Diss. ETH No. 23058

Scalable and Personalized Energy Efficiency Services with Smart Meter Data

A thesis submitted to attain the degree of

DOCTOR OF SCIENCES of ETH ZURICH

(Dr. sc. ETH Zurich)

presented by

Christian Beckel

Dipl.-Inf., Eberhard Karls University, Tübingen

born on 25 April 1983

citizen of Germany

accepted on the recommendation of

Prof. Dr. Friedemann Mattern, examiner

Prof. Dr. Karl Aberer, co-examiner

Prof. Dr. Silvia Santini, co-examiner

Prof. Dr. Thorsten Staake, co-examiner

2016

Abstract

Information and communication technology plays an important role in addressing the world's energy problem. Networked digital electricity meters (so-called smart meters), for instance, can provide households with real-time information on their electricity consumption and thus help them to conserve energy. Initial expectations on the saving potential of this technology were too optimistic, however. In fact, recent pilot studies conducted under realistic assumptions have shown that savings induced by plain electricity consumption feedback are often significantly lower than many have originally expected.

In this dissertation, we take smart metering to a new level as we explore a data analysis-driven approach to personalize energy efficiency services that may be offered at large scale. An example for such a service is automated energy consulting, which consists in automatically providing energy saving recommendations to households by taking into account their appliance stock and usage profiles. In addition, we provide the foundation for an electricity bill that is tailored to the household as it shows the contribution of individual appliances to the overall bill or compares a household's consumption with other households that have similar characteristics. Behavioral trials indicate that such consumption feedback is potentially more successful in motivating households to reduce their electricity consumption than plain consumption feedback or generic energy saving recommendations.

One contribution of this thesis is the design, development, and evaluation of a system that automatically estimates characteristics of a household (like its socio-economic status, dwelling properties, and appliance stock) from the household's electricity consumption data. We evaluate our approach on real world smart meter data collected from more than 4000 households over a period of 1.5 years. Our analysis shows that inferring household characteristics is feasible, as our method achieves an accuracy of more than 70% over all households for many of the characteristics and even exceeds 80% for some of the characteristics. For utilities, the system creates valuable customer insights

that—without having to perform costly and cumbersome surveys—help to run energy efficiency campaigns more efficiently by targeting each household with the adequate service (e.g., offering energy consulting for retired people and a smart heating system, which automatically controls the thermostat based on occupancy, for employed people). Furthermore, these insights can be used to realize automated peer group comparisons on the electricity bill or in an online portal.

Providing automated, household-specific energy saving recommendations requires more detailed information about a household than its high-level characteristics. In particular, it is important to know when individual appliances are running and how much they consume. To avoid measuring each appliance individually through a complex sensing infrastructure, we investigate inferring this information from the overall electricity consumption measured by a smart meter. To explore this concept (non-intrusive load monitoring, NILM), we developed an evaluation framework and analyzed the performance of several state-of-the-art NILM algorithms. To this end, we collected electricity consumption data in six Swiss households over a period of eight months and made it publicly available. Along with fine-grained smart meter data (collected at 1 Hz), our data set contains ground truth measurements of 47 selected appliances and each of the household's occupancy state. Our analysis shows that—through the enhancement of an existing NILM algorithm—it is possible to achieve recognition rates of more than 90% for some typical appliances. This is sufficient for energy consulting scenarios; its practical use is limited, however, since a training period is required.

Ultimately, deploying smart meters comes with a cost that—for some of the households—can be higher than the achievable savings given today's energy prices. Maximizing societal benefits thus requires a well-managed interplay between (1) regulators, which define rules for smart meter deployments and set penalties if saving targets are not reached, (2) utilities, which develop and run energy efficiency campaigns, and (3) households, which should invest in energy saving solutions or adapt their lifestyle in order to use energy more efficiently. This thesis copes with this challenge as it shows how to utilize Internet of Things technologies and machine learning methods to enable personalized energy efficiency services that scale to thousands or even millions of households. We develop methods, build open source evaluation frameworks, and collect and analyze real world consumption data in order to better understand residential electricity consumption and improve the effect (and thus the value) of smart meter deployments and feedback mechanisms.

Kurzfassung

Informations- und Kommunikationstechnik kann einen substantiellen Beitrag zur Lösung unseres Energieproblems leisten. Intelligente Stromzähler (Smart Meter) können beispielsweise Privathaushalte zeitnah über ihren Stromverbrauch informieren und die Bewohner somit beim Stromsparen unterstützen. Pilotstudien, die in den letzten Jahren unter realistischen Bedingungen durchgeführt wurden, haben allerdings gezeigt, dass die Einsparungen, die durch einfaches Verbrauchsfeedback erzielt werden, um einiges geringer sind als von vielen erwartet.

In der vorliegenden Dissertation gehen wir über einfaches Verbrauchsfeedback hinaus und erforschen Ansätze zur Analyse von Smart-Meter-Daten mit dem Ziel, personalisierte Energieeffizienzdienstleistungen in grossem Umfang zu ermöglichen. Ein Beispiel für eine solche Dienstleistung ist eine Energieberatung, die automatisch Energiesparempfehlungen auf Basis der im Haushalt verfügbaren Haushaltsgeräte und deren Nutzung erstellt. Darüber hinaus legen wir die Grundlagen für eine auf den Haushalt zugeschnittene Stromrechnung, die den Beitrag einzelner Geräte zum Gesamtstromverbrauch darstellt oder den Stromverbrauch des Haushalts mit dem ähnlicher Haushalte vergleicht. Verhaltensstudien deuten darauf hin, dass sich durch solch personalisiertes Feedback höhere Einsparungen erzielen lassen als durch generische Energiespartipps oder durch Verbrauchsfeedback, das aus blossen Zahlen besteht. Die Erzeugung dieser Art von Feedback erfordert jedoch Kenntnisse über den Haushalt wie beispielsweise die Zahl der Bewohner, Zahl und Art der Geräte sowie deren Nutzung. Diese Informationen liegen Anbietern von Energieeffizienzdienstleistungen in der Regel nicht vor beziehungsweise sind nur in Form von kostenintensiven Umfragen oder durch zusätzliche Messinfrastruktur erzielbar.

Einer der Beiträge dieser Dissertation besteht im Entwurf, der Entwicklung und der Bewertung eines Systems, das Charakteristiken eines Haushalts (z.B. den sozioökonomischen Status der Bewohner, Gebäudeeigenschaften oder die Zahl der Geräte) aus dessen Stromverbrauch schätzt. Wir evaluieren unseren Ansatz zur automatischen Haushaltsklassifizierung mit Hilfe von Stromverbrauchsdaten aus über 4000 Haushalten, die über einen Zeitraum von anderthalb Jahren erfasst wurden. Unsere Analyse zeigt, dass eine automatische Haushaltsklassifizierung möglich ist, da wir mit unserer Methode eine Genauigkeit von 70% für die meisten Haushaltseigenschaften und über 80% für manche Haushaltseigenschaften erzielen. Das System ermöglicht Energieversorgern, nützliche Kundeninformationen zu ermitteln, ohne aufwändige Umfragen durchzuführen. Diese können daher Energiesparkampagnen effizient gestalten, indem sie jeden Haushalt mit der für ihn am

besten geeigneten Dienstleistung ansprechen (z.B. für Rentner eine persönliche Energieberatung und für Berufstätige die Installation einer intelligenten Heizungssteuerung, welche die Temperatur automatisch auf Basis der An- und Abwesenheit der Bewohner regelt). Des Weiteren kann dieses Kundenwissen genutzt werden, um einen automatischen zielgruppenspezifischen Vergleich auf der Stromrechnung oder in einem Online-Portal anzubieten.

Die automatische Erzeugung personalisierter Energiesparempfehlungen erfordert detailliertere Informationen als die durch die Haushaltsklassifizierung ermittelbaren Charakteristiken. Es ist hierzu wichtig zu wissen, wann einzelne Geräte in Betrieb sind und wie viel Strom sie benötigen. Im Rahmen dieser Dissertation untersuchen wir den Ansatz, diese Informationen ebenfalls aus dem Gesamtstromverbrauch abzuleiten. Zur Bewertung dieses Konzepts (non-intrusive load monitoring, NILM) entwickelten wir ein Evaluationsframework und ermittelten die Genauigkeit mehrerer NILM-Verfahren. Für unsere Analyse erhoben wir Stromverbrauchsdaten aus sechs Schweizer Haushalten über einen Zeitraum von acht Monaten. Zusätzlich zu den hochauflösenden Messungen des Gesamtstromverbrauchs der Haushalte (mit einer Frequenz von 1 Hz) umfasst unser Datensatz auch Ground-Truth-Messungen von 47 ausgewählten Haushaltsgeräten sowie Informationen über die An- und Abwesenheit der Bewohner. Unsere Analyse zeigt, dass es mit Hilfe eines von uns erweiterten NILM-Verfahrens möglich ist, den Stromverbrauch und die Schaltzeitpunkte einiger typischer Geräte mit über 90% Genauigkeit zu schätzen. Dies ist ausreichend für eine automatisierte Energieberatung; allerdings ist der praktische Einsatz durch das erforderliche Training des Systems begrenzt.

Einbau und Betrieb von Smart Metern verursachen Kosten, die bei manchen Haushalten höher sind als die monetären Einsparungen, die (bei gegenwärtigen Strompreisen) durch die Verbrauchsreduktion erzielt werden. Die Maximierung des gesellschaftlichen Nutzens erfordert daher ein Zusammenspiel mehrerer Akteure, und zwar (1) der Regulatoren, die Vorschriften für den Smart-Meter-Ausbau erlassen und Bussen für Energieversorger festlegen können, falls Effizienzziele nicht eingehalten werden, (2) der Energieversorger, die Energieeffizienzprogramme entwickeln und umsetzen, sowie (3) der Haushalte, die in Energieeffizienzlösungen investieren oder ihren Lebensstil anpassen sollen, um Strom möglichst effizient zu nutzen. Die vorliegende Dissertation geht diese Herausforderung an, indem sie aufzeigt, wie unter Verwendung von Internet-der-Dinge-Technologien und Methoden aus dem maschinellen Lernen personalisierte Energieeffizienzdienstleistungen entwickelt werden können, die kostengünstig auf Tausende oder Millionen Haushalte skalieren. Hierzu entwickeln wir Methoden, Open-Source-Frameworks, sammeln Stromverbrauchsdaten und analysieren diese, um den Stromverbrauch von Haushalten besser zu verstehen und den Effekt (und damit den Wert) des Smart-Meter-Ausbaus und des Verbrauchsfeedbacks für die Gesellschaft zu erhöhen.

Acknowledgements

This dissertation was created between 2011 and 2015 during my time as a researcher in the Distributed Systems group at ETH Zurich and in the Bits to Energy Lab, a joint research initiative of ETH Zurich and University of St. Gallen. I also performed a part of the work during a research visit at the Stanford Sustainable Systems Lab in the US. At this point, I would like to thank the Hans L. Merkle foundation (funded by Robert Bosch GmbH) for providing me with a scholarship to pursue my PhD in such an inspiring environment.

Foremost, I would like to express my deep gratitude to my doctoral advisor Friedemann Mattern. With his vision, his ideas, and his special view on the world and its people, Friedemann has inspired me many times during my PhD. Next, I would like to thank Silvia Santini and Thorsten Staake for mentoring me throughout all stages of my PhD. Thanks to your support I learned how to identify and solve relevant research challenges, how to plan and perform projects, and how to present results to the scientific community and beyond. I would also like to thank Karl Aberer for being a co-examiner of my thesis and Elgar Fleisch for letting me participate twice a year in the doctoral seminars of his group. I would further like to thank the staff of ETH Zurich, in particular Barbara von Allmen Wilson and Denise Spicher, for providing such excellent working conditions.

My special appreciation goes to my colleagues and friends Wilhelm Kleiminger and Leyna Sadamori. Willi served as a strong companion during the whole adventure: We shared an office, we collaborated on many different smart energy projects, and he assisted me with numerous on- and off-topic issues. The work with Leyna has started three years ago with his Master's thesis, and I am extremely grateful that we could continue our collaboration and share several memorable conference trips and ski outings since then.

An inspiring work environment is the key to successful research. Therefore, I would like to thank each of the group members for all professional and social interactions we had during the last years, which helped me to learn and understand the world a bit better:

Gábor Sörös and Matthias Kovatsch for a great summer school in Oulu, Simon Mayer for his active nature and positive mindset, Anwar Hithnawi and Hossein Shafagh for their research spirit, Elke Schaper and Aurelia Tamò for bringing some fresh air into our group, Mihai Bâce for his sense of humor, Subho Basu for the Indian cultural experience, Marian George, and Hông-Ân Cao. I also thank our former members for sharing their experiences when I joined the group, notably Christian Flörkemeier, Robert Adelmann, Benedikt Ostermaier, Alexander Bernauer, Markus Weiss, and Dominique Guinard. I am particularly grateful that I had the opportunity to work with many talented students who directly or indirectly supported the development of this thesis.

Beyond the boundaries of our group, I would like to thank—among many others—Verena Tiefenbeck, Claire-Michelle Loock, Felix Lossin, and Vojkan Tasic from the Bits to Energy Lab, Dominik Wörner, Thomas von Bomhard, and Markus Weinberger from the Bosch IoT Lab as well as Paul Baumann from TU Darmstadt for discussions, cooperations, and social experiences. Special thanks go to Ram Rajagopal for having me as a visiting researcher in Stanford as well as to Amir Kavousian and Adrian Albert for our cooperation during that time.

This research would not have been possible without the strong support by Energie Thun: Big thanks to Martin Bühler, Christoph Woodtli, and the four other households who bravely let us equip their homes with sensors for more than 8 months. In addition, I would like to thank the Irish Commission for Energy Regulation for making the data collected in their smart metering field trial accessible to the public.

Finally, I would like to express my deepest gratitude to my friends and my family. To my brothers for being who they are and for their contagious humor; to my parents for always being there when I need(ed) them; and most important to Jasmin, my love, for what you took upon yourself to help me follow my dreams. Without your support, I would have never started this endeavor. Thank you for everything!

Contents

Acronyms **xiii**

1 Introduction **1**
1.1 Motivation . . . 3
1.2 Research goals and contributions . . . 8
1.2.1 Automated household classification . . . 9
1.2.2 Non-intrusive load monitoring . . . 9
1.2.3 Applications for smart meter data analytics . . . 9
1.3 Thesis outline . . . 10

2 Background and related work **11**
2.1 Smart metering . . . 12
2.2 Smart meter data analytics . . . 14

3 Automated household classification **27**
3.1 The CER data set . . . 29
3.2 System design . . . 37
3.2.1 Features . . . 38
3.2.2 Household characteristics and class labels . . . 40
3.2.3 Classifiers . . . 49
3.2.4 Multiple linear regression . . . 52
3.3 Evaluation process . . . 53
3.3.1 Performance measures . . . 53
3.3.2 Training, evaluation, and feature selection . . . 55
3.3.3 Implementation details . . . 57
3.4 Results . . . 57
3.4.1 Accuracy . . . 57
3.4.2 Matthews correlation coefficient . . . 59

3.4.3 Regression . . . 60
3.4.4 Stability of the results . . . 62
3.4.5 Multiple weeks . . . 63
3.4.6 Discussion and limitations . . . 66
3.5 Impact of outdoor temperature and daylight . . . 67
3.6 Features . . . 72
3.7 Data granularity . . . 74

4 Non-intrusive load monitoring 79
4.1 Design space . . . 80
4.2 Evaluated algorithms . . . 84
4.2.1 Description . . . 84
4.3 Evaluation methodology and the NILM-Eval framework . . . 89
4.4 The ECO data set . . . 92
4.4.1 Available NILM data sets . . . 93
4.4.2 Data collection and measurement infrastructure . . . 95
4.4.3 Detailed description of the data set . . . 104
4.4.4 ON/OFF events . . . 113
4.5 Results . . . 115
4.5.1 Parson's algorithm . . . 116
4.5.2 Baranski's algorithm . . . 120
4.5.3 Weiss' algorithm . . . 122
4.5.4 Kolter's algorithm . . . 125
4.5.5 Jung's algorithm . . . 126
4.6 Summary of the results . . . 128
4.6.1 Limitations and future work . . . 130
4.6.2 Conclusion . . . 131

5 Applications of smart meter data analytics 133
5.1 Customer selection . . . 133
5.2 Peer group comparison . . . 137
5.3 Consumption breakdown . . . 141
5.4 Energy consulting . . . 149
5.5 Implications for households, utilities, and policy makers . . . 151
5.5.1 Households . . . 152
5.5.2 Utilities . . . 152
5.5.3 Policy makers . . . 153

6 Conclusions and outlook 155
6.1 Automated household classification . . . 155

6.2 Non-intrusive load monitoring . . . 156
6.3 Applications and implications . . . 157
6.4 Limitations and outlook . . . 158
6.4.1 Household classification . . . 158
6.4.2 Non-intrusive load monitoring . . . 159

Bibliography **163**

Referenced Web resources **181**

A Interviews with energy consultants **187**
A.1 Interview 1 . . . 188
A.2 Interview 2 . . . 192
A.3 Interview 3 . . . 197
A.4 Interview 4 . . . 200

B Features and characteristics used for the SOM analysis **205**

C Detailed appliance information for household 5 **213**

Acronyms

AC alternating current.

AFAMAP Additive Factorial Approximate MAP.

AMPds Almanac of Minutely Power data set.

ARMA auto-regressive moving average.

AUC area under the curve.

BRG biased random guess.

CBT customer behavior trial.

CDF cumulative distribution function.

CER Commission for Energy Regulation.

CHP combined heat and power.

CIE chief income earner.

CRM customer relationship management.

CSV comma-separated values.

DINK double income – no kids.

DWT discrete wavelet transform.

ECO Electricity Consumption and Occupancy.

EMF electromagnetic field.

EU European Union.

FDR false discovery rate.

FFT fast Fourier transform.

FN false negative.

FP false positive.

FPR false positive rate.

FSM finite-state machine.

GDR German Democratic Republic.

GPRS General Packet Radio Service.

HES Household Electricity Use Study.

HMM hidden Markov model.

HTTP Hypertext Transfer Protocol.

HVAC heating, ventilation, and air conditioning.

iAWE Indian Dataset for Ambient Water and Energy.

ICT information and communication technology.

IoT Internet of Things.

ISODATA iterative self-organizing data analysis.

kNN k-Nearest Neighbors.

LDA linear discriminant analysis.

LLH load-based load hiding.

MAE mean absolute error.

MCC Matthews correlation coefficient.

NILM non-intrusive load monitoring.

NILMTK non-intrusive load monitoring toolkit.

NTP Network Time Protocol.

OBIS Object Identification System.

OCR optical character recognition.

OLR overall load reduction.

OLS ordinary least squares.

OPEC Organization of the Petroleum Exporting Countries.

PIR passive infrared.

PLC power-line communication.

RECS Residential Energy Consumption Survey.

REDD Reference Energy Disaggregation Data Set.

REST Representational State Transfer.

RFID radio-frequency identification.

RG random guess.

RMSE root-mean-square error.

ROC receiver operating characteristic.

SEAI Sustainable Energy Authority of Ireland.

SFFS sequential floating forward selection.

SML Smart Message Language.

SOM self-organizing map.

SSH Secure Shell.

SVM support vector machine.

TN true negative.

ToU time of use.

TP true positive.

TPR true positive rate.

TV television.

UK-DALE UK Domestic Appliance-Level Electricity.

US United States.

Chapter 1

Introduction

The need for a sustainable use of natural resources traces back many years. Already in the beginning of the 18th century, population growth and increasing energy needs of the mining industry had caused deforestation across Europe, which had been identified as a serious trouble for future generations and has lead to the creation of sustainable forestry [176]. Even though historians today argue about the severance of this energy crisis in Germany often referred to as *Holznot* (wood shortage), this topic lead to social and political tensions in local communities that stem from the need to ration the resource wood [142] and ultimately to the invention of innovative, more energy-efficient ovens.

Although the subsequent electrification and the rise of coal technology and hydropower as main resources to produce electricity had reduced the dependency on wood, the need for an efficient use of energy remained. In 1917, for instance, United States (US) artist Coles Philipp calls for an efficient use of lighting in the name of the United States Fuel Association (*cf.* figure 1.1). Later, in the 1940s during World War II, one of the first large-scale energy efficiency campaigns in history was performed in Germany: Through the *Kohlenklau*—a comic figure that steals coal—the National Socialist German Workers' Party propagated that everybody who wasted energy implicitly stole coal from the community [191]. After the war, a similar comic figure named *Wattfrass* (*cf.* figure 1.2)—a figure that eats watts—intended to motivate people in the German Democratic Republic (GDR) to use less energy during certain hours of the week in order to stabilize the electricity grid [192]. This was one of the first demand-side management campaigns in history.

A cornerstone in the era of energy efficiency was the first oil crisis in 1973. To sanction the support of Western countries for Israel in the Yom Kippur War fought between Israel and a coalition of Arab states, the Organization of the Petroleum Exporting Countries (OPEC) imposed an embargo to the US and tremendously cut exports to other

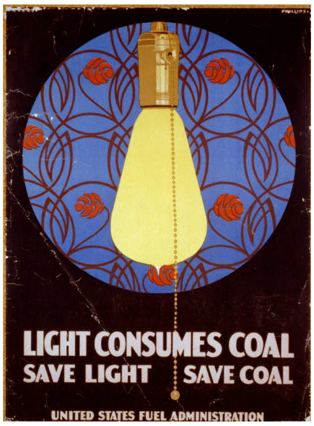


Figure 1.1: Artist Coles Philipp calls for energy efficiency (1917). [232]

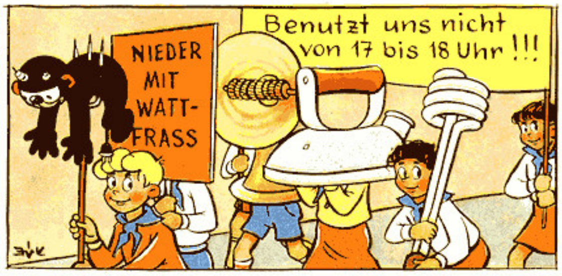


Figure 1.2: *Don't use us from 5 p.m. to 6 p.m.!* Demand-side management campaign in the GDR in the 1950s. [192]

Western states. After this step, which lead to an enormous increase in oil prices, oil-importing countries quickly ran into an economic crisis and suffered from stagnation and unemployment. As a consequence, governments reacted with different short-term energy saving measures such as introducing days with driving prohibition [207]. In addition, they set up a long-term strategy to reduce the dependency on OPEC through a variety of measures: First, governments invested heavily in alternative energy sources such as nuclear power and renewables. Second, they increasingly explored and exploited oil off-shore (*e.g.*, in the North Sea). Finally, since the oil crisis, governments have increasingly invested in research to develop energy-efficient technologies, understand how people consume energy, and learn how to stimulate energy-efficient behavior [73, 158].

Nowadays, the advent of information and communication technology (ICT) and the breakthrough of the Internet have changed the way governments or companies communicate with the masses in order to promote ideas or products. Driven by the possibility to collect and analyze large amounts of data, for instance, a whole new wave of *mass personalization* has emerged, which consists of automatically tailoring products or services to millions of individual customers based on their current needs and situation [152]. Amazon and Netflix are key drivers of this trend as they developed recommendation engines that automatically suggest products to customers depending on their purchasing and browsing behavior [226]. Other large companies in the service sector such as banks and insurance companies utilize data analytics to improve customer relationship management (CRM) for private customers [48]. The concept of mass personalization becomes even more important in the Internet of Things (IoT). As everyday objects are being connected to the Internet [118], household appliances, mobile phones, and other devices are increasingly capable of communicating their status information. Through the connection of these devices to the Internet, a plethora of services becomes possible in many different domains based on the analysis of real-time sensor information and product usage patterns.

In the energy domain, the advent of ICT, the IoT, and data analytics can lead to a mass personalization of the aforementioned energy saving efforts. Until today, giving people advice on how to reduce their energy consumption has faced the challenge that personal energy consulting sessions for every customer would be too costly, while generic energy saving programs do not adequately address each customer's individual situation. Like in other domains, the recent ICT developments can help to reach the sweet spot between these two extremes, as it now becomes possible to automatically tailor energy feedback to millions of individual households, *i.e.*, to create personalized energy feedback at large scale.

1.1 Motivation

The cleanest type of energy is the one that is not consumed at all. For this reason, governments worldwide have set quantitative targets on energy efficiency in order to reduce the amount of fossil fuels being burnt and thus the greenhouse gases being pumped into the Earth's atmosphere [184]. The European Union (EU), for instance, has recently reinforced the goal to reduce energy consumption by 20% until 2020 [44, 55] and by 27% until 2030 [45]—with respect to what was originally projected in 2007. Switzerland also decided in its Energy Strategy 2050 to save 13% of their yearly energy consumption until 2020 (with respect to what was consumed in 2010) and even 40% until 2035 [159].

In both the EU-28 states and Switzerland, the residential sector accounts for almost a third of the total energy consumption [32], [203]. Reducing residential energy consumption is thus crucial to reach the aforementioned saving targets. In the past, many efforts have been taken to improve energy efficiency in households: Space heating has become more efficient due to improved insulation, regulations now limit the standby power consumed by appliances, and household appliances are becoming more energy-efficient, which is communicated through mandatory energy efficiency labels and thus influences consumers' purchasing decisions [36].

Despite these measures to make appliances and houses more energy-efficient, reduction of residential energy consumption is still below expectations, or, as it is the case in Switzerland, the overall residential energy consumption has even increased by 11.3% from 2007 to 2013 [94]. One of the reasons is that certain barriers prevent consumers from investing in more energy-efficient equipment and therefore limit its penetration [49]: lack of information, split incentives (*e.g.*, between landlords and renters), high investment

ENERGY STAR Appliances	Rebate
Refrigerator	$100
Freezer	$50
Dishwasher	$50
Clothes Washer	$75
Heat Pump Water Heater	$300
HP Upgrade 14.5 SEER 8.2 HSPF	$300
Cental AC 14.5 SEER 8.2 HSPF	$300

Figure 1.3: The US Environmental Protection Agency ENERGY STAR offers rebates to promote purchase of energy-efficient equipment. [201]

costs, and a lack of technical expertise. To overcome these barriers, many countries increasingly incentivize purchase of new energy-efficient products through government agencies (*cf.* figure 1.3) or force energy providers to implement energy efficiency programs for households. Hence, more and more countries including Switzerland have introduced or plan to introduce mandatory saving goals for energy providers [49, 159]. This way, energy providers encourage households to change their energy-related behavior or they provide their customers incentives such as rebates to acquire energy-efficient equipment and technologies. Ultimately, in the context of the liberalization of the energy market, regulators expect a huge market for energy saving programs offered by energy providers or by third parties in both Europe and the US [14, 55].

In addition to promoting energy-efficient equipment, inducing behavior change is also a promising way to reduce energy consumption—and in particular electricity consumption—in the residential sector. Although most of the energy is consumed by space and water heating, electricity accounts for 26% of the residential energy consumption in Switzerland [94]. Figure 1.4 shows the electricity consumed by Swiss households from 2000 to 2013 broken down by its end use. The figure excludes electricity consumed by space and water heating, because the majority of energy consumed by space and water heating is not based on electricity. The graph shows that, even though household appliances have become more efficient, electricity consumption has increased by 15.4% from 2000 to 2013. Nonetheless, the effect of energy-efficient equipment is visible, since the overall consumption has remained almost constant since 2009, and it has even decreased in categories *lighting* (-14%), *ICT* (-8.4%), and *cooling & freezing* (-6.9%) since 2000. However, there is a strong increase recorded in categories *washing & drying* (+95%), *other electrical appliances* (76%), *cooking support* (+49%), and *dishwasher* (+8.5%). The effect that the overall consumption has increased despite the efficiency improvement of household appliances can be explained by the rising number of people living in Switzerland, the rising number of single-person households, the rising number of appliances per person, and an increase in comfort within the last 15 years. In order to reduce electricity consumption, it

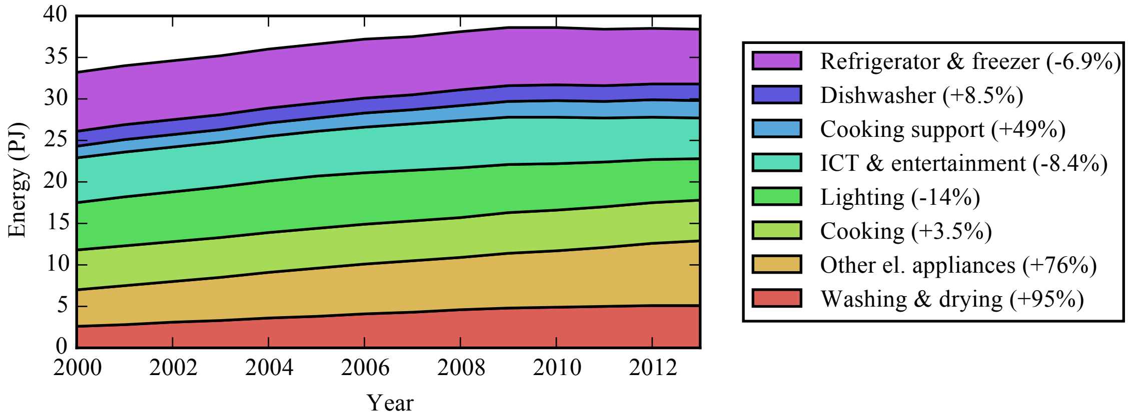


Figure 1.4: Electricity consumption of all Swiss households by end use (2000–2013) without space and water heating. Numbers taken from [18, 94].

is both necessary to increase penetration of energy-efficiency equipment and to motivate households to change their energy-related behavior.

ICT can play a significant role to motivate and help households conserve energy [119]. Digital and networked electricity meters (*smart meters*), for instance, can measure the electricity consumption of a household and provide feedback to its occupants, who can then adapt their behavior, identify energy guzzlers, or buy energy-saving equipment. Several different studies have investigated the effect of direct feedback on the electricity consumption of a household. Direct feedback is feedback that is provided timely, for instance in the form of an in-home display (*cf.* figure 1.5) or on a smartphone app (*cf.* figure 1.6). Early studies report that direct feedback can provide savings in the order of 5%–15% [46, 60]. More recent studies, however, find much smaller effects (*i.e.*, in the order of 2%–6%) as they are often performed on larger samples, which contain a higher proportion of people who are not interested in using the displays, or who lose interest in the course of the trial [47]. A large smart metering trial performed in Germany and Austria for instance reported 3.7% electricity savings by providing feedback through a Web portal [157]. In Switzerland, savings achievable through an in-home display are expected to be between 3% and 5% [51]. These findings go in line with a meta-analysis of 33 trials involving an in-home display: the analysis performed by McKerracher *et al.* suggests that *contrary to previous meta-analyses, 3%–5% is a more accurate expected conservation figure for a large-scale roll-out of in-home displays* [123].

In addition to stimulating energy-efficient behavior, smart meters play an important role in balancing demand and supply to facilitate the integration of renewable energy sources into the power grid [7]. They also enable direct cost savings for the energy provider as they allow for automated meter reading and billing [10]. For all these reasons, the number

Figure 1.5: Onzo's Smart Energy Kit for electricity consumption feedback. [221]

Figure 1.6: Feedback provided by the eMeter application. [180]

of smart meters installed in households is steadily increasing. In the US, for instance, more than 50 million smart meters have already been installed, which is a coverage of 43% nationwide [169]. In the EU, 45 million smart meters have been deployed in Finland, Italy, and Sweden. The EU aims at a coverage of 80% until 2020, provided that a cost-benefit analysis performed by each member state proves that the installment of smart meters is economically reasonable. After performing the cost-benefit analyses, 16 countries plan to perform a large-scale roll-out of smart meters by 2020 or earlier [54, 67]. Germany, however, which exhibits almost a fifth of all European households [204], is more reluctant and restricts mandatory roll-out to particular groups of customers, for instance to households that need a new electricity meter anyway or those with an annual consumption of 6000 kWh or more. According to the German cost-benefit analysis, only the latter group exhibits savings that are large enough to justify a replacement of current, analog electricity meters[1] [56]. Switzerland, albeit not a member state of the EU, is confident to reach 80% coverage until 2025 [31]. The Swiss cost-benefit analysis has indeed shown a positive outcome of 2.7% electricity savings induced by a nationwide roll-out [10].

The savings achievable through smart metering play an important role for the societal benefits of smart meter deployments, and ultimately for reaching global energy saving targets. To increase the 3%–5% savings reported above, however, energy feedback must go beyond merely visualizing the consumption of a household measured by its smart meter. Instead, consumption feedback should contain useful details on the household's energy use that are provided timely, in an appealing way, and easy to understand and act upon [60]. Many researchers therefore consider appliance-specific feedback as the

[1]The German cost-benefit analysis assumes savings of 0.5%, 1.0%, 1.5%, 2.0%, 2.5% for households with an annual electricity consumption of $<$ 2000 kWh, 2000 kWh–3000 kWh, 3000 kWh–4000 kWh, 4000 kWh–6000 kWh, and $>$ 6000 kWh, respectively.

"holy grail" of energy feedback [8, 60, 74]. Breaking down the overall consumption into the contribution of individual appliances (or activities) can help households understand electricity consumption in their particular home and help them to adapt their behavior accordingly. In contrast to generic energy saving recommendations, studies indicate that appliance-specific feedback leads to savings between 9% and 18% [60]. However, these numbers must be taken with care because the studies performed so far only had a small number of participants due to the large amount of sensing infrastructure required in each house.

The effect of feedback is also higher if advice and motivational cues are tailored to the recipient, for instance by comparing a household's electricity consumption to the consumption of households with similar characteristics [4, 9, 79]. In general, the characteristics of a household play an important role when designing energy efficiency programs. Households with high income, for instance, are more likely to invest in infrastructural changes, while people aged 65 years and older tend to be more critical towards technical changes and can be targeted by a behavior change program instead [139]. Among the households that are open for infrastructural investments, some might for instance be good targets for a heat pump marketing campaign, whereas others already have a heat pump installed and are therefore not worth contacting [71]. Similarly, single- or two-person households are typically working during the day—hence, they most often have a regular schedule and are suitable candidates for a smart thermostat, which automatically controls the heating system of a household depending on its occupancy state [104]. Families, on the other side, where someone is at home most of the time, should thus rather be motivated to change their behavior (*e.g.*, by reducing room temperature).

Personalizing energy efficiency programs to individual households is thus a necessity to improve energy savings achieved so far as it overcomes the aforementioned barriers that hinder households from saving energy. However, as we have seen, going beyond numeric consumption feedback or generic energy saving recommendations requires detailed knowledge on the characteristics of a household and its appliance stock and usage. On one side, enabling technologies such as sensors that observe the energy-relevant behavior of a household are increasingly becoming a part of our life [118]. On the other side, their deployment is still costly and cumbersome and the efforts often even out the savings achieved through the technology. Similarly, characteristics of a household like the number of persons living in it or the size of the dwelling could be obtained through customer surveys. Pursuing surveys to acquire customer information is yet time-consuming and expensive, and often only a small fraction of customers participates [165]. This thesis addresses these issues and suggests means to infer relevant information for personalized energy efficiency programs directly from the electricity consumption data collected by smart meters.

1.2 Research goals and contributions

The goal of the thesis is to provide the technical foundation for energy efficiency programs that are personalized to individual households and scale to thousands or millions of customers. This way utilities can improve the quality of energy-saving recommendations, increase participation rates and the specificity of behavioral interventions, and ultimately achieve higher savings and customer retention. In order to achieve personalized feedback at large scale, we develop and evaluate methods to automatically infer the information required for personalized consumption feedback from smart meter data, making expensive and cumbersome surveys and sensor deployments superfluous. We thereby address the following research questions:

What household characteristics can be revealed from smart meter data? Smart meters measure the electricity consumption of a household and can periodically send the aggregated value of the previous measurement period to the energy provider. The devices that are currently deployed typically provide the data at relatively coarse granularity of 15, 30, or 60 minutes. By analyzing the patterns of such coarse-grained consumption data, we investigate what relevant characteristics of a household we can infer with high confidence from the data. Household characteristics of interest include the socio-economic status of the occupants, information about the dwelling as well as appliance stock characteristics.

Can we observe the energy-related behavior of a household sufficiently well from its aggregate electricity consumption data? Most smart meters offer access to the data at a granularity of up to one sample per second through a customer interface. While this interface was originally planned for accessing the electricity consumption data to visualize it on a smartphone or in-home display, such fine-grained data also lends itself for further analysis: Through the analysis of the consumption pattern, we want to learn if it is possible to automatically find out when individual appliances are running and how much electricity they consume. Knowing the energy-related behavior of a household then serves as a basis for providing automated, personalized energy saving recommendations in order to stimulate the purchase of energy-efficient equipment or motivate occupants to change their behavior.

What applications are possible and what are potential constraints and implications? Inferring household characteristics and observing the energy-related behavior from electricity consumption data is subject to uncertainty, which depends on many factors such as the data granularity used for the analysis and the amount and quality of training data available. We investigate what applications are possible through smart meter data analytics under different constraints related to the measurement infrastructure and potential involvement of the users (*e.g.*, interactions to generate training data).

We evaluate our methods on real world consumption data. The following sections summarize our contributions to the state of the art of smart meter data analytics.

1.2.1 Automated household classification

The first contribution is a comprehensive system for automatically revealing household characteristics from smart meter data and an elaborate evaluation of our approach [22–24]. Our system uses supervised machine learning methods to automatically infer the value of the characteristics from features extracted from the electricity consumption data. We investigate 18 different characteristics which we have selected because of their relevance for utilities [21]. The evaluation is based on a publicly available data set, which contains smart meter data collected at a 30-minute granularity from more than 4000 households in Ireland over a period of 1.5 years. Along with the consumption data, the data set contains answers to more than 100 questionnaire items per household that serve as ground truth for our analysis. The source code of our analysis is publicly available [188].

1.2.2 Non-intrusive load monitoring

Non-intrusive load monitoring (NILM) refers to the analysis of fine-grained smart meter data to estimate when appliances are running and how much they consume.[2] To explore this concept, we developed the evaluation framework *NILM-Eval* and evaluated the performance of several state-of-the-art NILM algorithms [19, 20]. To this end, we collected the Electricity Consumption and Occupancy (ECO) data set, which contains fine-grained electricity consumption data of 6 Swiss households measured over a period of 8 months. Along with the smart meter data, the data set contains ground truth measurements of 47 selected appliances and each of the household's occupancy state [19, 103]. The ECO data set is now publicly available and widely used by researchers in the NILM domain [190].

1.2.3 Applications for smart meter data analytics

The third contribution consists of an investigation of different applications that can be realized with the methods developed in sections 1.2.1 and 1.2.2. In particular, we show how energy providers can use our system for automated household classification to select groups of customers based on their characteristics [24] and we explore the accuracy of

[2] Typical granularities range between one value per minute up to several kilohertz. In our case, we consider fine-grained consumption to be measured at 1 Hz.

an automated peer-group comparison. We further explore the potential of NILM for energy consulting and provide a classification of possible energy saving recommendations. Finally we discuss the implications of our findings for households, energy providers, and regulators.

1.3 Thesis outline

In chapter 2, we provide background information about smart metering and highlight existing work related to smart meter data analytics. Chapter 3 presents our system for automated household classification. We describe the features we compute on the data, the characteristics we infer based on the features, and the classifiers we utilize for classification. We further illustrate an alternative approach based on multiple linear regression and describe the performance measures used in our evaluation before presenting the results achieved on a large-scale data set.

In chapter 4, we provide a detailed overview of the ECO data set and introduce our evaluation framework *NILM-Eval*. We further present the results we obtained for six different NILM algorithms on the ECO data set. Chapter 5 presents our evaluation of potential applications based on automated household classification and NILM. We first show how to use the household classification system for customer selection, before we evaluate a way to perform automated peer group comparison. Finally, we investigate the potential application of NILM for automated energy consulting. In chapter 6, we discuss implications of our work for households, utilities, and regulators, before we conclude the thesis and give an outline on potential future research challenges.

Chapter 2

Background and related work

One of the biggest challenges of the 21st century is to provide electricity to the world's ever increasing population in a sustainable way. Burning fossil fuels such as oil, gas, or coal to create electricity leads to CO_2 emissions into the atmosphere and should thus be avoided in order to mitigate climate change. Nuclear power, on the other side, is considered one of the cleanest ways to create electricity. However, due to recent catastrophes, the safety of nuclear power plants was put into question by politicians, experts, and a broad share of the population. This development—together with technological advances—leads to a significant increase of renewable energy sources like photovoltaics and wind power. In Germany, for instance, the federal government plans to increase the share of electricity generated from renewable energy sources from 26.2% (in 2014) [193] to 35% in 2020, 50% in 2030, and 80% in 2050 [53].

This increasing penetration of renewable energy sources inevitably leads to a transition of the power grid. In contrast to electricity generated from fossil fuels and nuclear power, electricity obtained from renewable energy sources is subject to fluctuations depending on the momentary availability of wind and sun. In addition, electricity is increasingly generated decentrally in the low-voltage network because many solar panels and wind turbines directly feed into the distribution grid [7]. The increasing fluctuation and the decentralized generation lead to a bi-directional power flow that imposes multiple challenges to the grid compared to the previous situation, in which electricity generation was centralized and easier to plan and control. One way to deal with these challenges is making the grid "smart", which involves a combination of adapting demand, storing energy, and dynamically connecting or disconnecting generating sources. However, making the power grid more intelligent raises the need for fine-grained monitoring and control of power flows in the distribution grid. If possible, this enhanced infrastructure should reach up to the household level [68, 171]: With the advent of remotely controllable appliances, heating systems, and electric cars, households provide a large potential to contribute to the balance of demand and supply in the smart grid. This chapter is partially based on the contributions made in [24].

2.1 Smart metering

One way to connect households to the smart grid is to equip them with digital, networked electricity meters (*i.e.*, smart meters). In contrast to analog meters, those meters measure the consumption of a household at fine-grained intervals and make these measurements available over a communication network. Figures 2.1 and 2.2 show smart meters produced by Echelon and Elster for the European and the US market, respectively. Getting access to a household's consumption in real-time, thus gaining information on the electricity flows at the very edge of the grid, enables many opportunities for households, energy providers, distribution network operators, and other parties [77, 171].

Households, for instance, benefit from smart meters as they can receive fine-grained consumption feedback in real-time. Visualizing the consumption of a household may improve awareness and help occupants to save electricity, which reduces their electricity bill and ultimately lowers CO_2 emissions. In addition to consumption feedback, households benefit from a greater tariff variety and flexibility and from a better service quality (*e.g.*, fewer and shorter network outages) [171]. Automated meter reading further makes billing and change of suppliers easier for customers. As a consequence of the ongoing market liberalization and the possibility to provide new services based on the consumers' consumption data, smart metering thus increases competition among suppliers both at the price level and at the service level.

Energy providers benefit from smart meters in two ways [171]. First, smart meters make existing business easier as they simplify the billing process and enable sophisticated pricing schemes, in which for instance different electricity prices apply at different times. Second, having access to the customer's consumption data allows to increase the energy provider's service portfolio, for instance by offering new services that are based on the customers' consumption behavior. It further allows energy providers to optimize wholesale power purchases and participate in balance and reserve markets, for instance by getting a critical mass of demand responsive customers and provide those customers incentives to change the way they consume electricity (*i.e.*, shift consumption to off-peak times) [171].

Distribution network operators can utilize a wide-spread smart meter deployment to optimize the network's quality of service. They can quickly identify fault locations, detect

Figure 2.1: Smart meter from Echelon used in many smart metering pilots. [198]

Figure 2.2: US smart meter Elster REX type R1S. [200]

network losses or electricity theft[1], improve voltage stability, and inform concerned customers in cases of fault [171]. Furthermore, the availability of comprehensive information on the low-voltage network allows for improved, more informed investment planning with respect to new infrastructure and reinforcements of existing infrastructure.

A widespread deployment of smart meters may be beneficial for society as a whole, because smart meters help to improve the power grid's stability and to reduce CO_2 emissions, as a consequence of their potential to promote energy savings [8, 46, 60] and demand shifting [132]. However, deploying smart meters is costly and raises both security and privacy concerns [95, 121]. Including all those aspects in a cost-benefit analysis to decide whether or not to perform a wide-spread smart meter deployment is challenging, because both costs and benefits are difficult to quantify and differ from country to country. Overall, the potential of smart metering highly depends on how the technology is utilized by utilities, accepted by consumers, and supported by regulators [131]. For this reason, we will in the following give an overview of the services that may be possible based on the analysis of the data collected by smart meters.

[1] Reducing energy theft has been reported to be the biggest saving achieved by deploying 27 million smart meters in Italy. [220]

2.2 Smart meter data analytics

In the past years, an increasing number of researchers have applied machine learning and data mining techniques to model, analyze, and understand residential electricity consumption and realize many different applications based on smart meter data analytics [61]. Some of these applications for instance focus on forecasting electricity consumption or on segmenting customers into different groups depending on their consumption pattern. Customer segmentation includes identification of suitable customers for demand-response campaigns (including estimations on how much can be achieved by particular measures) or classification of customers to improve energy efficiency programs or optimize marketing campaigns. Other applications aim at disaggregating the aggregate household electricity consumption in order to learn more about the appliance stock of a household and the consumption behavior of its occupants. Detecting anomalies in the consumption data is also an interesting application, as outliers potentially reveal defects in the system or electricity theft. While smart meter data analytics enables many applications for those who have access to the data, means to manage privacy for the customers have also gained significant importance and are thus also mentioned in this section.

Forecasting

Forecasting residential electricity consumption has been an important topic for a long time as it helps utilities to perform short-term, mid-term, and long-term supply and infrastructure planning [112, 127, 133]. Early work focuses mostly on predicting electricity consumption of groups of households (*e.g.*, the ones within the same supply area) using a variety of modeling techniques [168]. To identify patterns in electricity consumption data, for instance, De Silva *et al.* propose a data mining framework and introduce an incremental learning algorithm that predicts future electricity usage of groups of households [50]. With the advent of smart metering, forecasting electricity demand has become possible on the level of individual households. In a smart grid, for instance, such advanced forecasting mechanisms support algorithms that optimize utilization of energy storage capabilities in the low voltage network [82]. This helps to balance demand and supply and thus to make efficient use of the increasing amount of fluctuating energy sources.

Customer segmentation

Customer segmentation approaches mostly focus on electricity consumption data recorded at intervals of the order of minutes or hours (typically 15, 30, or 60 minutes). To energy providers, applications based on such data are of particular interest as this is the type of data that was collected during most of the smart meter trials so far. In the following, we structure the work related to the analysis of coarse-grained consumption data by distinguishing between (1) approaches that analyze consumption data only and (2) approaches that estimate *side-information* from consumption data or find correlations between the consumption data and side-information. In this context, we refer to side-information as any information about the household in addition to its electricity consumption data, for instance the socio-economic status of the household, the geographic location of the dwelling, or the amount of energy attributed to heating and cooling.

Many authors have investigated unsupervised techniques such as clustering to detect patterns and usage categories in the consumption profiles [37]. Knowledge about the existence and characteristics of clusters that exhibit similar consumption patterns can be used to develop novel tariff schemes, improve network management, select the right customers for demand-side management campaigns, or to perform load forecasting. An early example of this class of approaches is provided by Chicco *et al.* [38]. The authors cluster electricity consumption data of 471 commercial customers. Evaluating the clusters along with the current tariffs of each of the customers, they detect examples of inefficient billing practices (*e.g.*, in case there is a poor correlation between discriminatory factors and actual load patterns) [38]. Similarly, Mutanen *et al.* cluster 660 customers of different types (*e.g.*, residential, industrial, public administration) into distinct clusters using the iterative self-organizing data analysis (ISODATA) algorithm [130]. The feature vector used to cluster the customers includes the customer's load profile (2016 hourly values per customer) as well as four temperature dependency parameters obtained through a linear regression analysis.

Several other related approaches that cluster consumers by their consumption pattern rely on a clustering technique called self-organizing maps (SOMs) [52, 72, 125, 155, 174]. A SOM is an unsupervised learning method based on neural networks that can be used to automatically extract clusters out of an otherwise unstructured (and unlabeled) set of data [105]. For instance, Figueiredo *et al.* use SOMs to identify groups of consumers with similar consumption behavior [72]. The authors further create rules that form a decision tree to automatically assign new households to one of the clusters by following that tree. Their results are based on electricity consumption traces from 165 households in Portugal collected at a 15-minute granularity. Dent *et al.* also utilize SOMs to separate households into clusters by their electricity consumption [52]. The authors rely their study on hourly consumption traces of 93 households in the UK. Verdu *et al.* leverage self-organizing maps and investigate the evolution of consumption patterns over time [174]. Their goal is to recognize consumption patterns that deviate from a "typical" behavior as well as identify new (commercial) customers [174]. McLoughlin *et al.* further investigated the problem of automatically clustering consumers with similar consumption patterns [125]. The data set used in this study is however significantly larger than others previously considered and—although this is not explicitly stated in the paper—is most likely to be the same data that we use for our investigation. The approaches described above have in common that they build clusters using only the plain electricity consumption data thus without computing complex features of the data itself. In contrast to that, Sanchez *et al.* first compute specific features of the data and then feed the SOM with these features [155]. In contrast to the other approaches, Sanchez *et al.* add questionnaire information collected from 625 Spanish households to the features derived from the electricity consumption data.

In a recent review paper [37], Chicco provides an overview of clustering techniques used to group residential or commercial customers according to their electricity consumption pattern. Using similar clustering techniques, both Kwac *et al.* [110, 111] and Cao *et al.* [33] have focused on identifying the "right" customers for demand-side management campaigns. Whereas Kwac *et al.* aim at detecting stable profiles over a certain time period, Cao *et al.* focus on identifying households with a similar time of peak usage. Finally, Haben *et al.* present a finite mixture model-based clustering and apply their method on smart meter data collected in more than 4000 Irish households [83]. The authors detected ten distinct behavior groups that describe customers based on their demand and variability. In contrast to previous work, Haben *et al.* test the robustness of the clustering method (*i.e.*, the certainty to which cluster individual households belong to) through a bootstrapping method, which means they apply the clustering method multiple thousand times to different configurations of the data and calculate particular variables to evaluate the robustness of the clusters [83].

The approaches described so far are solely based on the analysis of electricity consumption data, disregarding the correlation of a household's consumption with its socio-economic characteristics, dwelling properties, or appliance stock and usage. The influence of such characteristics on residential electricity consumption has been investigated in many different studies [27]. Heating, ventilation, and air conditioning (HVAC), for instance, accounts for a large portion of the energy consumed in the residential sector.[2] Hence, several approaches investigate the correlation between electricity consumed by HVAC systems and the overall electricity consumption of a household [3, 28]. If energy providers know when the HVAC systems of their individual customers are running and how much they consume, they can optimize the energy efficiency and demand-side management programs by tailoring advice to individual households. To this end, Birt *et al.* disaggregate the total electricity consumption into different load categories [28]. Evaluating hourly consump-

[2] In the US, HVAC accounts for 43% of the residential energy consumption [233], in Switzerland space heating accounts for 70% [94].

tion data from 327 households in Canada, the authors observe correlations between the electricity consumption of a household and the air temperature both in winter and summer. These correlations allow to determine the electricity consumed by heating and cooling systems, respectively. In cases where households are heated by gas furnaces, the heating systems use electric fans for air circulation, which also provide traces in the electricity consumption data. Albert and Rajagopal developed a method to automatically identify the thermal response of a home, which is the reaction of the electricity consumption to outdoor temperature changes. Given the thermal response, the authors compute a probability that for a given temperature the home will be heating, cooling, or not using HVAC at all [3]. The authors applied their thermal profiling method on hourly smart meter data collected in 1923 households over a period of one year in a hot climate zone in California [236]. They show that, for the majority of the users, the four states *heating*, *cooling*, *no HVAC*, and *bursty* are sufficient to explain 85% of the variance in the electricity consumption data. They further show the application of their method for user selection in a demand response program, in which the utility triggers demand response events asking users to reduce air conditioning depending on the temperature forecasts [3].

Several approaches aim at inferring activities performed by the occupants from coarse-grained smart meter data [35, 106]. Kolter *et al.* for instance apply a sparse coding algorithm to separate aggregate consumption data into the different electrical appliances being used [106]. Applying their approach on data collected at a 15-minute granularity, the authors achieve an accuracy of roughly 55%. Chen *et al.* simplify the problem by focusing on behavior related to water fixtures. They developed a statistical framework based on hidden Markov models (HMMs) to disaggregate coarse-grained water consumption data by modeling fixture characteristics, household behavior, and activity correlations [35]. Evaluating their approach on 15-minute water consumption data, Chen *et al.* show that disaggregating the water consumed by washing machines and showers is possible with high accuracy, whereas toilet flushs are more difficult to detect.

Beyond recognizing activities, many researchers have investigated the combination of electricity consumption with side-information in recent studies [2, 76, 90, 107, 124, 126, 144, 182]. Similar to the clustering approaches described above, Räsänen *et al.* also use SOMs to cluster households [144]. However, as input the authors rely on dwelling characteristics only, with the goal of providing personalized electricity use information to households within the same cluster. Sanquist *et al.* relate electricity consumption with household characteristics in order to identify which factors have the highest influence on residential electricity consumption [156]. Their study is based on the 2005 Residential Energy Consumption Survey (RECS) data, which contains information about lifestyle factors from a few thousand housing units statistically selected to represent 111.1 million US housing units. Correlating the factors extracted from the survey data from 2165 single houses with the households' annual electricity consumption, the authors found that five lifestyle factors account fore more than 40% of the variance in electricity consumption: *air conditioning usage*, *laundry usage*, *PC usage*, *television usage*, and *climate*. Income, on the other side, only adds 1% to the predictive power. Ghaemi and Brauner also investigate the correlation between annual electricity consumption and household characteristics [78]. Evaluating consumption and survey data from 51 households in Austria, they found a correlation between the annual electricity consumption and the floor area, house type (*e.g.*, apartment, detached), number of appliances in the household, and the number of occupants. Kolter *et al.* exploit the correlation between household characteristics and electricity consumption to provide feedback on electricity use: They estimate monthly consumption data from household characteristics derived from public databases in the US through linear regression [107]. Comparing the estimates with the actual consumption of a household enables personalized feedback for the inhabitants. Also relying on a regression model, Kavousian *et al.* analyze the effect of different determinants on the household electricity consumption [90]. In particular, the authors define four major categories of determinants that affect the overall consumption: (1) Weather and location, (2) dwelling characteristics, (3) appliance and electronics stock, and (4) occupancy and behavior. After applying their model on 1628 households in the US, the authors come to the conclusion that weather and dwelling characteristics have a larger influence on residential electricity consumption compared to the appliance stock and occupancy behavior. It is important to note, however, that the data used in their study also accounts for electricity consumed by heating and cooling, which in the US represent a large portion of the overall electricity consumption.

