*. 2011.

40. Congmiao Li, Dipti Srinivasan, and Thomas Reindl. "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)*. Nov. 2015, pp. 1–6. DOI: 10.1109/ISGT-Asia.2015.7387165.

41. J.A.P. Lopes, C.L. Moreira, and A.G. Madureira. "Defining control strategies for MicroGrids islanded operation". In: *IEEE Transactions on Power Systems* 21.2 (May 2006), pp. 916–924. ISSN: 1558-0679. DOI: 10.1109/TPWRS.2006.873018.

42. L.J.P. van der Maaten and G.E. Hinton. "Visualizing High-Dimensional Data Using tSNE". In: (2008).

43. Wes McKinney et al. "Data structures for statistical computing in python". In: *Proceedings of the 9th Python in Science Conference*. Vol. 445. Austin, TX. 2010, 51 to 56.

44. Hasan Mehrjerdi and Reza Hemmati. "Coordination of vehicle-to-home and renewable capacity resources for energy management in resilience and self-healing building". In: *Renewable Energy* 146 (2020), 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.

45. Andres F. Moreno Jaramillo et al. "Load modelling and non-intrusive load monitoring to integrate distributed energy resources in low and medium voltage networks". In: *Renewable Energy* 179 (2021), 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.

46. Khosrow Moslehi and Ranjit Kumar. "A Reliability Perspective of the Smart Grid". In: *IEEE Transactions on Smart Grid* 1.1 (June 2010), pp. 57–64. ISSN: 1949-3061. DOI: 10.1109/TSG.2010.2046346.

47. Yael Parag and Benjamin K. Sovacool. "Electricity market design for the prosumer era". In: *Nature Energy* 1.4 (Mar. 2016), p. 16032. ISSN: 2058-7546. DOI: 10.1038/nenergy.2016.32. URL: https://doi.org/10.1038/nenergy.2016.32.

48. Keon-Jun Park and Sung-Yong Son. "A Novel Load Image Profile-Based Electricity Load Clustering Methodology". In: *IEEE Access* 7 (2019), pp. 59048–59058. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2019.2914216.

49. The European Parliament and the Council of the European Union. "Directive 2010/31/EU of the European Parliament and of the Council of 19 May 2010 on the energy performance of buildings (recast)". In: (2021).

50. Luigi Patrono, Piercosimo Rametta, and Jochen Meis. "Unobtrusive Detection of Home Appliance's Usage for Elderly Monitoring". In: *2018 3rd International Conference on Smart and Sustainable Technologies (SpliTech)*. June 2018, pp. 1–6.

51. F. Pedregosa et al. "Scikit learn: Machine Learning in Python". In: *Journal of Machine Learning Research* 12 (2011), 2825 to 2830.

52. Elisavet Proedrou. "A Comprehensive Review of Residential Electricity Load Profile Models". In: *IEEE Access* 9 (2021), pp. 12114–12133. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2021.3050074.

53. Haroon Rashid. *Annotated load anomalies from the REFIT Dataset*. 2019. DOI: 10.15129/9729A2A0-11CE-4CCE-B0D0-144C483FCB33. URL: https://pureportal.strath.ac.uk/en/datasets/9729a2a0-11ce-4cce-b0d0-144c483fcb33.

54. Haroon Rashid et al. "Can non-intrusive load monitoring be used for identifying an appliances anomalous behaviour?" In: *Applied Energy* 238 (2019), 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.

55. Haroon Rashid et al. "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)*. May 2019, pp. 8325–8329. DOI: 10.1109/ICASSP.2019.8683792.

56. Carla B. Robledo et al. "Integrating a hydrogen fuel cell electric vehicle with vehicle-to-grid technology, photovoltaic power and a residential building". In: *Applied Energy* 215 (2018), 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.

57. Jingshuang Shen, Chuanwen Jiang, and Bosong Li. "Controllable Load Management Approaches in Smart Grids". In: *Energies* 8.10 (2015), pp. 11187–11202. ISSN: 1996-1073. DOI: 10.3390/en81011187. URL: https://www.mdpi.com/1996-1073/8/10/11187.

58. Mercy shuma-iwisi. "Domestic appliances standby power losses: The case of eleven suburbs in the greater Johannesburg," in: June 2009.

59. Yousra Sidqi et al. "Flexibility quantification in households: a swiss case study". In: *Energy Informatics* 3 (Oct. 2020). DOI: 10.1186/s42162-020-00126-4.

60. Catalina Spataru and Stephanie Gauthier. "How to monitor people 'smartly' to help reducing energy consumption in buildings?" In: *Architectural Engineering and Design Management* 10.1-2 (2014), 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.

61. Lukas G. Swan and V. Ismet Ugursal. "Modeling of end-use energy consumption in the residential sector: A review of modeling techniques". In: *Renewable and Sustainable Energy Reviews* 13.8 (2009), 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.

62. Kenneth Train, Joseph Herriges, and Robert Windle. "Statistically adjusted engineering (SAE) models of end-use load curves". In: *Energy* 10.10 (1985), 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.

63. Guido Van Rossum. *The Python Library Reference*. Python Software Foundation, 2020.

64. Geert P.J. Verbong, Sjouke Beemsterboer, and Frans Sengers. “Smart grids or smart users? Involving users in developing a low carbon electricity economy”. In: *Energy Policy* 52 (2013). 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.
65. Pauli Virtanen et al. “SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python”. In: *Nature Methods* 17 (2020), 261 to 272. DOI: 10.1038/s41592-019-0686-2.
66. Paolo Visconti et al. “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)*. June 2019, pp. 264–269. DOI: 10.1109/IWASI.2019.8791399.
67. C.F. Walker and J.L. Pokoski. “Residential Load Shape Modelling Based on Customer Behavior”. In: *IEEE Transactions on Power Apparatus and Systems* PAS-104.7 (1985), pp. 1703–1711. DOI: 10.1109/TPAS.1985.319202.
68. Michael L. Waskom. “seaborn: statistical data visualization”. In: *Journal of Open Source Software* 6.60 (2021), p. 3021. DOI: 10.21105/joss.03021. URL: https://doi.org/10.21105/joss.03021.
69. Sook-Chin Yip et al. “An anomaly detection framework for identifying energy theft and defective meters in smart grids”. In: *International Journal of Electrical Power and Energy Systems* 101 (2018), 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.
70. Alireza Zakariazadeh et al. “A new approach for real time voltage control using demand response in an automated distribution system”. In: *Applied Energy* 117 (2014), 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.