McLoughlin *et al.* also investigate the correlation between electricity consumption data and household characteristics [126]. Like Kavousian *et al.*, the authors use a multiple linear regression analysis to model the electricity consumption of households on the basis of their characteristics. Relying on the same data as our study (*cf.* section 3)—which does not account for thermal loads—the authors found a strong relationship between four electricity consumption parameters (*i.e.*, total consumption, maximum demand, load factor, and time of use) and different dwelling, household, and appliance stock characteristics. In his dissertation, McLoughlin further investigated methods to automatically cluster households in order to segment them into profile groups according to their electricity consumption [124]. McLoughlin then investigates the distribution of household characteristics over the clusters with the goal of characterizing electricity use depending on the customer characteristics. Similarly, Wijaya *et al.* propose a method to cluster consumers based on their electricity consumption [182]. Using the same data set, the authors further investigate the correlation between cluster membership and household characteristics such as the number of persons in the household or the floor area of the dwelling. In their evaluation, the authors found a strong correlation between the floor area of a household and the cluster the household resides in. On the other side, the year a house was built had no effect on the cluster membership, since the distribution of this variable is similar in all the clusters. In contrast to Kavousian *et al.*, to McLoughlin *et al.*, and to Wijaya *et al.*, we propose a method that utilizes the correlation between electricity consumption data and household characteristics to estimate the characteristics from the consumption data.

Albert and Rajagopal recently presented an approach that also infers household characteristics from smart meter data [2]. The authors first even out the impact of weather on a household's electricity consumption using a linear regression model. On the residuals they utilize a HMM to infer specific occupancy states per household. All parameters gained from this analysis then serve as input to an AdaBoost classifier in order to estimate specific household characteristics. To evaluate their work, the authors rely on the same data set as Kavousian *et al.* [90], which consists of smart meter data and household characteristics of 950 Google employees. The fact that this source of data is biased towards Google employees complicates a detailed analysis of the applicability of the approach to other settings. Fusco *et al.* also developed a method to "mine" household information from smart meter data [76]. They rely on a supervised machine learning approach and use different classifiers in their analysis. The features extracted by Fusco *et al.* include the raw electricity consumption data, the time of day, features obtained from applying fast Fourier transform (FFT) and discrete wavelet transform (DWT) on the consumption data, and information about when appliances are switched on or off. To this end, the authors model appliances as finite-state machines (FSMs) to find out when they are running, which has shown to be difficult given the low granularity of the consumption data used in the study. In fact, the study is performed on the Commission for Energy Regulation (CER) data set, which contains data at 30-minute granularity and which we also use for our investigation. The approach presented by Fusco *et al.* achieves a slightly increased accuracy (up to 11%) over the prior. In contrast to the work we present in chapter 3, Fusco *et al.* rely on different features, investigate fewer household characteristics, and use only accuracy and area under the curve (AUC) to investigate the performance of their approach.

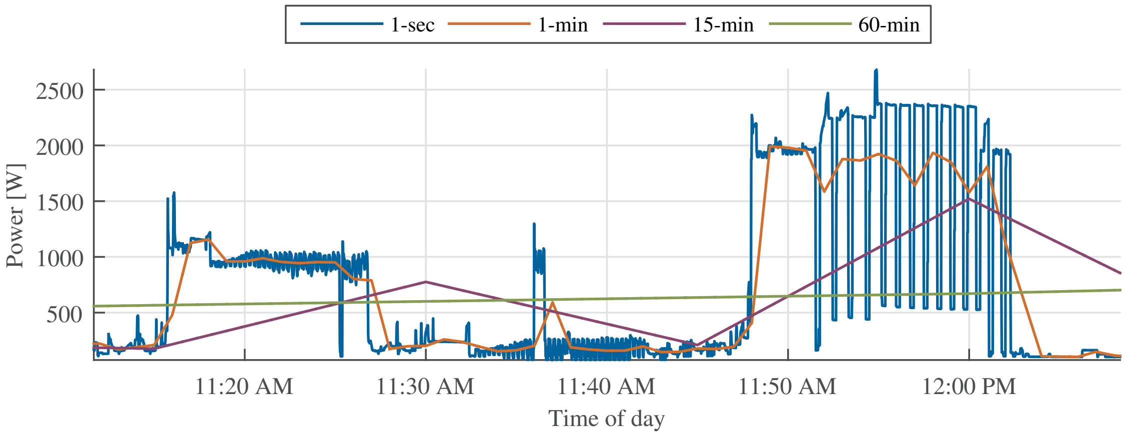
Finally, Fei *et al.* show a case how smart meter data analytics helps to identify the right customers for an energy efficiency marketing campaign [71]. They propose a method for automated heat pump classification based on electricity consumption data and weather information, which means they automatically identify those customers who do not have a heat pump installed and are thus suitable customers for a corresponding marketing campaign. Having access to daily smart meter data from roughly 300,000 customers and assuming a market penetration of heat pumps of 42%, filtering out those who already have a heat pump installed significantly improves the efficiency of the campaign. In the case of Fei *et al.*, 4565 households are known to own a heat pump according to prior sales data and the remaining households may or may not have a heat pump. The authors rely on a biased support vector machine (SVM) classifier, which is an instance of positive and unlabeled learning and therefore suitable for such a scenario in which no ground truth for negative cases is available. The features computed by the authors utilize the fact that the consumption of heat pumps depends on the outside temperature. Therefore, the features are temperature-dependent heating parameters like the average consumption in the heating period or the ratio between this value and the average consumption during non-heating and non-cooling period. As a result, the number of customers without a heat pump identified by Fei *et al.*'s approach complies with the market share of heat pumps. However, due to the lack of known negatives (*i.e.*, customers that are known not to own a heat pump), the precision of the result cannot be validated.

This section on customer segmentation described several approaches that analyze smart meter data to understand residential electricity consumption and help utilities to optimize their demand-side management or energy saving programs. Our approach for automated household classification presented in chapter 3 contributes to this field as we present—to the best of our knowledge—the first comprehensive system that automatically estimates household characteristics from electricity consumption data using supervised machine learning. We further utilize electricity consumption data and household characteristics of more than 4000 households to train our classifiers and evaluate our approach.

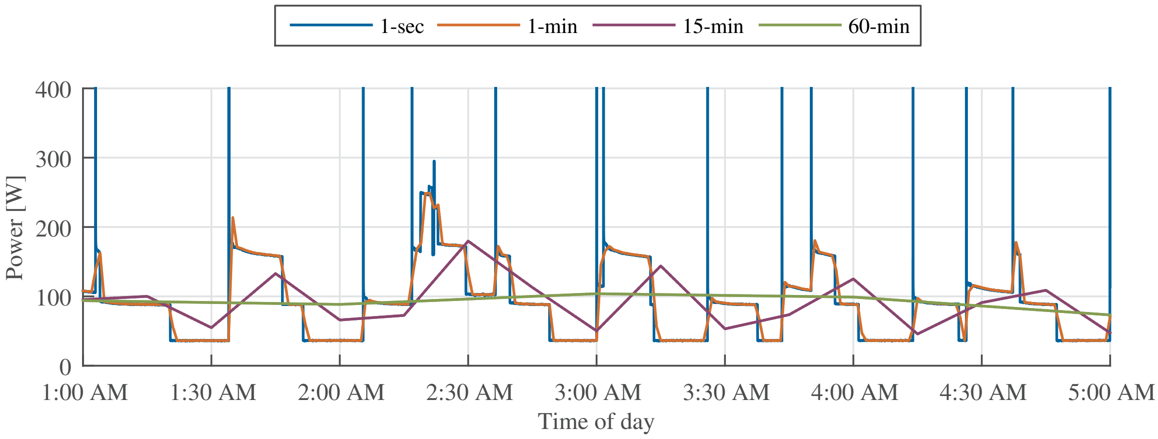
Load disaggregation

A popular line of research in the context of smart meter data analytics focuses on non-intrusive load monitoring (NILM) [114,186,187]. Using aggregate electricity consumption data of individual households (*e.g.*, measured at one value per second or millisecond), researchers have tackled the problem of disaggregating the consumption of individual appliances. This information allows in turn to provide detailed consumption feedback to the households [8,46,60]. It is further of interest to policy makers or utilities to learn about the number and type of appliances that contribute most to a household's electricity consumption. In the US, for instance, on average eight appliances account for 80% of a household's total consumption [34]. Since both the number and the type of appliances that contribute to these 80% of the total consumption differs from household to household, inferring such information from the household's aggregate consumption data would enable more personalized and thus more efficient demand-side management or energy saving programs.

Zeifman *et al.* [186], Zoha *et al.* [187], and Liang *et al.* [114] have recently provided a comprehensive overview of existing NILM approaches and the challenges researchers are facing when designing and evaluating their algorithms. An important aspect of NILM is the



(a) Consumption recorded during the run of a washing machine (7 July 2012, household 1).



(b) Consumption recorded during the night (*i.e.*, standby consumption, refrigerator, and freezer; 7 July 2012, household 2).

Figure 2.3: Smart meter data aggregated at different granularities. The data is part of the ECO data set [190].

granularity of the aggregate consumption data, which is analyzed to infer the consumption of individual appliances. Figure 2.3 shows the consumption pattern of a washing machine during spinning (*cf.* figure 2.3a) and the consumption pattern of a household during the night (*cf.* figure 2.3b) using different (aggregated) data granularities: one value per second (*1-sec*, blue line), one value per minute (*1-min*, orange line), one value per 15 minutes (*15-min*, purple line), and one value per hour (*60-min*, green line). On the data recorded at 1-second granularity, the spinning cycles of the washing machine are visible in the consumption pattern. 1-minute data shows the intensity of the spinning but hides the "lows" during calm periods. The plot of the 15-minute consumption data barely follows the spinning pattern, whereas the 60-minute data does not show significant changes at all. The cooling pattern in figure 2.3b includes the overlapping consumption of a refrigerator and a freezer. The 1-minute data follows the pattern of the 1-second data closely, except that the initial spikes at the beginning of the cooling cycles are missing due to the aggregation of the data. For the 15-minute and 60-minute data, the cooling cycles are only indicated by a zigzag pattern, and refrigerator and freezer are difficult to distinguish. Overall, the figure shows that data granularity is important when detecting the consumption of individual appliances in the aggregate consumption data, and that at least 1-minute or 1-second data should be collected.

The smart meters deployed in practice typically communicate data aggregated to one value for each 15-minute, 30-minute, 60-minute period. This granularity is however not sufficient for NILM applications, since the accuracy obtained by the algorithms suffers from the high level of aggregation [106]. Although smart meters exist that provide data at higher granularity (*e.g.*, the Landis+Gyr E750 presented in figure 4.7 in section 4.4), replacing the smart meter is a high barrier that may hinder the adoption of NILM in practice. Similarly, households with an old, analog meter installed may not install a smart meter for this application at all.

In practice, several solutions exist to mitigate this problem and lower the barrier for NILM [205, 206, 214, 234]. For households without a smart meter, interfaces to the analog meter exist that read the meter through a camera and optical character recognition (OCR) and transmit the reading over a wireless interface [205]. However, since the granularity of the analog reader is relatively low (*i.e.*, $\frac{1}{10}$ kWh), the data rate is accordingly low particularly when the consumption is low. Alternatively, customers can buy and install a *Fluksometer* [206], which provides current clamps that can be put on the electrical wire to determine the current that flows through the mains by measuring the surrounding electric field. This way, it is possible to obtain data at a 1-second granularity. For households that have a smart meter installed, retrofit solutions make use of an infrared interface located at the front side of many off-the-shelf smart meters in Europe. The infrared interface provides the detailed electricity consumption information (including real and reactive power or each of the three phases) roughly every two seconds through a standardized protocol [214, 229]. Figure 2.4 for instance shows an optical reader [214] that reads the data in 10-second intervals through the infrared interface and makes it available to an Ethernet module (which is not shown on the image). Figure 2.5 shows an optical reader that provides access to the data wirelessly via Bluetooth.

To be applicable for real-world applications, a NILM algorithm should satisfy the following requirements [185]:

- The features computed by the algorithm should be compatible with data acquired at maximum 1 Hz, because higher data rates impose high requirements on the measurement infrastructure.

Figure 2.4: Optical sensor that reads fine-grained consumption data from a smart meter through an infrared interface. [214]

Figure 2.5: Bluetooth version of the optical reader. [234]

- The disaggregation accuracy should be (at least) in the range of 80% and 90%.
- No training should be required, and the algorithm should automatically detect changes in the appliance setting (*i.e.*, when new appliances are plugged in or existing appliances leave the system).
- Computation should work (almost) in real-time to be able to provide timely feedback to the occupants.
- The algorithm should be scalable to a large number (*i.e.*, 10–20) of household appliances.
- It should be able to detect different appliances types [85]: on-off appliances (*i.e.*, appliances that consume constant power), finite-state appliances (*i.e.*, appliances with multiple state changes), variable-power devices (*e.g.*, dimmable lights), and permanent consumer devices (*e.g.*, a wired smoke alarm).

As these requirements are difficult to achieve, NILM researchers typically choose to focus on a subset of these, leaving the others for future work. In section 4.1, we present an overview and a classification of existing NILM algorithms and perform a detailed evaluation of five selected ones which we believe are a representative subset of the approaches presented in the literature.

Submetering

As prices for electronics are falling, the collection and the analysis of plug-level electricity consumption data has become more cost-effective. In contrast to consumption data collected at the household level, plug-level data reflects the consumption of a single appliance and is thus not prone to overlaps and noise caused by other appliances in the household. Reinhardt *et al.* utilize electricity consumption data measured by such smart plugs at 1 Hz in order to detect which appliance is connected to the plug [147]. In addition, they predict how much electricity the appliance connected to the plug will consume within the next time intervals (ranging between one minute and the time until deactivation of the appliance) [148]. Englert *et al.* also detect which appliance is connected to a plug [65]. To this end, the authors developed a plug that measures electricity consumption at 96 kHz and use a random forest classifier to identify the appliance connected to the plug with more than 99% accuracy.

Anomaly and theft detection

An alternative application of smart meter data analytics is the detection of anomalies in the electricity network [12, 113, 117]. The cause of anomalies could be meter malfunctioning, power leakage, reading errors, or intentional customer fraud. Potential detection algorithms compare a customer's usage pattern with the one of customers with similar demographics or with historical data of the same customer [113]. In an early work, Bandim *et al.* present a methodology to identify abnormal consumption behavior and thus detect theft or malfunctioning meters [12]. The authors evaluated their work on simulated data representing 12 smart meters with a relatively low resolution of 1 kWh. Mashima and Cárdenas developed an approach based on an auto-regressive moving average (ARMA) model, which represents a normal electricity consumption probability distribution and detects outliers as potential attacks [117]. The authors validated their approach on data collected from 108 customers.

Privacy

As extracting information from smart meter data potentially poses privacy threats for consumers [128], more and more approaches investigate smart metering systems that preserve user privacy and at the same time enable useful applications [66, 122]. Erkin *et al.* for instance propose to homomorphically encrypt the smart meter data of individual customers [66]. This way, utilities can compute aggregated values over the consumption data of several customers without gaining access to the data of individual customers. Similarly, such a system can be designed to provide utilities means to collect data of individual customers at a very coarse granularity (e.g., for billing purposes) without gaining knowledge about the actual consumption pattern, which would contain privacy-infringing information such as information about the lifestyle of the occupants. A similar approach called differential privacy has been presented by Ács and Castelluccia [1] as well as by Shi *et al.* [160]. Ultimately, using such techniques, designers of smart metering systems can strike a balance between preserving user privacy and providing legitimate applications on top of smart meter data analytics [122]. It is hereby most important to take the individual user into account when designing products and services [77].

Egarter *et al.* present a load-hiding approach to realize smart meter privacy [59]: Using a battery and a controllable household appliance (*i.e.*, a boiler), the authors present an approach to obfuscate the household's electricity demand. They evaluate their technique called load-based load hiding (LLH) on a real world data set showing an increase in appliance-level privacy as the performance of NILM deteriorates. Instead of using a battery, Reinhardt *et al.* provide a pre-processing framework that modifies the consumption data before reporting it to the energy provider [149]. To this end, they implement multiple methods to add noise and to down-sample, average, and quantize appliance-level data.

Chapter 3

Automated household classification

Customer insights help utilities to optimize their energy efficiency programs in many ways [139, 172]. With knowledge of the socio-economic characteristics of individual households, for instance, utilities can automatically tailor savings advice to specific addressees (*e.g.*, to families with children, or to retirees). Further, they can offer consumption feedback that includes references to similar households or consider the financial reach of their customers when suggesting improvements in the appliance stock. Many studies have shown that such specific approaches improve the performance of efficiency campaigns [4, 9, 79]. Yet, such targeted measures require detailed information on individual customers, which might be gathered for research studies and local saving campaigns, but which is often not available for large-scale, cost sensitive efficiency programs that are directed to millions of households.

In fact, utilities' knowledge about their customers is often limited to their address and billing information. This is particularly true in Europe, where open information repositories like public tax registers do not exist or cannot be easily accessed. On the other hand, conducting surveys to acquire customer information is typically time-consuming and expensive, and often only a small fraction of customers participate [165]. We argue that utilities can instead utilize the electricity consumption data of a household to reveal customer information that is relevant to optimize their energy efficiency programs. This is valuable for utilities, because they are already deploying millions of smart electricity meters in private households along with infrastructure to collect, process, and store their electricity consumption data. [54, 67, 171]. Currently, utilities use this data mainly to improve their meter-to-cash processes, to enable advanced tariff schemes, and to provide customers with detailed information on their electricity consumption. Analyzing smart meter data that is collected anyway can therefore contribute to the value of the metering infrastructure without requiring any changes to the smart meters that have already been deployed.

This chapter describes the development and evaluation of a system to automatically infer household characteristics from smart meter data. Examples of such characteristics include the household's socio-economic status, its dwelling properties, and information on the appliance stock. Our analysis takes as input the electricity consumption of a household and estimates the value of several characteristics of interest. Depending on the characteristic, this value is either the class to which the household most likely belongs to (*e.g.*, employment status) or a numerical value (*e.g.*, the number of persons living in the household). To infer the value of household characteristics from consumption data, we extract features from the data itself and pass them as input to a classifier or regression model. An example of such a feature is the average consumption of a household between 10 a.m. and 2 p.m. divided by its daily average consumption. This particular feature helps to reveal household occupancy during lunch time and thus contributes to the estimation of characteristics such as the employment status of the inhabitants. We investigate 18 different characteristics which we have selected because they are relevant to utilities. We have evaluated our system according to these characteristics using smart meter data available at a 30-minute granularity from more than 4000 Irish households over a period of 1.5 years. This data set is publicly available and has been collected in the context of a smart metering trial conducted by the Irish Commission for Energy Regulation (CER) [211]. In the following, we refer to this data set as the *CER data set*. Along with smart meter data, the data set contains information on the characteristics of each household collected through questionnaires before and after the study. This information is crucial for our work, because it represents ground truth data we can use to train our models and validate our findings.

The contribution of this chapter is a comprehensive system for automatically revealing household characteristics from smart meter data and an elaborate evaluation of our approach. The results show that revealing household characteristics from smart meter data is feasible with sufficient accuracy. This holds in particular for characteristics related to the number of persons living in a household and for characteristics related to the occupancy of the household (which also includes information on the employment status of the chief income earner). We show that it is possible to infer eight of the 18 characteristics with an accuracy between 72% and 82%. Overall, our approach performs roughly 30 percentage points better than assigning characteristics to the households at random. According to these results, we conclude that utilities can estimate household characteristics from smart meter data with good reliability. Thus, they will be able to improve their energy efficiency campaigns and make them applicable to the mass market as they scale to thousands or millions of customers with little additional effort. Ultimately, creating these services to help their customers to use energy more efficiently is crucial for utilities' attempts to comply with regulatory targets [26]. In addition, the system provided in this section allows utilities to improve customer retention, which is becoming more relevant in a liberalized energy market [163]. To the best of our knowledge, we performed the first study that provided a quantitative analysis of the possibility of revealing household characteristics from electricity consumption data on such a large data set and at such a high accuracy [21–24].

We begin this chapter in section 3.1 by providing an overview of the CER data set. Next, we describe the design of our system in section 3.2. This section includes a list of the features used in our analysis, a description of the classifiers, and an overview of the household characteristics investigated in detail. To identify relevant household characteristics, we show the results of interviews performed with energy consultants as well as a data-driven analysis based on SOMs. Ultimately, we describe an alternative approach based on multiple linear regression. In section 3.3 we show the evaluation process and the performance measures, before we present the results of the analysis in section 3.4. Section 3.5 shows how to improve the results by including information about outdoor temperature and daylight, before sections 3.6 and 3.7 analyze in detail the features used in our system as well the data granularity required to perform household classification. This chapter is partially based on the contributions made in [21], [22], [23], and [24]. We further make the source code of the analysis publicly available [188].

3.1 The CER data set

In 2007, the Commission for Energy Regulation in Ireland started a smart metering project to evaluate the performance of smart metering and the impact and economic benefits of a nationwide rollout [211]. In this context, the CER launched an electricity customer behavior trial (CBT) and a gas CBT, equipping several thousand households with electricity or gas meters to assess the impact of smart metering on consumers' electricity or gas consumption, respectively. Based on the results achieved during both the technical and behavior studies, the CER decided for a national rollout to be completed by 2019, in line with the EU guidelines [40].

Our study relies on the CER data set, which was collected during the electricity CBT and contains measurements of electricity consumption gathered from several thousand households between July 2009 and December 2010 (75 weeks in total). Although data collection started already in July 2009, the behavior trial officially started on 1 January 2010 and lasted until 31 December 2010. More than 6000 households and businesses participated in the trial, each of them had a smart meter installed, which measured the electricity consumption (in kW[1]) in 30-minute intervals and reported the readings to the utility via General Packet Radio Service (GPRS) [42]. Another 3400 households were equipped with smart meters that communicated through power-line communication (PLC) or formed a 2.4 GHz wireless mesh network. These households, however, were not part of the electricity CBT and thus their consumption data is not part of the CER data set.

Each household that participated in the electricity behavior trial was asked to fill out a questionnaire before and after the study. The questionnaire contained questions about the household's socio-economic status, appliance stock, properties of the dwelling, the consumption behavior of the occupants, and previous participation in energy efficiency programs [210]. Out of all households that filled out the pre-trial questionnaire, 4232 households had a smart meter installed in their homes and thus provided electricity consumption data. We discarded one household in our data cleaning process, because it was most likely not a private household (it stated that more than six adults and more than six children were living in the house). From the remaining 4231 households, 744 participants had left the trial for different reasons: some participants had technical difficulties during some weeks of the trial, while others left the trial but still had their consumption recorded. As there is consumption data available for these particular households for most weeks of the trial, we decided to include those 744 households in our analysis, although only 18 of those have remained until the end of the study and have filled out a post-trial questionnaire.

The electricity CBT aimed at evaluating the effect of different types of behavioral stimuli and tariff structures on residential electricity consumption. Hence, participating households were assigned to four different groups receiving time of use (ToU) tariffs, in-home displays, detailed bills (monthly and bi-monthly), or overall load reduction (OLR) incentives, with each group being split into subgroups receiving four different tariff structures. In addition, a small group received a weekend-tariff (with bi-monthly bill and energy statement), and another group acted as control group. Table 3.1 shows the assignment of the households we used in our analysis to those different groups. In our analysis, we treated all 4231 households equally independent of the type of treatment they received. The reason for not differentiating between different treatment groups is that the behavioral stimuli only caused little changes in the consumption patterns: On average, households reduced their electricity consumption by 2.52% (overall) and by 8.81% (peak usage) [41].

[1]The first version of the documentation stated that each measurement denoted the kWh consumed during the previous 30 minutes. This, however, was a mistake, as reported by the CER: *An amendment has been made to the Electricity CBT Manifest document, which describes and summarizes the contents of the data sent to end users: In the manifest, the line: "Electricity consumed during 30 minute interval (in* kWh*)" should read: "Electricity consumed during 30 minute interval (in* kW*)".*

Table 3.1: Number of households used in our analysis and the tariff structure and treatment they received. *No participation* means the households filled out a questionnaire but did not participate in the study due to attrition or technical difficulties.

Tariff	**Bi-monthly bill**	**Monthly bill**	**Bi-monthly bill & IHD**	**Bi-monthly bill & OLR**	**Total**
A	226	241	232	244	943
B	90	97	93	87	367
C	250	245	233	238	966
D	93	96	90	88	367
Weekend	-	-	-	-	76
Control group	-	-	-	-	768
No participation	-	-	-	-	744
Total	659	679	648	657	4231

Electricity consumption data

In the data set, erroneous data (*e.g.*, when communication to the meter was lost) is indicated through zero electricity consumption [124]. Since zero readings have significant impact on the households' load profiles, we removed for each of the households the traces of all weeks that contain more than ten zero readings.

Figure 3.1 illustrates the average consumption data of all 4231 households in a heat map. Each row in the graph represents the normalized average consumption of all households over the course of one day. The top row corresponds to the Monday of the first week of the trial (20 July 2009), and the bottom row to the Sunday of the last week (26 December 2010). As some of the households joined the trial in early 2010, not all of the rows contain consumption data from all 4231 households. The graph shows a difference from summer to winter consumption, most likely because people spend more time at home during winter. The graph also shows an increase of consumption during the Christmas holidays in both 2009 and 2010.

On average, the yearly electricity consumption per meter was 4465 kWh. Figure 3.2 shows how the electricity consumption distributes over the 4231 households during the first 52 weeks of the study. The left plot, figure 3.2a, shows the empirical cumulative distribution function (CDF), which indicates the probability that a (randomly selected) household has a value less than or equal to a given electricity consumption provided on the x-axis. The plot shows, for instance, that the yearly consumption of 64% of all households is smaller than 5000 kWh, while 36% of all households consume more than 5000 kWh per year. Only 1.4% of all households consume more than 10,000 kWh per year, whereas the maximum consumption is close to 25,000 kWh. Figure 3.2b shows the cumulative

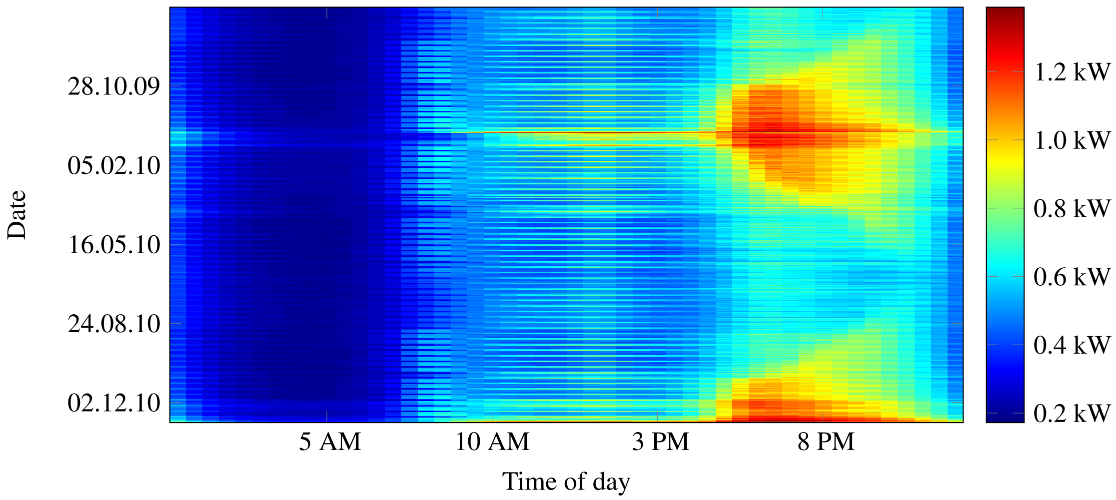


Figure 3.1: 30-minute aggregate consumption data of all 4231 households of the CER data set used in our study. Each horizontal line in the plot represents a single day during the study.

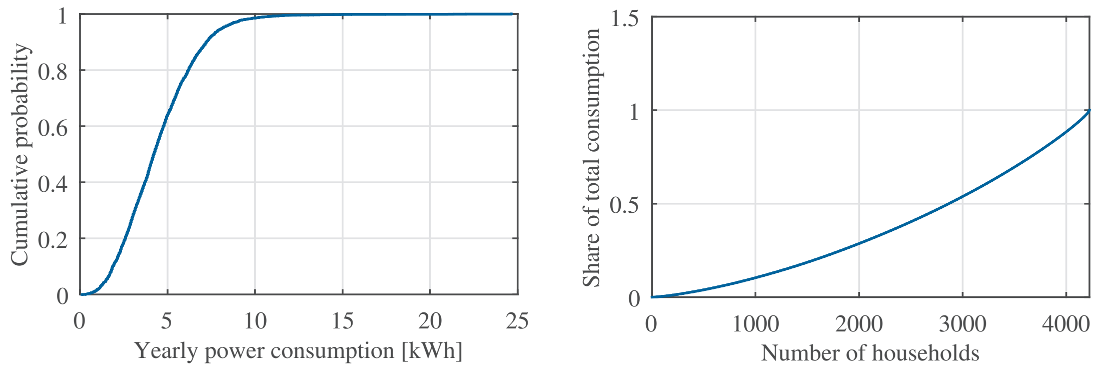
sum of each household's proportion of the total electricity consumption. The graph shows that the 1000 households with the lowest consumption (*i.e.*, 23.7% of all households) are responsible for roughly 10% of the overall electricity consumption. Similarly, the 1000 households with the highest consumption are responsible for roughly 40% of the overall consumption.

Figure 3.3 shows the power consumption averaged over all households during the course of one week. From Monday to Friday, one can see two small spikes during breakfast and lunch time. In the evening, the power consumption is almost twice as high as during the day, and more than four times higher than during the night. On the weekend, the spike during lunch time is higher, whereas the spike during breakfast time disappears (most likely because people are getting up at different times on the weekend). Overall, the average power consumption denotes roughly 500 W over all households.

To illustrate the difference in consumption between summer and winter, we plotted the weekly average consumption of all households over the course of the 75 weeks of the trial in figure 3.4. The plot shows that the consumption is up to 50% higher during winter weeks (particularly over Christmas), where days are shorter and colder.

Household characteristics

The pre-trial questionnaire and the post-trial questionnaire each contain more than 100 questions related to the socio-economic characteristics of the household, properties related



(a) Empirical CDF

(b) Cumulative sum

Figure 3.2: Distribution of the electricity consumption over the 4231 households during the first 52 weeks of the trial (20 July 2009 to 18 July 2010).

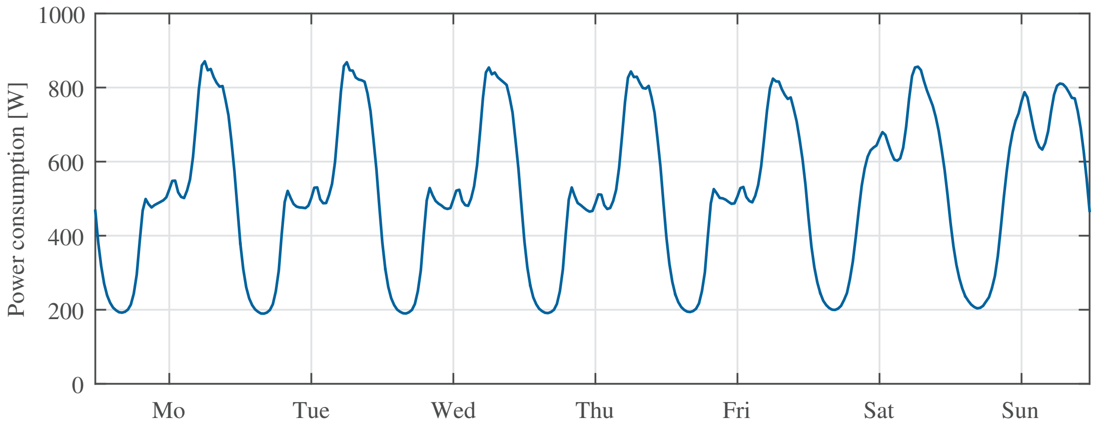


Figure 3.3: Average power consumption of all households and all 75 weeks of the trial plotted over the course of one week. The ticks on the x-axis mark the time period between 12:30 a.m. and 1 p.m. for each day.

to the dwelling, appliance penetration and usage, as well as self-assessment of the occupants' attitude towards energy saving. In the following, in order to get a brief overview of the data set, we will show the distribution of the answers to questions related to the dwelling type, the type of space and water heating, the type of cooking, and the penetration of household appliances. Showing the answers for all of the questions goes beyond the scope of this chapter. For a detailed investigation of the side-information we refer to the dissertation of Fintan McLoughlin [124], to the description of how we defined classes and class labels in section 3.2.2, and to the data itself [211].

Figure 3.5 shows the distribution of households for different characteristics related to space heating, cooking, water heating, and the type of dwelling. For space heating (*cf.* figure 3.5a), the plot shows that most of the households heat their homes using oil,

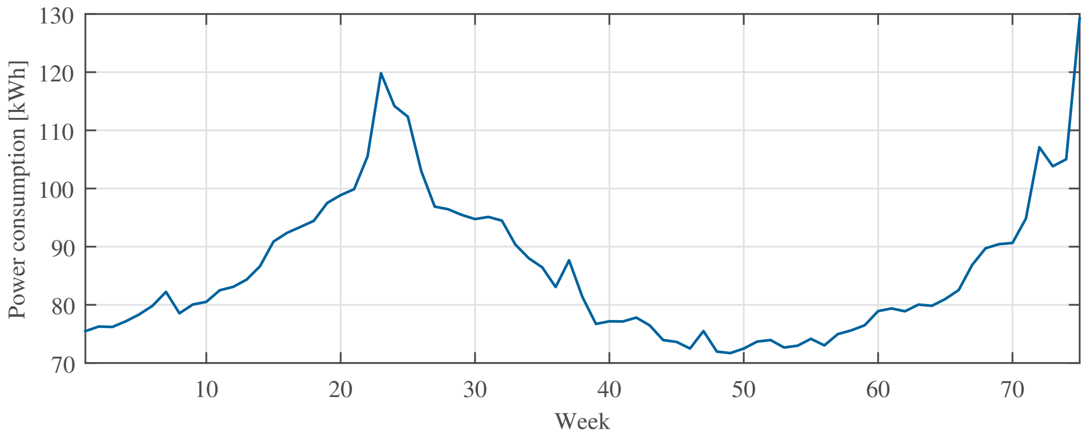
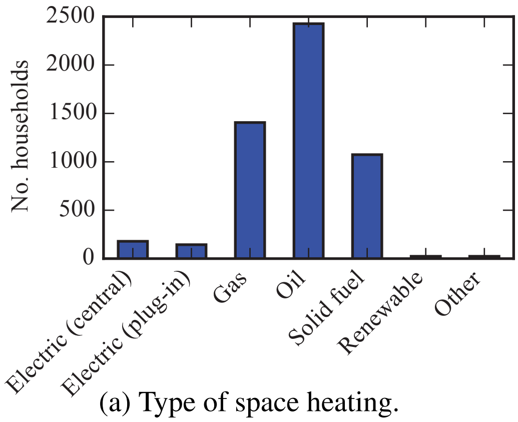


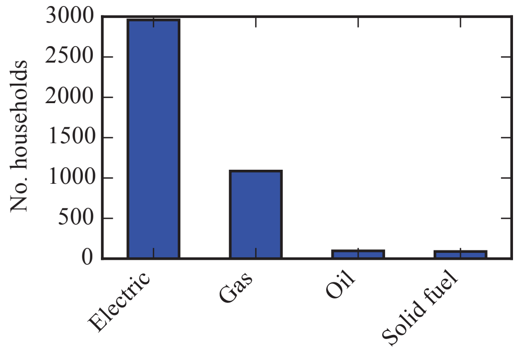
Figure 3.4: Average weekly power consumption of all households over the 75 weeks of the trial (from 20 July 2009 to 26 December 2010).

gas, or solid fuels (*e.g.*, coal, peat or wood). In contrast to other studies that investigated large-scale electricity consumption data (such as [2, 90, 107]), the energy attributed to heating is not reflected in the electricity consumption data provided by the CER data set. The penetration of electric heating is relatively small: Only 179 of the 4231 households use a central electric heating system, while 144 specified that they have a plug-in electric heater. However, since households were allowed to give multiple answers, these plug-in heaters are most likely used in addition to another, primary heat source. Similarly, it is not clear whether the electricity consumed by the central electric heating of the 179 households is covered by the data set. These numbers are in line with the information provided by the Sustainable Energy Authority of Ireland (SEAI), which states that the penetration of heating systems powered by electricity was only 3% in 2005 [166]. Among all households participating in the study, more than 80% control through a timer when the heating system is switched on and off.

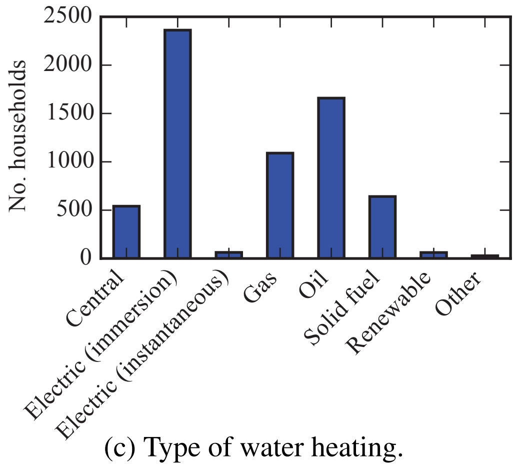
Cooking (*cf.* figure 3.5b) is mostly performed using electricity. Only 1086 households (roughly 25%) have a gas stove, while the number of households that use oil or solid fuel to heat their meals is negligible. The case of water heating (*cf.* figure 3.5c) is more interesting: Although most of the Irish households have an immersion system installed, only 10% use it as the primary method for heating water [124, 178]. This is consistent with the survey results of the CBT, where 517 of the 4231 households specified that they used electric immersion as the only source to heat water. Hence, in the CER data set, we can assume that water heating for most of the households is performed using gas, oil, solid fuel, or through a central heating system and therefore not directly reflected in the smart meter data.



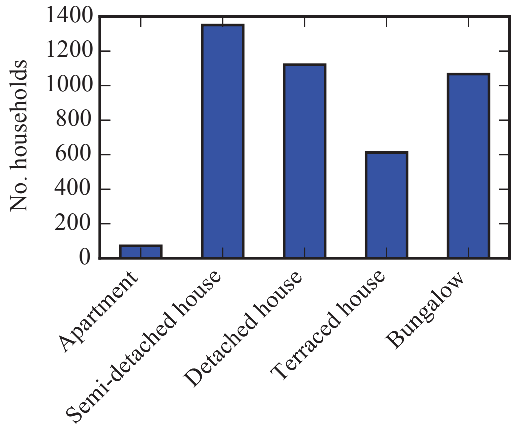
(a) Type of space heating.



(b) Type of cooking.



(c) Type of water heating.



(d) Type of dwelling.

Figure 3.5: Selected household characteristics based on the pre-trial questionnaire data. In case of space heating and water heating, participants could give multiple answers.

Most of the people who participated in the survey live in semi-detached, detached, or terraced houses or in bungalows (*cf.* figure 3.5d). Only 72 households stated that they lived in an apartment, which is a share of only 1.7%. According to national census data from 2011, apartments represent about 10% of the building stock in Ireland [194]. One explanation for this underrepresentation is that the CER potentially aimed at avoiding short-term tenancies when designing the smart metering trial to reduce the attrition [42, 124].

Table 3.2 gives an overview of the penetration of appliances such as washing machines, freezers, or computers. The questionnaires did not explicitly ask for the number of refrigerators, possibly because the authors assume that each household in Ireland owns exactly one refrigerator. While the penetration of washing machines is also close to 100%, the table shows that only two thirds of the households own a tumble dryer and also two thirds own a dishwasher. Finally, only 16% of households do not own a large TV,

Table 3.2: Penetration of appliances owned by the 4231 households that we used in our analysis. For each appliance type, the columns specify the percentage of households that own exactly 1, 2, 3, or 4 appliances of this type.

Appliance type	**1**	**2**	**3**	**4**	**Sum**
Washing machine	97.7%	0.6%	-	-	98.3%
Tumble dryer	68.1%	0.1%	-	-	68.3%
Dishwasher	67.0%	0.2%	-	-	67.2%
Electric shower (instant)	63.8%	5.1%	0.5%	-	69.4%
Electric shower (pumped)	26.8%	2.1%	0.4%	-	29.4%
Electric cooker	77.0%	0.3%	0.0%	-	77.3%
Electric heater	23.3%	5.2%	1.7%	-	30.2%
Stand-alone freezer	47.8%	1.7%	0.1%	-	49.7%
Water pump or electric well	19.0%	0.4%	0.0%	-	19.5%
Water heating	76.6%	0.3%	-	-	77.0%
TV ($<$ 21 inch)	39.6%	18.0%	5.9%	2.0%	65.4%
TV ($>$ 21 inches)	51.0%	25.0%	6.0%	2.3%	84.3%
Desktop computers	44.5%	2.3%	0.3%	0.1%	47.3%
Laptop computers	42.7%	8.4%	2.1%	0.8%	54.1%
Game consoles	22.6%	8.5%	2.2%	0.7%	34.0%

and—independent of the penetration of large TVs—35% of the households do not own a small TV. This means that at least 50% of the households own two TVs independent of their size. While the data set also contains information on usage of the appliances, we omit a detailed analysis and refer instead to [211].

The data that describes the household characteristics did not require any cleaning or transformation. An exception is the floor area of the dwelling, which could be specified either in square meters (1641 responses) or in square feet (140 responses). Figure 3.6 shows the distributions of the two types of responses (in square meters). The left part of the blue line, which indicates the responses that were given in square meters, shows a similar distribution to the red line, which specifies the distribution of the responses given in square feet. However, the right part of the blue line (starting at 600 m^2) seems to show the households that describe the floor area of their dwelling in square feet instead of square meters. The first reason is that these values are too high for floor areas to be specified in square meters, assuming typical household sizes. Second, no household that specified the floor area in square feet reaches values in this region. For this reason, we divided the floor area of those households by 10.764 (*i.e.*, the conversion factor from square feet to square meters) and obtained an overall floor area distribution similar to the one shown by the red line.

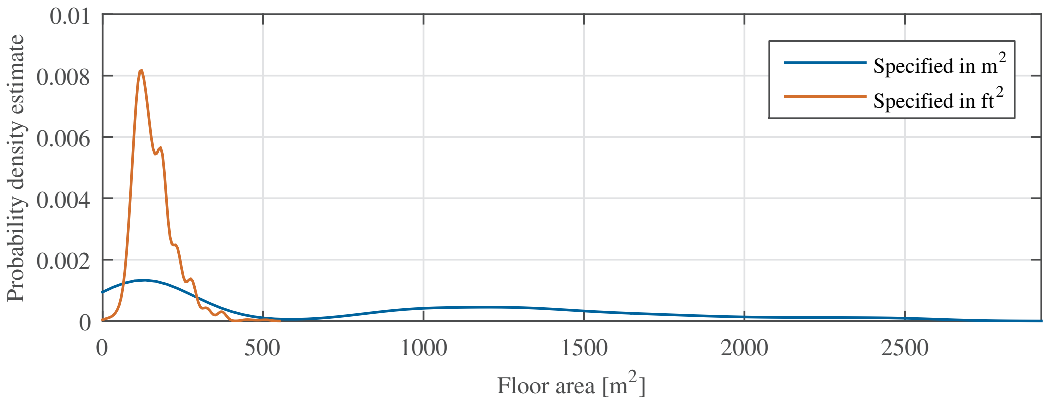


Figure 3.6: Probability density for the floor area of the households used in our analysis. In the questionnaire, the floor area could be specified either in square meters (blue line) or in square feet (red line).

To the best of our knowledge, the CER data set is the largest data set available that contains both electricity consumption data and side-information such as information related to the socio-economic characteristics, the dwelling, or on appliance stock and usage. For this reason, we consider the results presented in this thesis generalizable and assume our methods achieve similar results on other data sets. However, other large-scale data sets that include detailed information on electricity consumption data typically lack such detailed side-information. The PG&E data set, for instance, contains electricity consumption data from more than 30,000 customers in the US [236]. In addition to the consumption data, it only includes the ZIP-code of the households as well as information on their participation in energy efficiency and demand response programs.

3.2 System design

Our analysis relies on supervised machine learning techniques to infer a household's characteristics from its electricity consumption data. Figure 3.7 depicts the household characteristic estimation process. First, we compute a set of *features* on the electricity consumption records of a household. This is a typical step performed in supervised machine learning to obtain a set of discriminative values for each sample (*i.e.*, household). The features then serve as input to a *classifier* or *regression model*, depending on the characteristic. As output, our system provides an estimate of the class or the value of each characteristic.

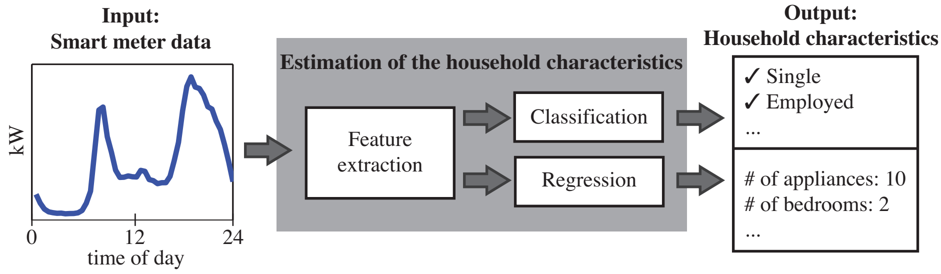


Figure 3.7: Overview of the household characteristic estimation.

3.2.1 Features

Table 3.3 lists the features we compute on the electricity consumption data. We divide the features into five groups: consumption figures (10 features), ratios of consumption figures (7 features), features related to temporal dynamics (4 features), statistical properties (3 features), and the first ten principal components [183]. Consumption figures correspond to simple aggregates of the actual consumption values of a household. For instance, the minimum or maximum consumption values of a day or the average consumption within a specific period (*e.g.*, in the morning or during the night) are referred to as consumption figures. Ratios are quotients of average consumption values of different periods of a day. An example is the ratio between the average consumption in the morning and that during lunch-time. Temporal properties describe how long the consumption is above a given threshold. Finally, statistical properties allow to capture qualitative characteristics of the consumption curve. For instance, in order to determine how consumption profiles (of the same household) correlate to each other over subsequent days we compute the cross-correlation between these profiles. The table also shows the labels (in the center column) we use to indicate the different features. The intervals *morning*, *noon*, *evening*, and *night* are defined as the time periods 6 a.m.–10 a.m., 10 a.m.–2 p.m., 6 p.m.–10 p.m., and 1 a.m.–5 a.m., respectively. Our system assumes the data to be available at a granularity of one measurement every 30 minutes and it computes each feature on one week of data. However, it can be easily adapted to cope with other data granularities and time periods. Part of the features have been used in previous work on the analysis of electricity consumption data [155].

Many statistical methods assume the input data to follow a normal distribution [134]. For this reason, researchers often apply a non-linear transformation (*e.g.*, a logarithmic or square root transformation) to each of the features if it improves normality [134]. To find the right transformation, we (visually) compare the distribution of the transformed feature with the normal distribution using a normal quantile plot [177] and choose the transformation that most closely approximates the normal distribution. The aforementioned table

Table 3.3: List of features that form the input vectors of the classifiers. $\bar{P}$ denotes the 30-minute mean power samples provided by the data set. Where not otherwise stated, the feature is computed over the weekdays only. The last column shows if a logarithmic (log) or square root (sqrt) transformation has been applied to the feature.

Description	Name	Transformation
(1) Consumption figures		
$\bar{P}$ (daily, week)	`c_total`	sqrt(x)
$\bar{P}$ (daily, weekdays)	`c_weekday`	sqrt(x)
$\bar{P}$ (daily, weekend)	`c_weekend`	sqrt(x)
$\bar{P}$ for (6 a.m. – 10 p.m.)	`c_day`	sqrt(x)
$\bar{P}$ for (6 p.m. – 10 p.m.)	`c_evening`	sqrt(x)
$\bar{P}$ for (6 a.m. – 10 a.m.)	`c_morning`	sqrt(x)
$\bar{P}$ for (1 a.m. – 5 a.m.)	`c_night`	log(x)
$\bar{P}$ for (10 a.m. – 2 p.m.)	`c_noon`	sqrt(x)
Maximum of $\bar{P}$, week	`c_max`	x
Minimum of $\bar{P}$, week	`c_min`	log(x)
(2) Ratios		
Mean $\bar{P}$ over maximum $\bar{P}$	`r_mean/max`	log(x)
Minimum $\bar{P}$ over mean $\bar{P}$	`r_min/mean`	sqrt(sqrt(x))
`c_morning` / `c_noon`	`r_morning/noon`	log(x)
`c_evening` / `c_noon`	`r_evening/noon`	log(x)
`c_noon` / `c_total`	`r_noon/day`	sqrt(x)
`c_night` / `c_day`	`r_night/day`	log(x)
`c_weekday` / `c_weekend`	`r_weekday/weekend`	log(x)
(3) Temporal properties		
Proportion of time with $\bar{P} > 0.5$kW	`t_above_0.5kw`	x
Proportion of time with $\bar{P} > 1$kW	`t_above_1kw`	x
Proportion of time with $\bar{P} > 2$kW	`t_above_2kw`	x
Proportion of time with $\bar{P} >$ mean	`t_above_mean`	x
(4) Statistical properties		
Variance	`s_variance`	sqrt(sqrt(x))
$\sum(\lvert\bar{P}_t - \bar{P}_{t-1}\rvert)$ for all t	`s_diff`	sqrt(x)
Cross-correlation of subsequent days	`s_x-corr`	x
$\#\bar{P}$ with $(\bar{P}_t - \bar{P}_{t\pm1} > 0.2$ kW)	`s_num_peaks`	x
(5) Principal components		
First ten principal components	`pca_i` *(i = 1..10)*	x

of features (*cf.* table 3.3) lists this transformation on the right column. Figure 3.8 shows the normal quantile plot for features `c_total` and `r_morning/noon` transformed by a logarithmic and a square root transformation, respectively. The linearity of the sample quantiles of the features (x-axis) versus the theoretical quantiles of a normal distribution (y-axis) implies that the transformed features are (roughly) normally distributed. After the transformation, we normalize each feature such that it has zero mean and unit variance. Data normalization is required by some of the classifiers we consider in our study, for example when their objective function calculates a distance between two samples based on their features.

3.2.2 Household characteristics and class labels

A classifier estimates a characteristic of a household by assigning the household to a specific *class* out of a set of classes. Instead of estimating the age of a person, for instance, a classifier estimates the person to be at young, medium, or old age. Alternatively, it could estimate whether the person is a teenager, in its twenties, in its thirties, *et cetera*. Hence, when designing a system, it is important to specify which characteristics should be classified (*e.g.*, the age of a person) and how to separate them into meaningful classes (*e.g.*, young, medium, old).

The definition of classes can for instance be done depending on the intended use of the classification outcome or by exploring the actually available data. We refer to these two approaches as the *application-driven* and *data-driven* class definition methods, respectively. In this section, we describe both our application-driven approach and our data-driven approach to identify relevant household characteristics and class labels. In the first case, we consider one of the target applications for our classification system and define a number of characteristics of the households that are relevant in order to support it. Examples for such applications are personal energy consulting, providing households with a personalized electricity bill, or customer segmentation for energy-efficiency and marketing campaigns. In the context of this section, the target application is personal energy consulting. Customers of several European energy providers can already request such a consulting service in order to survey their household and identify ways to reduce their energy consumption. This can often be achieved simply by replacing inefficient devices but also by discovering incorrect wirings that can make a customer pay for its neighbor's hyperactive boiler. As a consequence of the current political pressure towards an overall more thrifty usage of energy resources, providers are also starting to offer energy consulting services for free to their customers. To evaluate the use of a classification system for energy consulting, we present the results of four interviews conducted with energy consultants from different Swiss utilities. By analyzing the results of the interviews

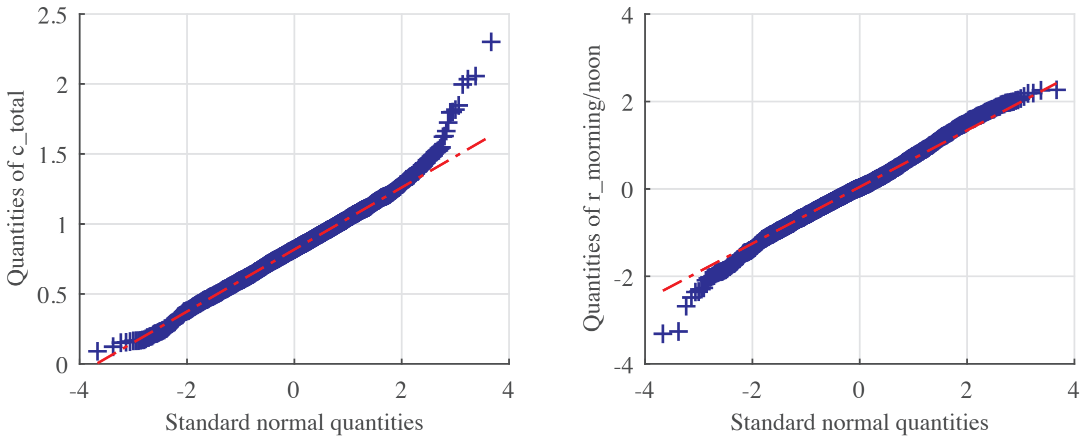


Figure 3.8: Normal quantile plots showing that features c_total (left) and r_morning/noon (right) are (roughly) normally distributed after applying the log and square root transformations, respectively.

we thus elicit a first significant set of household characteristics—and thus classes of our classification problem.

We then turn to a data-driven approach and analyze the CER data set. In particular, we define a number of potentially interesting features of the consumption profiles and investigate their correlation to several different household characteristics. To this end, we rely on self-organizing maps [105]. Our results show that, for instance, the size of a household and the income of its occupants are both interesting for the energy consulting application and most likely to be reliably identified by a common classifier. In the following, we first describe the application-driven and data-driven approaches, before we present the resulting list of household characteristics and class labels.

Application-driven analysis: Interviews with energy consultants

This section describes the results of interviews that we conducted with employees of four medium-sized Swiss energy providers. In the context of these interviews, we focused on a specific consumer-tailored service: personal energy consulting. The goal of the interviews is to elicit the characteristics of a household whose knowledge is of highest value to a provider that aims at offering an energy consulting service. We provide the complete interviews in the appendix A.

Each interview lasted about two hours and focused on five main topics: 1) Identification and selection of customers representing potential targets of energy consulting services; 2) Typical flow of an energy consulting session; 3) Assessment and determination of potential savings; 4) Use of outcome of the energy consulting session (both for the customer as well as the provider); 5) Energy consulting in the long-term.

According to the interviews, several energy providers already offer their customers an energy consulting service. However, the service is typically not free and only offered upon request of the customer. Energy providers are however subject to an ever increasing political and social pressure to contribute in reducing the overall energy consumption. As also outlined in our interviews, several providers see the possibility to offer free energy consulting services as a practical way to fulfill the mentioned political "mandate" as well as to please customers. Furthermore, in case of a rise in energy prices the providers also expect the number of customers that request an energy consulting service to increase.

The analysis of these interviews allows us to make the following qualitative considerations. First, the interviewed energy consultants believe that the availability of additional information about the characteristics of a household can significantly support the preparation and execution of an energy consulting session. Second, in order to offer energy consulting services on a large-scale and in an efficient way, providers must become able to automatically select customers that are most likely to benefit from the service. In the remainder of this section, we discuss these considerations in more detail.

Preparation and execution of energy consulting sessions: As mentioned above, the analysis of our interview data shows that energy consultants consider the availability of information about a household valuable when preparing or executing a consulting session. The respondents indicate as particularly valuable characteristics like the size of the household (*e.g.*, expressed in terms of floor area), the number of bedrooms, and the number of adults and children living in the household. All four respondents stress that even rough estimates of these values, combined with available consumption data, enable the consultant to gain a comprehensive picture of the efficiency of the household and to formulate customer-tailored recommendations. Further characteristics mentioned by the respondents as particularly valuable include the number and type of electrical appliances present in the household as well as the type of space and water heating.

One of the respondents also notes that hints at potential energy-wasting sources within the household are particularly valuable. Such hints can be obtained by comparing household characteristics estimated from consumption data with actual characteristics surveyed with household occupants at the beginning of a consulting session. Interestingly, the respondent also mentioned that about 10% of the private customers served by his company are likely to be affected by wiring errors that might make them pay, for instance, for their neighbor's water heating due to a wrongly connected boiler. The availability of information about household characteristics makes detection of such flaws much easier for the consultant.

Selection of "high-potential" customers: In order to make energy consulting services successful, energy providers must become able to identify customers that are likely to be pleased by—and can possibly benefit from—such services.

The analysis of our data shows that the interviewed consultants consider two groups of households as particularly interesting targets for an energy consultation: households with a large energy saving potential, and households occupied by certain types of consumers. The first category of households can be identified by looking at high average consumption (*e.g.*, caused by a high number of appliances) as well as the presence of inefficient appliances or of an old infrastructure. The type of heating or cooling used in a household is for instance relevant in order to select "high-potential" customers. As for the second category, one of our respondents points out that retired individuals represent an example of an interesting class of consumers. Indeed, many retired individuals are often ready to invest their time and engage in a consulting session and might be more keen in adapting their consumption behavior even without the promise of a corresponding financial compensation. One other of the respondents also indicates double income – no kids (DINK) households as particularly interesting as their occupants are more likely to invest in renovation measures to improve their energy efficiency.

In summary, our application-driven analysis allows to define the following characteristics of a household as particularly relevant in order to support energy consulting services: type of employment of the occupants, number of adults/children living in the household, type of space heating, type of water heating, total number of appliances in the household, age of the household dwelling. Whether these characteristics can actually be detected from typical electricity consumption data is discussed in the following section.

Data-driven analysis: Self-organizing maps

The CER data set contains answers to more than 100 questions. To identify suitable classes in such a scenario, when no "natural" definition of characteristics exists (*e.g.*, the floor area of a household), we perform an exploratory, unsupervised data analysis to identify the characteristics that we include in our household classification system. The goal of the data-driven analysis presented in this section is thus to provide a list of household characteristics that can likely be automatically detected through the analysis of electricity consumption data. The data-driven analysis also allows us to verify beforehand the existence of an adequately significant overlap between the properties that are interesting for realistic application scenarios—described in the previous section—and those that can most likely be discovered from the data.

Our analysis relies on the CER data set described in the previous section. For carrying out the analysis we utilize self-organizing maps (SOMs)—a well-known method to project high-dimensional data onto a 2-dimensional space [105]. A SOM is an artificial neural network that relies on unsupervised learning to group input vectors into *regions* of a map. Each vector is assigned to a specific region depending on its Euclidean (or other type of) distance to already mapped vectors. Clustering procedures can then be applied to group vectors within neighboring regions into *clusters*.

In our analysis, the input vectors consist of features we extract from electricity consumption traces. If a large number of households with the same value for a specific characteristic are mapped to the same cluster on the map and households with different values for this characteristic are mapped to a different cluster, we can conclude that it is also possible to classify households according to this characteristic using electricity data only. We should however point out that this procedure does not provide a classification of the households. Indeed, we use the SOM to explore the data set and to discover which classes are more meaningful to be included in our automated classification system. The implementation and evaluation of the classification system itself are described in the subsequent sections.

The features used to build the input vectors of the self-organizing map are similar to the ones described in table 3.3 in section 3.2.1, which describes the features we (later) use for classification. However, this list of features has been updated after we performed the SOM analysis. For the SOM analysis, we computed the features over a whole week, while the features listed in the table are computed only over the weekdays to capture weekly routines. Also, the temporal features used in the SOM analysis specify the time of the first occurrence of an event, while the temporal features we use for classification (*i.e.*, the ones listed in table 3.3) represent the time period where the consumption exceeds a certain threshold. Finally, we did not include the principal components in the SOM analysis. In appendix B, we provide the exact definition of the features used for the SOM analysis presented in this section.

To perform the SOM analysis we first compute all the features described above for each household. To this end, we use the consumption data corresponding to a single week of the trial (week 26). We then normalize the values of the features using the unit variance scaling method [170]. This normalization is necessary since we use the computed features as the components of the input vectors of a SOM. In particular, we use the Euclidean distance as the function that determines the distance between different input vectors and thus, their position on the SOM. Due to their different nature these features might however exhibit values of very different magnitude. Without normalization, features with large absolute values would thus bias the computation of the Euclidean distance between input vectors and "mask" the effect of features with small(er) magnitudes [170]. The experiments are based on a subset of 3488 out of all the traces available in the data set, because we limited the analysis to those households that have filled out questionnaires and are listed as *residential* in the documentation of the data set. This excludes households that have filled out a questionnaire but left the trial due to technical problems or for other reasons.

To implement the SOM we use the *SOM Toolbox 2.0* developed for MATLAB by researchers at the Helsinki University of Technology [215]. This tool automatically determines the number of clusters into which the input data is grouped. In particular, it determines a first set of regions on the map using all the input vectors and then applies a *k*-clustering filter to balance the map and reduce the total number of regions.

Figure 3.9 shows the map resulting from running the SOM analysis with input vectors that contain all the features as their components. The input vectors are grouped in 15 clusters. The bar on the left side shows the number of "hits" per cluster (*i.e.*, the number of samples assigned to each cluster). It shows that all clusters are of about the same size and thus contain roughly the same number of households.

While figure 3.9 displays the final output of the SOM, it does not allow to draw the conclusions we are actually interested in (*i.e.*, if the features computed on the electricity consumption data actually cause households with similar properties to get "naturally" grouped together on the map). To analyze this aspect we display the percentage of households that exhibit a specific property in each of the clusters identified by the SOM. Figure 3.10a, for instance, displays these percentages for the property *employment*, which describes the employment state of the chief income earner (CIE) of the household. The plot shows that the CIEs of about 80% of the households assigned to clusters 7, 9, and 12 are employed, while this percentage decreases to about 30% for clusters 15, 10, and 3. This means that the employment status of the CIE is a property that can likely be discovered from the data, as it is distributed unevenly over the different clusters. In contrast to this, the percentage of households that have a time-controlled water heating is nearly the same in all clusters, as shown in figure 3.10b. This means that using the set of features we have defined it is most likely not possible to determine automatically whether a household has a time-controlled water heating or not. Similarly, figure 3.10d also shows that the property *own_house*, which tells whether the occupants of the household are also its owners, can hardly be distinguished according to the clustering provided by the SOM. On the other side, figure 3.10c shows that detecting whether a household has a number of appliances higher than a given threshold might indeed be possible, although not straightforward. In particular, about 80% of the households included in cluster 15 have less than six appliances. In contrast to this, nearly all of the households included in cluster 13 have more than four appliances.

As an outcome of our data-driven analysis we can conclude that the following properties are most likely to be inferable from electricity consumption data: Employment of the CIE,

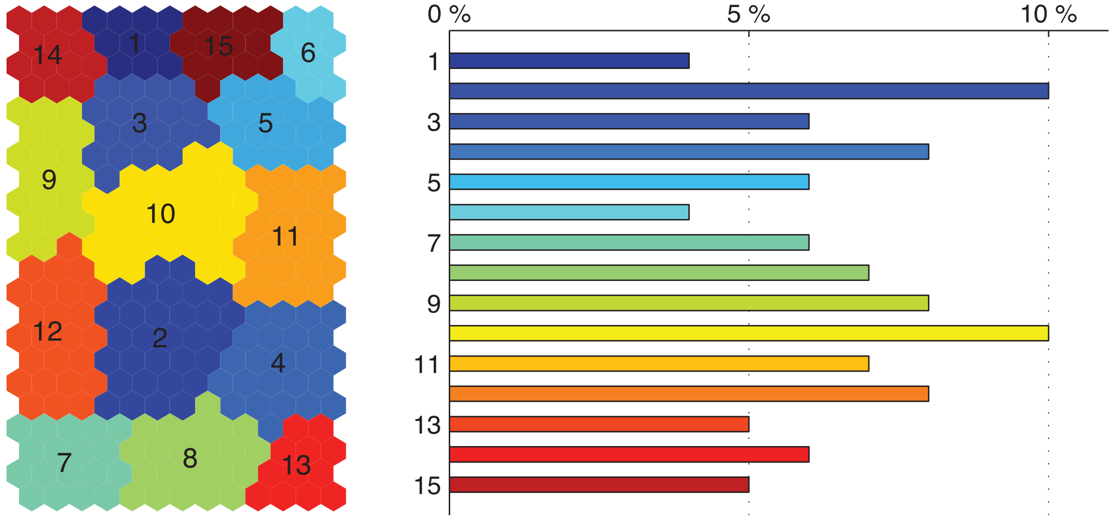
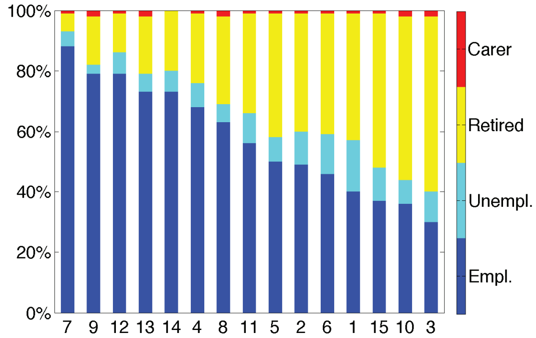


Figure 3.9: Clusters obtained when training the SOM using input vectors that contain the features described in section 3.2.2. The histogram on the right side indicates the proportion of "hits" per cluster. [153]

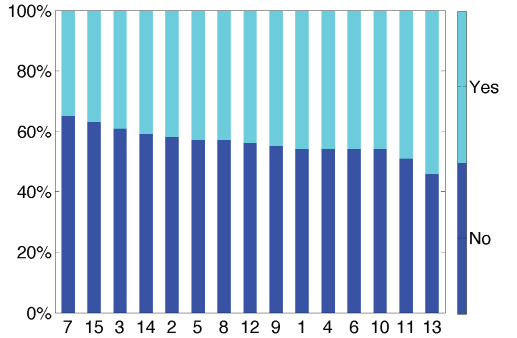
type of cooking, number of bedrooms, floor area, social class of the CIE, number of persons, number of persons that are at home during the day, and number of appliances. We thus argue that by applying standard classification techniques to electricity consumption data, it is possible to automatically classify private households according to these properties.

To further analyze the clustering performance of the SOM it is also interesting to investigate the *component planes*. A component plane displays how a single feature is distributed over the map or, equivalently, how the feature contributes to the final shape of the map. For instance, the upper left plot in figure 3.11 shows the component plane relative to the feature *c_day*, which represents the daily electricity consumption of a household. A different color corresponds to a different value of the feature (as indicated by the color bar on the right side of each plot). The plot shows that households with a low average consumption are assigned to the top of the map while households with high consumption tend to cluster on the lower right corner. A far less regular distribution is instead exhibited by the component plane relative to the feature *t_daily_max*—also shown in figure 3.11—which describes the time of the day at which the maximal value of electricity consumption is reached. Figure 3.11 displays other examples of features with fairly regular (*c_morning*, *c_weekday*, *c_weekend*, *s_variance*, *r_mean/max*, *r_min/mean*, *r_night/day*, *r_evening/noon*) or quite irregular (*s_x-corr*, *t_above_mean*) component planes.

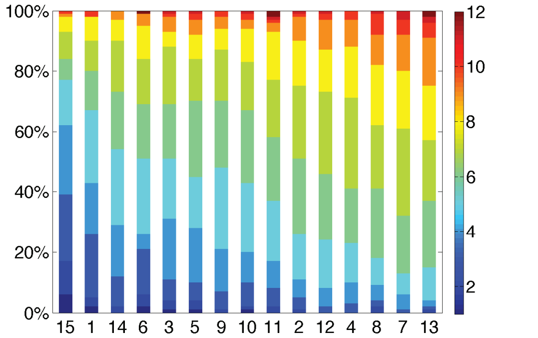
Features whose component planes exhibit a "regular" distribution over the map are likely to induce a similarly regular structure on the overall map, which is the map that results from a combination of all component planes. In contrast, an "irregular" component plane



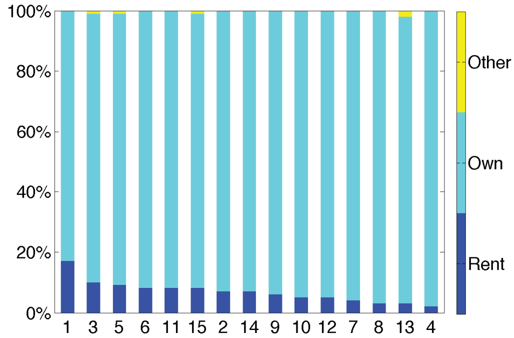
(a) Employment status of the chief income earner of the household.



(b) Time-controlled water heating.



(c) Number of appliances present in the household.



(d) Ownership statistics of the household.

Figure 3.10: Distribution of selected household characteristics over the different clusters. The distributions of the other household characteristics are shown in appendix B.

indicates a feature that does not succeed in inducing a regular clustering. This observation can be used to identify the set of features that can help improving the final classification results. Based on the results of this analysis, we updated the list of features before integrating them into our household classification system (*cf.* table 3.3 in section 3.2.1 for the final list of features).

Resulting household characteristics and class labels

Through the application-driven and data-driven analysis, we identified 18 different characteristics that we use to evaluate our household classification system. Table 3.4 shows the 18 characteristics along with the corresponding classes and class definitions for each characteristic. The characteristics capture socio-economic status of the household (*e.g.*, `age_person`, `employment`), dwelling properties (*e.g.*, `#bedrooms`, `floor_area`), or characteristics related to the behavior or appliance stock (*e.g.*, `#appliances`, `unoccupied`).

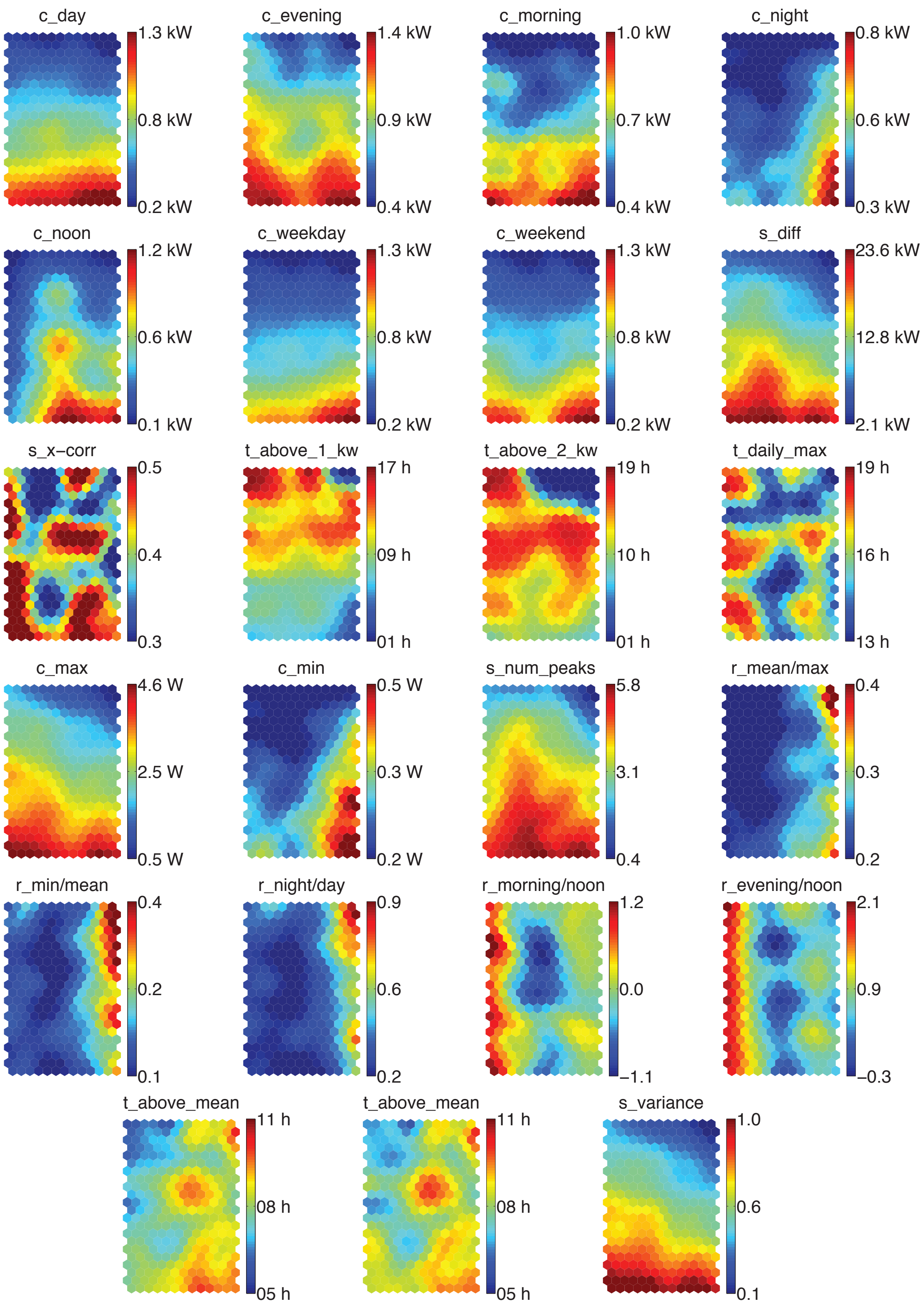


Figure 3.11: Component planes of the features used to train the SOM. The names of the features correspond to the description given in appendix B.

#adults and #children represent the number of adults and children in the household, respectively. The table also shows the number of samples for each class, where each sample corresponds to one household in the CER data set.

In our application-driven analysis, we identified the characteristics that are interesting for utilities by conducting interviews with four energy consultants. The interviews revealed, for instance, that knowing the composition of a household (*e.g.*, single, family) is particularly relevant to energy consultants, because families are potentially more interested than singles in receiving information about energy consulting services. Furthermore, through a data-driven analysis we selected characteristics with well-separable classes, which means that the samples from different classes have (on average) a high distance in the feature space. As an example, figure 3.12 illustrates class separability of the characteristic single for features c_total and r_evening_noon based on the empirical cumulative distribution (ECD) for each of the two features. The left plot shows that the ECD of the first class (*Single*) significantly differs from the ECD of the second class (*No single*) for feature c_total. This means that the classes *Single* and *No single* are well separable with respect to feature c_total. On the other side, the right plot shows that the ECD of the two classes are almost the same for feature r_noon/day. As a consequence, we say that single is well-separable because there is at least one feature that properly separates the classes.

In terms of class labels, there are natural definitions of class labels for some of the characteristics (*e.g.*, *Single/No single*, or *Family/No family*). For other characteristics (*e.g.*, age_person, #bedrooms, floor_area), we define the class labels (1) according to qualitative considerations gathered during the aforementioned interviews and (2) by adjusting the number and definition of class labels such that the number of households in each class is similar.

3.2.3 Classifiers

There exist several classifiers that can be used to perform supervised machine learning tasks [6, 29, 43, 75]. These classifiers typically differ in terms of implementation and computational complexity, or in the assumptions they make on the distribution of the data. For the study described in this section, we have selected five well-known classifiers: the k-Nearest Neighbors (kNN) classifier [29], the linear discriminant analysis (LDA) classifier [29], the Mahalanobis distance classifier [6], the support vector machine (SVM) classifier [43], and the AdaBoost classifier [75]. For a detailed description of the five classifiers, the reader is referred to [6, 29, 43]. Here, we outline the specific trade-offs

Table 3.4: List of household characteristics, their class labels, and the number of samples per class. The characteristics eligible for regression are marked with $^{(*)}$.

Characteristic	Description	Classes	Samples
age_person$^{(*)}$	Age of chief income earner	Young ($\texttt{age\_person} < 35$)	436
		Medium ($35 < \texttt{age\_person} \leq 65$)	2819
		High ($65 < \texttt{age\_person}$)	953
all_employed	All adults work for pay	Yes	1013
		No	2409
#appliances$^{(*)}$	Number of appliances	Low ($\texttt{\#appliances} \leq 8$)	1421
		Medium ($8 < \texttt{\#appliances} \leq 11$)	1479
		High ($11 < \texttt{\#appliances}$)	1332
#bedrooms$^{(*)}$	Number of bedrooms	Very low ($\texttt{\#bedrooms} \leq 2$)	404
		Low ($\texttt{\#bedrooms} = 3$)	1884
		High ($\texttt{\#bedrooms} = 4$)	1470
		Very high ($4 < \texttt{\#bedrooms}$)	465
cooking	Type of cooking facility	Electrical	2960
		Not electrical	1272
employment	Employment of chief income earner	Employed	2536
		Not employed	1696
family	Family	Family ($\texttt{\#adults} > 1$ and $\texttt{\#children} > 0$)	1118
		No family	3114
floor_area$^{(*)}$	Floor area	Small ($\texttt{floor\_area} \leq 100\ m^2$)	232
		Medium (floor_area from $100\ m^2$ to $200\ m^2$)	1198
		Big ($200\ m^2 < \texttt{floor\_area}$)	351
house_type	Type of house	Free (detached or bungalow)	2189
		Connected (semi-detached or terraced)	1964
income$^{(*)}$	Yearly household income	Low ($\texttt{income} < 50000$)	940
		High ($50000 \leq \texttt{income}$)	997
lightbulbs	Proportion of energy efficient light bulbs	Up to a half	2041
		About three quarters or more	2191
children	Children	Yes ($\texttt{\#children} \geq 1$)	1229
		No ($\texttt{\#children} = 0$)	3003
age_house	Age of building	Old ($30 < \texttt{age\_house}$)	2151
		New ($\texttt{age\_house} \leq 30$)	2077
#residents$^{(*)}$	Number of residents	Few ($\texttt{\#residents} \leq 2$)	2199
		Many ($3 \leq \texttt{\#residents}$)	2033
retirement	Retirement status of chief income earner	Retired	1285
		Not retired	2947
single	Single	Single ($\texttt{\#adults} = 1$ and $\texttt{\#children} = 0$)	859
		No single	3373
social_class	Social class of chief income earner according to NRS social grades	A or B	642
		C1 or C2	1840
		D or E	1593
unoccupied	Is the house unoccupied for more than 6 hours per day?	Yes	885
		No	3347

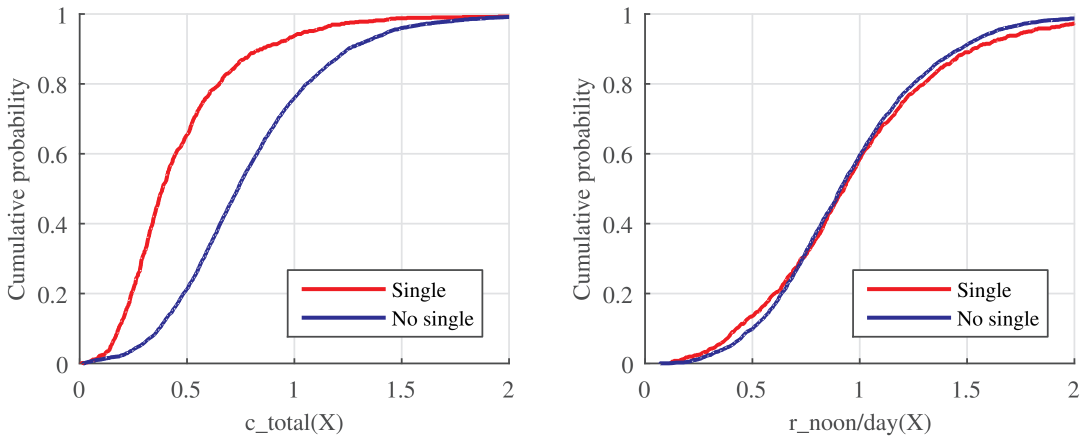


Figure 3.12: Empirical cumulative distributions of the (unscaled) features `c_total` (left) and `r_noon/day` (right) for characteristic `single`.

exposed by these classifiers and provide a few details about their implementation within our classification system.

Besides its simplicity, a main advantage of the kNN classifier is that it does not make any assumption on the distribution of the input data, which also does not need to be linearly separable. On the other side, the kNN classifier has high computational and memory requirements. For the LDA classifier, we assume a (multivariate) Gaussian distribution of the input data samples. This causes the parameters of the discriminant functions that partition the feature space to be dependent on the mean and covariance of the distributions for each class. The linear functions of the LDA classifier are obtained by assuming a common (pooled) covariance matrix for all classes, constructed by averaging the covariance matrices of each class. The need to assume Gaussianity of the input data is a major drawback of the LDA classifier. On the other side, it has very low requirements in terms of computation and memory usage. The Mahalanobis classifier is conceptually similar to the LDA classifier. One of the main differences is that the former relies on stratified covariance matrices, instead of a pooled covariance matrix. This results in quadratic discriminant functions, which typically allows the Mahalanobis classifier to have better classification performance than the LDA classifier. However, its performance is also more sensitive to the estimation accuracy of the stratified covariance matrices. SVMs are widely used in classification applications [29], which is due to their flexibility and thus, applicability to many classification problems of different natures. A major strength of SVMs is their ability to compute decision boundaries without assuming specific distributions of the input data (like the kNN classifier but unlike the LDA or Mahalanobis classifiers). Further, SVMs are able to cope with data that is not linearly separable, since they support non-linear decision boundaries. A major drawback is the computationally expensive training phase. Finally, the AdaBoost classifier is a powerful classification algorithm that combines many weak and possibly inaccurate classifiers (that are better than random guessing) to form one powerful classifier [75]. Its name stems from adaptive boosting: It boosts the results of the weak classifiers by performing multiple classification rounds, while it assigns a weight to each sample in the training set. The weight is adapted from run to run depending on the error obtained when classifying this particular sample. A strength of AdaBoost is that it does not require much configuration from the user: the user selects a weak classifier (or a set of weak classifiers) as well as the number of training rounds. AdaBoost has enjoyed high success in disciplines such as biology, computer vision, and speech processing [208]. The disadvantage of AdaBoost is that it is sometimes susceptible to noisy data and outliers [208].

The right column in table 3.4 shows that some of the characteristics are imbalanced in the CER data set. This means that some classes have a significantly higher number of samples than other classes. For example, there are 859 households for which the characteristic `single` takes value *Single* and 3373 for which it takes value *No single*. This bias affects the performance of some of the classifiers: Since the trained model of these classifiers is biased towards the class with the majority of samples, they often assign samples of the underrepresented classes to the majority class [22]. An effective method to deal with class imbalance consists in *undersampling* the data during the training process [86, 87]. By randomly removing samples from the overrepresented classes, undersampling creates evenly distributed classes (*i.e.*, classes having the same number of samples equal to the number of samples in the smallest class). In order to support applications that rely on identifying samples of underrepresented classes, our system can thus also perform undersampling.

3.2.4 Multiple linear regression

Some of the characteristics in table 3.4—namely `age_person`, `#bedrooms`, `#appliances`, `floor_area`, `income`, and `#residents`—take values in a continuous interval. For these characteristics we train a regression model in order to estimate the value of the characteristic. We model each of the characteristics individually and use a multiple linear regression model for its simplicity and interpretability of the parameters. The model is expressed as follows:

$$f_R : y_j = \beta_0 + \beta^T x_j + \varepsilon, \varepsilon \sim N(0, \sigma^2), \tag{3.1}$$

where y_j represents the j-th household's observed value in the training set and x_j denotes the feature vector computed for household j. The coefficients β are then estimated using ordinary least squares (OLS) regression [145].

3.3 Evaluation process

This section describes how we use the features and classifiers described above to derive quantitative results on the potential to reveal household characteristics from electricity consumption data.

3.3.1 Performance measures

The first step in determining the performance of a classification outcome is to count the number of correct classifications and the number of misclassifications for each class and thus derive the *confusion matrix* CM. Consider a classification with K classes $(1,...,K)$ and S samples. The confusion matrix consists of K rows and K columns. The element (i, j) of the confusion matrix represents the number of samples of class i that have been classified as class j. Therefore, the elements on the main diagonal of the matrix CM, indicated as CM_{ii} $(i = 1,...,K)$, represent the number of correctly classified samples for each class. If $K = 2$, the entries CM_{11}, CM_{22}, CM_{21}, CM_{12} denote the number of true positives (TPs), true negatives (TNs), false positives (FPs), and false negatives (FNs), respectively.

Sokolova and Lapalme provide an extensive overview of different performance measures for classification tasks [162]. A commonly used performance measure is the *accuracy* of a classifier, which is defined as the total number of the correctly classified samples divided by the total number of samples:

$$ACC = \frac{\sum_{k=1}^{K} CM_{kk}}{S}. \tag{3.2}$$

We compare the accuracy achieved by the five classifiers considered in this study with the accuracy of two random classifiers. The first is a random guess (RG) classifier, which randomly selects a class assuming equiprobable classes. This classifier achieves an accuracy of

$$ACC_{RG} = \frac{1}{K}. \tag{3.3}$$

To account for the fact that classes are not always equiprobable, we also consider a biased random guess (BRG) classifier. The BRG classifier uses knowledge of the proportion of samples of each class in the training data to perform a biased random decision. The accuracy obtained by the BRG classifier is

$$ACC_{BRG} = \sum_{k=1}^{K} (\frac{S_k}{S})^2, \tag{3.4}$$

where S_k denotes the number of samples of class k.

The accuracy measure treats all classes equally and is often a weak measure when dealing with imbalanced classes [120, 162]. For this reason, we also utilize the Matthews correlation coefficient (MCC) to quantify the performance of the considered classifiers [120]. The MCC ranges between -1 and 1, where 1 represents a perfect classification, 0 denotes a classification that is no better than a random classification, and -1 indicates complete disagreement between classification and observation. In case $K = 2$, the MCC is computed as

$$MCC = \frac{TP * TN - FP * FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}. \tag{3.5}$$

For $K > 2$, we use the generalization of the MCC to multi-class classifications as presented by Gorodkin in [80].

To evaluate the performance of the multiple linear regression, we first obtain the estimate $\hat{y}_j$ for each household j as

$$\hat{y}_j = \beta_0 + \beta^T x_j, \tag{3.6}$$

using the parameters β and the feature vector x_j. We then compare the estimation with the ground truth data y_j by computing the *coefficient of determination* (R^2) as a performance measure [145]:

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}}. \tag{3.7}$$

R^2 ranges between 0 and 1 and denotes the proportion of the variance of the estimation error

$$SS_{res} = \sum_j (y_j - \hat{y}_j)^2 \tag{3.8}$$

to the variance of the ground truth data

$$SS_{tot} = \sum_j (y_j - \bar{y}_j)^2. \tag{3.9}$$

We further compute the out-of-sample root-mean-square error (RMSE) to evaluate the deviation of the estimation $\hat{y}_j$ to the ground truth data y_j for each of the household characteristics [145].

3.3.2 Training, evaluation, and feature selection

Listing 3.1 illustrates the training and evaluation procedure we apply to reveal household characteristics from electricity consumption data. As input, we use a single week of consumption data for all households, which we divide into four disjoint subsets. One subset is used for training the classifiers, the others to validate their performance using a 4-fold cross validation. The performance metrics of interest (accuracy, MCC) are thus computed for each week of the data. In each fold of the cross-validation, a feature selection algorithm determines a subset of the features defined in section 3.2.1, which is then used by the classifiers. The output of the classifiers is used along with ground truth data to compute the confusion matrix for each classifier. The matrix is then in turn used to compute the performance measures described above. The performance measures are the only difference in this process when performing regression instead of classification.

As line 12 in listing 3.1 indicates, we rely on the feature selection method sequential floating forward selection (SFFS) [140]) to determine a suitable set of features $\bar{F} \subseteq F, \bar{F} = \cup_{i=1}^{|F|} c_i f_i, c_i = \{0,1\}$, where f_i is the i-th feature in F, $|F|$ denotes the size of F, and $c_i = 1$ indicates membership of f_i in $\bar{F}$. There is an optimal set of features F_{opt}, with which a classifier achieves the best value for a specific performance measure (*e.g.*, the highest accuracy). Since F_{opt} typically differs from F [181], feature selection methods approximate F_{opt} by iteratively running the classification (or regression) using different subsets of features. This type of feature selection is called a wrapper feature selector: Feature selection wraps around the classifiers, taking the classifier into the process and therefore taking into account its bias when selecting features [84]. In contrast to filter selectors, which identify a feature set without running the classifier (*e.g.*, by analyzing the correlation between individual features), wrapper selectors consume more run time since the classification runs multiple times depending on the strategy to identify the feature set. After each run these methods compute a *figure of merit*, which can be any of the performance measures described in the previous section. There are different strategies to maximize the figure of merit and thus optimize the feature set. SFFS is a method that starts with an empty set and consecutively adds the feature to the set that allows one to achieve the highest improvement of the figure of merit. In each step, SFFS also considers removing one or more features from the set, since removing a feature that has been added previously and adding a different one might lead to an increase of the figure of merit. We perform feature selection on the training set as described above. Since the feature selection itself requires both training and test data, we perform another cross-validation on the three subsets of the training set $D \setminus D_i$ (see listing 3.1).

Our implementation of SFFS relies on the code provided by the authors of [170]. As an improvement to this existing implementation, we also make the SFFS maintain a logbook of the states it reaches—where a state is represented by a (sub)set of features—in order to prevent it entering infinite loops. To limit the number of iterations and to avoid overfitting, we restrict the removal of features as follows: Assume feature f is added to state s and state $s' = s \cap f$ is reached. A feature $f' \neq f$ is them removed from s' only if the figure of merit of $s'' = s' \setminus f'$ is more than a threshold T higher than the figure of merit of s. We set $T = 0.005$ because, based on our experiments, differences of less than 0.005 in the figure of merit of two states are often due to random effects and thus do not necessarily imply a significant improvement of s'' over s. In this way we avoid overfitting, which could lead to reduced performance when finally using $\bar{F}$ on a new set of data (which we do in lines 13 and 14 of listing 3.1). Similarly, we limit the number of features selected by SFFS to $|\bar{F}| <= 8$, because adding more than eight features to $\bar{F}$ often leads to overfitting the feature set to the training data in our experiments.

```
1  In:  Consumption data D
2       Classifier Cl
3       Characteristic P, classes C = {C_1,...,C_n}, n: #classes in P
4  Out: Accuracy ACC_{P,Cl}
5       Matthews Correlation Coefficient MCC_{P,Cl}
6       Posterior probabilities P(C_j|D)_{P,Cl} ∀ C_j ∈ C
7  begin
8       Divide D into disjoint subsets: D = {D_1,D_2,D_3,D_4}
9       for each fold i = 1,...,4 do
10           test_set = D_i
11           training_set = D\D_i
12           F̄ = sffs(training_set)
13           model = train(training_set, F̄, Cl)
14           P(C|D_i) = test(model, test_set, F̄, Cl)
15           CM_i = compute confusion matrix from P(C|D_i) and D_i
16      end for
17      P(C|D)_{P,Cl} = {P(C|D_1),...,P(C|D_4)}
18      ACC_{P,Cl} = ACC(CM_1,CM_2,CM_3,CM_4)
19      MCC_{P,Cl} = MCC(CM_1,CM_2,CM_3,CM_4)
20 end
```

Listing 3.1: Evaluation process that is performed for each classifier and characteristic.

3.3.3 Implementation details

We implemented our system in MATLAB [218]. All results presented in the following section are obtained by performing independent experiments for each of the characteristics listed in table 3.4, each of the five classifiers (kNN, LDA, Mahalanobis, SVM, and AdaBoost) described in section 3.2.3, and two different performance measures (accuracy and MCC). For regression, we employ multiple linear regression and use the adjusted R^2 score as a figure of merit. To perform this large number of experiments we utilized ETH Zurich's computing cluster *Brutus* [202] and ran all experiments in parallel on different machines. We use the kNN, LDA, Mahalanobis, and AdaBoost classifiers from MATLAB's Statistics toolbox [219]. For kNN, we choose $k = 5$ and use the Euclidian distance as the distance metric. In case of AdaBoost, we use the AdaBoostM1 and AdaBoostM2 learners to classify characteristics with two classes and more than two classes, respectively. As for the SVM classifier, we rely on the publicly available implementation LIBSVM [195] with a radial basis function kernel.

3.4 Results

To quantify the performance of our system, we first consider a single week of data in the CER data set (week 26). We then repeat our analysis for each week of data, showing that these results are consistent irrespective of which week of data is used. Week 26 was chosen as an "exemplary" week since (1) there is no holiday during this week, (2) the week is not during vacation, and (3) it includes data from households that joined the trial in early 2010. As we show later in this section, experiments on all other weeks of the trial have shown no significant impact on the performance of our approach (2% standard deviation per characteristic on average). However, by combining the classification results over multiple weeks, the accuracy and the MCC can be improved on average by three and six percentage points, respectively, compared to a single week analysis.

3.4.1 Accuracy

Figure 3.13 shows the accuracy achieved by our system and the two baseline classifiers random guess (RG) and biased random guess (BRG) when estimating the 18 household characteristics of interest. For each characteristic, ACC_{C^*} denotes the highest accuracy value among each of the five classifiers kNN, LDA, Mahalanobis, SVM, and AdaBoost. Showing the accuracy of the best performing classifier allows us to outline the accuracy that can be obtained in principle when using our system for estimating household

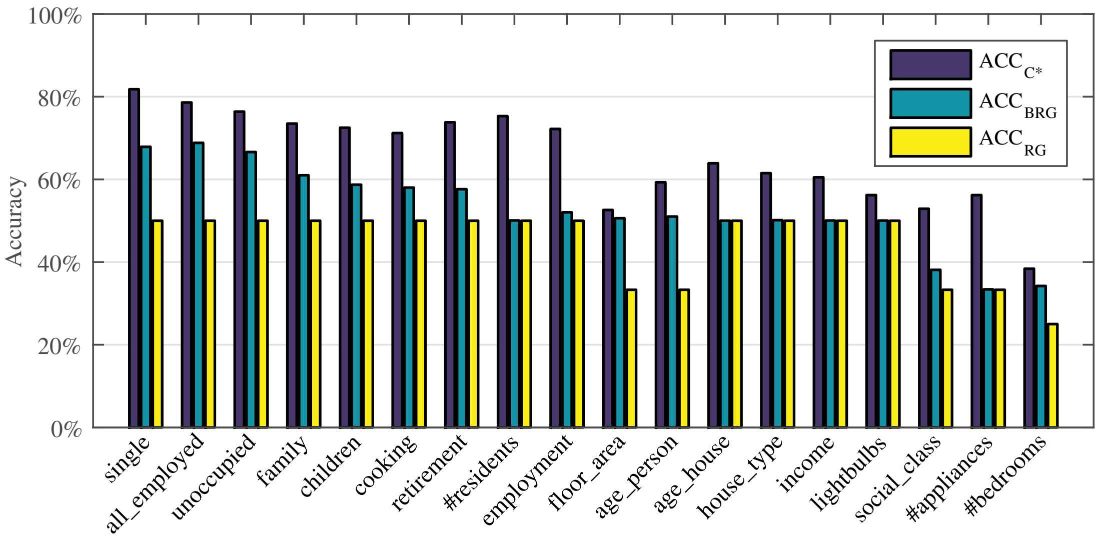


Figure 3.13: Accuracy of our system (ACC_{C^*}) compared to the random guess (ACC_{RG}) and biased random guess (ACC_{BRG}) classifiers based on week 26 of the trial.

characteristics. We leave it to our future work to explore solutions that allow us to maximize the performance of our system in practical scenarios. The results in figure 3.13 are obtained using the accuracy as a figure of merit during feature selection and without undersampling (*i.e.*, the classifiers have knowledge of the class distribution from the training data). The graph compares the accuracy of our system (ACC_{C^*}, left bars) with the accuracy of the biased random guess classifier (BRG, center bars) and the random guess classifier (RG, right bars). The characteristics on the x-axis are those listed in table 3.4 in section 3.2.2. Among the 18 characteristics, there are four three-class characteristics (`age_house`, `#appliances`, `social_class`, `floor_area`) and one four-class characteristic (`#bedrooms`). The remaining characteristics are two-class characteristics. For these characteristics, the accuracy of the RG classifier is 33%, 25%, and 50%, respectively. The accuracy of the BRG classifier is computed using equation 3.4 and the number of samples per class as listed in table 3.4. If the classes are balanced, ACC_{BRG} is equal to ACC_{RG}. Figure 3.13 shows that ACC_{C^*} exceeds both ACC_{RG} and ACC_{BRG} for all of the 18 characteristics. ACC_{C^*} is 4.0 to 33.8 percentage points higher than ACC_{RG} and 4.0 to 22.3 percentage points higher than ACC_{BRG}. Our system thus outperforms RG and BRG by 22.0 percentage points and 14.2 percentage points, respectively. In terms of individual characteristics, our approach achieves more than 80% accuracy for characteristics `single`, `all_employed`, and `unoccupied`. The worst accuracy of ACC_{C^*} compared to ACC_{BRG} is achieved when estimating the proportion of energy-efficient light bulbs (`lightbulbs`) in a household. In this case, ACC_{C^*} exceeds ACC_{BRG} by only 4.0 percentage points. We believe this results from the fact that the mere number of lightbulbs is not reflected in the electricity consumption. We expect this result to improve when classifying the actual usage of energy-efficient light bulbs. However, this data is not available in the CER data set.

Figure 3.14 illustrates the accuracy of all five classifiers. For most of the characteristics, the difference between the highest and lowest accuracy is ten percentage points or less. The SVM classifier achieves the highest accuracy for 13 of the 18 characteristics. The AdaBoost classifier performs one percentage point worse than the SVM classifier on average. The LDA and Mahalanobis classifiers show similar performance for some of the characteristics but have a low accuracy for the characteristics with imbalanced classes (*e.g.*, `floor_area`, `#bedrooms`). Among the considered classifiers, SVM is thus the one providing the overall best performance in terms of accuracy. If our system were to be used to maximize the estimation accuracy, we would thus use SVM as the default classifier. We leave it to our future work to verify whether this consideration can be generalized to other data sets.

3.4.2 Matthews correlation coefficient

As described in section 3.3.1, the MCC is typically a more suitable performance measure than accuracy when classes are imbalanced. The MCC "rewards" true positives of the underrepresented classes and thus "punishes" classifiers that bias their model too strongly towards the overrepresented classes. It is a performance measure that ranges between -1 (*i.e.*, total disagreement between the ground truth data and the estimation) and 1 (total agreement), whereas the MCC is 0 for random estimations [11]. Figure 3.15 shows the MCC for each of the classifiers considered in our system. In these experiments, we used the MCC as a figure of merit during feature selection and use undersampling to prevent classifiers from biasing their model towards the overrepresented classes. The plot shows that for characteristics related to the number of people in a household (*i.e.*, `single`, `#residents`, `family`, `children`), the considered classifiers achieve high values of MCC (up to 0.459). Figure 3.15 further shows that the MCC is also high (up to 0.346) for characteristics related to occupancy (*i.e.*, `employment`, `all_employed`, `unoccupied`, `retirement`). Finally, classifying characteristic `#appliances` provides an MCC of 0.31. The results indicate that for these characteristics, classification is feasible. Whether the results are good enough to provide energy efficiency services must be decided on a per-application basis.

With respect to the individual classifier performance, the kNN classifier performs worse than the LDA, Mahalanobis, SVM, and AdaBoost classifiers across (almost) all of the characteristics. Among the other four classifiers, neither of the four classifiers' performances dominates over all characteristics, which makes the choice of classifier dependent

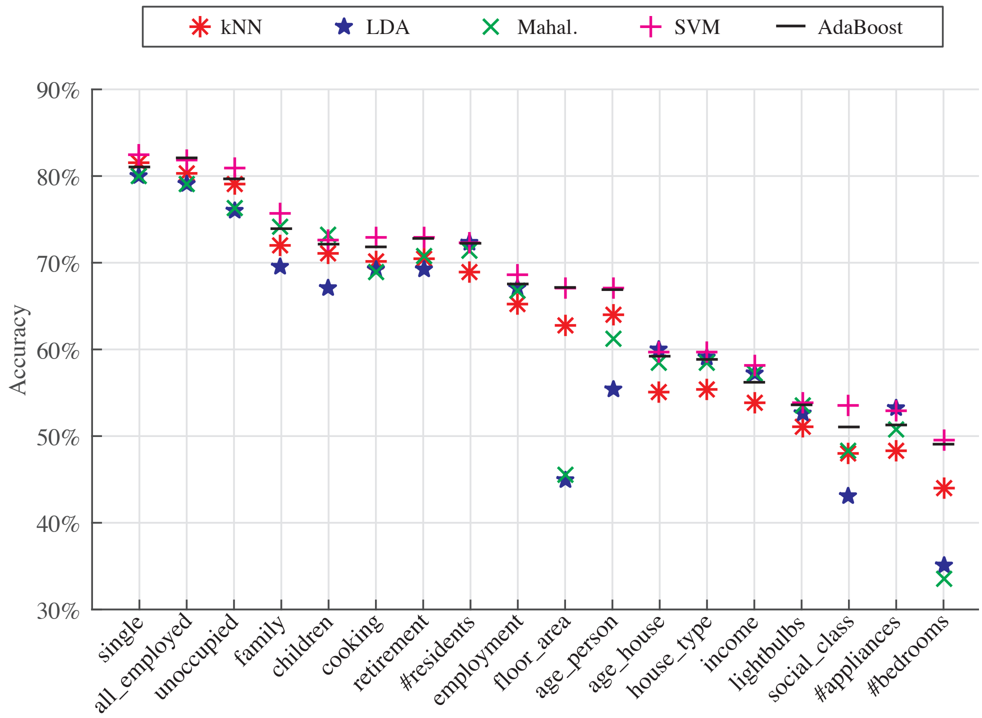


Figure 3.14: Accuracy of different classifiers for all household characteristics. The classifiers are trained using data from week 26 of the CER data set.

on the characteristic to be classified. We thus argue for the use of a comprehensive system that uses a particular classifier depending on the specific characteristic to be estimated.

3.4.3 Regression

The regression model described in section 3.2.4 allows us to estimate continuous values of selected characteristics instead of assigning the household to pre-defined, discrete classes. Figure 3.16 shows the results for characteristics `age_person`, `#bedrooms`, `#appliances`, `floor_area`, `income`, `#residents`. The x-axis of each plot shows the ground truth values for the characteristics. The reason for choosing a box plot instead of a scatter plot for some of the characteristics is that the ground truth data is binned (*e.g.*, `age_person`, `income`) or provides only a few discrete values (*e.g.*, `#bedrooms`, `#residents`). The y-axis shows the estimation obtained by applying f_R from equation 3.1 to the features of the test data. The subplot at the bottom right, for instance, shows one box per group of households with 1, 2, 3, 4, or 5+ (5 or more) residents. The red lines denote the median of the estimated number of residents for each of the five groups, and the top and bottom ends of the boxes denote the 25th and 75th percentiles, respectively. For the characteristic `floor_area`, each value is plotted individually in a scatter plot.

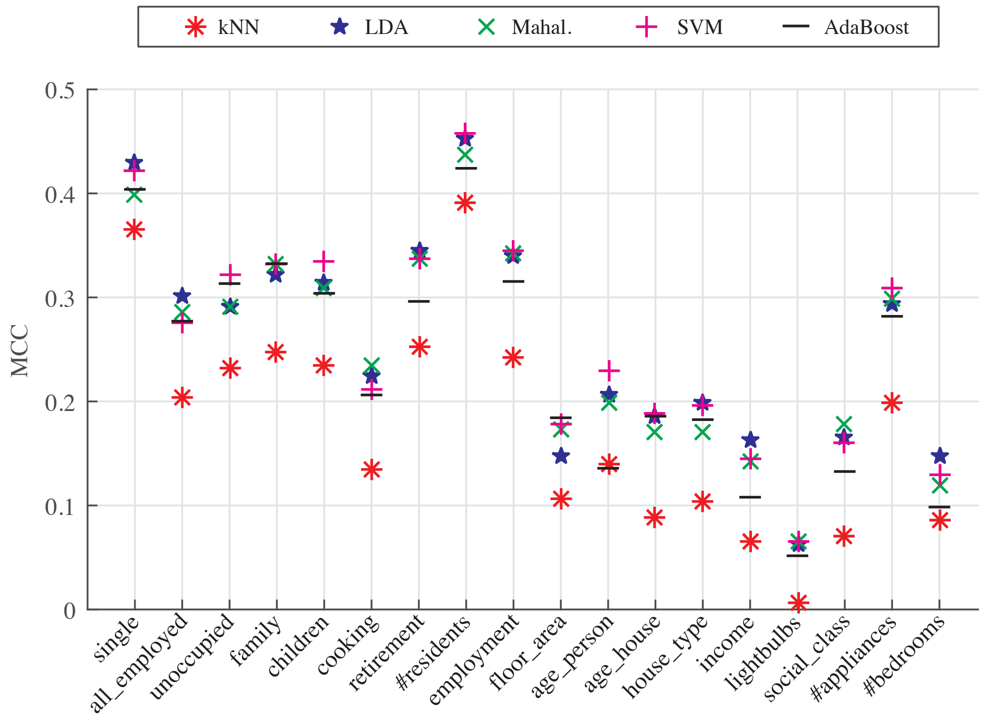


Figure 3.15: MCC of different classifiers for all household characteristics. The classifiers are trained using data from week 26 of the CER data set.

For each of the characteristics, the figure shows the RMSE and the coefficient of determination (R^2). The latter one ranges between -1 and 1, whereas 0 shows no correlation between the estimated and the ground truth data and 1 indicates a perfect estimation. The characteristics `#residents` and `#appliances` achieve the highest R^2 with 0.30 and 0.29, respectively. The characteristics `#bedrooms`, `age_person`, and `floor_area` follow suit with 0.14, 0.17, and 0.14, respectively. Finally, `income` is hardest to reveal with a very low R^2 of 0.083. Although the plots show a clear correlation between the estimated and the actual values, the R^2 score is overall relatively low. We assume this is due to the fact that the linear regression model is very sensitive to outliers. Examples of such outliers are households that have their ground truth incorrectly specified in the questionnaires. These results suggest that utilities should rely on the estimated class rather than striving for exact, continuous values. As a part of our future work, we aim at improving the R^2 scores and thus the applicability of the regression analysis for utilities by automatically identifying households or groups of households that negatively affect the performance of the analysis.

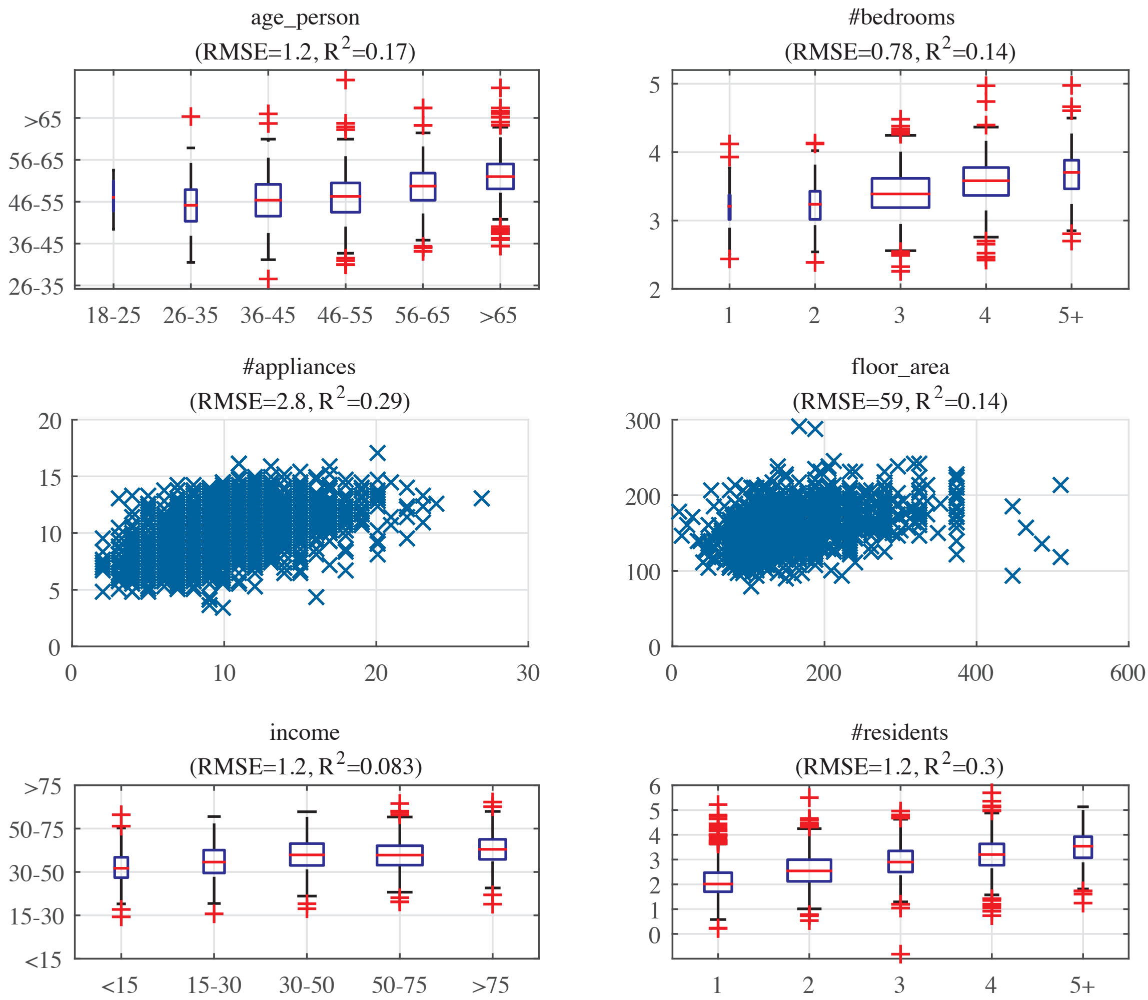


Figure 3.16: Regression analysis for selected characteristics and consumption data of week 26.

3.4.4 Stability of the results

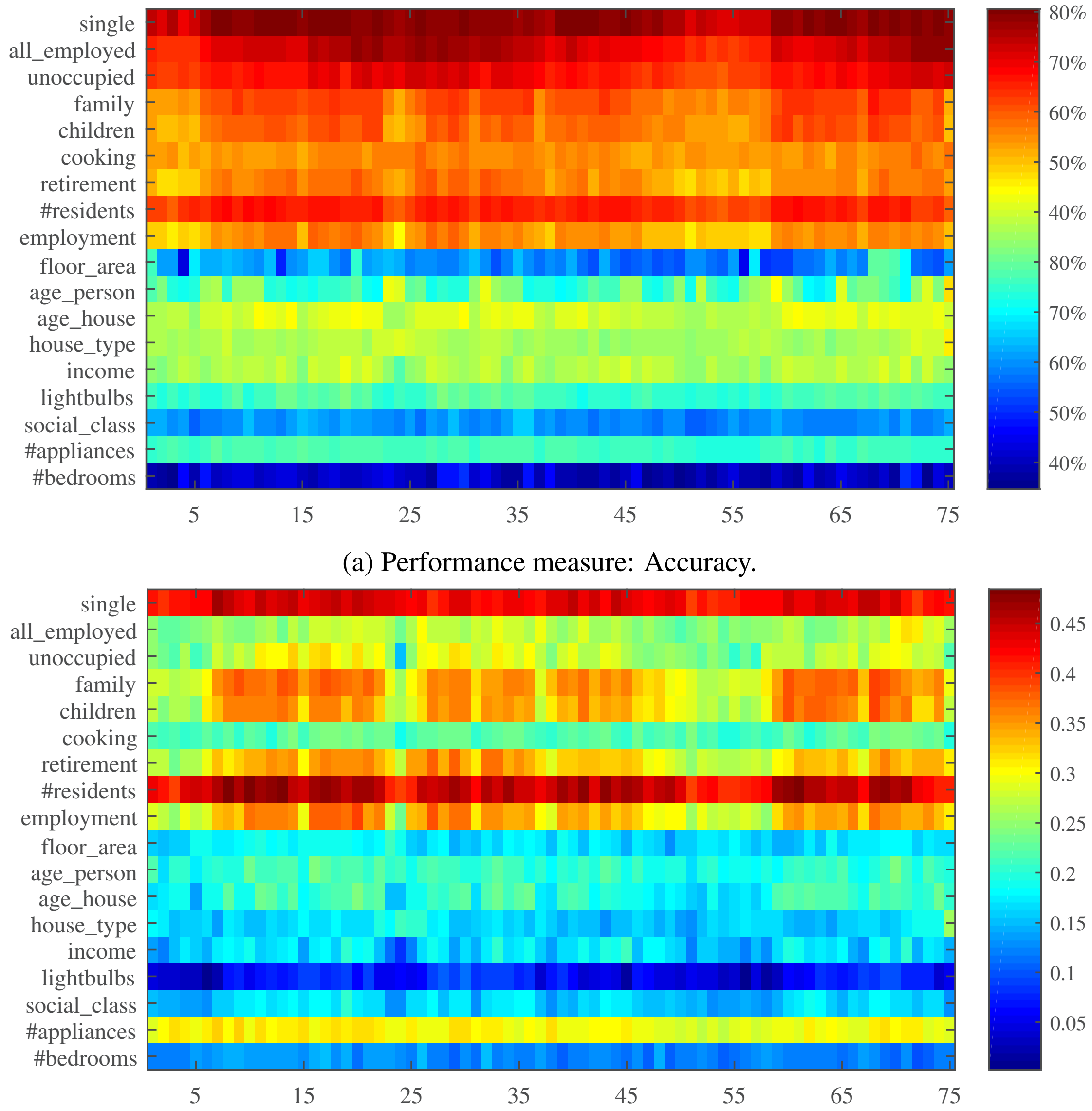
The results presented thus far are based on the analysis of a single week of consumption data (week 26). We then extended the experiments to the whole data set by classifying all of the characteristics on each week of the study, following the procedure described in listing 3.1. This means we evaluate each week separately by computing the features on this particular week and (as before) using different households for training and testing. For these experiments, we rely on the LDA classifier only, because it is the fastest of the five classifiers evaluated in this section. We assume the results can be generalized to the other classifiers, which is subject to future testing. Figure 3.17 shows the results of these experiments. The left plot illustrates the accuracy for each characteristic on each week encoded with colors ranging from dark blue (30% accuracy) to dark red (80% accuracy).

Similarly, the right plot shows the MCC ranging from dark blue (0) to dark red (0.5). The plots show that the difference in performance between weeks is relatively low for most of the characteristics: In terms of accuracy, the average standard deviation for all characteristics is 0.017. Only the characteristics `floor_area` and `#appliances` exhibit large variations with a standard deviation of 0.034 and 0.033, respectively. For these two characteristics, for instance, the difference between minimum and maximum accuracy is 13.7% and 18%, respectively, whereas for all other characteristics this difference is 6% on average. For the MCC, the standard deviation is below 0.04 for all characteristics. An interesting observation can be made for weeks 50 to 57 of the trial, where the classification performs slightly worse than for the rest of the weeks. This is particularly true for characteristics that are related to the number of persons in the household (*e.g.*, `single`, `#residents`) or related to the presence of people (*e.g.*, `retirement`, `employment`). We believe that this loss in performance is due to the fact that these weeks represent summer vacation in Ireland. Thus, in these weeks we observe non-usual consumption patterns, which cause the classification to perform less reliably than in other weeks.

The experimental results reported above show that it is possible to reliably extract household characteristics from electricity consumption data irrespective of the specific week of data used for the analysis. This implies that utilities might need to collect as little as one week of fine-grained data (*i.e.*, one sample every 30 minutes) to be able to extract household characteristics. In future work, we plan to investigate whether the results are also promising when we train the classifiers using a specific week of data and classify (different) households using a different week of data. This would allow utilities to include new households in the analysis (*e.g.*, households with a newly installed smart meter) without retraining the classifiers. Next, the stability of the results shows that utilities can build weekly customer profiles using our approach. For each customer, such a profile can for instance show when a family grows or goes on vacation. The fact that results are stable across multiple weeks further enables us to identify "atypical" weeks for individual households as outliers (*e.g.*, when the house is unoccupied because the family is on vacation). In the following section, we show that running our analysis on multiple weeks allows us to compensate for such outliers and improve the performance of the classification.

3.4.5 Multiple weeks

In this section, we combine the results of the classifiers over multiple weeks. We train the classifiers for each week separately using the households in the training set. Next, we estimate—for each household in the test set—the class of the household for each week and then assign the household to the class C that was estimated most often. We decided



(a) Performance measure: Accuracy.

(b) Performance measure: MCC.

Figure 3.17: Classification results for each of the characteristics on all 75 weeks of the trial using the LDA classifier.

for this majority vote for its simplicity; we leave alternative methods of performing the analysis on multiple weeks (*e.g.*, computing the features over a longer time period) for future work. Ultimately, we create a confusion matrix for each of the characteristics by comparing C with the actual ground truth. As the number of households varies over the weeks due to missing meter readings for some households, we only consider households in the analysis for which data over more than 50 weeks is available. Table 3.5 lists both accuracy and MCC of the classification using week 26 only (as described in detail above) and the results obtained when running the analysis over the whole period of the trial (*i.e.*, 75 weeks). The table shows that the accuracy increases by up to ten percentage points (three percentage points on average) and the MCC by up to ten percentage points (six percentage points on average). The performance of the classification only decreases for characteristic `house_type`; however this change is very low (one percentage point for the MCC). We thus argue that, in practical scenarios, utilities interested in maximizing classification performance should utilize several weeks of data to estimate household characteristics. However, the computational effort increases linearly with the number of weeks. It is a part of our future work to investigate this trade-off in more detail.

Table 3.5: Accuracy and MCC for each characteristic obtained by assigning to each household in the test set the majority of classifications over the whole trial. The results are based on the LDA classifier.

	Accuracy		**MCC**	
Characteristic	**Week 26**	**All weeks**	**Week 26**	**All weeks**
single	80%	82%	0.43	0.5
all_employed	79%	79%	0.3	0.32
unoccupied	76%	76%	0.29	0.38
family	69%	74%	0.32	0.42
children	67%	73%	0.31	0.41
cooking	69%	71%	0.22	0.29
retirement	69%	73%	0.35	0.43
#residents	72%	76%	0.45	0.51
employment	67%	72%	0.34	0.44
floor_area	45%	50%	0.15	0.21
age_person	55%	59%	0.21	0.3
age_house	60%	64%	0.19	0.28
house_type	59%	59%	0.2	0.19
income	57%	61%	0.16	0.23
lightbulbs	52%	55%	0.062	0.098
social_class	43%	53%	0.17	0.22
#appliances	53%	56%	0.29	0.34
#bedrooms	35%	39%	0.15	0.15
Mean	**62%**	**65%**	**0.26**	**0.32**

3.4.6 Discussion and limitations

The experimental results presented thus far show that three types of characteristics can be inferred particularly well from electricity consumption data. These are characteristics that reflect the occupancy state of the house (*e.g.*, `employment`, `unoccupied`), the number of persons in the house (*e.g.*, `single`, `#residents`, `family`), and the appliance stock (`#appliances`). On the other hand, characteristics related to the dwelling itself (*e.g.*, `floor_area`, `#bedrooms`) are more difficult to extract from electricity consumption data. This is due to the fact that heating and cooling only plays a minor role in the consumption data available for this study. The results show that the income of a household is also difficult to infer from electricity consumption data.

It is in general important to note that the results presented in this section might be affected by inaccuracies in the ground truth data. Questionnaire answers given by the participants in the CBT can be wrong, ambiguous, or based on estimations. For characteristic `all_employed`, for example, the questionnaires do not specify full-time or part-time employment. Characteristic `unoccupied` relies on the estimated absence rather than on actual measurements. For characteristic `income`, the process of extracting well-separated classes from the ground truth data was difficult due to the complex structure of questions that captures the income of the respondents. For instance, they could specify their income on a yearly or monthly basis as well as before or after tax.

A major challenge in applying this work in practice is to collect reliable ground truth data. This step typically requires surveys, which are costly and cumbersome to perform. Yet, even if only a small percentage of customers reply [165], this small amount can be used to train the classifiers and estimate the characteristics of the remaining households. There is also the possibility to use ground truth data from a different data set (*e.g.*, collected from a project performed in a different geographic region).

In our work, we evaluated each of the five classifiers and then decided which one performs best. To implement the approach in real scenarios, however, we must decide on the classifier on the basis of the training data only. Similarly to what we do for feature selection, we propose dividing the training data into two sets and using one of the sets to train the classifiers and the other set to evaluate their performance. This process also allows for fine-tuning of each of the classifiers for each of the characteristics, while in this thesis we only provide an overview of the potential of large-scale electricity consumption data analysis.

Performing the analysis on one week of consumption data leads to results that might be sufficient for utilities to estimate household characteristics. Further, extending the analysis to data collected during the whole trial improves on these results. One reason for this is that weeks with irregular consumption patterns (*i.e.*, outliers) can be overruled by the results obtained using data from other weeks.

3.5 Impact of outdoor temperature and daylight

The analysis presented thus far focusses on electricity consumption data as the single source of input to predict household characteristics. Weather information such as the outside temperature has a significant effect on a household's electricity consumption [2, 3, 28, 91]. In this section, we investigate the correlation between electricity consumption and weather and evaluate if it can be utilized to improve the performance of our classification.

Since the CER data set does not contain weather information, we extracted temperature readings from online weather data provided by *WeatherOnline* [235]. For each day, WeatherOnline provides weather parameters at a 30-minute granularity. These include wind speed, temperature, humidity/visibility, precipitation, clouds, and air pressure. WeatherOnline's data has been collected from different weather stations across Ireland (where not all stations provide data at all times). Since the CER data set does not provide information on the location of the dwellings of the participating households, we selected three weather stations across the country that provided weather information for almost all the 30-minute time slots during the trial. These stations are: Cork-Corcaigh, located in the south of Ireland at 162 m altitude; Dublin Airport, located in the East of Ireland at 85 m altitude; and Shannon, located in the West of the country at 20 m altitude. The distance between the three stations is approximately 200 km. From these three weather stations we extracted semi-hourly temperature readings and used their average for the rest of the analysis.

Figure 3.18 shows the correlation between electricity consumption and outside temperature. Each of the 525 points on the plot[2] corresponds to on the y-axis the average electricity consumption of all 4231 households on a particular day and on the x-axis the average outside temperature of that day. The red line shows the regression line computed using the OLS method [145]. The plot shows that electricity consumption increases as the outside temperature decreases. This is in part due to the fact that heating systems consume more electricity when it is cold outside. However, for the households in the CER data set, most of the dwellings are heated using fossil fuels, while only a small percentage of households uses a central electric heating system or electric plug-in heater (*cf.* figure 3.5a). Besides electricity consumed by space heating, we believe that lifestyle effects also contribute to the increase in electricity consumption when temperature decreases. For instance, people

[2] We included 525 days in our analysis, which corresponds to 75 weeks from 20 July 2009 to 26 December 2010.

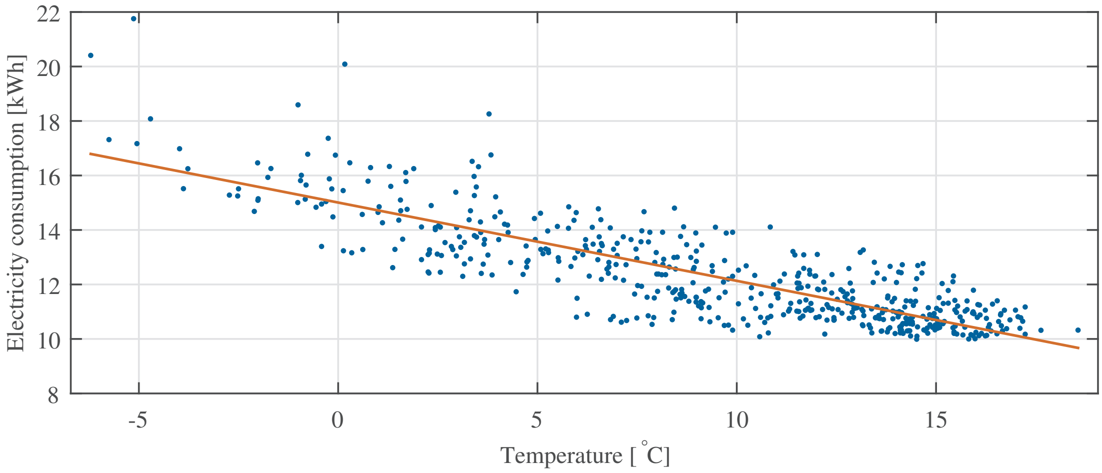


Figure 3.18: Correlation between electricity consumption (y-axis) and outside temperature (x-axis). Each dot represents the daily electricity consumption averaged over all households and the average outside temperature on that particular day.

spend more time at home—and thus consume more electricity—when it is cold or dark outside such as it is on winter days.

In addition to the sensitivity to outside temperature, we investigate whether the correlation between electricity consumption and daylight is a useful feature when classifying households by their characteristics. To this end, we extracted the times of sunrise and sunset for each day of the trial using the Astronomy API from *timeanddate.com* [231].

To extract the sensitivity to outside temperature and daylight, we use the linear regression model

$$y_{jt} = \alpha_j + \sum_{k=1}^{3} \beta_j^{(k)} W_t + \sum_{k=1}^{24} \gamma_{jk} \mathbb{1}\{\text{ToD}(t) = k\} + \sum_{k=1}^{2} \delta_{jk} \mathbb{1}\{\text{Weekday}(t) = k\} + \varepsilon_{jt}, \quad (3.10)$$

where y_{jt} represents the j-th household's electricity consumption at time $t \in T$, $T = 1, \ldots, 25200$. T thus represents all 30-minute time slots during the 75 weeks of the trial. The coefficient α_j denotes a constant term for each household, γ and δ are dummy variables for time of day (in hours) and the fact whether t is on a weekday or on the weekend, respectively. The coefficient ε represents the error term.

The three coefficients $\beta^{(1)}$, $\beta^{(2)}$, and $\beta^{(3)}$ account for the sensitivity to sunrise, sunset, and temperature, respectively. These values are incorporated in W_t, given by

$$W_t = \begin{pmatrix} \text{Sunrise}(t) \\ \text{Sunset}(t) \\ \text{Temperature}(t) \end{pmatrix}, t \in T. \quad (3.11)$$

Temperature(t) denotes the outdoor temperature at time t, which we compute on the data collected by the weather stations. Further, we determine in which time slot of the day sunrise and sunset occur. Using the number of minutes between midnight and sunrise (`min_sunrise`) and the number of minutes between midnight and sunset (`min_sunset`) from the astrology data, we compute

$$\texttt{sunrise\_slot} = \frac{\texttt{min\_sunrise}}{30} \tag{3.12}$$

and

$$\texttt{sunset\_slot} = \frac{\texttt{min\_sunset}}{30}. \tag{3.13}$$

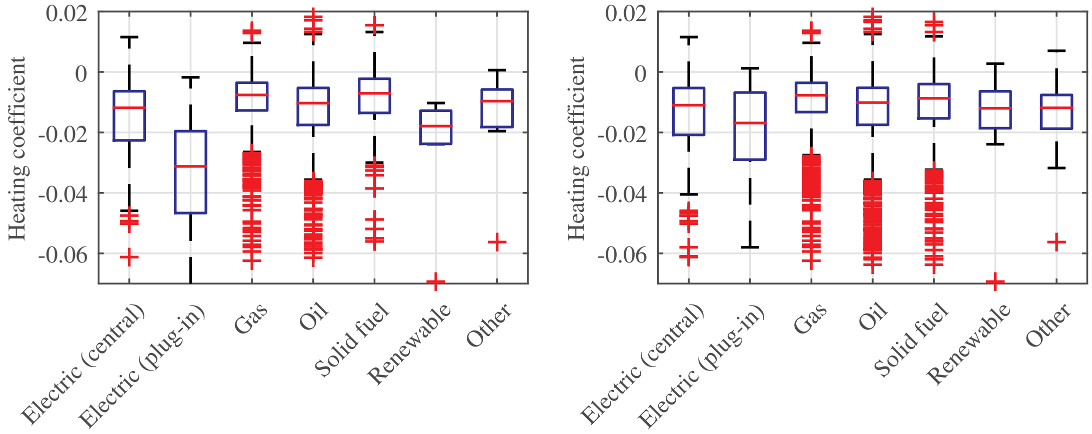
We then set Sunrise(t) and Sunset(t) to 1 if t lies within the next hour of the sunrise or within the hour that precedes sunset, respectively:

$$\text{Sunrise}(t) = \begin{cases} 1 & \text{if } t \pmod{\lceil \texttt{sunrise\_slot} \rceil + 1} = 1 \vee t \pmod{\lceil \texttt{sunrise\_slot} \rceil + 2} = 1, \\ 0 & \text{otherwise.} \end{cases} \tag{3.14}$$

$$\text{Sunset}(t) = \begin{cases} 1 & \text{if } t \pmod{\lceil \texttt{sunset\_slot} \rceil} = 1 \vee t \pmod{\lceil \texttt{sunset\_slot} \rceil - 1} = 1, \\ 0 & \text{otherwise.} \end{cases} \tag{3.15}$$

For instance, if sunrise is at 5:15 a.m. and sunset at 9:15 p.m., the 12th and the 13th time slots of the day indicate sunrise (*i.e.*, from 5:30 a.m. to 6:30 a.m.), and the 42th and 43rd time slots of the day (*i.e.*, from 8:30 p.m. to 9:30 p.m.) indicate sunset. By introducing dummy variables for time of day and weekday/weekend we ensure that we determine the effect of sunrise, sunset, and outdoor temperature on electricity consumption data using comparable time periods.

Figure 3.19 shows a box plot of the temperature coefficients $\beta^{(3)}$ for each household obtained from the regression analysis. The left plot only includes those households that specified a single heat source in the questionnaire. It shows the strongest correlation between electricity consumption and outdoor temperature for those households that have an electric plug-in heater installed in their household. The second strongest correlation is shown for those households that use renewable energy sources such as solar panels to heat their home. A potential reason for this is that those households generate most electricity during warm, summer days. The effect for households that heat with gas, oil, or solid fuels is also visible, although lower than for those households with electric heating. Most



(a) Only households with a single heat source. (b) All households including those with multiple heat sources.

Figure 3.19: Temperature coefficients obtained from the regression analysis categorized by the type of space heating installed in the household.

likely the sensitivity to temperature for those households that heat with fossil fuels stems from the fact that their lifestyle depends on the outdoor temperature. In the right plot, we show the temperature coefficients for all households including those that have multiple heat sources according to the answers to the questionnaires. The effects described above are still visible. However, they are much lower due to the fact that the data set does not specify which of the heat sources is mainly used in a household. For instance, a household that mostly heats by gas but also owns a (rarely used) plug-in heater contributes to both the second and the third column of the plot, evening out effects of households that only use a plug-in heater. In fact, only 108 households do not heat their house using oil, gas, or solid fuels, and from those households, 76 households rely on central electric heating. Only 16 households heat their house using a plug-in heater only. Overall, the analysis shows that energy consumed by space heating—at least as the primary heat source—is barely accounted for in the electricity consumption data of the CER data set.

To evaluate the use of weather coefficients for household classification, we run the analysis described in section 3.4.5 and include $\beta^{(1)}$, $\beta^{(2)}$, and $\beta^{(3)}$ as additional input features. In particular, we run the analysis two times for each of the 75 weeks in the data set using the LDA classifier. The LDA classifier achieved good results on a single week of data and it has low computational requirements, which facilitates performing the experiments on the whole data set. In the first run we use accuracy as figure of merit during feature selection. The second time we run the analysis using MCC as figure of merit instead. Table 3.6 shows the results of the analysis and compares it to the results obtained in section 3.4.5, which are based on the same analysis except that we did not include the weather coefficient as input features. The left side of the table shows the

Table 3.6: Results obtained by classifying the 4231 households using the features from the previous analysis (*Original*) and the new feature set including daylight and temperature coefficients (+ *Weather*). Again, the results are based on the LDA classifier and obtained by assigning to each household in the test set the majority of classifications over the 75 weeks. Column *Classes* denotes the number of classes for each characteristic.

		Accuracy			MCC		
Characteristic	**Classes**	**Original**	**+ Weather**	**Change**	**Original**	**+ Weather**	**Change**
single	2	82%	81.8%	-0.15	0.495	0.503	+0.0077
all_employed	2	78.6%	78.6%	0	0.324	0.313	-0.011
unoccupied	2	76.4%	76.4%	-0.025	0.376	0.38	+0.0039
family	2	73.7%	73.5%	-0.22	0.419	0.415	-0.0043
children	2	72.8%	72.5%	-0.35	0.408	0.407	-0.0015
cooking	2	71.2%	71.2%	0	0.286	0.29	+0.0041
retirement	2	73.5%	73.8%	0	0.43	0.427	-0.0036
#residents	2	75.5%	75.3%	-0.2	0.508	0.513	+0.0045
employment	2	72.3%	72.2%	-0.12	0.436	0.443	+0.0079
floor_area	3	50.5%	52.6%	**+2.1**	0.205	0.197	-0.0085
age_person	3	58.6%	59.3%	0	0.295	0.297	+0.0018
age_house	2	63.7%	63.9%	0	0.278	0.277	-0.0014
house_type	2	59.3%	61.5%	**+2.3**	0.191	0.236	**+0.046**
income	2	61.1%	60.5%	-0.54	0.229	0.214	-0.015
lightbulbs	2	55.1%	56.2%	**+1.1**	0.0976	0.116	**+0.018**
social_class	3	52.9%	52.9%	0	0.225	0.227	+0.0016
#appliances	3	55.7%	56.2%	0	0.343	0.354	**+0.011**
#bedrooms	4	38.7%	38.4%	-0.27	0.151	0.155	+0.0039
Mean		**65.1%**	**65.4%**	**+0.29**	**0.317**	**0.32**	**+0.0036**

accuracy obtained for each of the characteristics. In contrast to the results presented in section 3.4.5, adding daylight and temperature coefficients to the feature set particularly improved accuracy for characteristics `floor_area` (increase by 2.1 percentage points from 50.5% to 52.6%), `house_type` (increase by 2.3 percentage points from 59.3% to 61.5%), and `lightbulbs` (increase by 1.1 percentage points from 55.1% to 56.2%). The other characteristics exhibit a negligible increase, decrease, or no change. Even though we perform feature selection, a small decrease can occur in cases where the new feature set performs better than the previous feature set on the training data but worse on the test data. Overall, accuracy increased by 0.29 percentage points from 65.1% to 65.4%. In terms of MCC, we observe a large increase for characteristics `house_type` (increase by 0.046 points from 0.191 to 0.236), `lightbulbs` (increase by 0.018 from 0.0976 to 0.116), and `#appliances` (increase by 0.011 from 0.343 to 0.354), achieving an overall increase by 0.0036 points from 0.317 to 0.32.

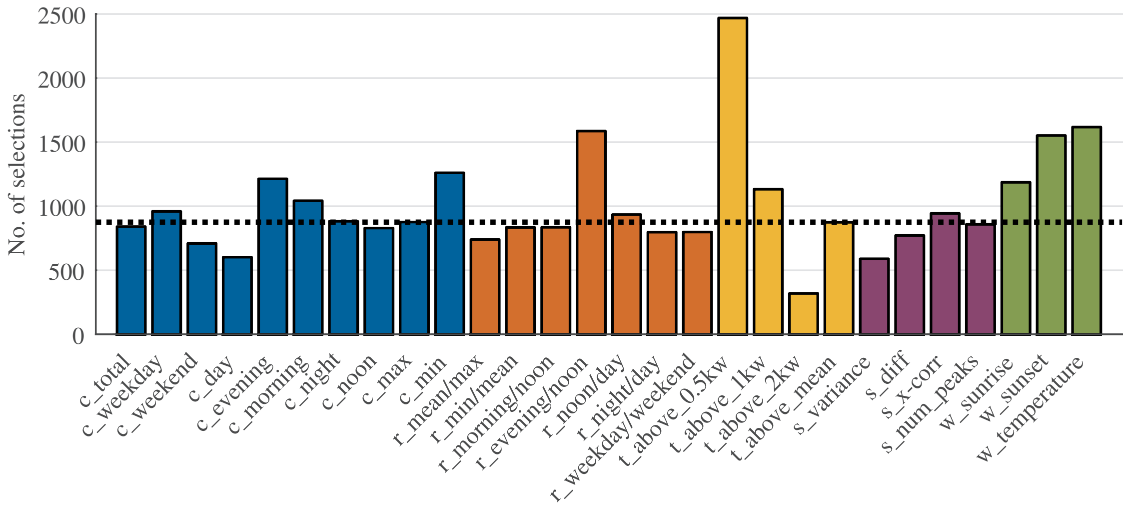
Overall, the weather coefficients (*i.e.*, daylight and temperature features) increase performance for characteristics related to the dwelling (*i.e.*, `floor_area` and `house_type`) or to the efficiency of lighting (*i.e.*, `lightbulbs`). While the former increase is possibly caused by the temperature coefficient, the latter can most likely be attributed to the inclusion of sunrise and sunset into the analysis. In the next section, we present a detailed analysis of which features are selected by the feature selection method for different household characteristics.

3.6 Features

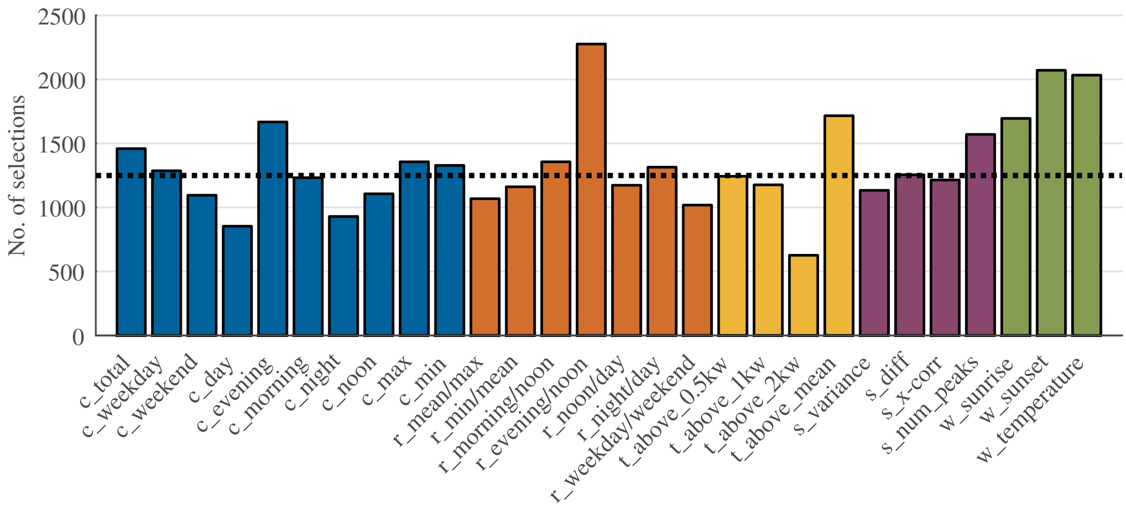
In this subsection, we evaluate the importance of each individual feature used in our analysis. To this end, we exclude the principal components from the feature set and run the analysis described in the previous section, which includes daylight and temperature coefficients in addition to the original features described in section 3.2.1. In particular, we run the analysis for each of the 18 characteristics and for all weeks of the trial using the LDA classifier and both accuracy and MCC as figure of merit.

The feature selection algorithm described in section 3.3.2 selects a maximum of eight features per run and only adds a feature to the feature set if the figure of merit improves significantly. Since we perform 4-fold cross validation, this results in up to 32 selected features for each characteristic, whereas each individual feature can be selected up to four times. Overall, since we run the classification separately on each of the 75 weeks' data, each feature can be selected up to 300 times for each characteristic, or up to 5400 times in total.

Figure 3.20 shows how often each feature has been selected by the feature selection algorithm in the experiments. The colors of the bars indicate the type of the feature: The first ten features (blue bars) represent consumption figures, the next seven features (orange bars) represent consumption ratios, the next four features (yellow bars) stand for the temporal properties, the next four features (purple bars) represent statistical properties and ultimately, the last three features (green bars) represent the weather coefficients described in the previous section. The top plot (figure 3.20a) shows the features selected when using accuracy as figure of merit. The black, dotted line indicates the median value of all features. The plot shows that the features that represent average consumption data are roughly equally distributed. Exceptions are features `c_evening` and `c_min`, which have been selected 1214 and 1261 times, respectively, and thus more often than the others. Similarly, the features that represent consumption ratios are roughly equally distributed, except `r_evening/noon`, which was selected 1587 times. The feature that indicates if a household's consumption reaches 0.5 kW throughout a day was selected most often: in 2469 out of the 5400 classifications, `t_above_0.5kw` was selected by the feature selection algorithm. Ultimately, weather coefficients, in particular `w_sunset` and `w_temperature`,



(a) Figure of merit: Accuracy.



(b) Figure of merit: MCC.

Figure 3.20: Number of times each feature has been selected by the feature selection algorithm.

have been selected more often than many other features with 1552 and 1618 selections, respectively.

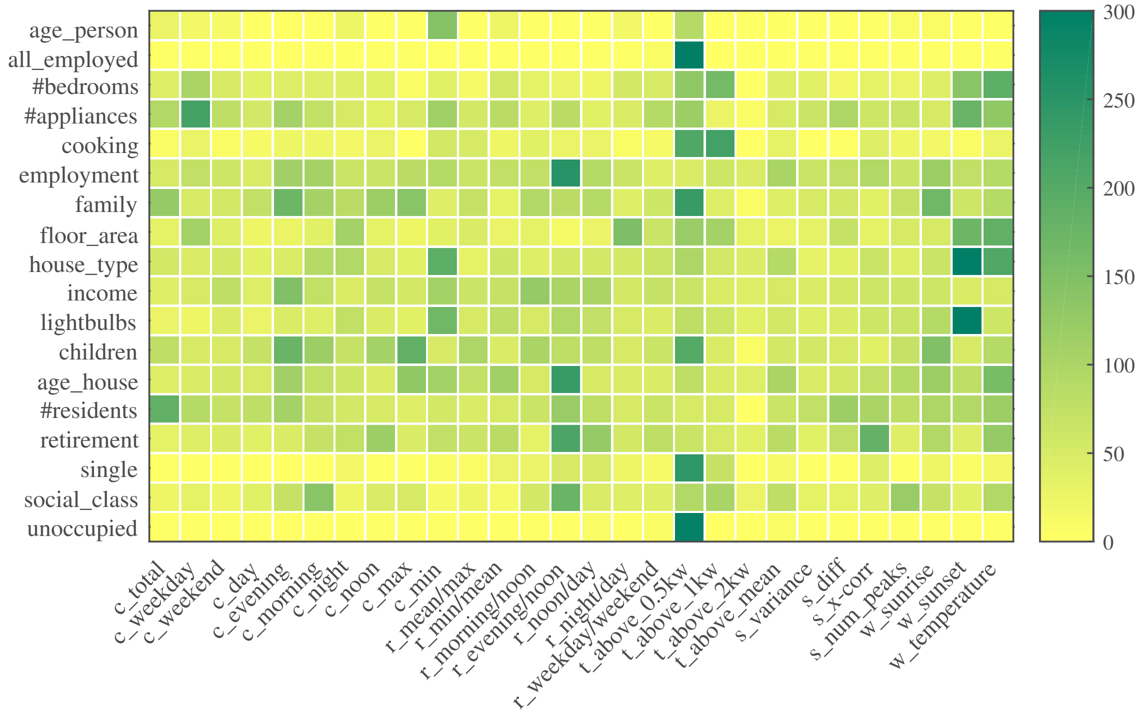
The bottom plot (figure 3.20b) shows the number of selections per feature when using the MCC as figure of merit. It shows that among the consumption figures, `c_evening` is highest with 1667 selections. The feature with the most selections is `r_evening/noon` with 2276 selections, followed by the two weather-related features `w_sunset` and `w_temperature` with 2071 and 2033 selections, respectively.

Overall, the most selected features are two different consumption features representing the evening and minimum consumption, a consumption ratio which divides the evening by the noon consumption, the fact whether a household's consumption reaches 0.5 kW throughout a day, and weather coefficients (sunset and temperature). The statistical properties have been rarely selected by the feature selection algorithm.

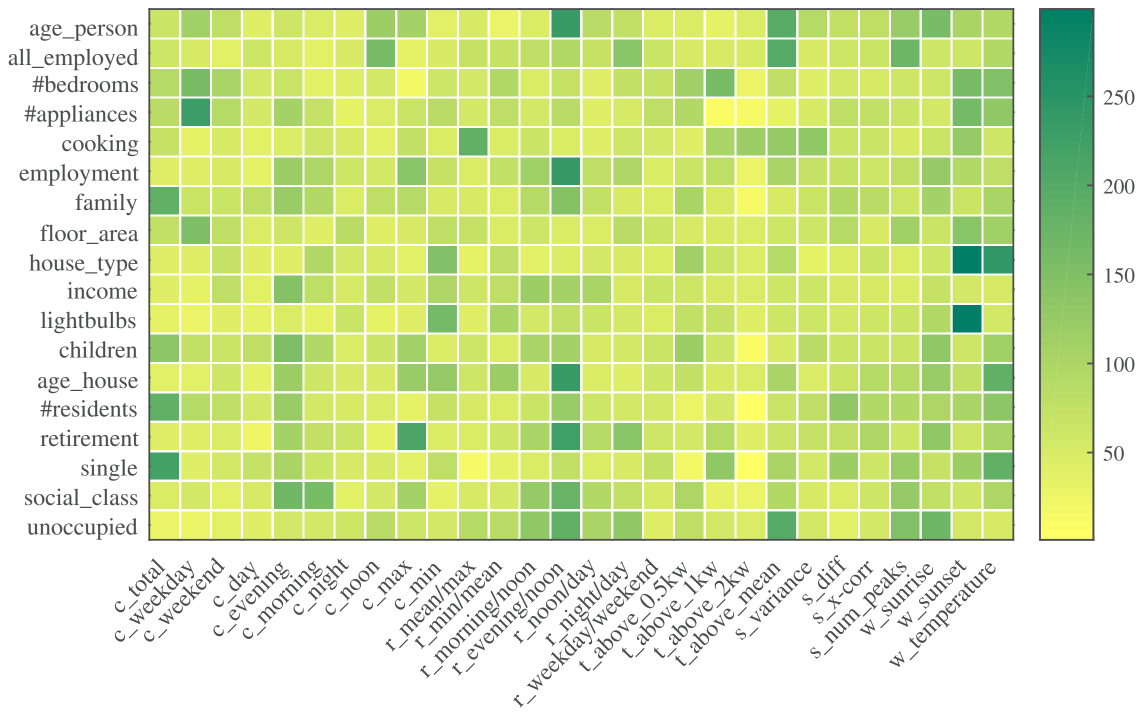
In figure 3.21, we show the number of selections per feature for each household characteristic. Again, the top plot (figure 3.21a) shows the experiments with accuracy as figure of merit, the bottom plot (figure 3.21b) shows the experiments with the MCC as figure of merit. The plots show—for each feature and household characteristic—the number of times the particular feature has been selected indicated by the color of the particular segment. The total average consumption of a household (`c_total`), for instance, plays a major role when classifying the characteristics related to the number of persons or appliances in the household (*i.e.*, `#children`, `#residents`, `family`, and `#appliances`). Consumption ratios are often included into the feature set when classifying characteristics related to the occupancy of the house (*i.e.*, `employment`, `retirement`). Finally, weather-related features play an important role when classifying characteristics related to the number of appliances (*i.e.*, `#appliances`) or the dwelling (*i.e.*, `#bedrooms`, `floor_area`, `house_type`). For those characteristics with a relatively uneven distribution of samples per class (*i.e.*, `age_person`, `all_employed`, `single`, and `unoccupied`), using accuracy as figure of merit leads to the fact that only a few distinct features are selected for classification. When using MCC as figure of merit, the observations described above hold for these features as well. In a practical setting, since the selected features differ for each characteristic, we recommend to implement all features and let the feature selection select the best feature set.

3.7 Data granularity

To evaluate the impact of smart meter data granularity on the performance of our household classification system, we transformed the data to obtain *30-minute data* (no averages), *60-minute data* (averages of two consecutive readings), and *daily data* (averages of 48 consecutive readings). For both, 60-minute data and daily data, we adapted the features accordingly. The former case only requires recomputing the existing features on the new data. For the daily data, however, many of the features such as the consumption ratios (except `r_weekday/weekend`) and statistical features (*e.g.*, variance and cross-correlation) must be omitted. To compute the sensitivity of each household to outdoor temperature and daylight given 60-minute data, we performed the regression analysis as described before except that we also aggregated the temperature readings accordingly. In the case of



(a) Figure of merit: Accuracy.



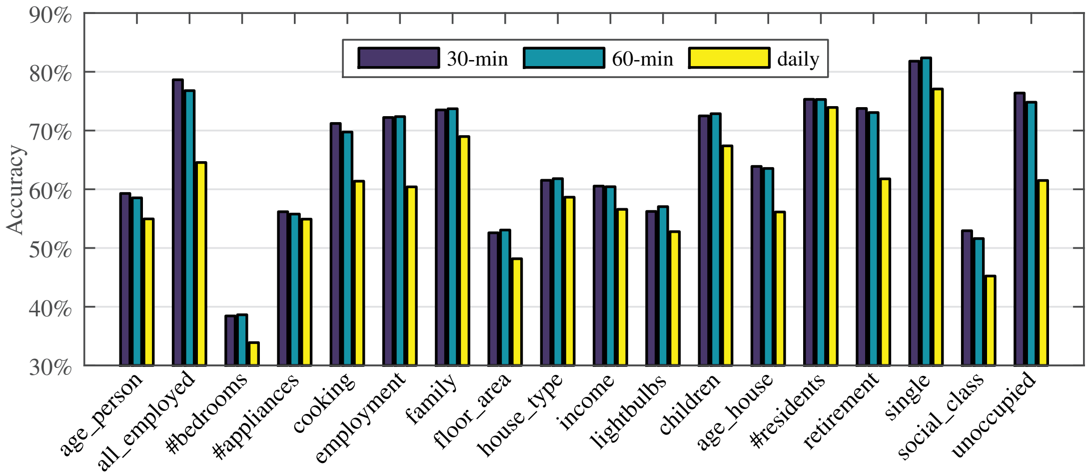
(b) Figure of merit: MCC.

Figure 3.21: Number of times each feature has been selected by the feature selection algorithm for each household characteristic.

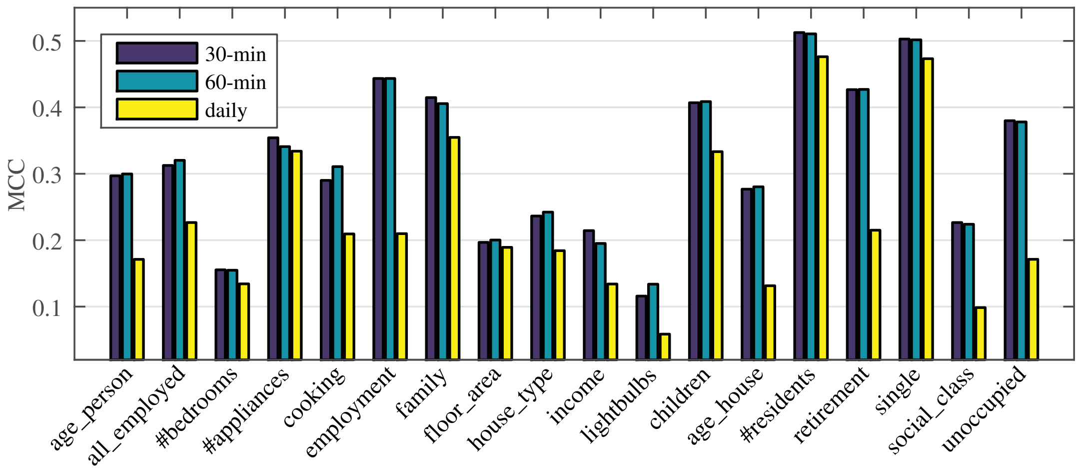
daily data, we omitted the daylight features and computed temperature sensitivity based on daily electricity consumption and temperature data. Using these modified feature sets, we performed the analysis as described above, using 75 weeks of data and both accuracy and MCC independently as figure of merit during feature selection.

Figure 3.22a shows the results obtained when using accuracy as figure of merit. Each set of bars illustrates the accuracy obtained when running the classification on 30-minute data (left bar), 60-minute data (center bar), and daily data (right bar). The plot shows that the difference between 30-minute data and 60-minute data is negligible. On average, accuracy is 0.3 percentage points higher when 30-minute data is available. However, for some of the characteristics, classification using 60-minute data shows slightly better results. While this seems counterintuitive, a potential reason is that training and testing performance can slightly differ because a different set of households is used during testing and the set of features determined by the feature selection is not necessarily the best performing feature set on the test set. Comparing the results obtained on 30-minute data with the results obtained when classifying on daily data shows that the accuracy using 30-minute data is much higher (*i.e.*, 6.6 percentage points on average). This is in particular visible for characteristics that make use of consumption ratios: Classifying characteristics `all_employed`, `unoccupied`, `retirement`, and `employment` shows a difference of 14.9, 14.1, 12.0, and 11.8 percentage points, respectively. Characteristics `#appliances` and `#residents`, however, correlate stronger with the overall electricity consumption, which is indicated by a relatively low difference between 30-minute data and daily data with 1.3 and 1.4 percentage points, respectively. These observations are in line with the results obtained when using MCC as figure of merit, as presented in figure 3.22b. The difference between 30-minute data and 60-minute data is negligible, as it is always below 0.02. On average, the MCC obtained when performing the analysis on 60-minute data is 0.0008 higher, which can be explained by particularities of the feature selection as described above. Running the analysis on daily data, however, performs significantly worse as the average MCC is 0.0921 lower than the MCC obtained using 30-minute data.

Overall, the analysis shows that when deploying smart meters in a real-world setting, collecting 30-minute data or 60-minute data is a necessity in order to infer a broad range of household characteristics from smart meter data. In our study, performance on such data is on average 6.6 percentage points better than the results obtained when inferring household characteristics from daily data.



(a) Figure of merit: Accuracy.



(b) Figure of merit: MCC.

Figure 3.22: Results obtained for each characteristic achieved when running the analysis on smart meter data with different temporal granularities.

Chapter 4

Non-intrusive load monitoring

Providing feedback on electricity consumption is a powerful way to induce a more energy-efficient behavior in households [46, 74]. In particular, feedback has shown to be effective when it is detailed and provided in a timely manner, tailored to individual households and contains information on the consumption of individual appliances [8, 46, 74]. Utilities, which are increasingly forced (and motivated) by policy makers to help their customers save electricity [49, 159], are thus highly interested in providing appliance-specific consumption feedback as a service to their customers (*e.g.*, in the form of automated saving recommendations). The data needed to provide such feedback could be obtained through sensors that monitor the consumption of individual appliances in the household. Deploying such a sensing infrastructure is however costly and cumbersome.

To avoid the need of monitoring individual appliances, non-intrusive load monitoring (NILM) algorithms have been proposed in the literature [114, 186, 187]. These algorithms analyze the *aggregate electricity consumption* of the household, *i.e.*, the total electricity consumption of the household measured using a single electricity meter. Through this analysis, the algorithms can identify which individual appliances are running and how much electricity they consume. NILM approaches may differ on several aspects, including the granularity at which they assume consumption data to be available or whether they apply supervised or unsupervised methods to learn typical consumption patterns of household appliances. NILM algorithms are often evaluated on single, possibly non-publicly available data sets and the parameters of the algorithms are tuned to operate on those data sets [186, 187]. Different underlying assumptions, tailored parameter settings, and lack of comprehensive data sets thus make the comparison and evaluation of NILM algorithms cumbersome, time consuming, and often non-exhaustive. This also hampers the possibility to derive general insights about which algorithms are best suited to be used in which scenario.

In this chapter, we address the problem described above and provide the following four contributions. First, we outline the key dimensions of the design space of these algorithms. Second, we describe a novel, comprehensive data set for non-intrusive load monitoring and occupancy detection research—the ECO data set—that we collected and made publicly available [190]. While we rely on this data set to evaluate an approach to detect household occupancy in another work [102, 103], we present here the data set in detail and utilize it for the evaluation of NILM algorithms. The ECO data set contains aggregate electricity consumption data—including real and reactive power for each of the three phases—and plug-level measurements of selected household appliances. The data has been collected at 1 Hz granularity over a period of eight months. Furthermore, the data set also contains occupancy information of the monitored households. The third contribution is the design and implementation of a comprehensive evaluation framework for NILM algorithms. The framework, called *NILM-Eval*, is similar in scope to the recently presented non-intrusive load monitoring toolkit (NILMTK) [17] and aims at allowing researchers to run comprehensive performance evaluations of NILM algorithms. Like for the ECO data set, we made the NILM-Eval framework publicly available [189]. The last contribution of this chapter consists of the evaluation of the performance of selected NILM algorithms. The algorithms are chosen so as to represent different sectors of the design space of NILM algorithms. We evaluate their performance using our NILM-Eval framework and rely on the ECO data set. The results yield insights about the performance of the selected algorithms and allow to outline their trade-offs and to discover potential for further improvements of the considered algorithms. This chapter is partially based on our contributions made in [19] and [20].

4.1 Design space

The first well-known NILM approach has been proposed by George Hart [85] in 1992. Hart's algorithm identifies step changes in the aggregate electricity consumption and matches them with a signature database to learn which appliance has been switched on or off. Building upon Hart's seminal work, several algorithms that rely on different principles (*e.g.*, combinatorial or probabilistic), utilize different learning methods, or rely on different data granularities have been proposed in the literature. For a comprehensive overview of the challenges in the field and existing NILM algorithms we refer to recent surveys by Zeifman *et al.* [186], Zoha *et al.* [187], and Liang *et al.* [114].

We outline the diversity of NILM approaches by classifying them in a three-dimensional design space. Each of the dimensions represents a parameter that must be considered when deciding which NILM algorithm to implement in a real scenario: *data granularity*, *learning methods* and *information detail*.

The first dimension, *data granularity*, represents the data granularity for which the algorithms were designed and optimized for—although most of the algorithms can potentially also run on data of a different granularity. The granularity typically ranges from $1/60$ Hz (*i.e.*, data aggregated to one value per minute) [135] to multiple kilohertz (*e.g.*, [25, 81, 136, 167]).

Second, NILM algorithms may utilize different *learning methods*. There exist unsupervised and supervised NILM algorithms as well as approaches that are classified as semi-supervised [135]. The notion of semi-supervised learning in the NILM domain slightly differs from the machine learning definition. In machine learning, the distinction between unsupervised, supervised, and semi-supervised derives from the type of labelling that is available [29]. Supervised learning implies that labelled samples can be used for training, whereas unsupervised learning achieves to detect structures in the input data without any label information. In semi-supervised learning, both labelled and unlabelled data is used for training, whereas typically a small set of labelled data and a large set of unlabelled data is available. In the NILM domain, supervised methods rely on appliance-level data from the test household for training [223]. Unsupervised learning implies that only aggregate consumption data is available for the test household without any prior knowledge on the number and type of appliances. Semi-supervised approaches are classified between supervised and unsupervised algorithms: They utilize labelled training data from non-test households and unlabelled data from test households to avoid the need to intrusively install sensors when installing the system in practice. However, to cope with the challenge that the number and type of appliances differ from household to household, semi-supervised algorithms often require high-level information on the composition of appliances in the test household.

Unsupervised approaches typically rely on low-frequency (*e.g.*, 1 Hz) aggregate consumption data [13, 97, 108]. Baranski and Voss detect switching events in the aggregate consumption data and use them as input to a genetic algorithm, which automatically creates event chains for different appliances [13]. Other authors utilize hidden Markov models (HMMs) to model the states of each appliance [97, 108]. A disadvantage of unsupervised methods is that they require manual labelling after detecting consumption patterns of appliances in the aggregate load. Parson *et al.* [135] developed an approach that is also based on HMMs and only requires data at a granularity of $1/60$ Hz. In contrast to the other approaches, Parson's algorithm is considered semi-supervised, as it utilizes generic appliance models as input to the disaggregation algorithm. This avoids the need to intrusively install sensors or use other training methods when installing the system in practice. The overhead of training the system can be reduced to filling out a questionnaire that asks for the number and type of appliances installed in the household. Similarly, Egarter *et al.* developed a method called *PALDi* that also utilizes appliance models and HMMs [57]. As the complexity of appliance disaggregation with factorial HMMs increases exponentially with the number of appliances, Egarter *et al.* utilize particle filtering to approximate appliance state estimation. Applying their approach on 1 Hz consumption data, the authors achieve an accuracy of more than 90%. However, the evaluation is based on aggregate consumption data that is computed by adding the consumption of six appliances. The performance of the approach on real world aggregate consumption data is still to be evaluated. In order to facilitate computation of the appliance states required for PALDi (*i.e.*, to avoid the need of a priori knowledge about appliances), Egarter *et al.* propose to learn the number of appliances and the corresponding appliance models from the aggregate consumption data [58].

Supervised approaches can be classified by the granularity of consumption data they are developed for. Gupta *et al.* [81], for instance, developed the algorithm *ElectriSense*, which detects consumer electronics devices and fluorescent lighting by their electromagnetic interference generated during operation. To this end, the authors rely on consumption data measured at 10 kHz. Similarly, Suzuki *et al.* also determine which appliances are running at what time, utilizing the fact that current waveforms are characteristic for each appliance. Measuring the current at 40 kHz and using as input the waveform of each state of each appliance in the household, the algorithm developed by Suzuki *et al.* solves an integer linear programming problem to identify the state of each appliance that is most likely. Berges *et al.* measure the electricity consumption of a household at 20 Hz and detect edges in the consumption data using features computed on both real and reactive power [25]. Farinaccio *et al.* developed a pattern recognition approach that applies rules to identify the consumption pattern of a refrigerator and a heater [69]. Both Weiss *et al.* [179] and Marchiori *et al.* [116] make use of real and reactive power measurements: The former approach is closely related to Hart's algorithm as it detects switching events of appliances in the consumption pattern. The latter creates two-dimensional histograms using the real and reactive power measurements and subsequently applies a clustering procedure to identify clusters that belong to individual appliances. Spiegel *et al.* [164] pursue a classification approach using features (*i.e.*, the first order difference of the consumption data) extracted from 1 Hz real power measurements. However, it is unclear if the algorithm can be applied in practical settings, because the authors evaluate only the accuracy of their approach and did not evaluate other performance measures. Even though they achieve more than 90% accuracy for most of the appliances, this value does not automatically imply a large number of true positives, which means that the algorithm successfully detected that the appliance was running. For an appliance that runs only a few minutes per day, for instance, the algorithm easily achieves a high accuracy when it always reports that the appliance is switched off. Elafoudi *et al.* developed a NILM algorithm based on dynamic time warping [62]. The authors report recognition rates of 85% including accuracy, precision, and recall. However, the data set collected and used by the authors is not publicly available, and their evaluation is only based on a few appliances that have very distinct consumption patterns. Lastly, Elhamifar and Sastry propose another supervised learning method [63]. First, the algorithm learns a dictionary of power consumption signatures from the training data. To this end, it learns the electricity consumption of each appliance using a mixture of dynamical systems, whereas each of these systems models the consumption of an operation mode of the appliance. It then selects subsets of these models to define each appliance's signature, which the authors call "powerlets". When disaggregating the aggregate consumption data, the algorithm searches for the combination of powerlets that best matches the aggregate consumption data. The authors ultimately evaluate their approach on the Reference Energy Disaggregation Data Set (REDD), achieving an accuracy of 81.8% (table 4.3 in section 4.4.1 provides for an overview of existing NILM data sets).

Finally, algorithms differ in *information detail*, which is the type of data they assume to be available. For instance, some of the aforementioned algorithms require real power consumption data only (*e.g.*, [13], [69], [97], [108], [135], [164]). Other algorithms rely on both real and reactive power (*e.g.*, [25], [85], [116], [179]) or make use of the fact that the consumption is split into individual phases (*e.g.*, [179]). Other approaches utilize information provided by other sensors as additional input to the algorithm [89, 98, 154], which can improve the estimation performance compared to analyzing the aggregate electricity consumption only. An example for such a sensor is an event detector developed by Rowe *et al.* [151], which detects state changes of appliances by sensing the electromagnetic field (EMF) in the surroundings based on magnetic and electric field fluctuations. Such cheap sensing approaches then provide relevant input for algorithms like ViridiScope [98] or the disaggregation algorithm developed by Jung and Savvides [89]. While ViridiScope also relies on other types of sensors (*e.g.*, light sensors), Jung *et al.* assume that each appliance in the household is equipped with a binary sensor that reports whether or not an appliance is running (*i.e.*, by monitoring when appliances are switched on and off). Instead of equipping each appliance with a sensor, Saha *et al.* combine smart meter data with sensor data observed by smartphones (*i.e.*, Wi-Fi scans and audio signals) [154]. Based on the analysis of the combined sensor data, the authors detect activities in homes, infer location of the occupants and are able to attribute electricity consumption to individual occupants. The authors evaluate their system called *EnergyLens* in a controlled, single occupant setting and achieved good results (up to 100% precision and recall) in case the phone is held outside the pocket. However, in a more realistic scenario, in which the phone is located in the user's pocket, precision and recall drop due to Wi-Fi positioning problems and misclassifications by the audio recognition algorithm.

4.2 Evaluated algorithms

Table 4.1 summarizes the main characteristics of the five NILM algorithms we implemented and evaluated for this study. The set of selected algorithms spans the design space discussed in the previous section. It includes supervised, unsupervised, and semi-supervised approaches as well as algorithms that require different levels of detail with respect to the measurements (*e.g.*, real power only vs. real and reactive power). In terms of aggregate consumption data, however, we include only algorithms that have been developed to operate on data measured at a frequency of at most 1 Hz. The reason for this restriction is that data at this granularity can be provided by most off-the-shelf electricity meters, for instance through the installation of a sensor that captures fine-grained consumption data from the smart meter's "user interface" (*cf.* section 2.2).

4.2.1 Description

The five algorithms are briefly described in the following subsections. More details about our implementation of these algorithms are provided in [39] and [99].

Parson's algorithm

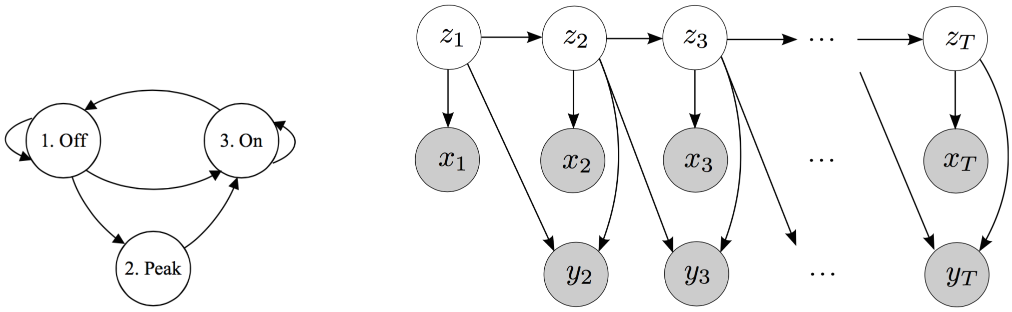
The algorithm of Parson *et al.* [135] relies on HMMs and the Viterbi algorithm [175] to disaggregate the electricity consumption of a household. For each appliance, it determines the most likely sequence of states (*i.e.*, operating states of an appliance), depending on the observed aggregate electricity consumption, state transition probabilities, and the estimated consumption of an appliance in each state. Using this state sequence, the algorithm estimates the consumption of the appliance, subtracts it from the aggregate consumption, and then iteratively estimates the consumption of other appliances in the household. Figure 4.1 shows the state transition model of a refrigerator (*cf.* figure 4.1b) and the difference HMM used to disaggregate the household's consumption (*cf.* figure 4.1a). In the model, the white nodes represent the hidden variables (*i.e.*, the state of the refrigerator) and the grey nodes represent the observed variables (*i.e.*, both the aggregate consumption x_t and the difference between consecutive measurements $y_t = x_t - x_{t-1}$ at time t).

Table 4.1: Overview of the five NILM algorithms evaluated in this study. *Data granularity* refers to the granularity of the data which the authors used to evaluate their algorithm in their original work.

<table>
<tr><th>Authors</th><th>Learning method</th><th>Data granularity</th><th>Information detail</th></tr>
<tr><td rowspan="2">Parson et al. [135]</td><td>Semi-supervised</td><td>1/60 Hz</td><td>Real power</td></tr>
<tr><td colspan="3">Trains factorial HMMs using prior knowledge of appliance types. Evaluated on REDD [109].</td></tr>
<tr><td rowspan="2">Baranski & Voss [13]</td><td>Unsupervised</td><td>1 Hz</td><td>Real power</td></tr>
<tr><td colspan="3">Clusters switching events and applies a genetic algorithm to assign events to appliances. Evaluated on simulated and real-world data (not publicly available).</td></tr>
<tr><td rowspan="2">Weiss et al. [179]</td><td>Supervised</td><td>1 Hz</td><td>Real and reactive power</td></tr>
<tr><td colspan="3">Based on Hart's algorithm [85]: Extracts switching events and finds best match in signature database. Evaluated in artificial lab setting (not publicly available).</td></tr>
<tr><td rowspan="2">Kolter & Jaakkola [108]</td><td>Unsupervised</td><td>1 Hz</td><td>Real power</td></tr>
<tr><td colspan="3">Generates HMMs from "snippets" identified in the aggregate consumption data. Evaluated on REDD [109].</td></tr>
<tr><td rowspan="2">Jung & Savvides [89]</td><td>Unsupervised</td><td>1 Hz</td><td>Real power and ON/OFF events</td></tr>
<tr><td colspan="3">Solves a linear optimization problem to estimate the contribution of each appliance to the overall power consumption. Evaluated on real-world data (not publicly available).</td></tr>
</table>

To determine the transition probabilities and power demand of each appliance, Parson *et al.* developed a semi-supervised training process. Instead of using sub-metered consumption data of an appliance, the algorithm utilizes a *generic appliance model*, which contains information on the characteristics of a certain appliance type. In case of a refrigerator, for instance, the algorithm incorporates information such as the average consumption of other refrigerators as a priori knowledge. On the basis of the generic appliance model, Parson's algorithm infers the parameters of a *specific appliance model* that describes the behavior of the appliance in the specific household.

The authors evaluated the performance of their approach on the Reference Energy Disaggregation Data Set (REDD) [109]. They estimated the consumption of four appliances (*i.e.*, refrigerator, microwave, clothes dryer, air conditioning) achieving a mean



(a) State transition model of a refrigerator modeled with three states.

(b) Difference HMM with hidden variables z and observed variables x, y.

Figure 4.1: State transition model and difference HMM used by Parson's algorithm. [135]

normalized error of 21% to 77% and a root mean squared error between 77 W and 477 W using aggregate data at a granularity of 1/60 Hz.

Baranski's algorithm

Baranski's algorithm [13] identifies recurring patterns in the aggregate electricity consumption data and attributes those patterns to individual appliances. Figure 4.2 shows the concept of the algorithm: First, it extracts events (*i.e.*, changes in electricity consumption over a given threshold) from the aggregate consumption and clusters those events, assuming that events in the same cluster belong to the same appliance. Next, a genetic algorithm creates a FSM and the most likely state sequence for each of the appliances.

Baranski's algorithm is unsupervised and thus can operate without knowing which appliances exist in the target household. The algorithm has been evaluated on both simulated data and real-world data, the latter one being collected with an optical sensor in one household over a time period of about five to ten days [13]. By inspecting the resulting clusters, the authors claim to confidently identify chief consumer load devices like refrigerators, heaters, or stoves. However, although the algorithm is unsupervised, it requires a manual assignment of the resulting clusters or FSMs to appliances in order to generate meaningful feedback for the occupants.

Weiss' algorithm

Weiss' algorithm [179] extracts switching events from the household's aggregate electricity consumption and assigns each event to the appliance with the best match in a signature

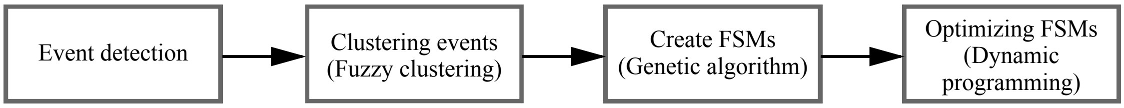


Figure 4.2: Structure of Baranski's algorithm. [13]

database. The algorithm is based on the approach developed by Hart [85], which clusters events by their real and reactive power in a training phase and assigns each event to the appliance with the best matching cluster during operation. The number of clusters is determined dynamically. In contrast to Hart, Weiss' algorithm relies on three-dimensional consumption data (*i.e.*, real power, reactive power, and distortion power; *cf.* section 4.4 for a detailed explanation of the physical units) and smoothes the power signal before extracting events. Weiss *et al.* also propose a novel method to unobtrusively generate signatures with the help of a smartphone application, which helps to identify switching events for each appliance during a training process. Up to now, the algorithm has not been evaluated on real world consumption data. Instead, the authors created a demo setting with eight different devices and performed a lab study, in which those eight devices were switched on and off 80 times in total. In this study, the algorithm identified 77 of 80 switching events correctly. Due to the lack of a large scale signature database, we treat Weiss' algorithm as a supervised approach that is trained using plug-level data.

Kolter's algorithm

Like Parson's algorithm, the algorithm developed by Kolter and Jaakkola [108] also models appliances as HMMs in order to disaggregate a household's electricity consumption. However, Kolter's algorithm is unsupervised as it only requires a household's aggregate electricity consumption data. To create an HMM for each appliance, the algorithm estimates the number of appliances and their consumption patterns from the aggregate consumption data. To this end, it extracts snippets of consumption data that likely correspond to an appliance's *ON cycle*, which is defined as the period between the appliance's start-up and shutdown. Next, it models each of the snippets as HMM (*i.e.*, it estimates the mean and variance of each state as well as the state transition probabilities) and identifies those snippets that most likely belong to the same appliance. This results in a factorial HMM (*i.e.*, a composition of several independent HMMs), which the authors then use to estimate the consumption of each individual appliance. To this end, Kolter and Jaakkola developed Additive Factorial Approximate MAP (AFAMAP), an approximate inference technique for factorial HMMs [108]. The authors evaluated their approach on REDD, analyzing 1 Hz aggregate consumption data using plug-level measurements of seven appliances for validation. Figure 4.3 shows the snippets that the algorithm extracted from the aggregate

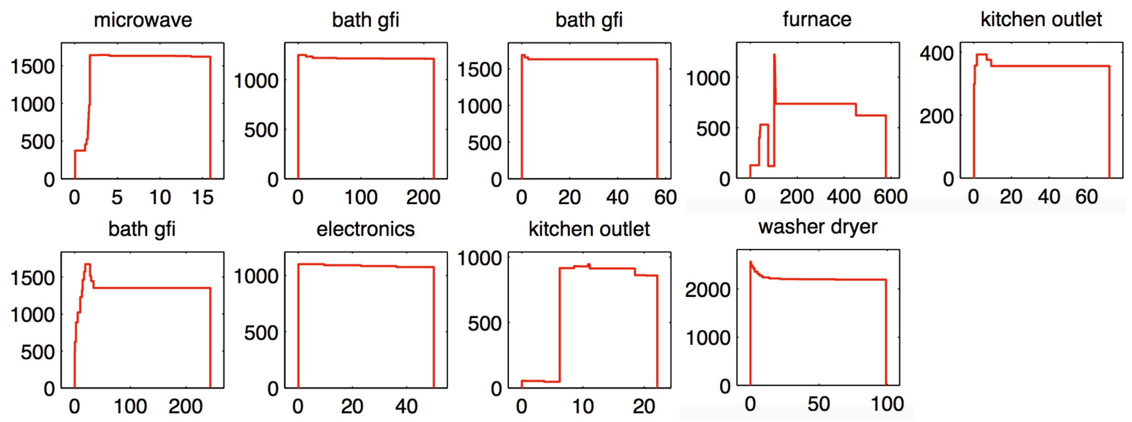


Figure 4.3: Signatures extracted by Kolter's algorithm. Each plot represents one signature and shows the time (in s) on the x-axis and the power consumption (in W) on the y-axis. [108]

consumption data in an unsupervised way. For the two plugs connected to the bathroom and kitchen outlets, the algorithm detected multiple snippets each, potentially because multiple appliances were connected to the each of the outlets. Each constant power segment in the signatures is then mapped to a state in the appliance's HMM. Using the HMMs as input to the AFAMAP algorithm, Kolter's algorithm achieved 87% precision and 60% recall.

Jung's algorithm

The algorithm developed by Jung and Savvides [89] combines smart meter data with switching events (*ON/OFF events*) of appliances. Having knowledge about the ON/OFF state of each appliance in the household, the algorithm solves a linear optimization problem to estimate the contribution of each appliance to the overall power consumption. To this end, it maintains a trace of the total electricity consumption as well as a state vector of active appliances. States with the same set of active appliances are merged on the fly by averaging the total power consumption and increasing a state counter. Using this data, the algorithm computes the average consumption of each appliance by minimizing the mean square error between the sum of estimates of all active appliances and the total electricity consumption. To improve the estimation accuracy, samples of the ON/OFF state vector that show fewer appliances in the ON state as well as samples that occur frequently are given higher weight in the estimation. Similarly, stationary loads are also given higher importance, as they can be estimated more accurately. The estimation procedure is performed over a specific time interval (*e.g.*, one hour) and then restarted, while estimations from previous intervals are "remembered" for each successive iteration.

Figure 4.4: RF Code dry contact sensing active RFID tag [227] used by Jung and Savvides to capture ON/OFF events of appliances.

Table 4.2: Samples of the data used as input for Jung's algorithm: At time t, the table shows the i^{th} appliance's ON/OFF state $x_i(t)$ and the total power consumption $y(t)$. [89]

t	$x_1(t)$	$x_2(t)$	$x_3(t)$	$y(t)$
1	0	0	1	62 W
2	0	0	1	60 W
3	1	0	0	120 W
4	0	1	1	380 W
5	1	0	1	160 W
6	0	1	1	371 W
7	1	1	1	469 W
8	0	0	1	56 W
9	0	1	1	357 W

Table 4.2 shows sample data that can serve as input for Jung's algorithm. The table shows each appliance's ON/OFF state at time t as well as the aggregate power consumption of the household at t. In this example, the samples in which appliances 1 and 2 are switched off and appliance 3 is running (*i.e.*, the 1st, 2nd, and 8th sample) gets a higher weight than the other samples, because they occur frequently and contain only a single running appliance.

Jung and Savvides evaluated their algorithm on real-world data they collected in a one-bedroom apartment over a period of three days. During this time period, they collected aggregate consumption data (*i.e.*, real power at 1 Hz) as well as information about ON/OFF events of 12 different appliances captured through radio-frequency identification (RFID) tags (*cf.* figure 4.4). Running their algorithm on the collected data, the authors achieved disaggregation with 9.46% relative error on average.

4.3 Evaluation methodology and the NILM-Eval framework

Performing a performance evaluation of a NILM algorithm is difficult, because there is no standard evaluation procedure to apply [187][1]. This is because composition and

[1] Even more challenging is performing a fair comparison between the performance of different algorithms, given that they all have different learning methods and different requirements to the input data.

usage of appliances differ significantly from household to household and over time. The performance of an algorithm thus highly depends on aspects such as the number of appliances running at the same time, the "noise" in the aggregate consumption data caused by unreported appliances, the performance metrics selected by the authors, and the input parameters they choose to tune their algorithm to the underlying data.

In order to gain a comprehensive view on the performance of a NILM algorithm it is thus necessary to run the algorithm in different scenarios (*e.g.*, using data from different households and from multiple time periods) and to experiment with different input parameters of the algorithm. To this end, we developed a MATLAB-based open source framework called NILM-Eval, which enables the evaluation of NILM algorithms on multiple data sets, households, data granularities, time periods, and specific algorithm parameters. By encapsulating those parameters in *configurations*, NILM-Eval allows the user to repeat experiments performed by others with little effort, to evaluate an algorithm on a new data set, and to fine-tune configurations to improve the performance of an algorithm in a new setting.

Figure 4.5 shows the NILM-Eval framework, which we made available to the public [189]. As input, a user provides (or selects) the implementation of an algorithm and specifies one or more *default configurations*. The default configurations provide means for the developer of the algorithm or for the user who evaluates the algorithm to adapt it to the corresponding household or data set. A user then creates *experiments* by selecting one or more default configurations and by optionally overwriting their parameters. NILM-Eval then evaluates all combinations of parameters specified by the user and thus supports evaluation of a broad range of parameter combinations. For each of the combinations, NILM-Eval creates a setup file, which then serves as input for the evaluation system. Since each run is performed on a separate MATLAB instance, NILM-Eval scales over many experiments (*e.g.*, by running it on a computing cluster). Ultimately, NILM-Eval provides as results for each experiment (1) the value of each of the performance metrics supported by the algorithm, (2) the estimated consumption of each appliance or, alternatively, labeled events, and (3) a series of plots illustrating the results.

To measure the performance of a NILM algorithm, NILM-Eval supports several metrics depending on the type of result provided by the algorithm. In case the algorithm returns the inferred electricity consumption of individual appliances, the framework computes for each appliance n the root-mean-square error (RMSE),

$$\mathrm{RMSE} = \sqrt{\frac{1}{T}\sum_{t}(y_t^{(n)} - \hat{y}_t^{(n)})^2}, \tag{4.1}$$

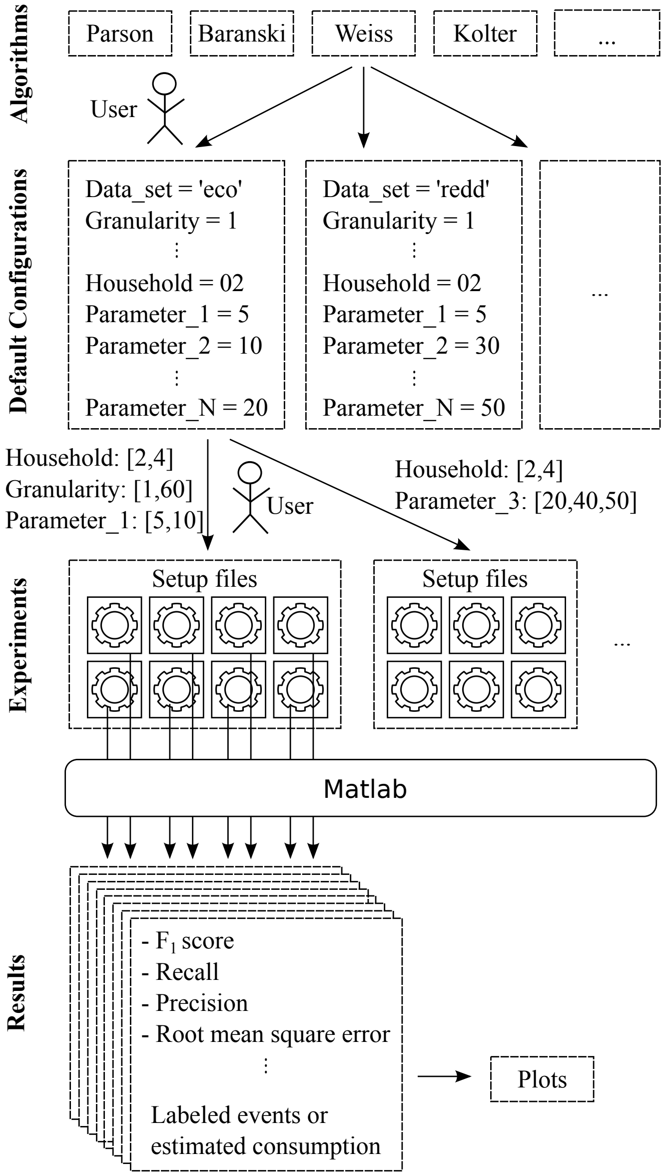


Figure 4.5: Evaluation framework NILM-Eval.

where $y_t^{(n)}$ denotes the actual electricity consumption of n at time t, $\hat{y}_t^{(n)}$ corresponds to n's inferred electricity consumption at time t, and T corresponds to the total number of time steps. NILM-Eval also determines the *deviation* of the inferred electricity consumption from the actual electricity consumption over a period of time [17, 135],

$$\text{Dev} = \frac{\left| \sum_{t=1}^{T} y_t^{(n)} - \sum_{t=1}^{T} \hat{y}_t^{(n)} \right|}{\sum_{t=1}^{T} y_t^{(n)}}. \tag{4.2}$$

Additionally, NILM-Eval determines the number of true positives (TPs), false positives (FPs), and false negatives (FNs) for each appliance. To this end, we define an appliance-specific threshold θ. If $\hat{y}_t, y_t > \theta$ we consider $\hat{y}_t$ a true positive, if $\hat{y}_t > \theta$ and $y_t < \theta$, a false positive, and if $\hat{y}_t < \theta$ and $y_t > \theta$, a false negative. NILM-Eval then computes the *F_1 score* as

$$\mathrm{F}_1 = 2 * \frac{\mathrm{PRC} * \mathrm{RCL}}{\mathrm{PRC} + \mathrm{RCL}}. \tag{4.3}$$

PRC and RCL denote the precision and recall, which are defined as

$$\mathrm{PRC} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{4.4}$$

and

$$\mathrm{RCL} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}. \tag{4.5}$$

In case the algorithm estimates switching events instead of the inferred electricity consumption at each time instant (like Weiss' algorithm), NILM-Eval computes only F_1 score, precision, and recall. In this case, for an appliance n, TP corresponds to the number of events correctly assigned to n and FP to those assigned to n even though the event was not caused by n. FN denotes the number of events missed by the algorithm.

Batra *et al.* recently published NILMTK, an open source framework for non-intrusive load monitoring [17]. Like NILM-Eval, NILMTK also provides functions that are useful when designing and implementing NILM algorithms. NILMTK has rich metadata support [92], preprocessing capabilities, and supports different statistics functions and performance metrics. For instance, it implements functionality such as data set diagnostics (*e.g.*, detection of gaps in the data), preprocessing of the data (*e.g.*, downsampling or voltage normalization), plotting functionality, or computation of different performance metrics. NILMTK is based on Python and supports numerous data sets and contains implementations of several algorithms such as Hart's algorithm [85]. With respect to NILMTK, NILM-Eval facilitates the design and execution of large experiments that consider several different parameter settings of various NILM algorithms. Furthermore, while NILMTK is written in Python, NILM-Eval is based on MATLAB.

4.4 The ECO data set

The analysis presented in this chapter is based on the ECO data set, which we collected for the purpose of this study, and which we made publicly available to the research community [190][2]. The ECO data set contains data collected from six households in Switzerland over a period of eight months (June 2012 to January 2013). In this section, we first describe alternative data sets that exist to evaluate non-intrusive load monitoring algorithms, and we argue why we decided to collect the ECO data set instead of using an existing data set. Next, we describe the data collection infrastructure, before we give information on the households that participated in the data collection and describe the data set in detail.

4.4.1 Available NILM data sets

In the past few years, several data sets collected for the purpose of evaluating NILM algorithms have been published. Each of those data sets exhibits different characteristics with respect to the number of households, the granularity of the collected data, the duration of the deployment, available contextual information (*e.g.*, coverage of appliances with smart plugs), and the level of detail of the smart meter data (*e.g.*, if it contains both real and reactive power).

Table 4.3 provides an overview of existing NILM data sets (including the ECO data set). The first data set that has been made available to the NILM research community is the Reference Energy Disaggregation Data Set (REDD) released by Kolter and Johnson in 2011 [109]. The initial release of the dataset (version 1.0) contains electricity consumption measurements from six households in the US collected in April and May 2011. There are up to 19 consecutive days of measurements available for each house. REDD provides data from the two main phases of each house at a granularity of one reading per second and measurements from 9 to 24 individual circuits—depending on the house—measured every 3–4 seconds. Some of the circuits contain a single appliance (*e.g.*, the dishwasher) and thus qualify for a device-level consumption breakdown. Other circuits contain multiple appliances (*e.g.*, lights, kitchen outlets), which can then only be treated as a group of devices by the NILM algorithm. Due to the early date of publication, REDD is to date the most used (and cited) dataset for NILM, even though the time of collected data is relatively short (*i.e.*, 19 days). REDD further contains high frequency measurements (*i.e.*, measured at 15 kHz), from which however only a small portion has been made available [213].

Since then, other data sets have been published such as *BLUED* [5], the *Smart** data set [15], *Tracebase* [147], the Household Electricity Use Study (HES) data set [196], the Almanac of Minutely Power data set (AMPds) [115], the Indian Dataset for Ambient Water

[2] At the time of writing, more than 50 researchers world-wide have downloaded the data set roughly six months after its publication.

Table 4.3: An overview of data sets that have been published for the analysis of NILM algorithms. The physical quantities represent current (I), voltage (V), phase shift (φ), real power (P), apparent power (S), reactive power (Q), and frequency (f).

Data set	**Date of publication**	**Location**	**Duration**	**No. of households**	**Features of smart meter data**	**Granularity of smart meter data**	**Submeters per household**	**Granularity of plug data**	**Additional information**
REDD [109]	2011	MA, USA	3 d to 19 d	6	S	1 Hz	9–24	3 s to 4 s	-
BLUED [5]	2012	PA, USA	8 d	1	I, V	12 kHz	$\tilde{5}0$	N/A	Event labels
Smart* [15]	2012	MA, USA	3 mon	1	P, S (25 circuits)	1 Hz	29	1 Hz	Motion detector, furnace, environmental data
Tracebase [147]	2012	Germany	N/A	N/A	N/A	N/A	158[1]	1 Hz	-
HES [196]	2013	UK	1 mon to 1 yr	251	I, V	10 min	13–51	10 min	Indoor/outdoor temperature
AMPds [115]	2013	BC, Canada	1 yr	1	I, V, P, S, Q, f	1 min	21	1 min	-
iAWE [16]	2013	India	73 d	1	I, V, ϕ, P, f	1 Hz	10	1 Hz	Many other sensors
UK-DALE [93]	2014	UK	3 mon to 17 mon	4	S	1 Hz & 16 kHz	5–53	6 s	Detailed metadata
GREEND [129]	2014	Austria, Italy	1 yr	9	N/A	N/A	9	1 Hz	-
ECO [19]	2014	Switzerland	8 mon	6	I, V, ϕ, P (for 3 phases each)	1 Hz	6–10	1 Hz	Occupancy information

[1] 158 plugs in total have been distributed to many different households.

and Energy (iAWE) [16], the UK Domestic Appliance-Level Electricity (UK-DALE) data set [93], and *GREEND* [129].

Our ECO data set extends these data sets on four aspects. First, it contains data collected over eight months. Only AMPds, UK-DALE, and the HES data set cover a comparably long time span. Second, the aggregate electricity consumption data provided with the ECO data set is very detailed as it contains measurements of voltage, current, and the phase shift between voltage and current for each of the three phases in the household. Of the other data sets, only the Smart* data set, AMPds, iAWE, and BLUED provide data that lends itself to compute both real and reactive power. Third, we collected plug-level data at 1 Hz frequency, which is otherwise only provided by the Smart* data set, by iAWE, by Tracebase, and by GREEND. Last but not least, the ECO data set is to the best of our knowledge the only data set that also includes ground truth occupancy information of the households[3].

4.4.2 Data collection and measurement infrastructure

We collected the ECO data set in the context of a project with a medium-sized Swiss energy provider. The purpose of the study was to learn what services utilities can in future—when each household is equipped with a smart meter—offer to their customers, and which measurement granularity is required to provide these services. To this end, we performed a deployment in six households that (voluntarily) participated in the project. In particular, we measured the aggregate electricity consumption and the occupancy status of each household as well as the consumption of selected appliances. In the following, we describe the measurement infrastructure we developed for each of the households in detail, before we provide an overview of the participating households, explain the data collection process, show how we cleaned and formatted the data, and give a detailed overview of the whole data set.

Figure 4.6 depicts the measurement infrastructure we installed in each of the six households. It consists of a smart electricity meter that measures the aggregate electricity consumption of the household, of several *smart plugs* to measure the consumption of selected appliances, and of a PIR sensor and tablet PC to capture occupancy information of the inhabitants. The data was sent to our aggregation server, from which we made it accessible to the tablet PC that also served as an in-home display to monitor current and past electricity consumption.

[3]The Smart* data set contains data from passive infrared (PIR) sensors, which indicates occupancy but is prone to errors, for instance when animals are present or when people are inactive or located in another room.

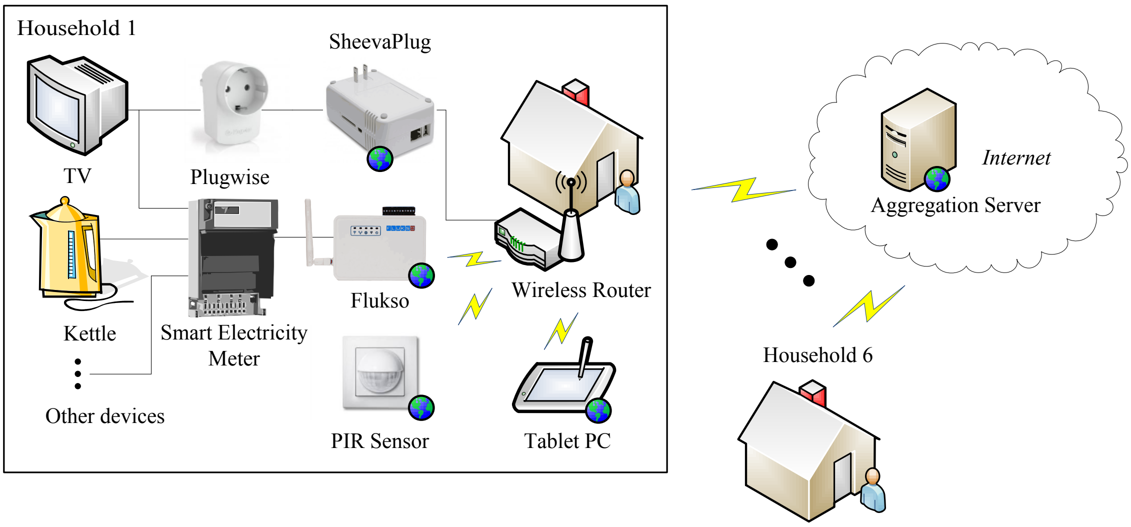


Figure 4.6: Overview of the measurement infrastructure deployed in each of the six households. [103]

Aggregate electricity consumption

To measure the aggregate electricity consumption of a household we used the E750 smart meter developed by Landis+Gyr [216]. The E750 (*cf.* figure 4.7) aggregates electricity consumption measured during one second and follows the Smart Message Language (SML) protocol to make the measurements available through an Ethernet interface. SML is a request-response protocol that specifies the format of requests and responses sent to and received from the smart meter. This way, it is possible to access the variables of the smart meter specified by their Object Identification System (OBIS) code. OBIS is an international standard to identify measurement values and abstract data types in communication systems and it is mainly used in the energy domain [199]. Table 4.4 shows the OBIS codes supported by the E750 and thus the data we collected from the smart meter. In our deployment, the E750 was installed in line with the original electricity meter. This way, the data neither accounts for the original meter's electricity consumption nor does it contain the electricity consumed by the E750 itself.

Figure 4.8 shows a Fluksometer [206], which is an embedded device we used to query the data from the smart meter. Originally, the Fluksometer connects three current clamps to the mains of a household and measures the magnetic field next to the wires. The strength of the magnetic field is proportional to the apparent power of the corresponding phase. In the deployment, we omitted the current clamps and instead use the Fluksometer to fulfill the following functions:

Table 4.4: Aggregate consumption data we collected through the SML interface of the Landis+Gyr E750 smart meter.

Variable	Description	OBIS code	Unit
powerallphases	Sum of real power over all phases	01 00 0F 07 00 FF	W
powerl1	Real power phase 1	01 00 23 07 00 FF	W
powerl2	Real power phase 2	01 00 37 07 00 FF	W
powerl3	Real power phase 3	01 00 4B 07 00 FF	W
currentneutral	Neutral current	01 00 5B 07 00 FF	A
currentl1	Current phase 1	01 00 1F 07 00 FF	A
currentl2	Current phase 2	01 00 33 07 00 FF	A
currentl3	Current phase 3	01 00 47 07 00 FF	A
voltagel1	Voltage phase 1	01 00 20 07 00 FF	V
voltagel2	Voltage phase 2	01 00 34 07 00 FF	V
voltagel3	Voltage phase 3	01 00 48 07 00 FF	V
phaseanglevoltagel2l1	Phase shift voltage phase 2 and 1	01 00 51 07 01 FF	°
phaseanglevoltagel3l1	Phase shift voltage phase 3 and 1	01 00 51 07 02 FF	°
phaseanglecurrentvoltagel1	Phase shift current/voltage phase 1	01 00 51 07 04 FF	°
phaseanglecurrentvoltagel2	Phase shift current/voltage phase 2	01 00 51 07 0F FF	°
phaseanglecurrentvoltagel3	Phase shift current/voltage phase 3	01 00 51 07 1A FF	°

Figure 4.7: Landis+Gyr E750 smart meter. [216]

Figure 4.8: Fluksometer we used to query the data from the smart meter. [206]

1. The Fluksometer requests the smart meter's measurements through SML and its Ethernet interface once per second and forwards the values over Wi-Fi to our server.

2. It adds a timestamp to each reading, because the E750 does not contain a real-time clock. The Fluksometer regularly synchronizes its time through Network Time Protocol (NTP).

3. It stores the readings to a local log file in case of a temporary loss of Internet connection. When Internet is available again, it synchronizes the readings with the server.

The software that runs on the Fluksometer is based on OpenWRT, a Linux distribution for embedded devices [222]. The software has been developed by Daniel Pauli in his Master's Thesis [137] and made available to the community as the Pylon project [212]. To communicate with the E750, Pylon integrates the open-source SML library *libSML* developed at the DAI-Labor at TU Berlin [209].

The power consumption in an alternating current (AC) circuit is computed as the product of the effective voltage of the source (V) and the effective current that flows through the circuit (I). If the circuit only consists of the source and a linear, resistive load, all power provided to the circuit is considered real power (P), which means it does "useful" work at the consumer. However, AC circuits typically also contain capacitive or inductive loads (*e.g.*, electrical motors), which store energy in electric or magnetic fields causing a delay between voltage and current. This delay is expressed by the phase angle φ, which ranges from $-90°$ to $90°$. If voltage and current are not in phase, the product of V and I, which defines the apparent power S, is greater than the real power. The vector difference between the apparent power (*i.e.*, the power given into the system) and the real power (*i.e.*, the power utilized by the consumers) is called reactive power (Q). It describes the energy that is stored and released by inductors or capacitors, moving back and forth in the circuit. The following formulas describe the relation between P, Q, S, and φ, whereas S_1 and Q_1 denote the fundamentals of S and Q, respectively [161]:

$$P = U * I_1 * cos(\varphi) \tag{4.6}$$

$$Q_1 = U * I_1 * sin(\varphi) \tag{4.7}$$

$$S_1 = \sqrt{P^2 + Q_1^2} \tag{4.8}$$

In practice, most AC circuits exhibit non-linear loads (*i.e.*, loads that draw nonsinusoidal currents), which distort the waveform of the current and make the computation of real, reactive, and apparent power more complex [30, 64, 161]. Examples for such non-linear loads are fluorescent lighting, computers and computerized controls, or rectifiers. The power of these non-sinusoidal currents is called *distortion power* and can be computed as

$$D = U * \sqrt{\sum_{v=2}^{\infty} I_v^2}, \tag{4.9}$$

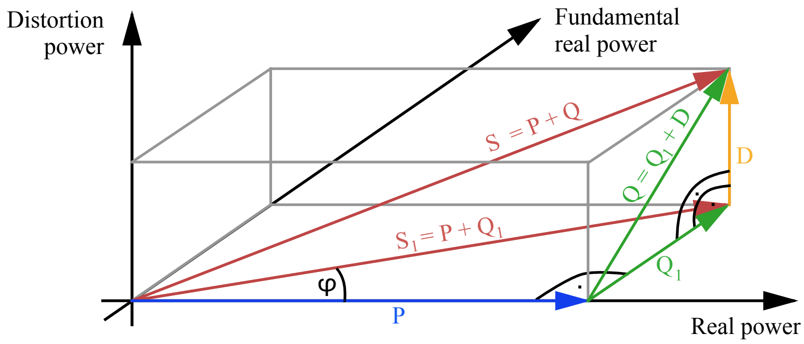


Figure 4.9: Vector representation of real power (P), reactive power (Q, Q_1), distortion power (D), and apparent power (S, S_1). "+" means addition in vector space. φ denotes the phase shift between the fundamentals of voltage and current.

where v iterates over the current's harmonics [161]. With the distortion power D we can compute the total reactive power Q by

$$Q = \sqrt{Q_1^2 + D^2} \tag{4.10}$$

and the total apparent power of the circuit by

$$S = \sqrt{P^2 + Q_1^2 + D^2}. \tag{4.11}$$

Figure 4.9 illustrates the relation between P, Q, S, and D. Since the computation of D given by equation 4.9 requires measurement of all harmonics, it is easier to compute D based on the fundamental reactive power Q_1 and the phase shift φ between the voltage and the current's fundamental. Therefore, we compute D by

$$D = \sqrt{S^2 - P^2 - Q_1^2} \tag{4.12}$$

with

$$S = U * I \tag{4.13}$$

and

$$Q_1 = P * tan(\varphi) \tag{4.14}$$

In a polyphase system that includes a neutral wire, the apparent power has a fourth component, A, which describes the asymmetry power that flows through the neutral wire [64]. A is nonzero when the power flowing through the phases is asymmetric. In this case, the apparent power consists of an additional component and can be computed as

$$S = \sqrt{P^2 + Q_1^2 + D^2 + A^2}. \tag{4.15}$$

As table 4.4 (described above) shows, the E750 smart meter provides—once per second and for each of the three phases—measurements of real power (P), voltage (V), current (I), and phase shift between voltage and current (φ). Although the E750 also provides the current that flows through the neutral wire, determining A used by equation 4.15 requires more information about the circuit such as the resistance of the three phases and of the neutral wire [64]. For this reason, in our analysis, we consider A^2 as a part of the distortion power and compute D using equation 4.12.

Figure 4.10 illustrates the different variables measured by the E750 smart meter (or computed thereof) by means of the consumption of a washing machine. The consumption was measured on phase 1 in household 1 on 7 July 2012. For space reasons, the plot does not show the consumption data collected from the plug that was connected to the washing machine. The plug data represents real power consumed by the washing machine and has the same pattern than the real power consumption (P) illustrated in the plot—except it is roughly 40 W lower. This observation shows us that P in the plot represents the real power consumption of the washing machine plus the consumption of one or more always-on consumers (*e.g.*, the router) that denote the standby consumption of phase 1. As expected, the apparent power (S) has the highest value in the plot, followed by the real power (P) and the distortion power (D). As mentioned above, the distortion power contains the power running through the neutral line, which is relatively high due to the asymmetry of the three phases when the washing machine is running. Finally, in the plot, the reactive power is most often negative while the motor is spinning, which indicates a negative phase shift on phase 1. However, this observation is misleading, because the reactive power was negative before and after the washing machine was running. The plot shows that during intense phases of spinning (*e.g.*, between 1:21 and 1:26), reactive power increases as the motor of the washing machine reduces the negative phase shift on phase 1.

Device-level electricity consumption

To record the electricity consumption of selected appliances (real power), we installed smart power outlets (*i.e.*, smart plugs) from Plugwise [225] in each home. Figure 4.11 shows the *Plugwise Sting*, which is a plugwise with connectors designed for Swiss power outlets. According to our investigation, the products from Plugwise are at the time of writing the smart plugs with the best trade-off between data granularity and ease of deployment. They are plugged between the appliance and the power outlet and can be accessed wirelessly via Zigbee [237]. When being plugged in, the plugs automatically establish a 802.15.4 mesh network among all smart plugs that are configured to be of the same group. The sink of the system is a USB dongle that can then be plugged into a host computer and interact with the plug network.

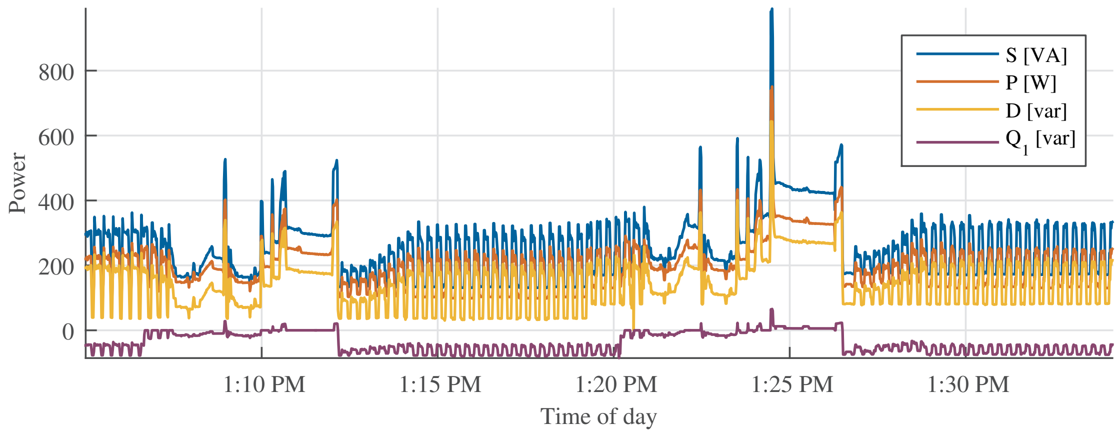


Figure 4.10: Different variables measured by the E750 smart meter in household 1 while the washing machine was running (7 July 2012).

Figure 4.11: *Plugwise Sting Type J*: Smart plug for Swiss power outlets. [225]

Figure 4.12: SheevaPlug mini computer to collect readings from the smart power outlets. [230]

Originally, each smart plug locally stores the total consumption of the appliance plugged into it. Through a proprietary software, it is then possible to access this aggregated value over the network. This way, Plugwise enables, for instance, to query the overall consumption of an appliance since initialization of the corresponding plug. Since our applications require consumption recorded at a much higher granularity, we utilize the open source library *python-plugwise* [224], which makes it possible to query the instantaneous electricity consumption of the appliance plugged into the plug. We wrote a Python script to query all plugs in the network for their current reading. This polling occurs sequentially, because sending queries to multiple plugs at the same time is not possible with the protocol implemented by the dongle. After sending a request to a plug, the script must therefore wait for the result before it can query the next plug [150]. While each query takes 80 ms to 120 ms, with the script we achieve to collect about one measurement per second per plug from approximately ten plugs.

However, problems that occur when querying a plug can lead to timeouts in the order of 5–10 seconds. Such timeouts occur, for instance, when there is network interference, when the connectivity is bad due to the physical infrastructure, or when the plug has been unplugged from the socket. While waiting for the response of a plug, no data can be collected from any of the other plugs. To avoid that such a timeout occurs in every single round when sequentially collecting data from all the plugs, the script skips the plug that caused the timeout and tries again after $T = 20\,\text{s}$. After each timeout it doubles T, whereas it resets T to its original value after a single successful query.

We use a *SheevaPlug* computer (*cf.* figure 4.12) to run our script and connect to the Plugwise network through the USB dongle. The SheevaPlug is then connected to the Internet and reports the measurements to our server located at ETH Zurich.

Occupancy information

Each household received an Android-based Samsung Galaxy Tab P7510 tablet computer with an application we developed for the purpose of the study [96]. The app provides real-time feedback on the aggregate electricity consumption measured by the smart meter and the consumption data recorded by the smart plugs. In addition, it shows the 7-day historical consumption and a historical chart with smooth zooming functionality. We utilized this app to collect ground truth data on the occupancy of each household. To this end, we implemented small buttons at the bottom of the screen and asked each person of the household to toggle the button when entering or leaving the house. The tablet computer was then installed near the entrance of each house to remind the occupants on pressing the button to indicate their change of absence or presence.

In addition, we deployed a PIR sensor mounted on a Roving RN-134 low-power Wi-Fi module [228] in each house. Yet, PIR sensors alone are—like the information obtained through the tablet PCs—not sufficient to provide ground truth for our purposes. Animals can for instance trigger the sensor when people are absent, or people could be at home but in another room. Instead, we use the PIR sensor data to complete the events obtained from the tablet PC in cases when occupants forget to specify that they enter or leave the house. When reporting the events, the round-trip time from the low-power Wi-Fi module to the server has significant impact on the module's energy consumption. For this reason, the module sends the readings to the SheevaPlug, from where we forward them to our server. This way, the module runs for approximately three months using two AA batteries.

Infrastucture

All the data collected in the households is sent through Hypertext Transfer Protocol (HTTP) to a RESTful web server located at ETH Zurich[4]. The web server runs a Java servlet that stores the data into a MySQL database. It also provides a user interface based on the Dojo JavaScript Toolkit [197] for administrative purposes such as checking the status of each sensor or adding, removing, and configuring sensors and residences. The web server further offers a RESTful interface to query the data, which is used by the app that runs on the tablet PCs.

To be independent of the unpredictability that comes with using existing Wi-Fi infrastructure, we brought a Linksys WRT54GL Wireless Router [217] into each of the households and connected it to the household's existing router over Ethernet. The Fluksometer, the tablet PC, and the low-power Wi-Fi modules log into the pre-configured Wi-Fi network which we set up on the Linksys router. The SheevaPlug connects to the Linksys router over Ethernet. Through reverse Secure Shell (SSH) tunnels, we then had constant access to every component of our setup.

Participating households

We deployed our infrastructure in six households and collected data from June 2012 to January 2013. The participating households were employees of our project partner, a medium-sized energy provider in Switzerland. The households were interested in smart metering in general and wanted to learn "how it feels like" to have a smart meter installed and to obtain real-time visualization of their electricity consumption. They participated on a voluntary basis: We searched for interested households by putting up a poster in the company's cafeteria. Out of ten households interested in the study, we selected six households to participate based on their characteristics. For instance, we searched for tech-savvy households due to the amount of sensing infrastructure we planned to install in the household.

Table 4.5 shows the characteristics of the participating households (*household 1* to *household 6*). We started the data collection in household 1 and household 2 to test our setup before we installed our deployment in households 3–6. The next section provides detailed information on the time periods with available measurements for each household. During the installation, we discovered that the E750 smart meters rounded values to the nearest 10 W. For this reason, we replaced those smart meters with new, re-calibrated

[4] Representational State Transfer (REST) is an architecture style for distributed systems, often used to create scalable Web applications.

ones. Also, we re-positioned some of the plugs because they were not able to form a mesh network in some places. After correcting these initial problems, we officially started our data collection.

Data cleaning and formatting

Before presenting some details of the ECO data set, we describe how we cleaned and formatted the data. For the analysis, we extracted the data from the MySQL database and stored it as comma-separated values (CSV) files—one file per day and per sensor. Since the smart meter measures data at 1 Hz, each file contains 86,400 rows. Each of the rows consists of the timestamp set by the Fluksometer as well as the values collected by the smart meter as shown in table 4.4. In case of data loss, we distinguish between two types depending on the amount of data that is missing: If measurements are missing for up to ten seconds, we fill the corresponding positions in the vector with the previous existing measurements. Typically, only few seconds are lost each day. In cases where more than ten consecutive readings are missing, the values are set to -1. These are rare cases, however, that occurred for instance when the Fluksometer crashed, restarted, or was turned off.

For the smart plugs, the frequency of the collected data varies because we had to read them sequentially from a central gateway. To be consistent with the aggregate consumption data, we resampled the plug measurements for each appliance at 1 Hz. If a small number of values is missing between two measurements (*i.e.*, less than 100), we replaced those missing values with the last existing measurement. If more than 100 consecutive measurements are missing, we assume that the plug has been removed and invalidated the missing values by setting them to -1.

The occupancy data collected from the tablet PC and the PIR sensors are stored in CSV files with 86,400 rows and N_i columns, where each row represents the second of the day and each column denotes the status of the i^{th} person or sensor, respectively. In case a person is home on a particular second of the day according to the data entered to the tablet PC, the corresponding entry in the matrix is set to 1. Otherwise, it is set to 0. Similarly, if a PIR sensor triggers an event, the 30 consecutive seconds after this event are set to 1, whereas the part of the data with no preceding event is set to 0.

4.4.3 Detailed description of the data set

The ECO data set contains the data collected in six Swiss households over a period of approximately eight months per household. For each of the households, it provides:

Table 4.5: Characteristics of the households that participated in our study.

No.	Tech savviness	Residence	Occupants (occupation, age)	Vacancy (h per day)[1]	Comments
1	7/7	House	Full-time, 33 yr Housemaker, 33 yr Kid, 3 yr Kid, 1 yr	3; 4	-
2	7/7	Apartment	Full-time, 34 yr Part-time, 32 yr	5; 4–6	-
3	7/7	House	Full-time, 40 yr Housemaker, 40 yr	3; 2	Electric water heating
4	4/7	House	Full-time, 55 yr Housemaker, 33 yr Kid, 17 yr Kid, 15 yr	2; 7	1 cat, 2 entrances, electric water heating
5	6/7	House	Full-time, 62 yr Housemaker, 64 yr	1; 2	1 dog,[2] 3 entrances, el. water heating
6	6/7	House	2 persons	n/a; n/a	-

[1] First value: On a weekday; Second value: On the weekend.
[2] Until 20 August 2012.

- 1 Hz aggregate consumption data. Each measurement contains data on current, voltage, and phase shift for each of the three phases in the household (*cf.* table 4.4).
- 1 Hz plug-level data measured from selected appliances.
- Occupancy information measured through a tablet PC (manual labeling) and a passive infrared sensor in some of the households.

In total, we collected more than 800 million measurements from the smart meter, the 45 smart plugs, the PIR sensors, and the tablet PCs deployed into the six households. Table 4.6 summarizes the device types, the quantity of devices installed in the households, as well as the number of readings per device type.

Table 4.6: Overview of the ECO data set: Number of readings per sensor type.

Sensor device	**Quantity**	**Type**	**Readings**
Landis+Gyr E750	6	Smart electricity meters	125,987,285
Plugwise Sting	45	Smart power outlets	686,655,790
Roving RN-134	6	PIR sensors	563,758
Samsung Galaxy Tab P7510	6	Occupancy (ground truth)	6396

We used the data collected by the PIR sensors to validate and if necessary filter the ground truth occupancy obtained from the tablet PC as described in [103]. As a result, we published the resulting occupancy data, neglecting the raw data collected from the PIR sensors and from the tablet PCs.

In the following, we provide details on the data collected in each household. Table 4.7 shows—for each household—the measurement periods of both the smart meter and the plug data collection. It also lists the appliances that we equipped with a plug and the number of days for which smart meter, plug, or occupancy measurements exist. In addition, the table illustrates the proportion of valid smart meter and plug measurements collected on these days. The occupancy data contains 90 days of data for households 1–5 on average. Household 6 did not provide occupancy information. For more information related to the occupancy data collected in the five households, we refer to [100] and [103].

Figure 4.13 illustrates—again for each household—the aggregate electricity consumption measured by the smart meter (on the left side) and a breakdown of consumption as measured by the smart plugs (on the right side). To compute the probability distribution of the aggregate consumption, we divide each day into 96 time slots (15 minutes each) and create 50 bins on a logarithmic scale from 10 W to 6025 W. We then count the number of times (*i.e.*, the number of seconds) the consumption lies within each bin for each of the 96 time slots over the whole measurement period. Dividing this number by the overall number of times per time slot (*i.e.*, the number of days $\times$ 15 minutes $\times$ 3600 measurements per minute) provides the probability distribution for each of the time slots, which we then plot on the x-axis along with the binned power consumption on the y-axis. The consumption breakdown consists of the monthly electricity consumption of each appliance covered by the smart plugs. As we equipped only 6–10 appliances per household with a plug, each of the charts shows a significant portion named "Other" that is measured by the smart meter but not attributed to any of the appliances.

Table 4.7: Summary of the aggregate consumption data, plug-level consumption data, and (cleaned) occupancy data for each of the households in the ECO data set.

(a) Household 1

	No. days	Coverage
Smart meter data (1 Jun 2012 – 31 Jan 2013)		
	245[1]	99.64%
Plug data (1 Jun 2012 – 23 Jan 2013)		
Refrigerator	231	98.53%
Dryer	231	98.56%
Coffee machine	113	85.36%
Kettle	203	77.65%
Washing machine	231	98.56%
PC[2]	66	84.77%
Freezer	231	98.56%
Occupancy		
Summer	39 d	
Winter	46 d	

[1] From 1 Jun 2012 to 29 Jun 2012, each power measurement has been rounded to 10 W precision.
[2] Includes the router.

(b) Household 2

	No. days	Coverage
Smart meter data (1 Jun 2012 – 31 Jan 2013)		
	244	98.58%
Plug data (1 Jun 2012 – 31 Jan 2013)		
Tablet	240	97.43%
Dishwasher	240	97.09%
Refrigerator	240	98%
Freezer	240	96.39%
Kettle	240	88.5%
Lamp[1]	240	90.21%
Laptops	240	83.36%
Stove[2]	28	100%
TV[3]	240	100%
Stereo[3]	240	95.95%
Occupancy		
Summer	83 d	
Winter	45 d	

[1] Dimmable lamp.
[2] Instead of the stove, the plug was connected to the air exhaust system, which is located above the stove and indicates when inhabitants are cooking.
[3] TV and stereo system were connected to the same plug.

(c) Household 3

	No. days	Coverage
Smart meter data (26 Jul 2012 – 31 Jan 2013		
	138	98.89%
Plug data (23 Oct 2012 – 31 Jan 2013)		
Tablet	104	94.5%
Freezer	104	90.71%
Coffee machine	67	70.79%
PC	42	64%
Refrigerator	47	56%
Kettle	42	67.82%
Entertainment[1]	48	67.65%
Occupancy		
Summer	57 d	
Winter	21 d	

[1] Entertainment consists of TV and stereo.

(d) Household 4

	No. days	Coverage
Smart meter data (26 Jul 2012 – 31 Jan 2013)		
	219	99.39%
Plug data (26 Jul 2012 – 23 Jan 2013)		
Refrigerator	194	97.01%
Kitchen appliances[1]	194	96.81%
Lamp[2]	170	93.54%
Stereo and laptop	169	90.98%
Freezer	192	93.08%
Tablet	189	93.6%
Entertainment[3]	186	94.69%
Microwave	194	97.08%
Occupancy		
Summer	38 d	
Winter	48 d	

[1] Coffee machine, bread baking machine, toaster.
[2] Lamp in the basement triggered by presence detector.
[3] Entertainment consists of TV and stereo.

(e) Household 5

	No. days	Coverage
Smart meter data (27 Jun 2012 – 31 Jan 2013)		
	215	99.05%
Plug data (27 Jun 2012 – 31 Jan 2013)		
Tablet	218	97.87%
Coffee machine	218	95.16%
Fountain[1]	71	99.43%
Microwave	218	97.87%
Refrigerator	218	97.87%
Entertainment[2]	192	89.14%
PC[3]	218	97.87%
Kettle	25	76.64%
Occupancy		
Summer	43 d	
Winter	31 d	

[1] The fountain ran a pump throughout the day and was illuminated from 7 p.m. to 10 p.m.
[2] Entertainment consists of TV and stereo.
[3] Includes router, SheevaPlug computer, and printer.

(f) Household 6

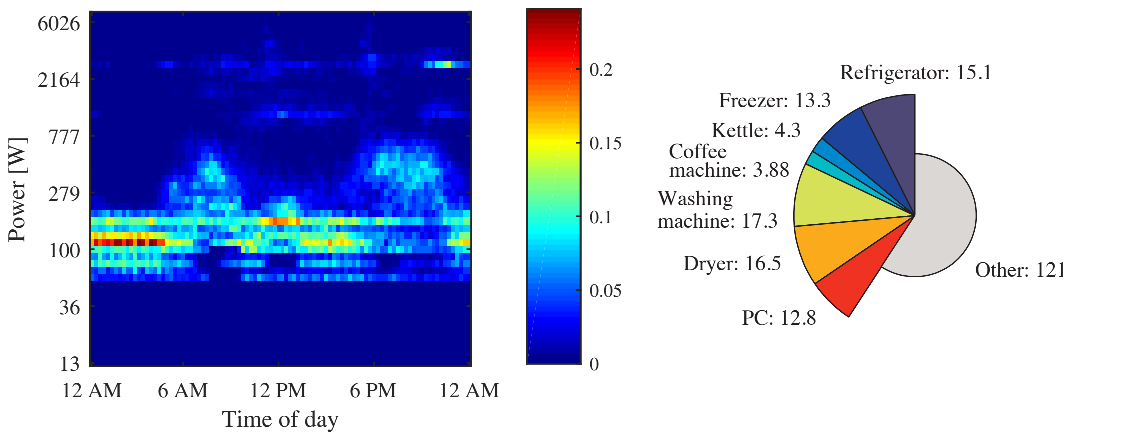
	No. days	Coverage
Smart meter data (27 Jun 2012 – 31 Jan 2013)		
	166	99.67%
Plug data (27 Jun 2012 – 31 Jan 2013)		
Lamp[1]	166	67.2%
Laptop	185	97.3%
Router[2]	88	96.73%
Coffee machine	179	86.03%
Entertainment[3]	181	95.86%
Refrigerator	179	95.78%
Kettle	147	82.54%
Occupancy		
No occupancy data available.		

[1] Includes printer.
[2] Includes two routers and a SheevaPlug computer.
[3] Entertainment consists of TV and stereo.

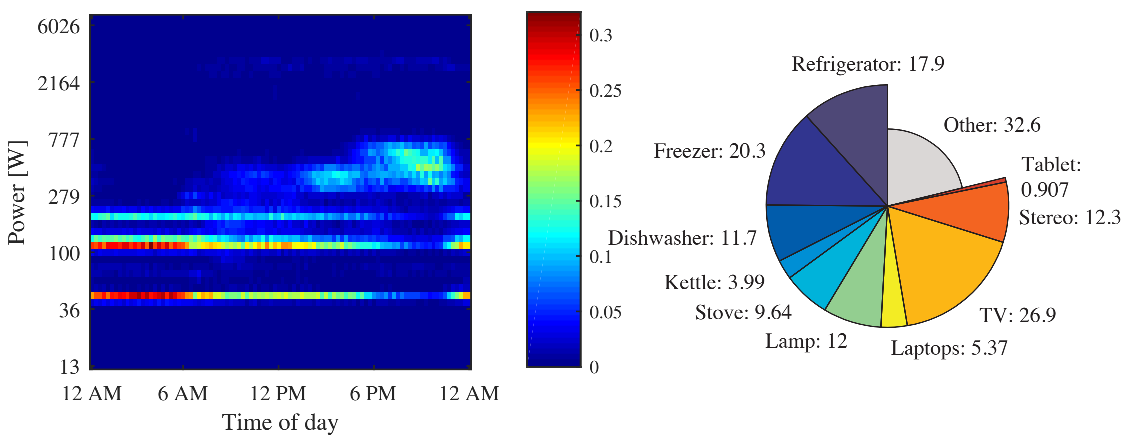
Household 1

Table 4.7a shows that the plugs connected to the refrigerator or to the freezer show both a large number of days with valid measurements as well as high coverage during the days measured. The number of days measured is lower for the coffee machine and the PC, because the corresponding plug was changed from the PC to the coffee machine in the middle of the study. Overall, the coverage is high (*i.e.*, $> 90\%$) for most of the plugs and a bit lower (*i.e.*, 75%–85%) for the plugs connected to the coffee machine, the PC, and the kettle. The house was quite large and concrete walls caused problems for the 802.15.4 connection, which resulted in timeouts.

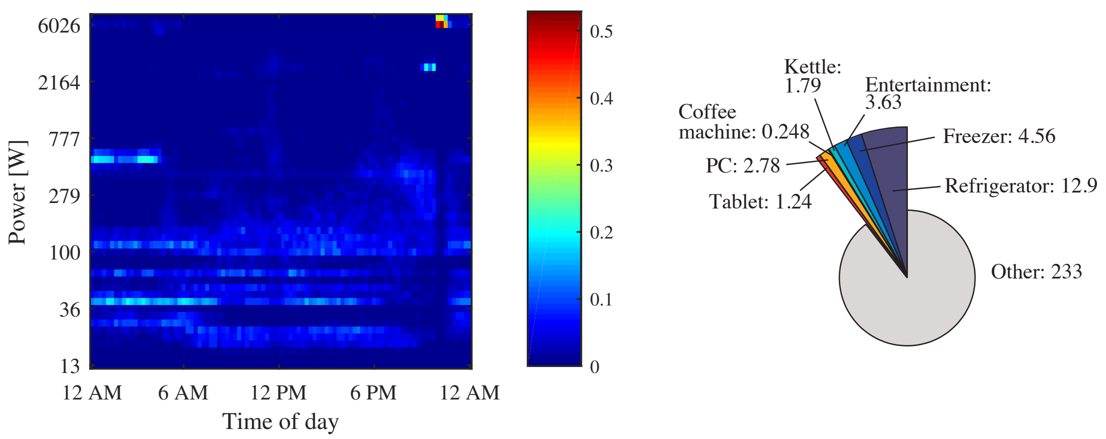
Monitoring the electricity consumption throughout the day (*cf.* figure 4.13a) shows a high morning and evening consumption as well as an increased consumption around noon. The latter one is probably caused by cooking activity, given that household 1 contains two small kids (aged 1 and 3) and a housemaker. The consumption breakdown shows that roughly 40% of household 2's consumption is covered by smart plugs. This is due to the fact that the house is relatively large and we focused on instrumenting certain areas of the house such as the basement and the kitchen, thus ignoring appliances like lamps and the television (TV).



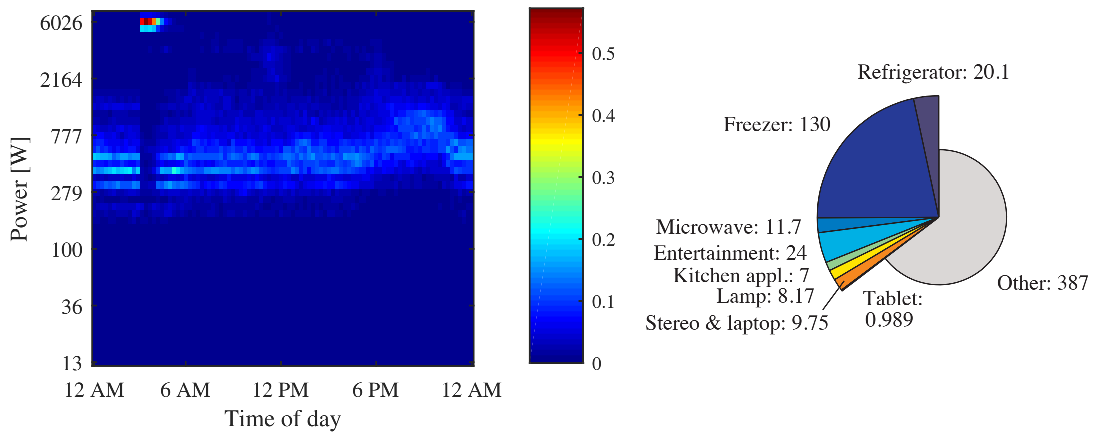
(a) Household 1.



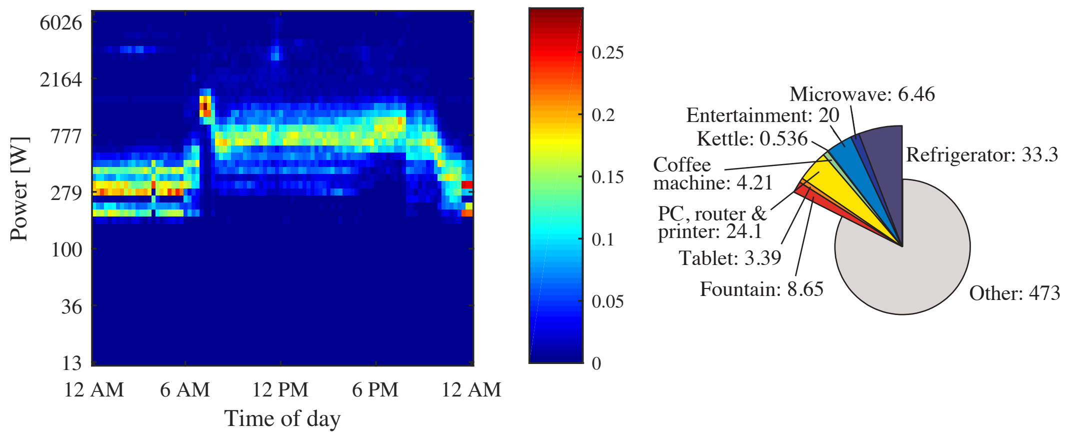
(b) Household 2.



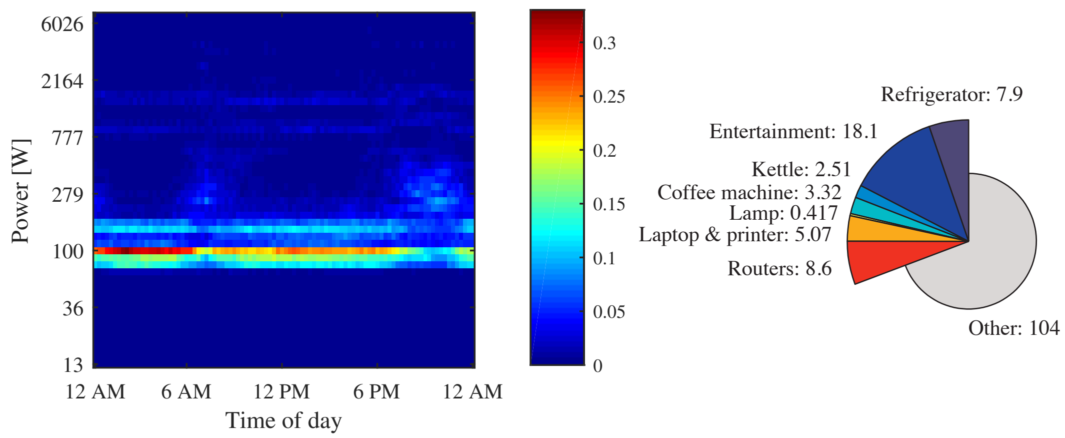
(c) Household 3.



(d) Household 4.



(e) Household 5.



(f) Household 6.

Figure 4.13: *Left side*: Probabilistic electricity consumption over 15-minute time slots for each of the six households of the ECO data set. *Right side*: Electricity consumption covered by the smart plugs. Each of the values represents kWh per month consumed by the appliance.

Household 2

Household 2 shows a high coverage for the smart meter and nearly all the smart plugs (*cf.* table 4.7b). The laptops and the kettle, which stand out in the table with 83% and 90% coverage, respectively, were connected to a switchable socket that was switched off from time to time. The stove in household 2 was not directly connected to a smart plug. To improve the overall coverage of household 2, we "created" plug-level data for the stove by manually estimating its consumption based on the aggregate electricity consumption data. We identified the stove's switching events by searching for time instants with switching events that occur on two phases simultaneously, utilizing the fact that the stove is the only appliance in the household that consumes electricity from more than one phase. To avoid assigning events caused by two appliances that run at the same time on different phases, we observed those time periods when occupants were cooking by collecting consumption data from the air exhaust located above the stove. The measurements of the stove are nearly complete from 2 June 2012 until 20 July 2012. Similarly, the entertainment system consists of a stereo system and of a TV. We "manually" disaggregated those two into plugs, whereas the consumption data attributed to the stereo system also contains the consumption data from the Blu-ray player and the video recorder. For more details, we refer to the helper functions of the NILM-Eval project [189] and to the documentation of the data set [190].

In household 2, as shown in figure 4.13b, roughly 80% of the electricity consumption is covered by the smart plugs. The remaining 20% probably stems from lighting (except the floor lamp, which we instrumented with a smart plug) and from mobile appliances such as the vacuum cleaner. We also did not instrument the appliances in the bathroom and sleeping room with smart plugs. The probability distribution of the aggregate consumption data shows two distinct consumption values at roughly 50 W and 120 W. These two patterns are probably caused by the refrigerator and the freezer (in addition to standby power) during absence of the occupants, when no other appliance is running. The plot further shows that the occupants were mainly active in the evening (between 6 p.m. and 11 p.m.). Indicators for this observation are the high aggregate consumption during this time period as well as the fact that the TV consumes 18% of the overall electricity. In addition to being the household with the highest plug coverage, household 2 was most active in specifying their occupancy. Overall, they specified occupancy on 128 days of the study.

Household 3

Household 3, in contrast, exhibits low coverage for all of the appliances both with respect to the number of days for which measurements are available and with respect to the coverage per day (*cf.* table 4.7c). This is due to the concrete ceiling in the basement, which disturbs the 802.15.4 connection between the SheevaPlug and the smart plugs. For this reason, we omit household 3 in the rest of our study and focus on the remaining five households instead.

Figure 4.13c shows a slight increase of consumption during lunch and dinner hours. Most noticeable in the figure that shows the probabilistic consumption is the spike at 6000 W at around 11 p.m., which is caused by the boiler that heats water during the night. The boiler is also one of the reasons for the large amount of unattributed consumption shown on the right side of the figure.

Household 4

As table 4.7d shows, coverage of the deployed plugs is relatively high in household 4 (*i.e.*, almost 200 days and more than 90% per plug). Figure 4.13d shows a similar consumption pattern than the one we observed for household 3: It shows a consumption increase during lunch and dinner time as well as a spike caused by the boiler, which in this household runs at roughly 4 a.m. The consumption breakdown shows that the freezer consumes almost a fourth of the household's electricity consumption. It consumes 130 kWh, which is about ten times as much as the freezers in the other households consume. Household 4 runs another refrigerator in their basement, which contributes to the aggregate electricity consumption, which we however were not able to equip with a smart plug.

Household 5

Except for the kettle and the fountain—the plug connected to the fountain was switched to the kettle on 6 September 2012—coverage per plug is relatively high for household 5 with more than 200 days and more than 95% coverage per day for most of the plugs (*cf.* table 4.7e). However, the consumption breakdown provided by figure 4.13e shows that a high proportion of the aggregate consumption is not attributed to by any of the plugs. The non-attributed consumption for household 5 is high because, for instance, the household uses a time-triggered water pump for the fountain located in the garden, which is not covered by a plug and which consumes 300 W during daytime. This is also reflected in the probabilistic aggregate consumption pattern of household 5. In addition, the pattern shows a spike at roughly 1000 W at 7 a.m., which is most likely caused by the coffee machine when the occupants are getting up in the morning. The consumption breakdown does not show a freezer. This is because the freezer was located in the basement without a 802.15.4 connection to the SheevaPlug.

For household 5, we received detailed information from the home owner about the number, type, and consumption of electrical appliances, which we did not or only partially measure with the plugs. We compiled a version translated from German in appendix C. The fountain in the house, for instance, is illuminated using a timer, which is active daily from 7 p.m. to 10 p.m. The fountain in the garden runs a pump, which consumes 300 W from 8 a.m. to 8 p.m., and which is illuminated, which consumes 70 W from sunset to 8 p.m. and from 6 a.m. to sunrise. The household further runs a thermostat-controlled heated water bed, which is switched on during the day and switched off from 10 p.m. to 7 a.m. Water heating is performed using either a pump (95 W) for the solar system or a boiler that consumes 6000 W. Both the wood-based and the sun-based space heating system run several pumps that add to the aggregate electricity consumption and—together with the other appliances—explain the high baseload of the household.

Household 6

Household 6 exhibits high coverage for all of the plugs except for the plug connected to the lamp. This plug provided measurements for 166 days but was only running two thirds of the time, probably due to a light switch that also switched off power of the plug (*cf.* table 4.7f). The probabilistic aggregate consumption (*cf.* figure 4.13e shows an increased consumption in the morning at around 7 a.m. and during the evening from 7 p.m. to 12 a.m. Due to the lack of cooking activity during lunch time we conclude that both occupants are working during the day. The consumption breakdown on the right side of the plot shows only a single cooling appliance (*i.e.*, the refrigerator). Although we connected a plug to the freezer as well, we excluded the data from the ECO data set because it was not a part of the aggregate consumption measured by the smart meter.

4.4.4 ON/OFF events

Some of the NILM algorithms require as input binary information on when individual appliances are running. To this end, we use the appliances' consumption data collected by the smart plugs to identify when appliances have been switched on or off (*i.e.*, their ON/OFF events). Intuitively, an appliance is running when its consumption is higher than a given threshold. However, appliances such as the dishwasher (*cf.* figure 4.14a) or the washing machine (*cf.* figure 4.14b) have idle times during a run, which are time periods in which they do not consume electricity.

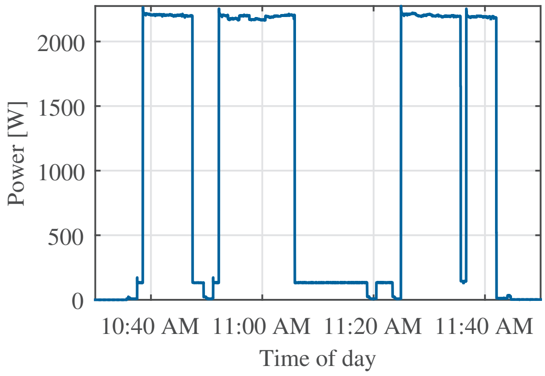
It is our goal to enclose each run of an appliance with an ON event at the beginning of the run and an OFF event at the end, which for such appliances includes idle times within a run. As the consumption patterns and the notion of ON/OFF events differ from appliance to appliance, we define four parameters for each appliance to automatically annotate the consumption data with 1 (*i.e.*, the appliance is running) or 0 (*i.e.*, the appliance is not running):

- *Power threshold* defines the minimum consumption increase (or decrease) between two consecutive measurements that is required such that the second of the two measurement counts as an ON event (or OFF event).

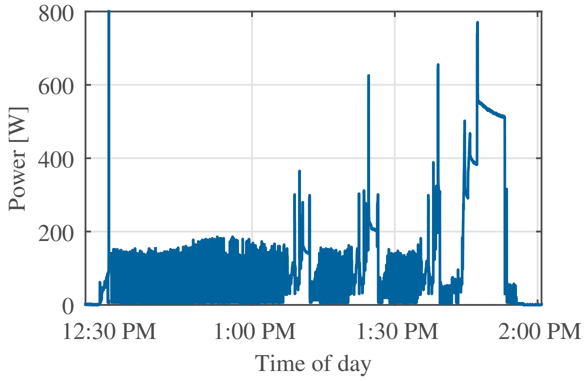
- *On time threshold*: The minimum time an appliance stays in ON state until it goes back into OFF state. This threshold avoids that drops in the consumption, for instance caused by the spikes that occur at the beginning of a refrigerator's cooling cycle (*cf.* figure 4.15), are classified as OFF events. It further avoids assigning OFF events to drops that occur before idle periods, such as the ones shown by the consumption pattern of the dishwasher (*cf.* figure 4.14a).

- *Idle time* and *idle power* are parameters to identify ON and OFF events for appliances with very fluctuating consumption patterns such as the washing machine (*cf.* figure 4.14b). A potential ON event (or OFF event) is only classified as such if the average consumption in the time window between the potential event and *idle time* before (or after) the event is lower than *idle power*.

Since both the consumption pattern of different appliances and the consumption pattern of different appliance types varies, these thresholds must be set for each household and for each appliance individually. Table 4.8 shows the thresholds we defined for household 2 in the ECO data set.



(a) Dishwasher.



(b) Washing machine.

Figure 4.14: Consumption pattern of the dishwasher (household 2) and the washing machine (household 2) on 23 June 2012 and 7 July 2012, respectively.

Table 4.8: Thresholds for the automatic detection of ON/OFF events from appliance-level consumption data.

Appliance	**Thresholds**			
	Power	**On time**	**Idle time**	**Idle power**
Dishwasher	30 W	1800 s	300 s	11 W
Air exhaust	30 W	10 s	600 s	7 W
Refrigerator	40 W	300 s	180 s	7 W
TV	120 W	600 s	300 s	70 W
Stereo	20 W	600 s	300 s	8 W
Freezer	30 W	240 s	120 s	7 W
Kettle	70 W	10 s	180 s	12 W
Lamp	20 W	20 s	120 s	9 W
Laptops	14 W	180 s	180 s	7 W

4.5 Results

In this section we present the results of the performance evaluation of the five NILM algorithms described in section 4.2. We compute the performance measures for each algorithm based on the algorithm's output, which can be switching events of appliances or their estimated consumption. The results show that (semi-)supervised algorithms perform better than unsupervised ones. This is mainly because the latter fail to identify consumption patterns of individual appliances in the aggregate consumption data. We further show that a data granularity of 1 Hz is required to reliably detect switching events of appliances. Weiss' algorithm, for instance, achieves F_1 scores up to 0.92 when detecting events of cooling appliances or appliances with high changes in their consumption patterns.

The F_1 scores obtained by Parson's algorithm, which relies on data at $1/60\,\text{Hz}$ frequency, are much lower and range from 0.51 to 0.80.

A comparison of the performance of the algorithms is however difficult—and even unfair—due to the differences on input data and training methods they rely upon as well as due to different types of output they provide. Weiss' algorithm, for instance, returns event labels, and Parson's and Kolter's algorithms return an estimate of each appliance's electricity consumption. In order not to compare apples with bananas, we present the results of the five algorithms individually and present a summary of the evaluation at the end of this section. For implementation details and parameter specifications other than the ones listed in this section, we refer to [39], to [99], and to the NILM-Eval project [189].

4.5.1 Parson's algorithm

Like Parson *et al.* [135], we aggregated the data to a granularity of $1/60\,\text{Hz}$. We first evaluate Parson's algorithm by inferring the electricity consumption of the refrigerator for each of the five households, because the refrigerator is the only appliance that was measured by a plug in every household. Next, we evaluate the microwaves in households 4 and 5 representing appliances with switching events that change the power consumption by at least 500 W.

Refrigerators

Figure 4.15 shows the consumption patterns of each of the five household's refrigerator on 3 July 2012. According to those patterns, we model a refrigerator as an appliance with two states (*i.e.*, ON and OFF states). To define the generic model required by Parson's algorithm, we performed different initial experiments evaluating the effect of the required parameters. Ultimately, we assume the emission probabilities for each state to be Gaussian distributed and describe the ON state with $\mu_{on} = 60\,\text{W}, \sigma^2_{on} = 40\,\text{W}^2$ and the OFF state with $\mu_{off} = 2\,\text{W}, \sigma_{off} = 5\,\text{W}^2$. We further define the transition probabilities as $\varphi_{on,off} = 0.2$ (*i.e.*, the probability that the state changes from ON to OFF) and $\varphi_{off,on} = 0.05$. To adapt the generic model to specific appliance models, we use ten training days and a training window length of 3600 s. The experiments are then performed five times (using different training periods) over 90 days of consumption data. Finally, to increase robustness of the algorithm to "noise" caused by unmodeled appliances, we set the likelihood threshold introduced by Parson *et al.* to 0.0001.

Table 4.9 shows the disaggregation results we obtained by inferring the electricity consumption of the refrigerator from the aggregate electricity consumption of five of the

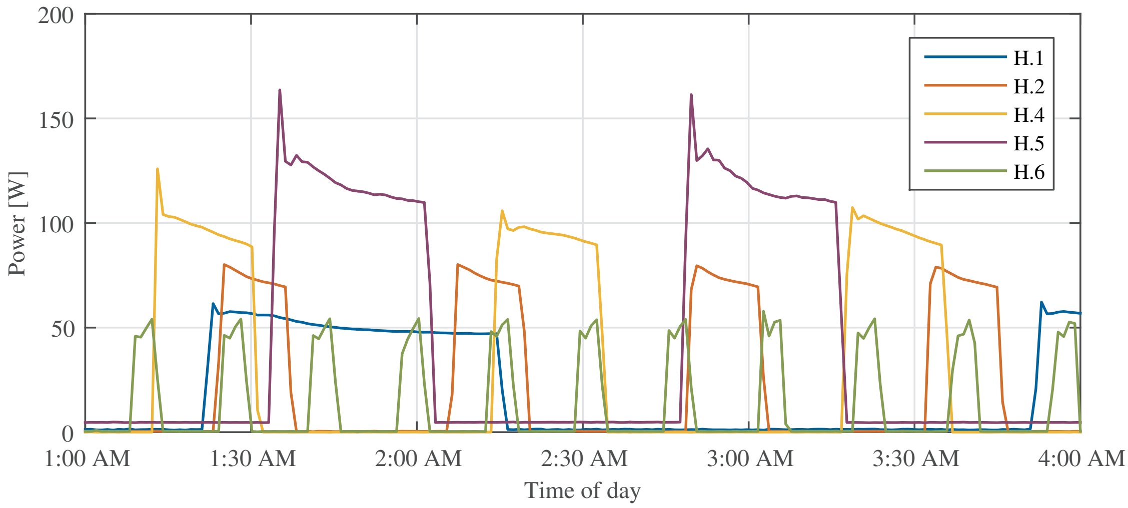


Figure 4.15: Ground truth consumption patterns of the five refrigerators in the ECO data set on 3 July 2012 (aggregated to 1-minute averages). Labels *H.1* to *H.6* indicate households 1–6.

households using the configuration described above. Training type *none* means that the generic model is used as input to the Viterbi algorithm. *Plug* training uses ground truth data to build a specific appliance model from the general model, whereas *aggregate* training aims at building the specific appliance model from the aggregate electricity consumption. Each value denotes the average of the five runs we performed. In terms of F_1 score, Parson's algorithm performs best (0.80) for household 6. However, this includes training on the sub-metered data. In real world settings (*i.e.*, training with aggregate consumption data), the algorithm performs slightly worse with $F_1 = 0.77$. Household 6 also exhibits the lowest RMSE (17 W) compared to the other households (which range from 29 W to 62 W). However, the refrigerator of household 6 is the most energy-efficient among all refrigerators in the data set, which explains that there is still a relatively large deviation of the estimation compared to the ground truth (62%).

Overall, the estimation performs better for household 6 compared to the other households. We believe this is due to the fact that the aggregate consumption in household 6 does not contain the electricity consumed by the household's freezer. As figure 4.16 shows, freezers have a consumption pattern that is difficult to distinguish from a refrigerator's consumption pattern at a 1-minute granularity. Training on aggregate data performs slightly worse than training on plug-level data. A possible explanation is that the generic model defined above is already close to the optimal model, because we performed initial experiments to carefully define the generic model.

One of the parameters we set based on these initial experiments is μ_{on}, which defines the mean consumption of the refrigerator's ON state in the HMM. To this end, we dis-

Table 4.9: Performance of Parson's algorithm on the ECO data set (disaggregating the consumption of the refrigerators). Labels *H.1* to *H.6* indicate households 1–6.

Metric	**Training**	**H.1**	**H.2**	**H.4**	**H.5**	**H.6**
	none	**0.65**	**0.64**	**0.51**	**0.52**	0.77
F_1 score	plug	0.42	0.60	0.47	0.47	**0.80**
	aggregate	0.42	0.46	0.40	-	0.71
	none	33 W	**37 W**	**48 W**	**62 W**	23 W
RMS	plug	**29 W**	41 W	62 W	75 W	**17 W**
	aggregate	34 W	45 W	64 W	-	20 W
	none	0.61	0.50	**0.23**	0.30	0.99
Dev	plug	**0.31**	**0.46**	0.75	**0.27**	**0.62**
	aggregate	0.48	0.49	0.52	-	0.77

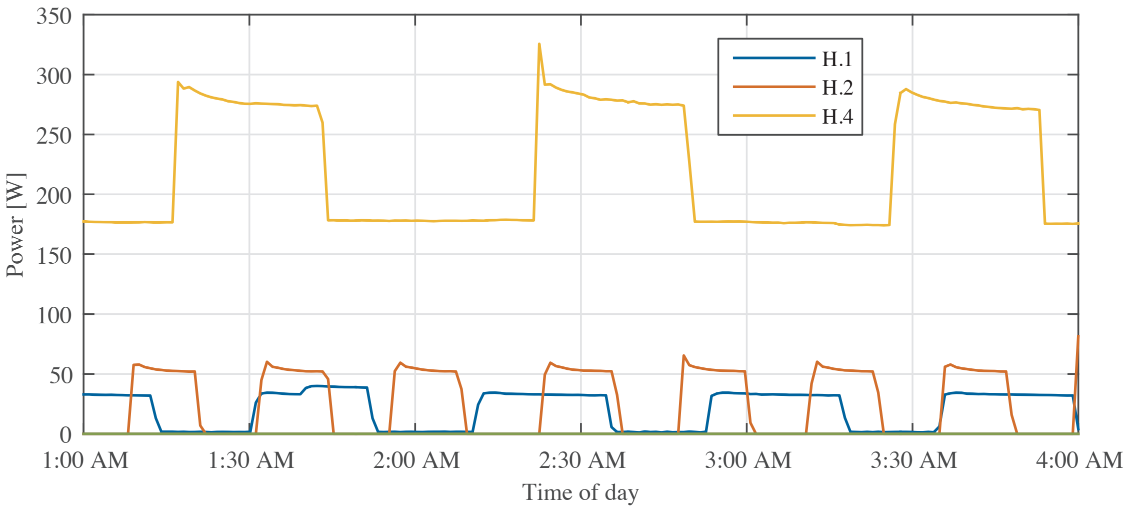


Figure 4.16: Ground truth consumption patterns of the three freezers in the ECO data set on 3 July 2012 (aggregated to a 1-minute resolution). Labels *H.1* to *H.4* indicate households 1–4.

aggregated the consumption of the refrigerator using different values for μ_{on} from 20 W to 150 W, leaving the other parameters (*i.e.*, μ_{off}, σ, φ, and the likelihood threshold) constant as described above. We performed similar experiments for the other parameters, but we omit the description of these results for space reasons and refer to [39] instead. Figure 4.17 shows the results of the experiments for parameter μ_{on}. It shows that disaggregation of the refrigerator in household 6 provides best performance for $\mu_{on} = 30$ W, whereas other households reach their maximum at around 70 W to 80 W. These results illustrate the importance of the generic model used as input for the algorithm. In practice, it is thus a challenge to identify parameters such as μ_{on} without a training process.

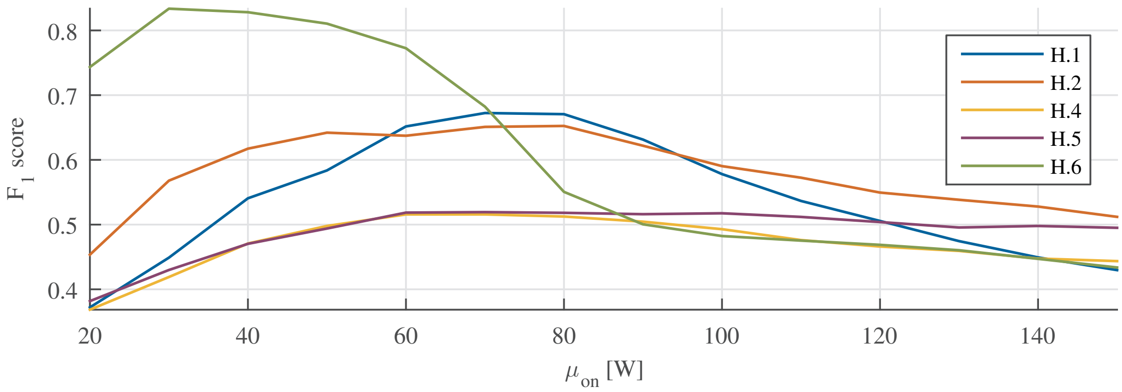


Figure 4.17: F_1 scores obtained when disaggregating the consumption of the refrigerators in five different households using a variety of μ_{on} parameters. μ_{on} defines the mean consumption of the refrigerators' ON state in the HMM of the generic appliance model. Labels *H.1* to *H.6* indicate households 1–6.

Microwave

To evaluate the disaggregation of the microwave's electricity consumption, we rely on the default configuration provided by Parson *et al.*'s implementation (*i.e.*, two states, $\mu_{on} = 1700\,\mathrm{W}$, $\sigma^2_{on} = 1000\,\mathrm{W}^2$, $\mu_{off} = 4\,\mathrm{W}$, $\sigma^2_{off} = 100\,\mathrm{W}^2$, $\varphi_{on,off} = 0.3$, $\varphi_{off,on} = 0.01$, and a likelihood threshold of 0.00001). In addition—in order to improve the performance of the estimation—we pre-processed the aggregate consumption data by replacing those edges that span over more than two time steps by "sharp" edges that span exactly two time steps. With this configuration we performed five runs using aggregate training. On average, Parson's algorithm achieves F_1 scores of 0.14 and 0.031 for households 4 and 5, respectively. These low F_1 scores are caused by very low precision values, which are 0.10 and 0.017 for households 4 and 5, respectively. Parson's algorithm overestimates the consumption of the microwave as it often infers the microwave is running even though it is switched off. By analyzing only the consumption on the phase on which the microwave is running, the F_1 scores improve to 0.18 and 0.055 for households 4 and 5, respectively.

There are two reasons for the low F_1 scores obtained when disaggregating the consumption of the microwave: lack of training and data granularity. Figure 4.18 illustrates the consumption of the microwave in households 4 and 5 with both 1-second and 1-minute granularity. First, the plot shows that the consumption pattern of the microwave in household 4 (blue and orange lines) differs significantly from the pattern of the microwave in household 5 (purple and green lines). This means that a training phase or a household-specific description of the appliance model is required. Second, under 1-minute granularity, some characteristics of the patterns are not visible compared to the consumption patterns for the 1-second data. For this reason, the precision of the results is low because the consumption patterns of other appliances with similar mean consumption can be wrongly attributed to the microwave.

We repeated the experiments for both refrigerator and microwave, testing a variety of appliance models in order to separate the effect of training the appliance models from the actual inference. The best F_1 scores achieved by the algorithm when disaggregating the consumption of the refrigerator range from 0.54 (household 4) to 0.84 (household 6). In case of the microwaves in households 4 and 5, the algorithm achieved maximum values of 0.29 and 0.14, respectively.

Overall, we see the following challenges: First, consumption patterns of appliances differ considerably, which makes it difficult to define a general model that represents all appliances of a certain appliance type. Second, disaggregating each appliance in isolation leads to errors due to overlapping consumption patterns. Thus it would be interesting to apply Kolter's AFAMAP algorithm [108] to infer the consumption of multiple appliances simultaneously. Finally, due to aggregation of the consumption data to $1/60$ Hz, lots of details in the consumption patterns are lost. Applying Parson's algorithm on 1 Hz data, however, requires significant changes to the algorithm as well as pre-processing of the data to avoid long periods without state changes.

4.5.2 Baranski's algorithm

We applied the unsupervised algorithm of Baranski and Voss on 30 days of aggregate 1 Hz consumption data from household 2. We set the number of resulting clusters to 20 and specified that each appliance consists of two states with a maximum length of the ON state set to 3600 s. Based on the results of our initial experiments, we decided not to assign weights to the length of a switching event as well as to the boost in electricity consumption that can occur when an appliance is switched on.

Table 4.10 shows the clusters that result from the experiment. Each cluster denotes a set of switching events that have a similar increase (or decrease) in electricity consumption. Column *Size* shows the number of events in a cluster. By comparing the timestamps of the events with the plug-level consumption data, we assigned each event to an appliance if possible. Column *Prop.* in the table shows the proportion of events assigned to the appliance named in the previous column divided by the overall number of events in the cluster. Columns *Appliance 1* and *Appliance 2* illustrate which appliances have the highest and the second highest number of assigned events in a cluster, respectively. The laptop, freezer, refrigerator, and stove are appliances that are often represented in the clusters. Clusters C_1 and C_2 almost exclusively contain start and stop events of the laptop. The events of the stove are spread across multiple clusters (*i.e.*, clusters C_6, C_7, C_8, C_{10}, C_{11},

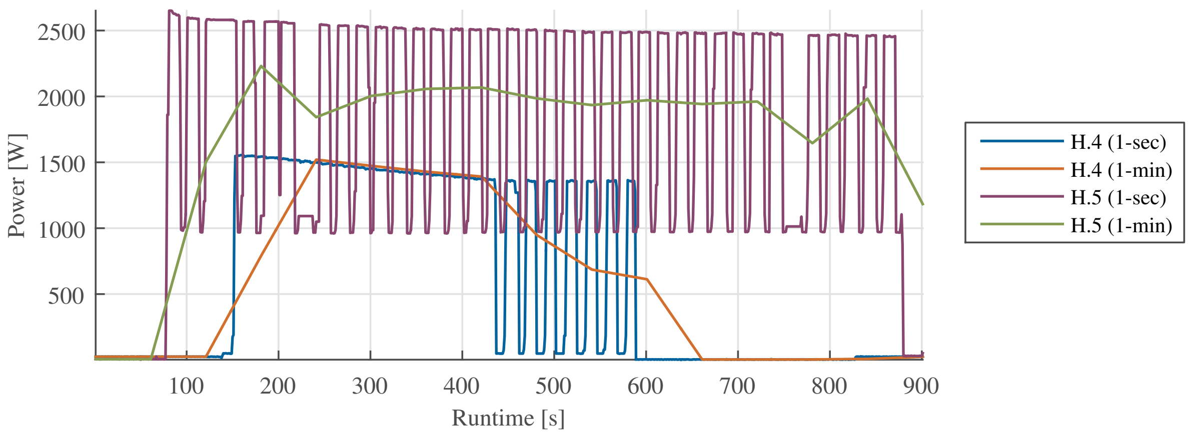


Figure 4.18: Illustration of the microwave's electricity consumption from household 4 (*H.4*) and household 5 (*H.5*) under two different data granularities.

Table 4.10: Event clusters in household 2 provided by Baranski's algorithm.

Cluster	ΔP	Size	Appliance 1	Prop.	Appliance 2	Prop.
C_1	−11 W	8963	Laptops	28%	Refrigerator	3%
C_2	11 W	8724	Laptops	31%	Refrigerator	3%
C_3	−58 W	3009	Freezer	41%	Refrigerator	28%
C_4	73 W	1960	Freezer	51%	Refrigerator	5%
C_5	93 W	1003	Refrigerator	69%	Freezer	2%
C_6	−1837 W	260	Stove	21%	Kettle	15%
C_7	1857 W	253	Stove	21%	Kettle	14%
C_8	1249 W	225	Stove	26%	Laptops	2%
C_9	−176 W	210	TV	5%	Freezer	4%
C_{10}	−1235 W	199	Stove	35%	Laptops	3%
C_{11}	2425 W	187	Stove	17%	Laptops	3%
C_{12}	−2365 W	155	Stove	26%	Dishwasher	6%
C_{13}	−509 W	122	Freezer	14%	Refrigerator	9%
C_{14}	−783 W	102	Refrigerator	18%	Freezer	18%
C_{15}	596 W	97	Freezer	10%	Refrigerator	5%
C_{16}	850 W	88	Freezer	16%	Refrigerator	11%
C_{17}	375 W	83	Freezer	12%	Refrigerator	4%
C_{18}	1064 W	60	Refrigerator	13%	Stove	5%
C_{19}	−1023 W	56	Refrigerator	9%	Laptops	5%
C_{20}	−3391 W	39	Stove	5%	Refrigerator	3%

and C_{12}), because the change in electricity consumption of the stove events varies. Note that clusters C_6 and C_7 are also populated to a large extent by switching events of the kettle, which makes forming a state machine for the stove and for the kettle difficult. The events of the refrigerator and the freezer are also spread over multiple clusters.

Based on the clustering results, Baranski's algorithm generated the FSMs shown in table 4.11. The second and third columns denote the power steps of the centroids of the two clusters that form the FSM. The next two columns list the number of sequences represented by the FSM as well as their average duration. Column *Appliance* denotes the appliance that is most likely represented by the FSM. Note that this labeling has been performed manually. The first FSM consists of events from clusters 1 and 2 and therefore represents the stove or the kettle. The third FSM represents the laptop, which is the only appliance that is clearly separable from the other appliances. FSMs 2, 4, and 10 also represent the stove, whereas FSMs 6, 8, and 11 represent the freezer. In case of the freezer, the first two FSMs exhibit high consumption and last only shortly, which is why we assume they are caused by the initial spike in the consumption pattern at the beginning of a cooling cycle. The refrigerator is not represented in the list of FSMs. The reason is that, although the ON event of the refrigerator is well-represented in cluster C_5, the corresponding OFF events are spread over multiple clusters. Baranski's algorithm computed a quality score for each FSM and thus discarded the FSM(s) that represent(s) the refrigerator due to a low quality score.

In practice, Baranski's algorithm requires the user to manually label the resulting FSMs without the assignments of events to appliances as provided in table 4.10. Even so, the algorithm generates multiple FSMs for some of the appliances due to the fact that some of the clusters contain events from multiple appliances. Therefore, there is a large ambiguity in the assignment of appliances to the FSMs. Possible improvements include (a) to allow creating an FSM using events from different clusters to reduce the number of FSMs, and (b) to improve the clustering procedure. Possible improvements of the clustering include using real and reactive power to make events of different appliances more distinguishable, or to apply post-processing that divides or combines clusters (*e.g.*, on the basis of the number of events in each cluster).

4.5.3 Weiss' algorithm

We use the 1 Hz consumption data of household 2 including real power, fundamental reactive power, and distortion power split into individual phases to evaluate Weiss' algorithm. We investigate appliances of the three categories (1) cooling appliances, (2) appliances with high consumption (*i.e.*, dishwasher, kettle, stove), and (3) remaining appliances (*i.e.*, lamp, laptop, TV, stereo system). The analysis is based on 90 days of consumption data plus 15 days of data used for training. In the training process, we extract timestamps of

Table 4.11: FSMs provided as a result by Baranski's algorithm for household 2.

FSM	**ΔP(C$_1$)**	**ΔP(C$_2$)**	**Sequences**	**Duration**	**Appliance**
1	1857 W	−1837 W	276	115 s	Stove or kettle
2	1249 W	−1235 W	312	14 s	Stove
3	11 W	−11 W	12,066	14 s	Laptops
4	2425 W	−2365 W	260	10 s	Stove
5	1064 W	−1023 W	56	63 s	?
6	850 W	−783 W	144	12 s	Freezer
8	596 W	−509 W	138	17 s	Freezer
10	1249 W	−1023 W	30	78 s	Stove
11	73 W	−58 W	3032	627 s	Freezer

switching events (*i.e.*, changes in power consumption above 5 W) from the plug data and extract the signature from the smart meter data at these timestamps.

Table 4.12 shows the signatures of household 2's appliances extracted from the aggregate consumption data. For each appliance, the table shows the change in real power (Δ Real) and reactive power (Δ Reactive)[5] at ON or OFF events. Column *Phase* illustrates on which phase the appliance is running. For cooling cycles, a switching event denotes the beginning or the end of the cooling cycle. For appliances with switching events of more than 500 W real power, only these events are considered during training and recognition phase. For the other appliances, each event with more than 5 W is considered a switching event. The table shows that the events of the refrigerator and of the freezer have differences in both real power (16 W to 18 W difference on average) and reactive power (22 VA to 24 VA), which is a good property in order to be distinguished by Weiss' algorithm. The stove runs both on phase 1 and on phase 2, the TV and stereo system run on phase 2, and the other appliances run on phase 1.

Table 4.13 illustrates the results achieved by Weiss's algorithm for each of the appliances. Identifying the switching events of the refrigerator and the freezer is possible with F_1 scores of 0.92 each. In case of the freezer, the algorithm misses only 196 out of 5992 switching events, which is a precision of 0.98. Events from appliances with high consumption, namely dishwasher, kettle, and stove, are recognized with almost no false positives, leading to a precision of 0.95, 0.95, and 1.0, respectively. However, the algorithm misses a large number of events for these appliances. This is why the F_1 scores are 0.56, 0.75, and 0.25, respectively. The remaining appliances exhibit relatively low F_1 scores. In case of the lamp, this is due to the fact that household 2 has a dimmable lamp, which means that the power steps caused by switching events vary. The laptop and the stereo system are

[5] Total reactive power as computed by equation 4.10 in section 4.4.2.

Table 4.12: Signatures of cooling appliances (top), appliances with high consumption (center), and other appliances (bottom) in household 2.

Appliance	**Event**	**Δ Real**	**Δ Reactive**	**Phase**
Refrigerator	OFF	−69 W	−6 VA	1
Refrigerator	ON	80 W	4 VA	1
Freezer	OFF	−52 W	17 VA	1
Freezer	ON	64 W	−20 VA	1
Dishwasher	OFF	−2058 W	4 VA	1
Dishwasher	ON	2060 W	−18 VA	1
Kettle	OFF	−1881 W	2 VA	1
Kettle	ON	1853 W	−4 VA	1
Kettle	ON	1884 W	4 VA	1
Stove	OFF	−903 W	−519 VA	1&2
Stove	ON	626 W	315 VA	1&2
Lamp	OFF	−185 W	−111 VA	1
Lamp	OFF	−185 W	−216 VA	1
Lamp	ON	222 W	91 VA	1
Lamp	ON	127 W	87 VA	1
Laptops	OFF	−20 W	−3 VA	1
Laptops	ON	23 W	10 VA	1
TV	OFF	−166 W	−36 VA	2
TV	ON	159 W	30 VA	2
TV	ON	161 W	33 VA	2
Stereo	OFF	−17 W	−12 VA	2
Stereo	ON	56 W	49 VA	2

difficult to reliably identify because their power consumption is very low (*i.e.*, 38 W on average for two laptops and 54 W on average for the stereo system including the Blu-ray player and video recorder) and can be easily confused with switching events or variations caused by other appliances.

Overall, Weiss' algorithm performs well for cooling appliances and for appliances with high power consumption. In the latter case, the precision is very high, but the algorithm misses many events and thus exhibits low recall values. The algorithm includes a scaling parameter r to control the maximum distance of a switching event to the (possibly) corresponding signature. Increasing r leads to a higher recall value, because the algorithm identifies more switching events of a particular appliance. However, this also results in a higher number of false positives. The optimal value of r can be determined per appliance as a part of the training process. In [39], we show the effect of r when disaggregating the refrigerator, the laptops, and the lamp in household 2. To further reduce the number

Table 4.13: Performance results achieved by Weiss' algorithm on consumption data from household 2.

	F_1 score	Precision	Recall	TP	FP	FN
Refrigerator	0.92	0.93	0.91	4855	385	477
Freezer	0.92	0.98	0.86	5948	137	947
Dishwasher	0.56	0.95	0.39	115	6	178
Kettle	0.75	0.95	0.62	122	6	74
Stove	0.24	1.0	0.14	28	0	209
Lamp	0.30	0.37	0.25	23	39	68
Laptops	0.11	0.10	0.12	63	593	498
TV	0.37	0.89	0.24	90	11	291
Stereo	0.10	0.23	0.06	144	471	2148

of false positives, we recommend including additional features such as time of day or the relationship between certain appliances (*e.g.*, the dryer often runs after the washing machine).

4.5.4 Kolter's algorithm

As described in section 4.2, Kolter's algorithm automatically identifies and clusters snippets (*i.e.*, consumption patterns of appliances) before it disaggregates the consumption data. Using the data from household 2, we analyzed seven days of 1 Hz real power consumption data searching for three types of snippets (*i.e.*, snippets with one, two, and three ON states).

The algorithm detected 399 snippets with one ON state, 221 snippets with two ON states, and 136 snippets with three ON states. Most of the snippets with one ON state have a mean power consumption between 0 W and 200 W. Whereas we can attribute many of those snippets to the freezer, the number of snippets we can attribute to the refrigerator (*i.e.*, snippets with mean power consumption between 60 W and 85 W) is relatively low. The reason is that the frequency of the freezer's cooling cycle is almost twice as high as the frequency of the refrigerator's cooling cycle. Thus almost all cooling cycles of the refrigerator interfere with the cooling cycles of the freezer. There are also 44 snippets above 200 W (*i.e.*, snippets with a mean power consumption of 1200 W, 1800 W, and 2400 W). Most of these snippets represent the stove. Thus the stove and the freezer are the only appliances with a single ON state for which we can reliably identify snippets in the consumption data using Kolter's algorithm.

Table 4.14: Centroids of clusters of the snippets with two ON states. P_1 and P_2 denote the mean power consumption of the two ON states of the snippets.

Cluster	P_1	P_2	Snippets
C_1	1263 W	21 W	2
C_2	320 W	733 W	3
C_3	65 W	42 W	23
C_4	7 W	52 W	12
C_5	1278 W	1278 W	2

For the snippets with two or three ON states, we applied k-means clustering to obtain cluster centroids that potentially represent the consumption pattern of individual appliances. Table 4.14 shows the five resulting cluster centroids for the snippets with two ON states. We compared all cluster centroids with the consumption patterns measured by the smart plugs, but we could not find a match between any of the cluster centroids and the consumption pattern of an appliance. The same holds for the snippets with three ON events. Since the application of the spectral clustering method did not result in HMMs that represent individual appliances, we decided to omit Kolter and Jaakkola's second step. Instead, we propose to extract snippets using the plug-level data in order to evaluate the AFAMAP algorithm developed by Kolter and Jaakkola.

4.5.5 Jung's algorithm

For the evaluation of Jung's algorithm we focus on consumption data of a six week period (*i.e.*, from 1 July 2012 to 11 August 2012 from household 2. In addition to the aggregate consumption data, we include ON/OFF events from the following nine appliances: refrigerator, freezer, kettle, laptops, dishwasher, air exhaust, lamp, TV, and stereo. These nine appliances attribute for 62% of the total electricity consumption measured by the smart meter.

Jung's algorithm assumes that ON/OFF states are known for all appliances. However, in the ECO data set, only a portion of the aggregate consumption data is covered by the smart plugs. For this reason, we evaluate three different settings:

- *Original*: Both smart meter data and plug data are used as input to the algorithm, ignoring the fact that the consumption data measured by the plugs does not account for 100% of the household's aggregate consumption.

- *Ghost power*: In this setting, we introduce a virtual consumer that is always running. It is modeled as an additional plug, accounting for the portion of the aggregate consumption that is not covered by other plugs.
- *Aggregated plugs*: In this setting, we compute the aggregate electricity consumption of the household by adding the measurements of the plugs deployed into the household. This artificial setting simulates a scenario in which all appliances in the household are equipped with ON/OFF sensors.

Jung's algorithm operates in rounds and estimates the consumption of each appliance per round, which we set (like the authors) to 3600 s. Over a long time period, over- and underestimations therefore even themselves out to a certain degree. Figure 4.19 shows the consumption breakdowns estimated by Jung's algorithm in each of the three settings described above along with the ground truth data measured by the plugs. For each setting, the consumption displayed in the plot is the sum of the estimated consumption of all the 3600-second slots during the six week period. The reason that the first bar is slightly smaller than the second and the right bars—although the height of all three bars is determined by the consumption measured by the smart meter—lies in the fact that the algorithm (per definition) does not cope with time periods in which no appliance is running but the smart meter data is positive. In the *original* setting, which assumes that all appliances are equipped with ON/OFF sensors, we therefore ignore these time periods, which were rare in our experiments. The results show that the original version of the algorithm overestimates the consumption of most of the appliances. The freezer, for instance, consumes 48.2 kWh according to the algorithm's estimation related to 30.8 kWh measured by the plug. These overestimations are due to the fact that the algorithm distributes 100% of the smart meter data to the appliances equipped with ON/OFF sensors, although other appliances contribute to the smart meter data as well.

The right bar shows the results of the experiments, in which we replaced the smart meter data by the sum of the consumption of all appliances equipped with ON/OFF sensors, which we measured with smart plugs. In this case, the consumption estimated for each appliance is close to the ground truth for many of the appliances. There is however a significant underestimation of the consumption of the stereo system (which—due to restrictions when installing the plugs—also contains the Blu-ray player and the video recorder) as well as an overestimation of the consumption of the TV. The reason is that those two appliances often run together, which makes it difficult for the algorithm to disaggregate their consumption. To avoid relying on plug-level measurements to compute the aggregate consumption, we perform the experiments on smart meter data and introduce a virtual appliance that accounts for the consumption of all the appliances that are not instrumented with ON/OFF sensors. We refer to this experiment as *ghost power* setting and classify the consumption estimated for the virtual appliance as *other* in figure 4.19.

The results show that the estimated consumption is close to the ground truth for most of the appliances. Some portion of the ghost power consumption is attributed to the stereo and the laptops, which are both appliances that run for a long time but do not consume much electricity.

Overall, the results show that Jung's algorithm performs well if all or a high fraction of appliances are equipped with ON/OFF sensors. In the latter case, introducing a virtual appliance that is always switched on improves the performance. With the advent of the IoT, we can assume that more and more appliances report their operating state. However, in current scenarios, these assumptions do not hold, because such information is either not collected or cannot be accessed due to infrastructural restrictions. To this end, it is required to improve Jung's algorithm such that it only assumes a low fraction of ON/OFF events or learns to deal with ON/OFF events that are lost. For instance, it can use an event detection approach like the approach designed by Weiss *et al.*, which we evaluated in the previous section, or utilize *ElectriSense* citegupta10eletrisense proposed by Gupta *et al.* to detect events from high frequency consumption data. In addition, to further improve the performance of the algorithm, the algorithm can take into account more details provided by the smart meter data in addition to the variance. The consumption increase (or decrease) when appliances are being switched on (or off), for instance, are good indicators for the overall consumption of the appliance and helps to separate the consumption data from appliances that run during similar times such as the stereo system and the TV.

As described in section 4.2, the algorithm proposed by Jung and Savvides "remembers" estimations from previous intervals. In each round it decides on the fly (depending on the variance of the data) if it utilizes the aggregated estimations starting from the previous round or from the beginning of the analysis. We observed that in our implementation, the algorithm almost always relied on the estimations from the previous round. This is probably due to the fact that the time window of our analysis is six weeks and thus much longer than the study performed by Jung and Savvides, which only lasted for three days. For this reason, a potential optimization of the algorithm is to utilize a sliding window instead of aggregating the variance of the estimations from the beginning of the analysis.

4.6 Summary of the results

The results show that supervision is required to achieve reasonable performance. Weiss' algorithm and Parson's algorithm perform better than the unsupervised approaches investigated in our study, which do not reliably identify appliances in the aggregate consumption data. An exception to this is Jung's algorithm, which however requires additional information about ON/OFF events of appliances. One reason for the need of supervision is

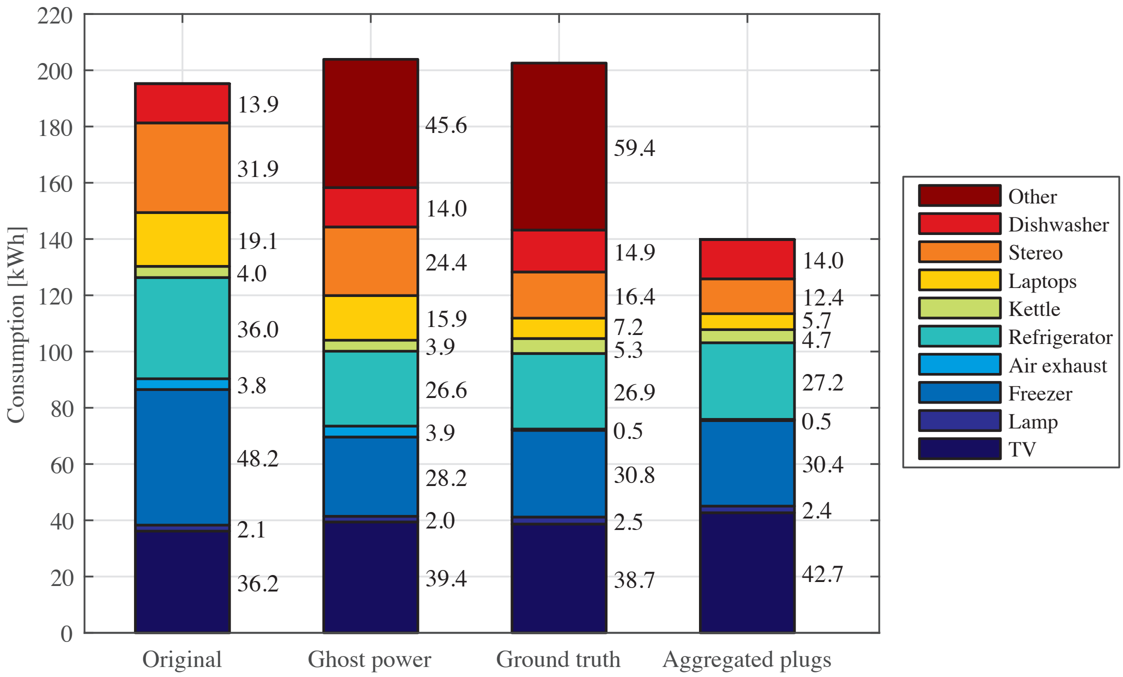


Figure 4.19: Consumption breakdowns (in kWh) estimated by Jung's algorithm analyzing six weeks of data in different settings as well as ground truth data measured by the plugs.

that time periods in which only a single appliance is running are rare. In household 2, for instance, 98.4% of the refrigerator's cooling cycles are overlapped by a cooling cycle of the freezer. As an additional constraint, each of the unsupervised approaches requires manual labeling after recognizing the appliances. To overcome the need to train the algorithm on plug data of the same household, Parson's semi-supervised approach utilizes generic appliance models and fine-tunes them given the aggregate consumption data. We further observe that Weiss' algorithm performs better than Parson's algorithm. We showed that this is due to the fact that Weiss's algorithm utilizes fine-grained consumption data (*i.e.*, measured at 1 Hz on multiple phases including real and reactive power), whereas Parson's algorithm uses data aggregated over a period of one minute. Using Weiss' algorithm, we can reliably identify events from cooling appliances as well as from appliances with high electricity consumption such as the stove or the dishwasher. The algorithm developed by Jung and Savvides targets future scenarios in which information about ON/OFF events is available along with the smart meter data. In this case—given some potential improvements to the algorithm as described above—it is possible to accurately identify the consumption of individual appliances.

In addition to optimizing these five algorithms as proposed in each of the subsections, we see potential in combining the algorithms. Using Parson *et al.*'s generic appliance model to generate HMMs followed by Kolter's algorithm, for instance, combines the advantages of both approaches. Leveraging the strength of Weiss *et al.*'s algorithm in recognizing switching events could further provide valuable input to the HMM-based approaches by providing information on appliance state changes. Similarly, having reliable information on switching events could remove the burden of deploying ON/OFF sensors as required as input for Jung's algorithm.

4.6.1 Limitations and future work

A limitation of each NILM evaluation—as with many data-driven approaches—is that the results highly depend on the data used and on the configuration of the algorithm parameters. For this reason, we collected consumption data over a particular long time frame and developed our evaluation system NILM-Eval to test a variety of combinations of parameters for each of the algorithms. We still observe a high variance in the results depending on the configuration of the household. For a faithful comparison, it is therefore necessary to extend the analysis to a larger number of households and compare the stability of the results on different data sets.

We evaluated each of the algorithms using the same data granularity and learning method as the authors in their original evaluation. However, the performance found in our analysis is likely insufficient for real-world applications. For this reason, it is important to investigate the performance of the algorithms under different requirements. For instance, we plan on applying Parson's algorithm on 1 Hz consumption data, and aim at training Kolter's algorithm on plug-level data rather than identifying the number and type of appliances in an unsupervised way.

Finally, the algorithms evaluated in this section do not incorporate information about the household's occupancy state. Knowing when occupants are at home and when they are away potentially provides valuable information about the state of individual appliances such as the stove, the kettle, or the TV. In practice, occupancy can be sensed by a smartphone and ultimately used to either improve the performance of NILM algorithms or—if NILM algorithms can be relied upon—to provide the occupants with information when appliances are running but they are not at home. In the next chapter, we discuss potential applications of NILM based on the results obtained through our evaluation.

4.6.2 Conclusion

In this chapter we presented both a comprehensive data set and an evaluation framework to analyze the performance of NILM algorithms and demonstrate the use of the framework on selected NILM approaches. Our results show that thanks to the use of our framework the suitability of selected approaches to be used in real scenarios as well as their limitations can be assessed. We make the ECO data set and the NILM-Eval framework including all implementation details and configurations available to the public [189, 190]. This way, we enable other researchers to specify configurations for their algorithms or to reuse and improve on existing configurations. Our approach therefore enables the consolidation of NILM evaluation efforts and makes a step towards identifying the best NILM algorithm given a specific scenario (*i.e.*, training method, measurement granularity and information detail of the collected data).

Chapter 5

Applications of smart meter data analytics

The household classification system presented in chapter 3 and the NILM analysis presented in chapter 4 investigate what type of information can be inferred from a household's electricity consumption data. In this chapter, we evaluate the applicability of these techniques for real world applications that aim at increasing the energy efficiency of households. To this end, we first evaluate the application of the household classification system to identify customers with specific characteristics (*e.g.*, customers that form a good target group for an energy-efficiency campaign). Next, we investigate to what extent households can be automatically compared to their peer group, given that the characteristics of the households are inferred from the consumption data. Next, we evaluate the potential of NILM to compute a consumption breakdown, which can then be provided on a household's electricity bill. Finally, we investigate the potential of NILM for energy consulting. In this case, we outline the types of saving advice that may be possible given the accuracy we achieved in our NILM analysis under different requirements. This chapter is partially based on the contributions made in [24].

5.1 Customer selection

In our analysis in chapter 3, we evaluated the accuracy and MCC of our system. While accuracy and MCC allow to describe the overall performance of a classifier, utilities are often interested in selecting a specific group of customers, which we call the *target group*. This is for instance necessary when a group of households (*e.g.*, those inhabited by a single person) are the target of an energy-efficiency program or a marketing campaign. Here, reducing the number of false positives (*i.e.*, of the cases in which a household is erroneously estimated to belong to the target class) is crucial. We show that by exploiting the confidence of the estimation obtained from the classifiers, it is possible to reduce the number of false positives significantly. To this end, we compute the true positive rate (TPR) and false positive rate (FPR) [80] as

$$TPR = \frac{TP}{TP+FN} \tag{5.1}$$

and

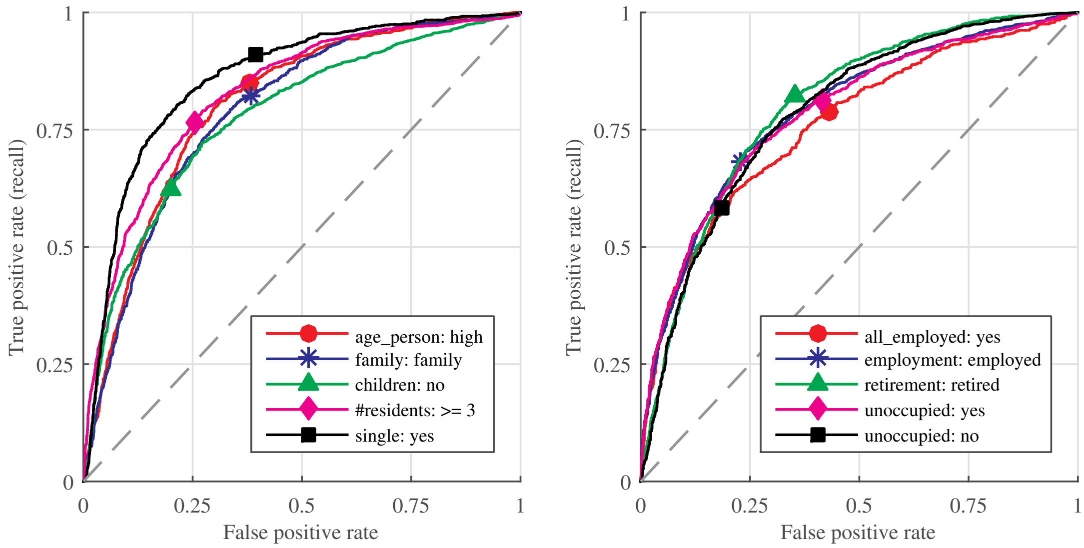
$$FPR = \frac{FP}{TN+FP}, \tag{5.2}$$

given the true positives (TPs), true negatives (TNs), false positives (FPs), and false negatives (FNs) of the classification.

In the example above, the TPR (or recall) indicates the number of correctly estimated *Single* households out of all households that belong to the class *Single*. The FPR indicates how many samples that do not belong to class *Single* were incorrectly classified as such. The receiver operating characteristic (ROC) curve relates these two metrics to each other, illustrating the trade-off between the benefits (true positives) and costs (false positives) of a classification. We implement the method described by Fawcett [70] to create the ROC curve for each target group, or target class, C. The method requires as input the posterior probability $P(C|x)$ for each sample, which is the probability that a sample belongs to C given the feature vector x. For $K > 2$ (*i.e.*, a characteristic with more than two classes), we create two groups based on the classes that are part of the target group and those that are not.

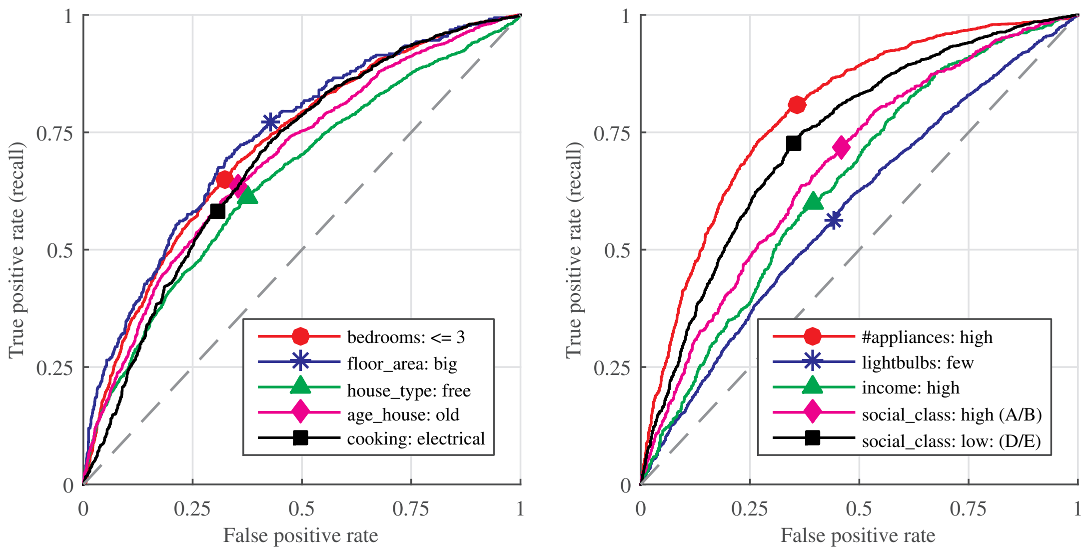
We compute the ROC curves on the basis of the experiments on the CER data set that include undersampling and use the MCC as a figure of merit during feature selection. We chose this particular configuration over the one that optimizes for accuracy because, based on our experiments, the latter one results in poor performance when targeting the underrepresented class of a characteristic. We further rely on the enhanced version of the household classification system presented in section 3.5, which contains features that represent a household's sensitivity to outdoor temperature and daylight. Classification is performed over all 75 weeks, whereas we compute $P(C|x)$ for each household by computing the average of the posterior probability obtained for each week.

Figure 5.1 shows the ROC curves for different target groups. The legend in each of the subplots indicates the characteristics as defined in table 3.4 in section 3.4. The diagonal, dashed line denotes the ROC curve of the RG classifier. For each target group, the graph shows the TPR (y-axis) that can be achieved at a given FPR (x-axis) and vice versa. Each point on the ROC curve represents a household h with posterior probability $P = P(C_i|x_h)$, which indicates the confidence that h belongs to class C_i given its feature vector x_h.



(a) Target groups related to the number and age of the occupants.

(b) Target groups related to employment and occupancy.



(c) Target groups related to the dwelling.

(d) Target groups related to the occupants' standard of living and energy efficiency.

Figure 5.1: ROC curves that show the trade-off between true positive rate and false positive rate for 20 different target groups. In each subplot, the diagonal line denotes the ROC curve of a random guess and the markers indicate the decision boundary of the classifier.

When selecting customers of a certain target group, utilities can choose a desired value for λ_{FPR} or λ_{TPR} such that $0 \leq \lambda_{FPR}, \lambda_{TPR} \leq 1$ under the condition that $\lambda_{TPR} = ROC(\lambda_{FPR})$ is a point on the ROC curve. This point p then implicitly defines a subset of households—the ones with a higher posterior probability than the household at p. The subset exhibits an FPR and a TPR of λ_{FPR} and λ_{TPR}, respectively, and the number of households in the subset depends on the selection of λ_{FPR} or λ_{TPR}.

Figure 5.1a illustrates the ROC curves for the target groups related to the number of occupants and the age of the household's chief income earner (CIE). It shows that when selecting single-person households (black line), for instance, a utility can identify 50% of all single-person households in their supply area (*i.e.*, $\lambda_{TPR} = 0.5$) with only 7.2% of the non-single households in the selected set. Similarly, in case 50% FPR is acceptable for the application envisioned by the utility (*i.e.*, $\lambda_{FPR} = 0.5$), the resulting subset of households contains 93.6% of the single-person households in the supply area. The age of the CIE as well as the other target groups related to the number of persons in the household also perform well with FPRs between 9% and 14% given that 50% of all households in the supply area that belong to the target group should be in the selected set of customers.

Figure 5.1b shows the ROC curves for the target groups related to the employment status of the CIE and to the occupancy state of the household. Similar to the examples related to the number of persons, FPRs are relatively low (*i.e.*, between 11.2% to 13.8%) in case a selection that contains 50% of the households in the target group is sufficient for the application envisioned by the utility. Increasing the "positive" number of households in the selection however comes with a cost: Lowering the barrier such that 75% of all households in the target group are included in the selection leads to FPRs between 28% and 38%.

Target groups related to the dwelling (*cf.* figure 5.1c) exhibit FPRs from 18.9% to 28.3% in case $\lambda_{TPR} = 0.5$. Finally, figure 5.1d shows a more diverse setting since FPRs for target groups related to the standard of living and energy efficiency range between 14% and 38%. In detail, the FPRs for the target groups in this last category are 14% (`#appliances: high`), 19% (`social_class: low`), 26% (`social_class: high`), 31% (`income: high`), and 38% (`lightbulbs: few`).

Overall, for 16 of the 20 groups it is possible to identify 50% of the households of the group with a false positive rate lower than 25% (18% on average). This shows that, when applying household classification to select households out of a target group (*i.e.*, for a marketing or energy efficiency campaign), subgroup selection with a carefully chosen FPR or TPR is important. The appropriate strategy is to create a ranking of customers by the posterior probability obtained from the classifiers. With this ranking, utilities can then assign customers to the group starting with the ones with the highest confidence.

5.2 Peer group comparison

When providing feedback to people, it is often recommended to relate their performance to the performance of their peers [4, 9, 79]. To this end, we evaluate the potential of the household classification system presented in chapter 3 to automatically compare a household's consumption to the consumption of households with similar characteristics (*i.e.*, households of the same target group). An application we envision is an automatically generated message such as "your electricity consumption is 20% higher than the consumption of other families", which can then be shown on a household's electricity bill or on a Web portal. To be applicable, this type of feedback must fulfill two important requirements: It requires an accurate estimate of the peer group's consumption (*i.e.*, compared to the consumption of the true group) and a low number of false positives. In this context, false positives are households that receive the wrong type of comparison because they are wrongly assigned to the target group by the household classification system.

Figure 5.2 shows the probability density functions of the mean electricity consumption of the 20 target groups we used in the previous analysis. Each subplot shows the probability density function of the households that belong to the target group according to the ground truth (*true group*, blue curve) and of the households that are predicted to belong to the target group (*predicted group*, orange curve). For comparison, the black curve shows the probability density function of all households, which is the same in each subplot. Table 5.1 lists the mean consumption of the true group, the predicted group, the difference between those two values, and the mean absolute error (MAE) between the two groups' probability density functions. The MAE illustrates the area under the curves covered by only one of the two groups.

On average, the mean consumption of the true group differs from the mean consumption of the predicted group by 0.0726 kW, which adds up to an error of 636 kWh over one year. The MAE is on average 0.1045. For individual target groups, we observe that some target groups have a relatively low MAE such as `age_person: high` (0.065), `age_house: old` (0.077), and the target groups related to the household's occupancy, which are `unoccupied: yes` (0.048), *unoccupied: no* (0.044), `employment: employed` (0.077), and `retirement: retired` (0.064). Target groups with high MAEs include `social_class: high (A/B)` (0.18), `income: high` (0.12), and `#appliances: high` (0.16). In a real-world setting, it is possible to reduce the difference between the consumption of the true group and the predicted group by incorporating a bias learnt during the training process. Therefore, we conclude that the first requirement for peer group comparisons—an accurate estimate of the peer group's consumption—holds in practice, which means that all households that are predicted correctly may receive a meaningful comparison with one or more of the target groups they belong to.

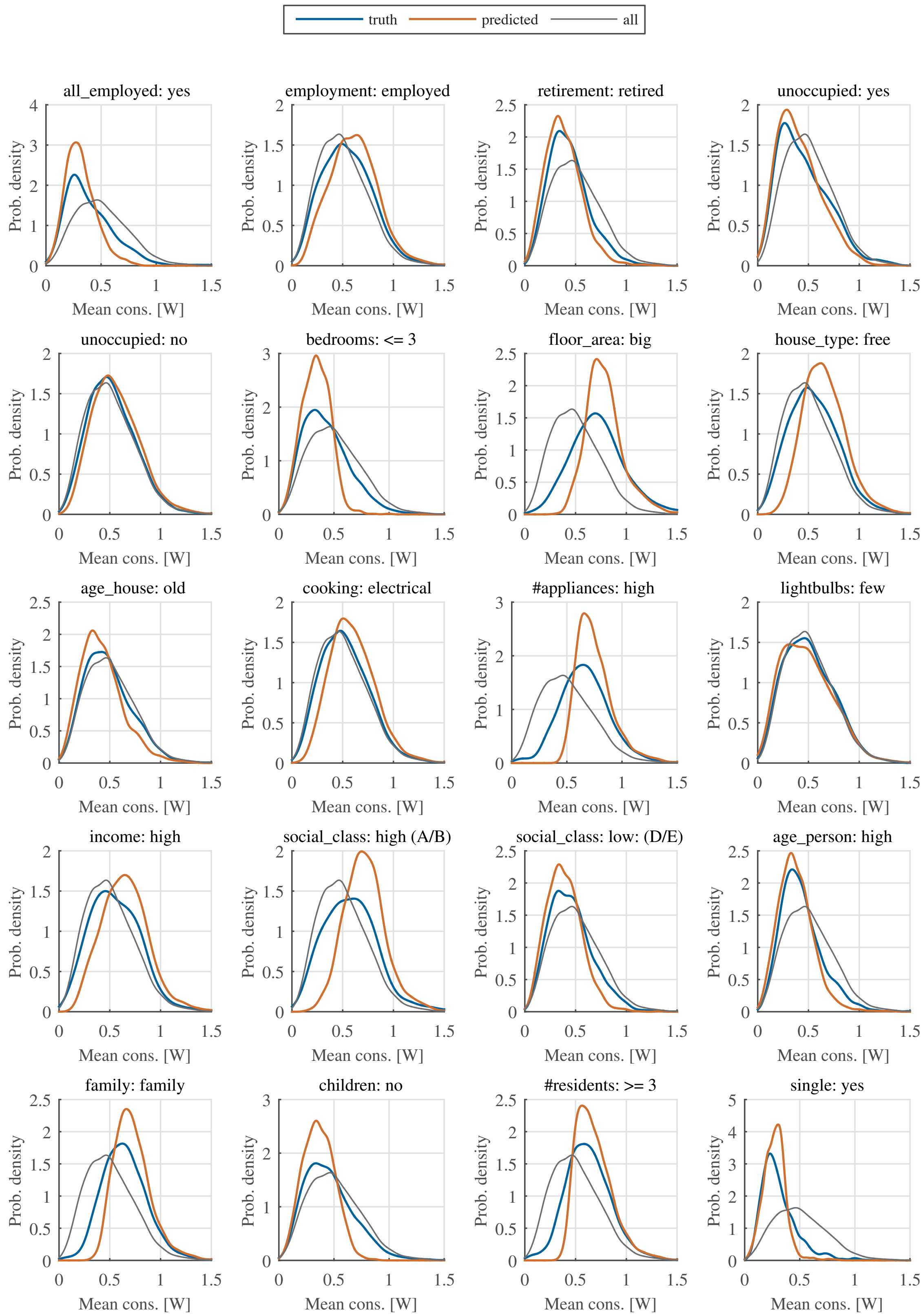


Figure 5.2: Probability density functions of the mean electricity consumption of all households that belong to a specific target group (blue), of all households that are predicted to belong to this group (orange), and of all households (black).

Table 5.1: Comparison of the mean consumption of the predicted and true samples for each target group. The mean absolute error (MAE) shows the difference between the probability density functions of those two groups (between 0 and 1).

Target group	**Mean cons. (true)**	**Mean cons. (predicted)**	**Difference**	**MAE**
lightbulbs: few	0.51 kW	0.49 kW	-0.017 kW	0.031
unoccupied: no	0.53 kW	0.57 kW	+0.041 kW	0.044
unoccupied: yes	0.46 kW	0.42 kW	-0.045 kW	0.048
retirement: retired	0.44 kW	0.38 kW	-0.053 kW	0.064
age_person: high	0.43 kW	0.38 kW	+0.057 kW	0.065
employment: employed	0.55 kW	0.62 kW	-0.064 kW	0.077
age_house: old	0.49 kW	0.43 kW	+0.063 kW	0.077
social_class: low: (D/E)	0.46 kW	0.39 kW	+0.068 kW	0.086
cooking: electrical	0.53 kW	0.6 kW	-0.071 kW	0.088
#residents: ≥ 3	0.63 kW	0.69 kW	+0.059 kW	0.11
single: yes	0.32 kW	0.27 kW	+0.043 kW	0.11
income: high	0.55 kW	0.66 kW	-0.1 kW	0.12
family: family	0.65 kW	0.73 kW	+0.082 kW	0.12
all_employed: yes	0.4 kW	0.31 kW	+0.084 kW	0.13
house_type: free	0.56 kW	0.67 kW	-0.11 kW	0.14
children: no	0.46 kW	0.36 kW	-0.099 kW	0.14
floor_area: big	0.72 kW	0.79 kW	+0.072 kW	0.15
bedrooms: ≤ 3	0.43 kW	0.34 kW	-0.093 kW	0.16
#appliances: high	0.67 kW	0.76 kW	+0.093 kW	0.16
social_class: high (A/B)	0.59 kW	0.73 kW	-0.14 kW	0.18

For any household, a comparison to other households of the same target group only makes sense if the household's target group is predicted correctly. For this reason, we evaluate the false discovery rate (FDR), which describes the number of false positives in the predicted group and is defined as

$$FDR = \frac{FP}{FP+TP}. \tag{5.3}$$

Table 5.2 shows the FDR for three different groups. Column *All* shows the FDR computed on the whole predicted group. Columns *Top 30%* and *Bottom 30%* focus on those households of the predicted group with an average consumption among the highest or lowest 30% of the predicted group, respectively. The reason for including these two groups is that those households are particularly interesting when designing consumption feedback, because they are the ones that should receive nudges (or rewards) as their consumption stands out. Column *Share* shows the proportion of households in the target group compared to the overall number of households for this characteristic. A low proportion makes it difficult for the classifier to identify households of this group and thus often leads to false positives, increasing the FDR accordingly.

The table shows that the FPR is relatively high (0.41 on average). In particular, estimation of `floor_area: big` and `social_class: high (A/B)` have a high FDR with 0.65 and 0.75, respectively. Depending on the target group, the FDR of the group of households with the highest average consumption per predicted target group is lower or higher than the FDR computed on the whole group of predicted households. On average, the MAE between those two FPRs is 0.0475. Similarly, the MAE between the group of predicted households with the lowest average consumption and the whole group of predicted households is 0.0463 on average.

There are two reasons for the high number of false positives shown in table 5.2. First, the target groups are small compared to the overall number of households (*i.e.*, the average of column *Share* is 0.42). Finding rare samples is difficult, which explains a high number of false positives in the predicted group. To reduce the FPR, we thus recommend a selection strategy such as proposed in section 5.1, which selects the households in the order of the confidence, starting with the households with the highest confidence. The second problem is that the features of the households in the groups top 30% or bottom 30% often overlap with features from other classes of the same characteristic. This is in particular a problem for characteristics for which the consumption-based features are important to the classifiers. Target group `#appliances: high`, for instance, has a high number of false positives among the households with the lowest 30% average consumption. Similarly, `bedrooms:` ≤ 3 has a high number of false positives among the households with a high average consumption, which often have more than three bedrooms. The difficulty to correctly assign households that are close to the decision boundary—and thus the difficulty to reliably identify outliers of a class—is part of the nature of classification problems in general. To mitigate this problem, we recommend to include new features in addition to those computed on the electricity consumption. Examples for such features are the location of the house, information obtained from public tax records, or information obtained by directly asking the households (*e.g.*, through an online portal). In addition, in case the consumption feedback is given through an online portal, the user could be provided with the opportunity to correct wrongly estimated characteristics.

Overall, in a practical setting, utilities are able to compare a households's consumption to households that belong to the same target group. However, they must either allow for a backchannel through which users can correct wrong predictions, or they must make sure the number of false positives is low. The latter approach requires restricting the selection to those households with high confidence or the inclusion of additional information (*i.e.*, beyond plain electricity consumption data) to the classification system.

Table 5.2: FDR of all households predicted to belong the target group. Columns *Top 30%* and *Bottom 30%* show the FDR computed on the 30% of households with the highest or lowest mean consumption, respectively. *Share* denotes the proportion of all households that belong to the target group.

Target group	**Top 30%**	**Bottom 30%**	**All**	**Share**
all_employed: yes	0.69	0.61	0.65	0.3
employment: employed	0.19	0.16	0.2	0.6
retirement: retired	0.51	0.51	0.48	0.3
unoccupied: yes	0.67	0.5	0.59	0.21
unoccupied: no	0.13	0.059	0.096	0.79
bedrooms: ≤ 3	0.41	0.18	0.31	0.54
floor_area: big	0.52	0.7	0.65	0.2
house_type: free	0.29	0.37	0.34	0.53
age_house: old	0.38	0.34	0.35	0.51
cooking: electrical	0.2	0.19	0.2	0.7
#appliances: high	0.33	0.5	0.42	0.31
lightbulbs: few	0.45	0.45	0.45	0.48
income: high	0.37	0.4	0.38	0.51
social_class: high (A/B)	0.76	0.75	0.75	0.16
social_class: low: (D/E)	0.45	0.4	0.43	0.39
age_person: high	0.59	0.58	0.56	0.23
family: family	0.49	0.52	0.52	0.26
children: no	0.17	0.07	0.14	0.71
#residents: ≥ 3	0.19	0.37	0.27	0.48
single: yes	0.63	0.36	0.51	0.2

5.3 Consumption breakdown

An important application of NILM is to provide households with consumption feedback that is disaggregated to the appliance-level. To this end, we investigate the use of Weiss' algorithm evaluated in chapter 4 to infer the consumption of individual appliances from the aggregate consumption data. The algorithm detects changes (*i.e.*, events) in the aggregate consumption data and identifies the appliance that is most likely responsible for the event. Unlike Hart's approach [85], which connects detected events in order to infer the consumption of individual appliances, Weiss *et al.* only provides the estimated events as output [179]. For this reason, we build upon Weiss *et al.* 's work and introduce an appliance-specific approach to combine events of appliances in order to estimate their consumption.

We rely on the supervised approach developed by Weiss *et al.*, because the performance of the unsupervised or semi-supervised algorithms evaluated in chapter 4 (*i.e.*, Baranski's algorithm, Parson's algorithm, and Kolter's algorithm) is too low to create a meaningful consumption breakdown. The algorithm of Jung and Savvides, which performs well in our analysis (*cf.* section 4.5.5), requires knowledge about ON/OFF switching events of appliances, which is typically not given in a real world scenario. Therefore, in this section, we assume that a training phase for the supervised algorithm is possible and refer to section 5.4 to evaluate what is possible in an unsupervised scenario without the burden of a training phase.

As a first step, we modified Weiss' algorithm: In the original version of the algorithm, an event is assigned to the appliance that represents the best matching cluster given that the distance between the event and the cluster center lies within a pre-defined range r. To avoid that events are wrongly assigned to an appliance that is closer to the cluster center, we run the algorithm separately for each appliance, thereby reducing the recall (*i.e.*, the number of missed events for each appliance). In addition, we modified the algorithm such that it assigns the event to the stove only if an event occurs on multiple phases simultaneously, because the stove was the only appliance connected to multiple phases in all households of our study. Table 5.3 shows the new results achieved on 90 days of data in household 2. The table shows the F_1 score, precision, and recall for each of the appliances as well as the scaling parameter r, which determines the maximum distance of an event to the cluster center. The table shows that events that belong to the refrigerator, freezer, dishwasher, kettle, and stove are detected with high precision. The TV and stereo system exhibit relatively high precision, while the events caused by both lamp and laptops can hardly be detected in the consumption data.

The results show that events from cooling appliances can be reliably detected by analyzing the aggregate consumption data. Other appliances, however, such as the dishwasher or the kettle, have high precision but exhibit many events that are missed. This observation—along with the fact that some of the appliances have a fluctuating consumption pattern—makes it difficult to apply the method based on combinatorial optimization proposed by Hart *et al.* to infer the consumption of appliances based on the detected events. Figure 5.3 shows, for instance, the events which the algorithm detected for the refrigerator (left plot) and the dishwasher (right plot). Solid lines indicate ON events, while dashed lines indicate OFF events. Events that represent true positives are colored green, missed events are colored gray. The blue graph shows the smart meter consumption and the black graph the consumption measured by the plug. In case of the refrigerator, the graph shows many true positives and true negatives as well as a few missed events. It is therefore necessary to take into account the possibility that individual ON or OFF events are not detected by the algorithm. In case of the dishwasher, the detected ON and OFF events barely match the consumption of the appliance measured by the smart plug, thus estimating the consumption by connecting ON and OFF events may lead to large

Table 5.3: Results obtained when applying our modified edge detection approach on the consumption data in household 2.

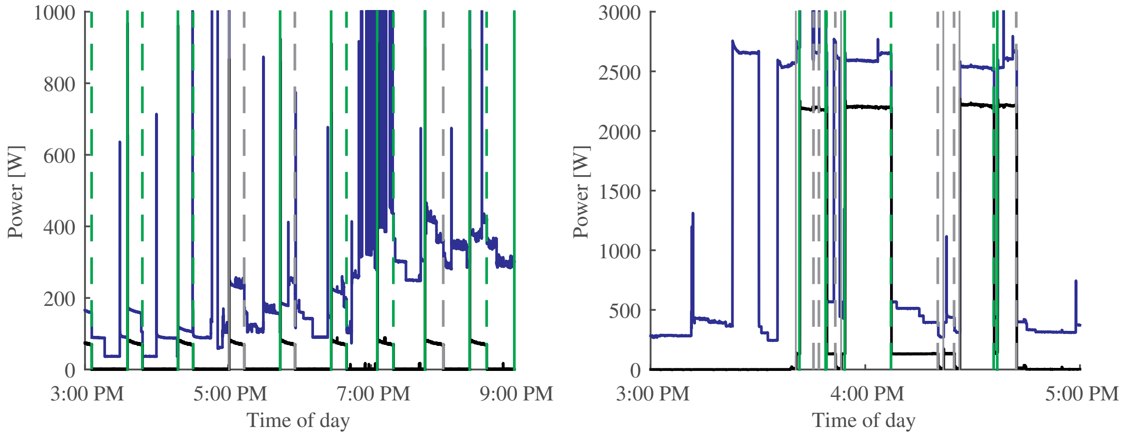
	F_1 score	Precision	Recall	r
Refrigerator	0.93	0.99	0.88	0.1
Freezer	0.90	0.99	0.83	0.2
Dishwasher	0.50	0.99	0.34	0.1
Kettle	0.82	0.98	0.70	0.1
Stove	0.998	1.0	0.996	1
Lamp	0.06	0.04	0.15	0.2
Laptops	0.037	0.066	0.026	0.2
TV	0.60	0.82	0.48	0.2
Stereo	0.14	0.47	0.08	0.2

estimation errors. For this reason, we developed a rule-based approach that takes into account appliance-specific characteristics in order to estimate when the appliances are running and how much they consume.

Our algorithm applies the following appliance-specific rules to detect when appliances have been switched on or off in cases where events have not been detected, and ultimately to infer the consumption of an appliance. The underlying assumption is that false negatives are more prevalent than false positives, which is based on the fact that precision is higher than recall for most of the appliances.

Refrigerator, freezer, TV, and stereo: These appliances have a start event, a stop event, and their consumption is almost constant. For each ON event detected by the first step of the algorithm, it searches for a succeeding OFF event. If no corresponding OFF event exists within the average runtime of the appliance, which we obtain from the training data, the algorithm assumes that the OFF event is missing and adds it at the end of the average runtime. Similarly, if the algorithm detects an OFF event without a corresponding ON event, it adds the ON event before the OFF event. Like the average runtime of the appliance, the average consumption of each appliance is obtained from the training data.

Dishwasher: The dishwasher exhibits multiple events during a single run. Due to its fluctuating consumption pattern, it is difficult to model the consumption of the appliance on the basis of the detected events. For this reason, our algorithm determines the average usage duration from the training data and assumes 1000 W average consumption per run [173]. It then infers the runs of a dishwasher based on the detected events.



(a) Refrigerator in household 2 (20 July 2012). (b) Dishwasher in household 2 (8 August 2012).

Figure 5.3: Illustration of the events detected by the algorithm. Solid lines indicate ON events, dashed lines OFF events. True positives are colored green, missed events gray. The blue graph shows the smart meter consumption and the black graph the consumption measured by the plug.

Water kettle: In case of the water kettle, the algorithm obtains the average consumption from the training data and searches for pairs of ON/OFF events within a time period of 1000 s.

Stove: The stove runs on multiple phases. For this reason, the algorithm searches for events that exhibit a change in consumption of more than a pre-defined threshold (*i.e.*, 200 W) on multiple phases simultaneously.

Laptop and lamp: Due to the low precision and recall of the detected events, we omit determining when the laptops and lamp were running and estimate the consumption based on the training data instead. For the laptops, we assume that they were running 4 h per day consuming 40 W on average. The lamp runs for 1 h per day and consumes 100 W on average.

Standby consumption: Albeit not covered by ground truth data, it is useful to provide households with an estimate of their standby consumption. To this end, we define the standby consumption to be the median of the daily minimum consumption values measured between 1 a.m. and 5 a.m.

Using these appliance-specific rules, we ran the algorithm on 90 days of consumption data in household 2 (15 days used for training and 75 days for testing). Figure 5.4a shows the ground truth obtained from the plug data, figure 5.4b shows the results of the estimation. For each appliance, the graph shows the monthly consumption in kWh as well as the share of the appliance's consumption related to the total consumption. Overall, the estimated

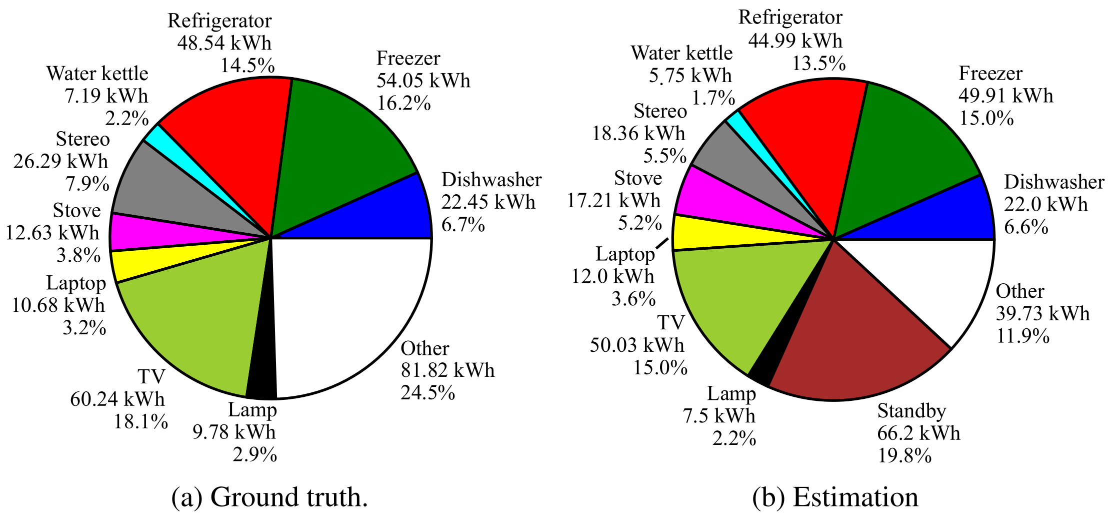


Figure 5.4: Consumption breakdown in household 2. For each appliance, the graph shows the share of the overall consumption as well as the appliance's monthly electricity consumption.

consumption of most of the appliances is close to the ground truth. The estimation errors of the dishwasher, the laptops, the water kettle, the lamp, the refrigerator, and the freezer are relatively low with 0.1, 0.4, 0.5, 0.7, 1.0, and 1.2 percentage points, respectively. The consumption of the TV and the stereo system are underestimated by 3.1 and 2.4 percentage points, respectively. Finally, the consumption of the stove is overestimated by 1.4 percentage points. The standby consumption accounts for 19.8% of the overall consumption, which means that 66.2 kWh per month are consumed by what we define as standby power.

Table 5.4 lists the results obtained when comparing the estimated consumption to the ground truth consumption over time for each appliances for which the algorithm estimated ON and OFF events. In contrast to the monthly consumption breakdown, estimation errors do not even out over time using this metric. In particular, we compute F_1 score, deviation, and RMSE based on the actual and estimated consumption values at each second using equations 4.3, 4.2, and 4.1, respectively. The results show that the consumption of the cooling appliances and the kettle can be estimated with high F_1 score and low deviation and RMSE. The dishwasher, stereo system, and TV have lower F_1 scores (and thus a higher deviation and higher RMSE). In case of the dishwasher, this is due to the fact that we assume a constant consumption instead of modeling a consumption pattern that encompasses multiple states. In case of the TV and stereo system, the algorithm sometimes wrongly assigns an OFF event after the average runtime, which leads to errors when the TV or stereo are running for a longer time than the average runtime.

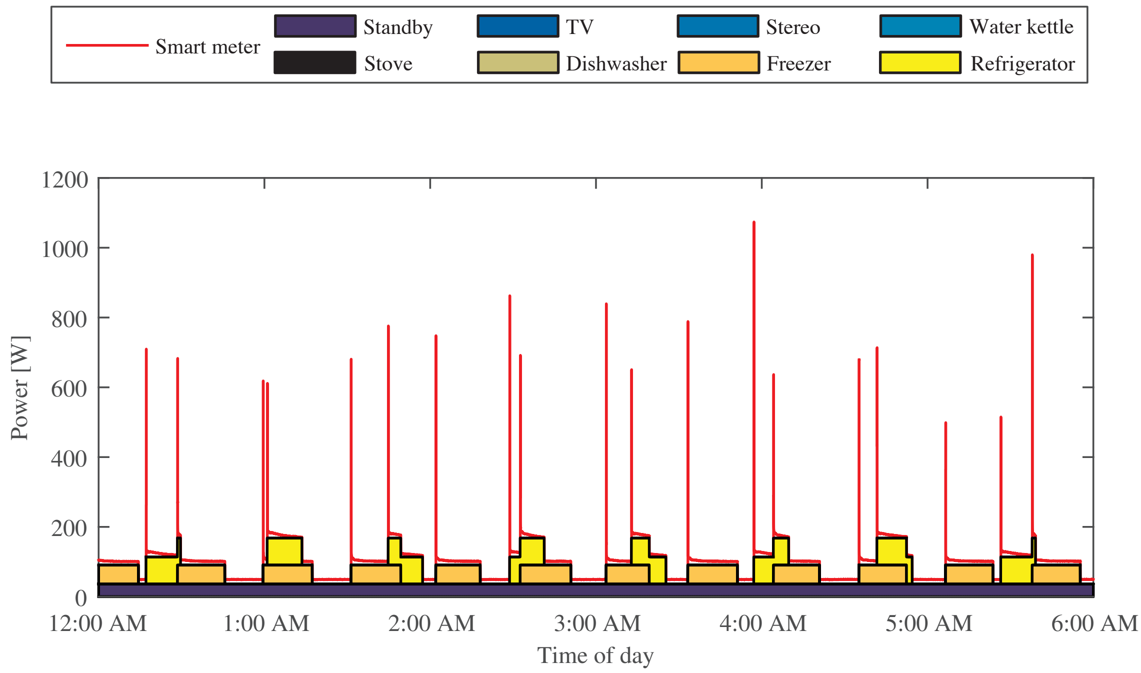
Table 5.4: Performance of the disaggregation algorithm comparing the actual and estimated consumption over time.

Appliance	**F_1 score**	**Deviation (equ. 4.2)**	**RMSE (equ. 4.1)**
Refrigerator	0.91	0.073	28 W
Freezer	0.94	0.059	28 W
Dishwasher	0.68	0.0015	134 W
Stereo	0.55	0.29	25 W
Stove[1]	0.74	0.14	88 W
TV	0.73	0.16	52 W
Kettle	0.92	0.076	33 W

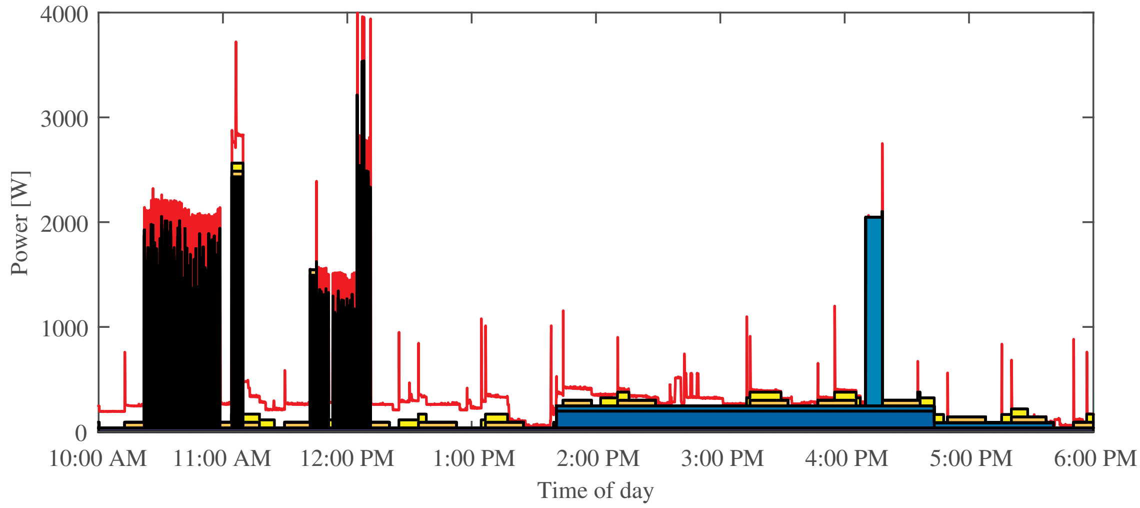
[1] Estimating the events of the stove instead of its consumption, the algorithm achieves a much higher performance (*i.e.*, $F_1 = 0.9981$).

Figure 5.5 shows the estimated consumption of individual appliances as well as the electricity consumption measured by the smart meter in household 2 on selected days. Figure 5.5a shows the consumption on 15 July 2012 during the night, which mainly consists of the consumption of the refrigerator (yellow area), the freezer (orange area), and the standby consumption (purple area). The plot shows that the algorithm deals well with overlapping consumption patterns of refrigerator and freezer. Figure 5.5b shows that—in addition to the cooling appliances—the consumption of the stove, TV, stereo, and water kettle are captured by the algorithm. Between 10 a.m. and 1 p.m., another appliance is running which is not detected by the algorithm. This appliance can either be a non-instrumented appliance or it can be the TV given the shape of the pattern. Figure 5.5c shows that the occupants were cooking during the evening as well as a subsequent run of the dishwasher. Finally, in figure 5.5d, it seems that the TV continued to run while the algorithm estimated an OFF event shortly after 5 p.m.

Determining the exact time when appliances have been switched on or off is prone to errors, because often the algorithm detects events when the appliance is running and misses the initial ON event or the final OFF event, as indicated by the low recall in table 5.3. In the following, we lower the requirements such that we estimate how often a particular appliance was running during a day. Table 5.5 shows the results of the analysis for those appliances in household 2 that have a user-induced start and stop event—unlike appliances that are always on such as cooling appliances. These appliances are the dishwasher, stereo system, stove, TV, and water kettle. Column *No. events* shows the number of times each appliance was switched on and off during the 75 days of the analysis. F_1 score, recall, and precision are computed based on how often the algorithm correctly estimated the number of times the appliance was running per day. Finally, column *Correct days* shows for how many days the algorithm correctly identified whether or not the appliance was

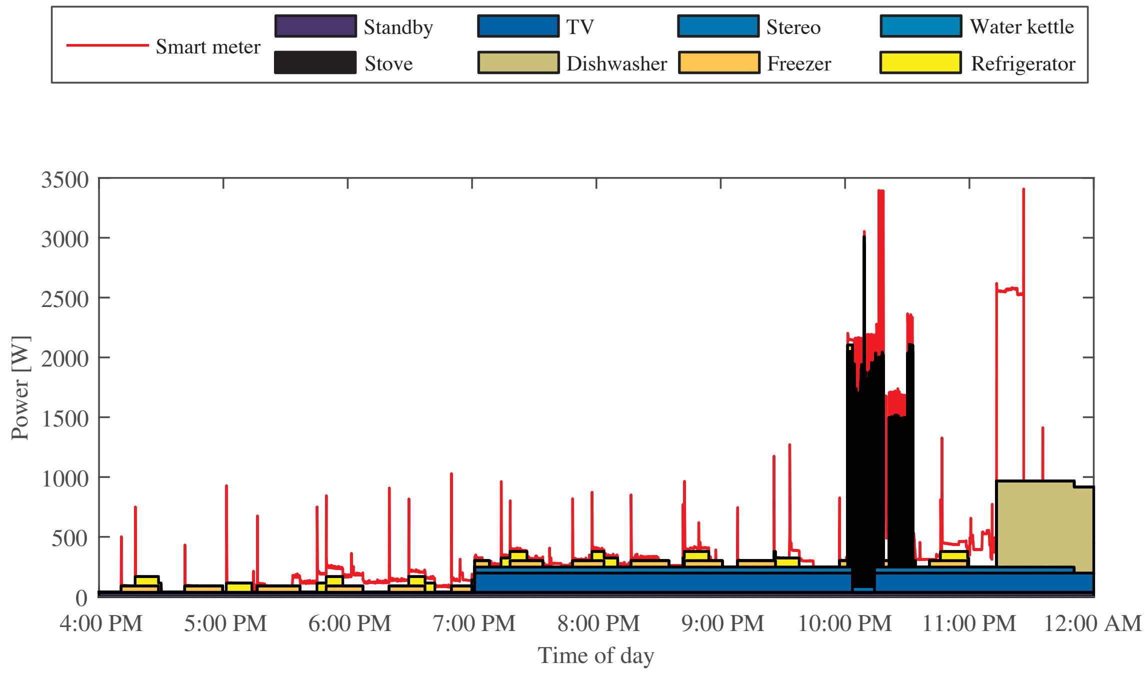


(a) Household 2, 15 July 2012 (night).

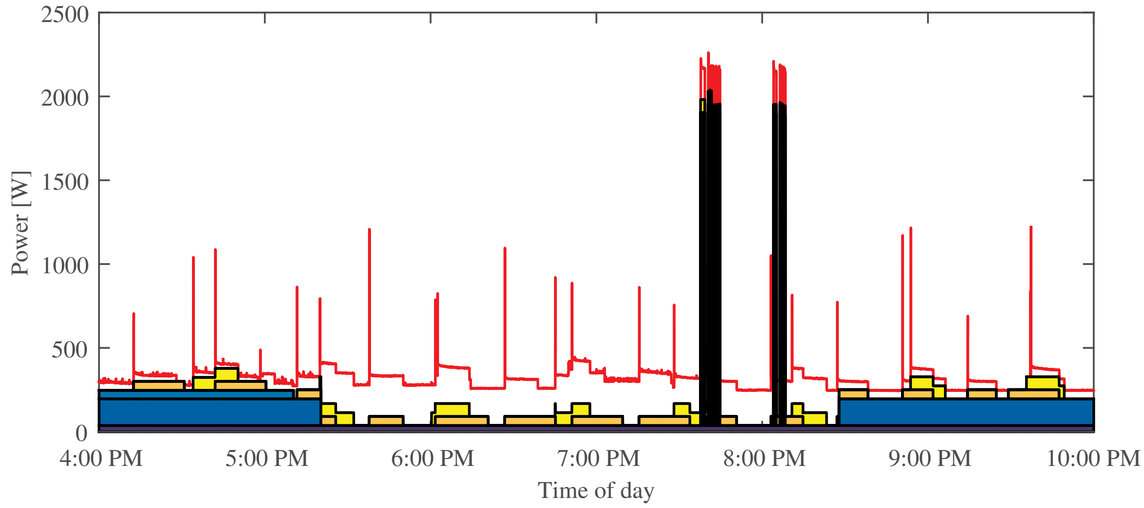


(b) Household 2, 4 July 2012.

running divided by the overall number of days. The results show that daily usage for the dishwasher, water kettle, and stove can be estimated almost perfectly. The error rates for the TV and stereo system are higher, but the algorithm still achieves F_1 scores of 0.82 and 0.77, respectively. The reason that the performance is lower is that both TV and stereo system were typically switched on and off multiple times per day, and missing a single event already invalidates the estimation for the whole day. When estimating whether or



(c) Household 2, 13 July 2012.



(d) Household 2, 15 July 2012.

Figure 5.5: Aggregate electricity consumption and estimated consumption of appliances on selected days in household 2.

not the TV or stereo system were running on a particular day, the algorithm achieved accuracies of 92% and 95%, respectively.

Table 5.5: Estimation if (and how often) appliances were in use on a daily basis.

Appliance	No. events	F_1 score	Recall	Precision	Correct days
Dishwasher	24	0.93	0.91	0.96	0.99
Stereo	161	0.77	0.67	0.90	0.95
Stove	5	1.0	1.0	1.0	1.0
TV	119	0.82	0.76	0.89	0.92
Water kettle	54	0.98	0.96	1.0	1.0

5.4 Energy consulting

Ideally, a disaggregation algorithm provides results at 80%–90% accuracy, achieves these results without training and in real-time, and it scales up to 20–30 appliances in a household [57, 185]. It should further work for various appliance types such as ON/OFF appliances (*e.g.*, a kettle), multi-state appliances (*e.g.*, a dishwasher), variable-power appliances (*e.g.*, a dimmable light), and continuous consuming appliances (*e.g.*, a router) [185]. Our analysis presented in chapter 4 and in section 5.3 shows that current disaggregation algorithms do not satisfy these requirements. The performance of the unsupervised or semi-supervised approaches investigated is too low to be useful for an energy consulting application that provides feedback to people by analyzing their electricity consumption. Using a supervised approach, as the disaggregation results presented in section 5.3 show, the overall consumption can be estimated with high accuracy—however, the algorithm requires a training phase and fine-tuning to appliance-specific characteristics. In the following, we outline the potential of NILM for energy consulting under different requirements (*e.g.*, with and without training).

Without training: In an unsupervised setting, it is possible to infer a household's standby consumption, the consumption of the cooling appliances (*i.e.*, refrigerator and freezer), and the time when occupants are cooking as well as how much electricity their stove consumes. For this analysis, 1 Hz consumption data measured individually for each phase is required. Instead of applying one of the unsupervised or semi-supervised algorithms evaluated in chapter 4, we propose to infer the standby consumption by analyzing the consumption data collected during the night between 1 a.m. and 5 p.m. as described in section 5.3. Similarly, we recommend analyzing the nightly consumption data to identify switching events of cooling appliances using the edge detection algorithm described in section 4.2.1 (without training). In a scenario in which the signatures of the cooling appliances are similar and the appliances run on the same phase, the algorithm can only infer the consumption of all cooling appliances combined.

With this knowledge, an energy consulting application application can compare the standby consumption of households with each other and issue a warning to those with a high consumption. It can further provide customers with energy-hungry cooling appliances a recommendation (or a coupon) to purchase a new refrigerator or freezer, or a recommendation to change settings of the existing ones to save electricity. The amount of electricity used for cooking is often negligible compared to the consumption of other appliances (*cf.* figure 5.4). For this reason, knowing the consumption of the stove is of little use for an energy consulting application that aims at reducing a household's electricity consumption. However, as cooking events can be detected with almost 100% accuracy (*cf.* table 5.4), knowing when and how often occupants are cooking reveals precious information about the occupants' lifestyle. This information can in turn be used by other applications such as demand-side management applications, by applications that need to know the occupancy state of the household (a cooking event most often indicates occupancy), and by assisted living applications that analyze the behavior of people over a long time period.

Supervised: The event-based detection evaluated in 5.3 can reliably identify usage of appliances such as the TV, dishwasher, or washing machine. For consumers with constant power consumption (*e.g.*, the TV), the algorithm provides both consumption and runtime. This information can be used by an energy consulting application, for instance, to notify users when their TV consumes more electricity than other TVs, when their TV usage increases over time[1], or when the TV is running while occupants pursue other tasks such as cooking. For consumers with multiple states or with a fluctuating consumption pattern (*e.g.*, the dishwasher or the washing machine), the NILM algorithm can learn about the number of runs of the appliance as well as the time of each run. This information can be useful an energy consulting application as it allows to compare the number of runs against households with similar characteristics. In addition, demand-side management applications can give recommendations on when it is most efficient to run these appliances.

Training a NILM algorithm by installing sensors in the household imposes a high burden on the occupants. An energy consulting application can therefore provide "user generated" events used to train the system. When the smart meter is newly installed, for instance, people can—and are often willing to—walk around the house, switch on and off appliances, and specify the type of appliance they triggered through a smartphone application [179]. This way it is possible to learn appliance signatures without installing additional sensors.

Occupancy monitoring: Information about a household's occupancy state is also useful for energy consulting. For instance, it is possible to provide the occupants with detailed information on how much electricity their household consumes while no one is

[1] One of the participants of our study mentioned that he switched off the TV when the electricity consumption of the household exceeds the consumption of the previous day.

at home. Also, occupants can receive information about the consumption of individual appliances they left on while being away, which is useful to increase energy efficiency as well as safety, for instance when making occupants aware that their stove is still running.

There are many ways to capture the occupancy state of a household [101]. For instance, it is possible to use dedicated sensors (*e.g.*, PIR sensors), to analyze GPS data obtained from the occupants' smartphones, or to capture if the occupants' smartphones are connected to the home router's Wi-Fi. Alternatively, it is possible to infer the occupancy state directly from the smart meter data [102, 103]. However, in this case, it is difficult for an energy consulting application to decide whether an appliance has been left on while the occupants are away or whether they are at home using the appliance.

Additional infrastructure: Our analysis in section 4.5.5 shows that using information obtained from ON/OFF sensors can significantly improve the disaggregation performance for appliances that are difficult to detect in the consumption data otherwise (*e.g.*, the laptops, or the lamp). Similarly, installing a device that measures the aggregate consumption at high frequency (*i.e.*, in the order of multiple kilohertz) allows to detect switching events of those appliances that are otherwise difficult to detect [81, 136]. With such an enhanced measurement infrastructure, energy consulting applications can provide the user with fine-grained saving advise, for instance by notifying them when they leave the lights running in unoccupied rooms or when consumer electronics are running while they are away.

NILM also enables applications beyond energy consulting. Ambient assisted living, for instance, is a domain that analyzes the lifestyle of occupants to draw conclusions on their physical and mental condition. Using NILM techniques to track cooking activities, the runtime of the TV, as well as reading and outgoing habits can provide a valuable contribution to assisted living systems, which currently depend on complex sensor deployments. Augmenting these systems with information provided by NILM algorithms can help to reduce the number of sensors that must be installed, to detect erroneous sensor values, and it may allow analyzing past behavior when it is necessary to justify the installment of an assisted living system.

5.5 Implications for households, utilities, and policy makers

The approaches presented in this thesis automatically reveal household characteristics and occupant behavior by analyzing a household's electricity consumption. They utilize data from smart meters that capture consumption information at fine granularity (*e.g.*, aggregated at 30-minute or 1-second intervals) and thus can be used in combination with many smart meters that are currently being rolled out throughout the world. On the basis of our findings, we see several (positive and negative) implications for households, utilities, and policy makers, which we outline in this section.

5.5.1 Households

Households can benefit from more informative and precise, more enjoyable, and motivating energy efficiency campaigns: As our analysis allows for identifying energy-relevant household characteristics, it becomes possible for utilities to benchmark households based on similar demographics and household type [9,36]. It also enables assessing a household's energy efficiency using an energy-efficiency label that is easy to understand. Utilities can further group similar households for engagement campaigns, summarize information or provide energy-saving recommendations that are relevant for each group, or define peer households to realize concepts based on games to increase user engagement. It has been shown in the past that efficiency campaigns benefit from tailoring advice and motivational cues to the recipient [4,9,79] and from providing personalized, disaggregated consumption feedback [8, 46, 74]. Overall, targeted campaigns might help to make efficiency-related topics interesting and therefore win the attention of more households. Ultimately, information can be directed in a way that triggers savings in terms of both electricity and money.

At the same time, the findings have strong implications for consumer privacy as well: Our system makes it possible to extract information that consumers may prefer to keep private, including data related to income, employment status, status of the relationship, or social class. Thus, households should engage in a discussion with those who capture and want to use the data, urging them to make techniques for privacy protection an inherent feature of the emerging smart metering infrastructure [122].

5.5.2 Utilities

Utilities benefit from insights into customers that they can reveal using the proposed system to estimate household characteristics. The same information that helps to make energy saving campaigns more effective can help to better market products and services. The latter includes identifying target households for specific offerings, *e.g.*, promoting solar panels only to mid to high-income customers who live in a house rather than in an apartment and offering green tariffs preferentially to families with young children. The revealed household characteristics might also help to lower the cost of efficiency campaigns by targeting households with a high potential such as those that show a mismatch between household characteristics and energy demand. Other actions that become possible include tailoring behavioral campaigns to retired individuals or young professionals, or to concentrate on behavioral campaigns rather than on triggering investments in low-income households [172]. Utilities we have collaborated with [21] also hypothesize that more directed customer interaction and better savings advice will improve customer satisfaction and ultimately customer retention. The latter is relevant in particular in energy markets that are liberalized or that face liberalization such that customers can freely choose their utility [138]. In short, the customer insight gained by applying the proposed techniques helps utilities to better allocate their budget, offer directed savings advice, and ultimately boost the impact of their sales and efficiency campaigns.

5.5.3 Policy makers

Given the advent of the outlined approach, policy makers need to define the rules that govern the use of metering data. It is crucial to promote the beneficial effects including increased energy efficiency and more targeted energy consulting services, and yet to limit the undesired uses of these techniques. As it is at this stage still unclear what "undesired uses" comprises, the stakeholders need to investigate the utilities' and private individuals' interests and find a compromise [143]. Whereas the former probably have an interest in leveraging the retrieved information for marketing campaigns of all sorts, private persons may demand varying levels of privacy protection [122]—depending on the culture they are embedded in.

Consequently, policy makers need to strike a balance between a regime that allows the full materialization of the benefits of smart meter data analytics and regulation that entirely protects privacy. Here, it is important to define who can access the data and what it is used for. One viable approach is to let individual households decide who has access and what they can do with the data [141]. Alternatively, it is technologically also feasible to design solutions in which data is collected and analyzed in the realm of the household (*e.g.*, on their router or on a dedicated computer). The utility only receives information that is necessary for billing purposes or for services that the household explicitly agreed upon. Households may also provide their data for training purposes, which is then rewarded by the utility. Beyond technical solutions, it is also possible to ensure that the information can be used at a large scale by designing an opt-out regime that by default grants utilities access to the data unless an individual actively decides against it. Experience shows that the number of customers who opt out is relatively low, as would be the number of individuals who actively decided for making their data available in an opt-in regime [88].

Chapter 6

Conclusions and outlook

The goal of this thesis is to provide the technical foundation for energy efficiency programs that are personalized to the individual household and at the same time scale to thousands or millions of customers. The key to this “mass personalization” lies in the automated analysis of household electricity consumption measured by smart meters.

In this final chapter, we outline our findings and summarize the contributions of the thesis along the three research questions: (i) What household characteristics can be revealed from smart meter data? (ii) Can we observe the energy-related behavior of a household sufficiently well from its aggregate electricity consumption data? (iii) Based on these findings—what applications are possible and what are potential constraints and implications?

6.1 Automated household classification

Having knowledge on their customers’ characteristics allows utility companies to provide personalized feedback and target the right households for energy-efficiency campaigns. Examples of such characteristics include a household’s socio-economic status, its dwelling properties, and information on the appliance stock. We selected 18 different household characteristics based on relevance to utilities identified through interviews as well as based on the outcome of a data exploration we performed using self-organizing maps. We then developed a comprehensive system that automatically infers those characteristics from electricity consumption data using supervised machine learning methods. It therefore reduces the need for utilities to perform costly and cumbersome surveys in order to gain access to such valuable customer insights. Our household classification system utilizes five different classifiers: The kNN classifier [29], the LDA classifier [29], the Mahalanobis distance classifier [6], the support vector machine (SVM) classifier [43], and the AdaBoost classifier [75]. In addition, we use linear regression to infer continuous values for some of the characteristics.

We evaluated our household classification system on the CER data set [211], which contains electricity consumption data from more than 4000 households collected at 30-minute granularity over a period of 1.5 years. Along with the consumption data, the data set contains answers to questionnaires each of the participants filled out before and after the study. The evaluation shows that most of the characteristics can be estimated with 70%–80% accuracy. Overall, our system performs roughly 30 percentage points better than assigning household characteristics at random. Three types of characteristics can be inferred particularly well from electricity consumption data. These are characteristics that reflect the occupancy state of the house, the number of persons in the house, and the appliance stock. However, characteristics related to the dwelling itself (*e.g.*, the floor area or the number of bedrooms) are more difficult to extract from electricity consumption data. This is due to the fact that heating and cooling only plays a minor role in the consumption data available for this study. We then added new features that take into account the sensitivity of households to the outdoor temperature and daylight. With this optimization, we can improve accuracy by up to 2.3 percentage points. Finally, we identify the features that are particularly relevant for our household classification and we evaluate the impact of different data granularities on the performance. We show that 30-minute data or 60-minute data is a necessity compared to performing the analysis on daily averages.

6.2 Non-intrusive load monitoring

In chapter 4 we evaluated the potential of non-intrusive load monitoring (NILM), which aims at inferring appliance-specific information from a household's aggregate electricity consumption data. We first describe the design space of NILM and categorize existing algorithms using the three dimensions *data granularity*, *learning methods*, and *information detail*. We then select five algorithms that cover the design space and have been developed to operate on data measured at a frequency of at most 1 Hz, which can be provided by most off-the-shelf electricity meters. To evaluate the suitability of these five approaches to be used in real scenarios and to assess their limitations, we developed the MATLAB-based evaluation framework NILM-Eval [189]. We further present the ECO data set [190], which contains electricity consumption data from six Swiss households collected over a period of eight months. The data set consists of (1) aggregate consumption data measured by a smart meter, (2) plug-level consumption data from selected appliances, and (3) information about the occupancy state of each of the household's inhabitants. The data set has several advantages over other NILM data sets and is already being used by many researchers worldwide. Our evaluation of the five algorithms on the ECO data set shows that training is required to achieve reasonable performance. Out of the algorithms we evaluated, Weiss' algorithm [179] and Parson's algorithm [135] perform better than the unsupervised approaches (*i.e.*, Baranski's algorithm [13] and Kolter's algorithm [108]), which do not reliably identify appliances in the aggregate consumption data. An exception is Jung's algorithm [89], which is unsupervised and provides a consumption breakdown close to the ground truth data. However, the algorithm requires additional information about ON/OFF events of appliances. We further observe that Weiss' algorithm performs better than Parson's algorithm, because the latter operates on 1-minute consumption data. Using Weiss' algorithm, we can reliably identify events from cooling appliances as well as from appliances with high electricity consumption such as the stove or the dishwasher.

6.3 Applications and implications

In chapter 5, we investigated the applicability of automatic household classification and NILM for real-world applications that aim at increasing the energy efficiency of households. First, we investigated the use case in which a utility aims at identifying households of a specific target group. We utilized the confidence of the classification to build ROC curves, which allow to define a subset of the target group that exhibits a lower false positive rate (FPR) than the whole target group. Overall, for 16 out of 20 target groups it is possible to identify 50% of the households of the group with a false positive rate lower than 25% (18% on average). Next, we explored the potential of the household classification system to realize automated peer group comparisons. The analysis shows that utilities are able to compare a households's consumption to households that belong to the same target group. However, the rate of wrong predictions (*i.e.*, the false discovery rate (FDR)) is relatively high on average, which means utilities should create a backchannel through which users can correct wrong predictions. When the feedback is provided on an online portal, for instance, families that are compared to single-person households can easily correct this mis-prediction. Alternatively, utilities can restrict this type of feedback to those households with high confidence of the classification.

To investigate the application of NILM in practical settings, we enhanced Weiss' algorithm such that it infers the consumption of each appliance based on the events detected for this appliance. Our algorithm achieves a consumption breakdown that is close to the ground truth measurements provided by the plug data. Further, the algorithm reliably detects whether or not certain appliances were running on a particular day. However, the algorithm assumes that training data is available for each of the households whose consumption should be disaggregated. Next, we explored the potential of NILM for an energy consulting application. In an unsupervised setting, the application can provide recommendations with respect to the standby consumption, cooling appliances, and the stove. In a scenario where the algorithm can get training data (*e.g.*, assisted by user input through a smartphone or by additional sensors), more recommendations are possible, addressing usage of all types of appliances.

At the end of chapter 5, we discussed implications of the work presented in this thesis. On one side, smart meter data analytics opens new opportunities for households, utilities, and policy makers. Households, for instance, receive personalized energy efficiency programs that can be offered using the information gained from the analysis of their smart meter data. Utilities also benefit when their customers save electricity, for instance because it helps utilities to reach saving targets imposed by the government. In addition, they gain valuable customer insights that have high relevance to improve customer retention and marketing campaigns. On the other side, inferring household characteristics and the consumption-related behavior of occupants raises privacy concerns. For this reason, it is important to define who can access the data and what it is used for.

By showing what is possible at large scale, we aim to help utilities to provide personalized and scalable energy-efficiency programs as well as intensify the discussion on the rules and regulations that will be needed to govern data analytics applications.

6.4 Limitations and outlook

In the following, we discuss potential limitations of the approaches developed in this thesis and outline possibilities for future work.

6.4.1 Household classification

Weekly profiling: The current implementation classifies each week individually and ultimately assigns a household to the class that has been classified most often. An application that analyzes the weekly classification results over time without selecting the majority class can profile customers to detect changes in the household, for instance if the occupants get a baby, go on vacation, or install new appliances.

Correlations: The current household classification system classifies one characteristic at a time. In addition, the system could exploit correlations between individual characteristics to realize use cases in which target households should match multiple characteristics (*e.g.*, families that live in a large house). By building classifier chains [146], for instance, it is possible to learn correlations of characteristics during the training process and potentially increase performance over classifying each of the characteristics individually.

A priori knowledge: In some cases, utilities have a priori knowledge about the household such as its location, or, in some states in the US, information that is available from public tax registers. Combining the data from different input sources potentially improves performance over using only electricity consumption data. In other cases, utilities have the possibility to interact with their customers (*e.g.*, through an online portal) and to ask for information directly. In this case, it would be interesting to investigate which questions utilities should ask their customers: What type of information are households most willing to provide, and knowledge about which characteristics helps most to improve the classification performance of the other characteristics?

Outlier detection: Our classification system is based on the assumption that similar types of households have similar consumption patterns. Comparing the classification result with the actual characteristics of a household can therefore reveal outliers that are worth a detailed investigation (*e.g.*, a single-person household that is classified as a family). This way, it could be possible to identify customers with high saving potential or to detect faults in the system such as leakage, energy theft, or boilers that are wrongly connected to the neighbor's electricity meter.

6.4.2 Non-intrusive load monitoring

The role of training: The most important barrier when it comes to the adoption of NILM algorithms in practical settings is that they require training in order to achieve a reasonably high performance. For this reason, improving unsupervised methods is one of the most important research challenges in the field of NILM. Alternatively, it is an interesting challenge to find ways to simplify the training process. For instance, a household could be asked for a list of appliances running in its house. The algorithm could then use signatures of these appliances collected in other households of similar type and perform training using those signatures.

Sensor fusion: Another way to improve the performance of NILM algorithms is to include information from other sources than the smart meter. Examples of these sources include appliances that have a network connection (*e.g.*, an Internet-connected TV), sensors that monitor switching events of appliances or movement of occupants, or smartphones, which can indicate activities of the occupants or their current location.

Upper bound: It is an interesting research challenge to evaluate if there is an upper bound for the performance of NILM algorithms. Among other things, the upper bound depends on the granularity of the consumption data provided by the smart meter, on the number and type of appliances in the household, and on appliance usage patterns.

Comparison to simple approaches: Simple disaggregation approaches can already provide a good estimate on a household's standby consumption and the consumption of its cooling appliances (*i.e.*, by investigating the consumption during the night). If an algorithm uses empirical values to determine the consumption of the other appliances in the household (*e.g.*, the average consumption of a dishwasher in similar households), it may obtain a good consumption breakdown. This consumption breakdown could already be useful for practical applications and serve as a lower bound for more complex disaggregation approaches.

High-frequency sensing: Certain types of appliances leave a characteristic trace on the power line that can be measured at high frequency. Switched electrical loads, for instance, produce noise in the form of a transient or continuous noise [136]. Similarly, appliances with switch mode power supplies (*e.g.*, consumer electronics or fluorescent lighting) cause electromagnetic inference [81]. As the noise propagates through the wire, it can be measured at any socket in the household. A device that senses the noise at high frequency (*i.e.*, in the order of kilohertz or megahertz) can therefore provide information about potential ON/OFF switching events to a NILM algorithm, which then disaggregates the aggregate electricity consumption measured at lower granularity (*e.g.*, 1-second granularity or 15-minute granularity).

When does NILM become obsolete? As more and more appliances are being equipped with a network connection, they will be able to report their current state and in some cases even their electricity consumption. Similarly, smart power outlets and other sensors will become cheaper and may find their way into households. While this additional infrastructure supports NILM algorithms in the short term (and thus provides new research challenges), it may make them obsolete eventually. However, integrating information from different sensors and systems has always been a challenge. In the long term, a NILM algorithm could still be a valuable module in each home automation system: It could incorporate information collected from different systems and vice versa support those systems with information it inferred from the aggregate consumption data (*e.g.*, to detect faulty sensors or appliances).

From perfect NILM to energy consulting: Assuming a NILM algorithm perfectly disaggregates a household's electricity consumption, how could that be useful for energy consulting? It is a challenge to quantify the costs and benefits of an appliance in order to provide meaningful recommendations. A big refrigerator, for instance, often consumes more electricity than a small refrigerator, but it also provides a higher benefit for the owners. Running the washing machine during high tariff is more expensive, but it may better suit the occupants' plans when they want to hang up the laundry. An energy consulting application has to learn the preferences of the occupants in order to distinguish between energy waste and intentional usage to increase comfort.

Improve forecasts: An accurate prediction of a household's future electricity consumption has become increasingly valuable in recent years. It can for instance help a control algorithm with the decision whether electricity produced by the solar panels should be consumed, stored, or fed into the grid. NILM can support algorithms that forecast a household's future electricity consumption with information about the occupants' lifestyle and therefore potentially improve the performance of such algorithms.

Bibliography

[1] Gergely Ács and Claude Castelluccia. I have a DREAM! (DiffeRentially privatE smArt Metering). In Tomáš Filler, Tomáš Pevný, Scott Craver, and Andrew Ker, editors, *Information Hiding*, volume 6958 of *LNCS*, pages 118–132. Springer, May 2011.

[2] Adrian Albert and Ram Rajagopal. Smart meter driven segmentation: What your consumption says about you. *IEEE Transactions on Power Systems*, 28(4):4019–4030, June 2013.

[3] Adrian Albert and Ram Rajagopal. Thermal profiling of residential energy use. *IEEE Transactions on Power Systems*, 30(2):602–611, March 2015.

[4] Hunt Allcott. Social norms and energy conservation. *Journal of Public Economics*, 95(9):1082–1095, October 2011.

[5] Kyle Anderson, Adrian Ocneanu, Diego Benitez, Derrick Carlson, Anthony Rowe, and Mario Bergés. BLUED: A fully labeled public dataset for event-based non-intrusive load monitoring research. In *2nd KDD Workshop on Data Mining Applications in Sustainability (SustKDD)*, August 2012.

[6] Theodore W. Anderson. *An introduction to multivariate statistical analysis*. John Wiley & Sons, September 1984.

[7] Hans-Jürgen Appelrath, Henning Kagermann, and Christoph Mayer (Eds.). Future energy grid. Migration to the Internet of energy. acatech STUDY, October 2012.

[8] K. Carrie Armel, Abhay Gupta, Gireesh Shrimali, and Adrian Albert. Is disaggregation the holy grail of energy efficiency? The case of electricity. *Energy Policy*, 52:213–234, January 2013.

[9] Ian Ayres, Sophie Raseman, and Alice Shih. Evidence from two large field experiments that peer comparison feedback can reduce residential energy usage. *Journal of Law, Economics, and Organization*, 29:992–1022, October 2013.

[10] Michael Baeriswyl, André Müller, Reto Rigassi, Christof Rissi, Simon Solenthaler, Thorsten Staake, and Thomas Weisskopf. Folgeabschätzung einer Einführung von Smart Metering im Zusammenhang mit Smart Grids in der Schweiz, June 2012. Bundesamt für Energie (BFE), Publikation 290600, http://www.bfe.admin.ch/php/modules/enet/streamfile.php?file=000000010824.pdf.

[11] Pierre Baldi, Søren Brunak, Yves Chauvin, Claus A. F. Andersen, and Henrik Nielsen. Assessing the accuracy of prediction algorithms for classification: An overview. *Bioinformatics*, 16(5):412–424, May 2000.

[12] Cesar J. Bandim, Jose Eduardo R. Alves Jr., Ary V. Pinto Jr., Fabio C. de Souza, Mauro R. B. Loureiro, Christiane A. Magalhães, and Federico Galvez-Durand. Identification of energy theft and tampered meters using a central observer meter: A mathematical approach. In *IEEE PES Transmission & Distribution Conference and Exposition*, volume 1, pages 163–168, Sept 2003.

[13] Michael Baranski and Jürgen Voss. Genetic algorithm for pattern detection in NIALM systems. In *4th International Conference on Systems, Man and Cybernetics (SMCS)*, pages 3462–3468. IEEE, October 2004.

[14] Galen L. Barbose, Charles A. Goldman, Ian M. Hoffman, and Megan Billingsley. The future of utility customer-funded energy efficiency programs in the United States: Projected spending and savings to 2025. *Energy Efficiency*, 6(3):475–493, January 2013.

[15] Sean Barker, Aditya Mishra, David Irwin, Emmanuel Cecchet, Prashant Shenoy, and Jeannie Albrecht. Smart*: An open data set and tools for enabling research in sustainable homes. In *2nd KDD Workshop on Data Mining Applications in Sustainability (SustKDD)*, August 2012.

[16] Nipun Batra, Manoj Gulati, Amarjeet Singh, and Mani B. Srivastava. It's different: Insights into home energy consumption in India. In *5th ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings (BuildSys)*, pages 3:1–3:8. ACM, November 2013.

[17] Nipun Batra, Jack Kelly, Oliver Parson, Haimonti Dutta, William Knottenbelt, Alex Rogers, Amarjeet Singh, and Mani Srivastava. NILMTK: An open source toolkit for non-intrusive load monitoring. In *5th International Conference on Future Energy Systems (e-Energy)*, pages 265–276. ACM, June 2014.

[18] Pia Baumann. Bundesamt für Energie. Personal communication on 31 March 2015 and 9 April 2015.

[19] Christian Beckel, Wilhelm Kleiminger, Romano Cicchetti, Thorsten Staake, and Silvia Santini. The ECO data set and the performance of non-intrusive load monitoring algorithms. In *1st ACM International Conference on Embedded Systems for Energy-Efficient Buildings (BuildSys)*, pages 80–89. ACM, November 2014.

[20] Christian Beckel, Wilhelm Kleiminger, Thorsten Staake, and Silvia Santini. Improving device-level electricity consumption breakdowns in private households using ON/OFF events. *ACM SIGBED Review*, 9(3):32–38, July 2012.

[21] Christian Beckel, Leyna Sadamori, and Silvia Santini. Towards automatic classification of private households using electricity consumption data. In *4th ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings (BuildSys)*, pages 169–176. ACM, November 2012.

[22] Christian Beckel, Leyna Sadamori, and Silvia Santini. Automatic socio-economic classification of households using electricity consumption data. In *4th International Conference on Future Energy Systems (e-Energy)*, pages 75–86. ACM, May 2013.

[23] Christian Beckel, Leyna Sadamori, Silvia Santini, and Thorsten Staake. Automated customer segmentation based on smart meter data with temperature and daylight sensitivity. In *6th IEEE International Conference on Smart Grid Communications (SmartGridComm)*. IEEE, November 2015.

[24] Christian Beckel, Leyna Sadamori, Thorsten Staake, and Silvia Santini. Revealing household characteristics from smart meter data. *Energy*, 78:397–410, December 2014.

[25] Mario Berges, Ethan Goldman, H. Scott Matthews, Lucio Soibelman, and Kyle Anderson. User-centered nonintrusive electricity load monitoring for residential buildings. *Journal of Computing in Civil Engineering*, 25(6):471–480, November 2011.

[26] Paolo Bertoldi, Silvia Rezessy, Eoin Lees, Paul Baudry, Alexandre Jeandel, and Nicola Labanca. Energy supplier obligations and white certificate schemes: Comparative analysis of experiences in the European Union. *Energy Policy*, 38(3):1455–1469, March 2010.

[27] Suchismita Bhattacharjee and Georg Reichard. Socio-economic factors affecting individual household energy consumption: A systematic review. In *5th*

International Conference on Energy Sustainability (ES). ASME, August 2011.

[28] Benjamin J. Birt, Guy R. Newsham, Ian Beausoleil-Morrison, Marianne M. Armstrong, Neil Saldanha, and Ian H. Rowlands. Disaggregating categories of electrical energy end-use from whole-house hourly data. *Energy and Buildings*, 50:93–102, July 2012.

[29] Christopher M. Bishop. *Pattern recognition and machine learning*. Springer, August 2006.

[30] Constantin Budeanu. Puissances réactives et fictives. Romanian National Institute, Bucuresti, 1927.

[31] Bundesamt für Energie (BFE). Grundlagen der Ausgestaltung einer Einführung intelligenter Messsysteme beim Endverbraucher in der Schweiz – Technische Mindestanforderungen und Einführungsmodalitäten, November 2014.

[32] Bundesamt für Energie (BFE). Schweizerische Gesamtenergiestatistik 2013.

[33] Hông-Ân Cao, Christian Beckel, and Thorsten Staake. Are domestic load profiles stable over time? An attempt to identify target households for demand side management campaigns. In *39th IEEE Conference of the IEEE Industrial Electronics Society (IECON)*, pages 4733–4738. IEEE, November 2013.

[34] Derrick R. Carlson, H. Scott Matthews, and Mario Bergés. One size does not fit all: Averaged data on household electricity is inadequate for residential energy policy and decisions. *Energy and Buildings*, 64:132–144, May 2013.

[35] Feng Chen, Jing Dai, Bingsheng Wang, Sambit Sahu, Milind Naphade, and Chang-Tien Lu. Activity analysis based on low sample rate smart meters. In *17th International Conference on Knowledge Discovery and Data Mining (KDD)*, pages 240–248. ACM, August 2011.

[36] Marshini Chetty, David Tran, and Rebecca E. Grinter. Getting to green: Understanding resource consumption in the home. In *10th International Conference on Ubiquitous Computing (UbiComp)*, pages 242–251. ACM, September 2008.

[37] Gianfranco Chicco. Overview and performance assessment of the clustering methods for electrical load pattern grouping. *Energy*, 42(1):68–80, June 2012.

[38] Gianfranco Chicco, Roberto Napoli, Petru Postolache, Mircea Scutariu, and Cornel Toader. Customer characterization options for improving the tariff offer. *IEEE Transactions on Power Systems*, 18(1):381–387, February 2003.

[39] Romano Cicchetti. NILM-Eval: Disaggregation of real-world electricity consumption data. Master's thesis, ETH Zurich, 2014.

[40] Commission for Energy Regulation. Decision on the national rollout of electricity and gas smart metering, CER12008, July 2012.

[41] Commission for Energy Regulation. Electricity smart metering customer behaviour trials (CBT) findings report, CER11080a, May 2011.

[42] Commission for Energy Regulation. Electricity smart metering technology trials finding report, CER11080b, May 2011.

[43] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine Learning*, 20(3):273–297, September 1995.

[44] Council of the European Union. 8/9 March 2007 Presidency Conclusions.

[45] Council of the European Union. 23/24 October 2014 Conclusions.

[46] Sarah Darby. The effectiveness of feedback on energy consumption. A review for DEFRA of the literature on metering, billing and direct displays. Environmental Change Institute, University of Oxford, April 2006.

[47] Sarah J. Darby. Metering: EU policy and implications for fuel poor households. *Energy Policy*, 49:98–106, October 2012.

[48] Thomas H. Davenport and Jill Dyché. Big Data in Big Companies, SAS Institute Inc., May 2013.

[49] Stephane de la Rue du Can, Greg Leventis, Amol Phadke, and Anand Gopal. Design of incentive programs for accelerating penetration of energy-efficient appliances. *Energy Policy*, 72:56–66, September 2014.

[50] Daswin De Silva, Xinghuo Yu, Damminda Alahakoon, and Grahame Holmes. A data mining framework for electricity consumption analysis from meter data. *IEEE Transactions on Industrial Informatics*, 7:399–407, August 2011.

[51] Kathrin Degen, Charles Efferson, Fabian Frei, Lorenz Goette, and Rafael Lalive. Smart Metering, Beratung oder sozialer Vergleich: Was beeinflusst den Elektrizitätsverbrauch? Technical Report 290850, Bundesamt für Energie (BFE), June 2013.

[52] Ian Dent, Uwe Aickelin, and Tom Rodden. Application of a clustering framework to UK domestic electricity data. *ArXiv e-prints*, July 2013. arXiv:1307.1380.

[53] Deutsche Bundesregierung. Das Energiekonzept – Beschluss des Bundeskabinetts vom 28. September 2010. Presse- und Informationsamt der Bundesregierung, Berlin. http://www.bundesregierung.de/ContentArchiv/DE/Archiv17/_Anlagen/2012/02/energiekonzept-final.html.

[54] Directive 2009/72/EC of the European Parliament and of the Council of 13 July 2009 concerning common rules for the internal market in electricity and repealing Directive 2003/54/EC.

[55] Directive 2012/27/EU of the European Parliament and of the Council of 25 October 2012 on energy efficiency, amending Directives 2009/125/EC and 2010/30/EU and repealing Directives 2004/8/EC and 2006/32/EC.

[56] Ernst & Young GmbH. Kosten-Nutzen-Analyse für einen flächendeckenden Einsatz intelligenter Zähler, July 2013.

[57] Dominik Egarter, Venkata Pathuri Bhuvana, and Wilfried Elmenreich. PALDi: Online load disaggregation via particle filtering. *IEEE Transactions on Instrumentation and Measurement*, 64(2):467–477, February 2015.

[58] Dominik Egarter and Wilfried Elmenreich. Autonomous load disaggregation approach based on active power measurements. *ArXiv e-prints*, December 2014. arXiv:1412.2877.

[59] Dominik Egarter, Christoph Prokop, and Wilfried Elmenreich. Load hiding of household's power demand. In *5th International Conference on Smart Grid Communications (SmartGridComm)*, pages 854–859. IEEE, November 2014.

[60] Karen Ehrhardt-Martinez, Kat A. Donnelly, and John A. Laitner. Advanced metering initiatives and residential feedback programs: A meta-review for household electricity-saving opportunities. Technical Report E105, American Council for an Energy-Efficient Economy, June 2010.

[61] Frank Eichinger, Daniel Pathmaperuma, Harald Vogt, and Emmanuel Müller. Data analysis challenges in the future energy domain. In Ting Yu, Nitesh Chawla, and Simeon Simoff, editors, *Computational Intelligent Data Analysis for Sustainable Development*, pages 1–55. Taylor & Francis, April 2013.

[62] Georgia Elafoudi, Lina Stankovic, and Vladimir Stankovic. Power disaggregation of domestic smart meter readings using dynamic time warping. In *6th International Symposium on Communications, Control and Signal Processing (ISCCSP)*, pages 36–39. IEEE, May 2014.

[63] Ehsan Elhamifar and Shankar Sastry. Energy disaggregation via learning ‘powerlets’ and sparse coding. In *29th Conference on Artificial Intelligence (AAAI)*, pages 629–635. AAAI Press, January 2015.

[64] Alexander E. Emanuel. Energetical factors in power systems with nonlinear loads. *Archiv für Elektrotechnik*, 59(3):183–189, May 1977.

[65] Frank Englert, Till Schmitt, Sebastian Kößler, Andreas Reinhardt, and Ralf Steinmetz. How to auto-configure your smart home? High-resolution power measurements to the rescue. In *4th International Conference on Future Energy Systems (e-Energy)*, pages 215–224. ACM, May 2013.

[66] Zekeriya Erkin, Juan Ramón Troncoso-Pastoriza, R. Inald L. Lagendijk, and Fernando Pérez-González. Privacy-preserving data aggregation in smart metering systems: An overview. *IEEE Signal Processing Magazine*, 30(2):75–86, February 2013.

[67] European Commission. Benchmarking smart metering deployment in the EU-27 with a focus on electricity, COM (2014), 356, June 2014.

[68] Xi Fang, Satyajayant Misra, Guoliang Xue, and Dejun Yang. Smart grid – The new and improved power grid: A survey. *IEEE Communications Surveys & Tutorials*, 14(4):944–980, October 2012.

[69] Linda Farinaccio and Radu Zmeureanu. Using a pattern recognition approach to disaggregate the total electricity consumption in a house into the major end-uses. *Energy and Buildings*, 30(3):245–259, August 1999.

[70] Tom Fawcett. An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8):861–874, June 2006.

[71] Hongliang Fei, Younghun Kim, Sambit Sahu, and Milind Naphade. Heat pump detection from coarse grained smart meter data with positive and unlabeled learning. In *19th ACM International Conference on Knowledge Discovery and Data Mining (KDD)*, pages 1330–1338. ACM, August 2013.

[72] Vera Figueiredo, Fátima Rodrigues, Zita Vale, and Joaquim B. Gouveia. An electric energy consumer characterization framework based on data mining techniques. *IEEE Transactions on Power Systems*, 20(2):596–602, May 2005.

[73] Aurelia Figueroa. From energy crisis to energy transition? German Development Institute / Deutsches Institut für Entwicklungspolitik (DIE) – The Current Column, October 2013.

[74] Corinna Fischer. Feedback on household electricity consumption: A tool for saving energy? *Energy Efficiency*, 1(1):79–104, May 2008.

[75] Yoav Freund and Robert E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. In *2nd European Conference on Computational Learning Theory (EuroCOLT)*, pages 23–37. Springer, 1995.

[76] Francesco Fusco, Michael Wurst, and Ji Won Yoon. Mining residential household information from low-resolution smart meter data. In *21st International Conference on Pattern Recognition (ICPR)*, pages 3545–3548, November 2012.

[77] Daphne Geelen, Angèle Reinders, and David Keyson. Empowering the end-user in smart grids: Recommendations for the design of products and services. *Energy Policy*, 61:151–161, 2013.

[78] Sara Ghaemi and Günther Brauner. User behavior and patterns of electricity use for energy saving. In *6. Internationale Energiewirtschaftstagung an der TU Wien (IEWT)*, February 2013.

[79] Noah J. Goldstein, Robert B. Cialdini, and Vladas Griskevicius. A room with a viewpoint: Using social norms to motivate environmental conservation in hotels. *Journal of Consumer Research*, 35(3):472–482, March 2008.

[80] Jan Gorodkin. Comparing two K-category assignments by a K-category correlation coefficient. *Computational Biology and Chemistry*, 28(5-6):367–374, December 2004.

[81] Sidhant Gupta, Matthew S. Reynolds, and Shwetak N. Patel. ElectriSense: Single-point sensing using EMI for electrical event detection and classification in the home. In *12th International Conference on Ubiquitous Computing (UbiComp)*, pages 139–148. ACM, September 2010.

[82] Stephen Haben, Jonathan Ward, Danica Vukadinovic Greetham, Colin Singletona, and Peter Grindroda. A new error measure for forecasts of household-level, high resolution electrical energy consumption. *International Journal of Forecasting*, 30:246–256, 2014.

[83] Steven Haben, Colin Singleton, and Peter Grindrod. Analysis and clustering of residential customers energy behavioral demand using smart meter data. *IEEE Transactions on Smart Grid*, PP(99), March 2015.

[84] Mark A. Hall. *Correlation-based Feature Selection for Machine Learning*. PhD thesis, The University of Waikato, 1999.

[85] George W. Hart. Nonintrusive appliance load monitoring. *Proceedings of the IEEE*, 80(12):1870–1891, December 1992.

[86] Haibo He and Edwardo A. Garcia. Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, 21(9):1263–1284, September 2009.

[87] Nathalie Japkowicz. The class imbalance problem: Significance and strategies. In *International Conference on Artificial Intelligence (ICAI)*, pages 111–117. CSREA Press, 2000.

[88] Eric J. Johnson and Daniel Goldstein. Do defaults save lives? *Science*, 302(5649):1338–1339, November 2003.

[89] Daekwoo Jung and Andreas Savvides. Estimating building consumption breakdowns using ON/OFF state sensing and incremental sub-meter deployment. In *8th ACM Conference on Embedded Networked Sensor Systems (SenSys)*, pages 225–238. ACM, November 2010.

[90] 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. *Energy*, 55:184–194, May 2013.

[91] Amir Kavousian, Ram Rajagopal, and Martin Fischer. Ranking appliance energy efficiency in households: Utilizing smart meter data and energy efficiency frontiers to estimate and identify the determinants of appliance energy efficiency in residential buildings. *Energy and Buildings*, 99:220–230, July 2015.

[92] Jack Kelly and William Knottenbelt. Metadata for energy disaggregation. *ArXiv e-prints*, May 2014. arXiv:1403.5946.

[93] Jack Kelly and William Knottenbelt. UK-DALE: A dataset recording UK domestic appliance-level electricity demand and whole-house demand. *ArXiv e-prints*, March 2015. arXiv:1404.0284.

[94] Andreas Kemmler, Alexander Piégsa, Andrea Ley, Philipp Wüthrich, Mario Keller, Martin Jakob, and Giacomo Catenazzi. Analyse des schweizerischen Energieverbrauchs 2000 – 2013 nach Verwendungszwecken. Technical report, Bundesamt für Energie, Bern, Schweiz, September 2014.

[95] Himanshu Khurana, Mark Hadley, Ning Lu, and Deborah A. Frincke. Smart-grid security issues. *IEEE Security & Privacy*, 8(1):81–85, January/February 2010.

[96] Sara Kilcher and Andreas Dröscher. Smart meters in the field – A sensor framework for a real world deployment. distributed systems laboratory report, eth zurich, February 2013.

[97] Hyungsul Kim, Manish Marwah, Martin Arlitt, Geoff Lyon, and Jiawei Han. Unsupervised disaggregation of low frequency power measurements. In *11th SIAM International Conference on Data Mining (SDM)*, pages 747–758. SIAM, April 2011.

[98] Y. Kim, T. Schmid, Z.M. Charbiwala, and M.B. Srivastava. ViridiScope: Design and implementation of a fine grained power monitoring system for homes. In *11th International Conference on Ubiquitous Computing (UbiComp)*, pages 245–254. ACM, September 2009.

[99] Manuel Kläy. Sensor-assisted device-level electricity consumption breakdowns in private households. Master's thesis, ETH Zurich, 2013.

[100] Wilhelm Kleiminger. *Occupancy Sensing and Prediction for Automated Energy Savings*. PhD thesis, ETH Zurich, 2015.

[101] Wilhelm Kleiminger, Christian Beckel, and Silvia Santini. Opportunistic sensing for efficient energy usage in private households. In *Smart Energy Strategies Conference*, September 2011.

[102] Wilhelm Kleiminger, Christian Beckel, and Silvia Santini. Household occupancy monitoring using electricity meters. In *ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp)*. ACM, September 2015.

[103] Wilhelm Kleiminger, Christian Beckel, Thorsten Staake, and Silvia Santini. Occupancy detection from electricity consumption data. In *5th ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings (BuildSys)*, pages 10:1–10:8. ACM, November 2013.

[104] Wilhelm Kleiminger, Friedemann Mattern, and Silvia Santini. Predicting household occupancy for smart heating control: A comparative performance analysis of state-of-the-art approaches. *Energy and Buildings*, 85:493–505, December 2014.

[105] Tuevo Kohonen. *Self-organizing maps*. Springer, 3rd edition, 2001.

[106] J. Zico Kolter, Siddharth Batra, and Andrew Y. Ng. Energy disaggregation via discriminative sparse coding. In *24th Annual Conference on Neural Information Processing Systems (NIPS)*, pages 1153–1161, December 2010.

[107] J. Zico Kolter and Joseph Ferreira, Jr. A large-scale study on predicting and contextualizing building energy usage. In *25th Conference on Artificial Intelligence (AAAI)*, pages 1349–1356. AAAI Press, August 2011.

[108] J. Zico Kolter and Tommi Jaakkola. Approximate inference in additive factorial HMMs with application to energy disaggregation. In *15th International Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 1472–1482. JMLR, April 2012.

[109] J. Zico Kolter and Matthew J. Johnson. REDD: A public data set for energy disaggregation research. In *1st KDD Workshop on Data Mining Applications in Sustainability (SustKDD)*, August 2011.

[110] Jungsuk Kwac, June Flora, and Ram Rajagopal. Household energy consumption segmentation using hourly data. *IEEE Transactions on Smart Grid*, 5(1):420–430, January 2014.

[111] Jungsuk Kwac, Chin-Woo Tan, Nicole Sintov, June Flora, and Ram Rajagopal. Utility customer segmentation based on smart meter data: Empirical study. In *4th International Conference on Smart Grid Communications (SmartGridComm)*, pages 720–725. IEEE, October 2013.

[112] K. Y. Lee, Y. T. Cha, and J. H. Park. Short-term load forecasting using an artificial neural network. *IEEE Transactions on Power Systems*, 7(1):124–132, 1992.

[113] Hongfei Li, Dongping Fang, Shilpa Mahatma, and Arun Hampapur. Usage analysis for smart meter management. In *8th International Conference & Expo on Emerging Technologies for a Smarter World (CEWIT)*, pages 1–6. IEEE, November 2011.

[114] Jian Liang, Simon K. K. Ng, Gail Kendall, and John W. M. Cheng. Load signature study – Part I: Basic concept, structure, and methodology. *IEEE Transactions on Power Delivery*, 25(2):551–560, April 2010.

[115] Stephen Makonin, Fred Popowich, Lyn Bartram, Bob Gill, and Ivan V. Bajic. AMPds: A public dataset for load disaggregation and eco-feedback research. In *Electrical Power & Energy Conference*, pages 1–6. IEEE, August 2013.

[116] Alan Marchiori, Douglas Hakkarinen, Qi Han, and Lieko Earle. Circuit-level load monitoring for household energy management. *IEEE Pervasive Computing*, 10(1):40–48, September 2010.

[117] Daisuke Mashima and Alvaro A. Cárdenas. Evaluating electricity theft detectors in smart grid networks. In Davide Balzarotti, Salvatore J. Stolfo, and Marco Cova,

editors, *Research in Attacks, Intrusions, and Defenses*, volume 7462 of *LNCS*, pages 210–229. Springer, 2012.

[118] Friedemann Mattern and Christian Flörkemeier. From the Internet of computers to the Internet of Things. In Kai Sachs, Ilia Petrov, and Pablo Guerrero, editors, *From Active Data Management to Event-Based Systems and More*, volume 6462 of *LNCS*, pages 242–259. Springer, 2010.

[119] Friedemann Mattern, Thorsten Staake, and Markus Weiss. ICT for green – How computers can help us to conserve energy. In *1st International Conference on Energy-Efficient Computing and Networking (e-Energy)*, pages 1–10. ACM, April 2010.

[120] Brian W. Matthews. Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochimica et Biophysica Acta (BBA)-Protein Structure*, 405(2):442–451, October 1975.

[121] Patrick McDaniel and Stephen McLaughlin. Security and privacy challenges in the smart grid. *IEEE Security & Privacy*, 7(3):75–77, May/June 2009.

[122] Eoghan McKenna, Ian Richardson, and Murray Thomson. Smart meter data: Balancing consumer privacy concerns with legitimate applications. *Energy Policy*, 41:807–814, February 2012.

[123] Colin McKerracher and Jacopo Torriti. Energy consumption feedback in perspective: Integrating Australian data to meta-analyses on in-home displays. *Energy Efficiency*, 6(2):387–405, May 2013.

[124] Fintan McLoughlin. *Characterising Domestic Electricity Demand for Customer Load Profile Segmentation*. PhD thesis, Dublin Institute of Technology, 2013.

[125] Fintan McLoughlin, Aidan Duffy, and Michael Conlon. Analysing domestic electricity smart metering data using self-organising maps. In *CIRED Workshop*, May 2012.

[126] Fintan McLoughlin, Aidan Duffy, and Michael Conlon. Characterising domestic electricity consumption patterns by dwelling and occupant socio-economic variables: An Irish case study. *Energy and Buildings*, 48:240–248, May 2012.

[127] K. Metaxiotis, A. Kagiannas, D. Askounis, and J. Psarras. Artificial intelligence in short term electric load forecasting: A state-of-the-art survey for the researcher. *Energy Conversion and Management*, 44(9):1525–1534, June 2003.

[128] Andrés Molina-Markham, Prashant Shenoy, Kevin Fu, Emmanuel Cecchet, and David Irwin. Private memoirs of a smart meter. In *2nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings (BuildSys)*, pages 61–66. ACM, November 2010.

[129] Andrea Monacchi, Dominik Egarter, Wilfried Elmenreich, Salvatore D'Alessandro, and Andrea M. Tonello. GREEND: An energy consumption dataset of households in Italy and Austria. In *5th International Conference on Smart Grid Communications (SmartGridComm)*, pages 511–516. IEEE, November 2014.

[130] Antti Mutanen, Maija Ruska, Sami Repo, and Pertti Järventausta. Customer classification and load profiling method for distribution systems. *IEEE Transactions on Power Delivery*, 26(3):1755–1763, July 2011.

[131] Bernard Neenan and Ross C. Hemphill. Societal benefits of smart metering investments. *The Electricity Journal*, 21(8):32–45, October 2008.

[132] Guy R. Newsham and Brent G. Bowker. The effect of utility time-varying pricing and load control strategies on residential summer peak electricity use: A review. *Energy Policy*, 38(7):3289–3296, July 2010.

[133] Francisco J. Nogales, Javier Contreras, Antonio J. Conejo, and Rosario Espínola. Forecasting next-day electricity prices by time series models. *IEEE Transactions on Power Systems*, 17(2):342–348, May 2002.

[134] Jason Osborne. Notes on the use of data transformations. *Practical Assessment, Research & Evaluation*, 8(6):1–8, December 2002.

[135] Oliver Parson, Siddhartha Ghosh, Mark Weal, and Alex Rogers. Nonintrusive load monitoring using prior models of general appliance types. In *26th Conference on Artificial Intelligence (AAAI)*, pages 356–362. AAAI Press, July 2012.

[136] Shwetak N. Patel, Thomas Robertson, Julie A. Kientz, Matthew S. Reynolds, and Gregory D. Abowd. At the flick of a switch: Detecting and classifying unique electrical events on the residential power line. In *9th International Conference on Ubiquitous Computing (UbiComp)*, pages 271–288, September 2007.

[137] Daniel Pauli. Open metering. Master's thesis, ETH Zurich, 2012.

[138] Adrian Payne and Pennie Frow. Relationship marketing: Key issues for the utilities sector. *Journal of Marketing Management*, 13(5):463–477, 1997.

[139] Wouter Poortinga, Linda Steg, Charles Vlek, and Gerwin Wiersma. Household preferences for energy-saving measures: A conjoint analysis. *Journal of Economic*

Psychology, 24(1):49–64, February 2003.

[140] Pavel Pudil, F. J. Ferri, Jana Novovicova, and Josef Kittler. Floating search methods for feature selection with nonmonotonic criterion functions. In *Pattern recognition, 1994. Vol. 2 - Conference B: 12th IAPR International Conference on Computer Vision & Image Processing*, volume 2, pages 279–283. IEEE, October 1994.

[141] Elias L. Quinn. Smart metering and privacy: Existing laws and competing policies. A report for the Colorado public utilities commission, May 2009.

[142] Joachim Radkau. Zur angeblichen Energiekrise des 18. Jahrhunderts: Revisionistische Betrachtungen über die "Holznot". *VSWG: Vierteljahrschrift für Sozial- und Wirtschaftsgeschichte*, 73(H.1):1–37, 1986.

[143] S. Raj Rajagopalan, Lalitha Sankar, Soheil Mohajer, and H. Vincent Poor. Smart meter privacy: A utility-privacy framework. In *2nd International Conference on Smart Grid Communications (SmartGridComm)*, pages 190–195. IEEE, October 2011.

[144] Teemu Räsänen, Juhani Ruuskanen, and Mikko Kolehmainen. Reducing energy consumption by using self-organizing maps to create more personalized electricity use information. *Applied Energy*, 85(9):830–840, February 2008.

[145] John O. Rawlings, Sastry G. Pantula, and David A. Dickey. *Applied regression analysis: A research tool*. Springer, April 1998.

[146] Jesse Read, Bernhard Pfahringer, Geoff Holmes, and Eibe Frank. Classifier chains for multi-label classification. *Machine Learning*, 85(3):333–359, December 2011.

[147] Andreas Reinhardt, Paul Baumann, Daniel Burgstahler, Matthias Hollick, Hristo Chonov, Marc Werner, and Ralf Steinmetz. On the accuracy of appliance identification based on distributed load metering data. In *2nd IFIP Conference on Sustainable Internet and ICT for Sustainability (SustainIT)*, pages 1–9. IEEE, October 2012.

[148] Andreas Reinhardt, Christin Delphine, and Salil S. Kanhere. Can smart plugs predict electric power consumption? A case study. In *11th International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services (Mobiquitous)*, pages 257–266. ACM, December 2014.

[149] Andreas Reinhardt, Frank Englert, and Delphine Christin. Averting the privacy risks of smart metering by local data preprocessing. *Pervasive and Mobile Computing*, 16, Part A:171–183, January 2015.

[150] Jeroen Reynders, Mathias Spelier, Nathalie Maes, Geert Van Ham, Bert Vande Meersche, and Geert Deconinck. Practical use of energy management systems. In *10th International Symposium on Ambient Intelligence and Embedded Systems (AmiEs)*, September 2011.

[151] Anthony Rowe, Mario Berges, and Raj Rajkumar. Contactless sensing of appliance state transitions through variations in electromagnetic fields. In *2nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings (BuildSys)*, pages 19–24. ACM, November 2010.

[152] Roland T. Rust and Ming-Hui Huang. The service revolution and the transformation of marketing science. *Marketing Science*, 33(2):206–221, January 2014.

[153] Leyna Sadamori. Automatic classification of private households using electricity consumption data. Master's thesis, TU Darmstadt, 2012.

[154] Manaswi Saha, Shailja Thakur, Amarjeet Singh, and Yuvraj Agarwal. Energylens: Combining smartphones with electricity meter for accurate activity detection and user annotation. In *5th International Conference on Future Energy Systems (e-Energy)*, pages 289–300. ACM, May 2014.

[155] Ignacio Benítez Sánchez, Ignacio Delgado Espinós, Laura Moreno Sarrión, Alfredo Quijano López, and Isabelle Navaló Burgos. Clients segmentation according to their domestic energy consumption by the use of self-organizing maps. In *6th International Conference on the European Energy Market (EEM)*, pages 1–6. IEEE, May 2009.

[156] Thomas F. Sanquist, Heather Orr, Bin Shui, and Alvah C. Bittner. Lifestyle factors in U.S. residential electricity consumption. *Energy Policy*, 42:354–364, March 2012.

[157] Joachim Schleich, Marian Klobasa, Marc Brunner, Sebastian Gölz, and Konrad Götz. Smart metering in Germany and Austria: Results of providing feedback information in a field trial. Working Paper Sustainability and Innovation S6/2011, Fraunhofer ISI, Karlsruhe, 2011.

[158] Sam H. Schurr. Energy efficiency and economic efficiency: A historical perspective. *Energy, productivity, and economic growth*, pages 203–214, 1983.

[159] Schweizer Bundesrat. Botschaft zum ersten Massnahmenpaket der Energiestrategie 2050 (Revision des Energierechts) und zur Volksinitiative "Für den geordneten Ausstieg aus der Atomenergie (Atomausstiegsinitiative)", September 2013.

[160] Elaine Shi, T.-H. Hubert Chan, Eleanor Rieffel, Richard Chow, and Dawn Song. Privacy-preserving aggregation of time-series data. In *18th Annual Network & Distributed System Security Symposium (NDSS)*, February 2011.

[161] Siemens. Einfluss moderner Technik auf Oberschwingungen im Verteilungsnetz. Technische Schriftenreihe, Ausgabe 14, http://w3.siemens.com/powerdistribution/global/DE/consultant-support/download-center/tabcardseiten/Documents/Technische-Schriften/Ausgabe_14_Einfluss_moderner_Technik_auf_Oberschwingungen_im_Verteilungsnetz.pdf.

[162] Marina Sokolova and Guy Lapalme. A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, 45(4):427–437, July 2009.

[163] José Luís Sousa, António Gomes Martins, and Humberto Jorge. Dealing with the paradox of energy efficiency promotion by electric utilities. *Energy*, 57:251–258, August 2013.

[164] Stephan Spiegel and Sahin Albayrak. Energy disaggregation meets heating control. In *29th Annual ACM Symposium on Applied Computing (SAC)*, pages 559–566. ACM, March 2014.

[165] Ineke A.L. Stoop. *The hunt for the last respondent: Nonresponse in sample surveys*. Sociaal en Cultureel Planbu, May 2005.

[166] Sustainable Energy Authority of Ireland (SEAI). Energy in the residential sector, June 2008.

[167] Kosuke Suzuki, Shinkichi Inagaki, Tatsuya Suzuki, Hisahide Nakamura, and Koichi Ito. Nonintrusive appliance load monitoring based on integer programming. In *SICE Annual Conference*, pages 2742–2747. IEEE, August 2008.

[168] Lukas G. Swan and V. Ismet Ugursal. Modeling of end-use energy consumption in the residential sector: A review of modeling techniques. *Renewable and Sustainable Energy Reviews*, 13(8):1819–1835, 2009.

[169] The Edison Foundation. Utility-scale smart meter deployments: Building block of the evolving power grid, September 2014.

[170] Sergios Theodoridis, Aggelos Pikrakis, Konstantinos Koutroumbas, and Dionisis Cavouras. *Introduction to pattern recognition: A MATLAB approach*. Elsevier, April 2010.

[171] Jorge Vasconcelos. Survey of regulatory and technological developments concerning smart metering in the European Union electricity market. *EUI RSCAS Policy Papers*, 1:1–64, 2008.

[172] Iana Vassileva and Javier Campillo. Increasing energy efficiency in low-income households through targeting awareness and behavioral change. *To appear in Renewable Energy*, 67:59–63, July 2013.

[173] Verbraucherzentrale Rheinland-Pfalz, Öko-Institut e.V.. Energieverbrauch von Spülmaschinen. http://www.oeko.de/uploads/oeko/forschung_beratung/themen/nachhaltiger_konsum/infoblatt_spuelmaschine.pdf.

[174] Sergio Valero Verdú, Mario Ortiz García, Carolina Senabre, Antonio Gabaldón Marín, and Francisco J. García Franco. Classification, filtering, and identification of electrical customer load patterns through the use of self-organizing maps. *IEEE Transactions on Power Systems*, 21(4):1672–1682, November 2006.

[175] Andrew J. Viterbi. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Transactions on Information Theory*, 13(2):260–269, April 1967.

[176] Hannß Carl von Carlowitz. *Sylvicultura Oeconomica – Haußwirthliche Nachricht und naturmäßige Anweisung zur wilden Baum-Zucht*. Braun, 1732.

[177] Morgan C. Wang and Brad J. Bushman. Using the normal quantile plot to explore meta-analytic data sets. *Psychological Methods*, 3(1):46–54, March 1998.

[178] Dorothy Watson and James Williams. Irish national survey of housing quality 2001-2002, November 2003.

[179] Markus Weiss, Adrian Helfenstein, Friedemann Mattern, and Thorsten Staake. Leveraging smart meter data to recognize home appliances. In *10th International Conference on Pervasive Computing and Communications (PerCom)*, pages 190–197. IEEE, March 2012.

[180] Markus Weiss, Friedemann Mattern, and Christian Beckel. Smart energy consumption feedback – Connecting smartphones to smart meters. *ERCIM News*, 92:14–15, January 2013.

[181] A. Wayne Whitney. A direct method of nonparametric measurement selection. *IEEE Transactions on Computers*, 100(9):1100–1103, September 1971.

[182] Tri Kurniawan Wijaya, Tanuja Ganu, Dipanjan Chakraborty, Karl Aberer, and Deva P. Seetharam. Consumer segmentation and knowledge extraction from smart

meter and survey data. In *14th SIAM International Conference on Data Mining (SDM)*, pages 226–234. SIAM, April 2014.

[183] Svante Wold, Kim Esbensen, and Paul Geladi. Principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, 2(1-3):37–52, August 1987.

[184] World Energy Council. World Energy Perspective, 2013.

[185] Michael Zeifman. Disaggregation of home energy display data using probabilistic approach. *IEEE Transactions on Consumer Electronics*, 58(1):23–31, February 2012.

[186] Michael Zeifman and Kurt Roth. Nonintrusive appliance load monitoring: Review and outlook. In *29th International Conference on Consumer Electronics (ICCE)*, pages 239–240. IEEE, Januaray 2011.

[187] Ahmed Zoha, Alexander Gluhak, Muhammad Ali Imran, and Sutharshan Rajasegarar. Non-intrusive load monitoring approaches for disaggregated energy sensing: A survey. *Sensors*, 12(12):16838–16866, December 2012.

Referenced Web resources

[188] Christian Beckel. CLASS – A system for automatically revealing household characteristics from smart meter data. https://github.com/beckel/class.

[189] Christian Beckel and Romano Cicchetti. NILM-EVAL: An evaluation framework for non-intrusive load monitoring algorithms. https://github.com/beckel/nilm-eval.

[190] Christian Beckel and Wilhelm Kleiminger. ECO (Electricity Consumption & Occupancy) data set. http://www.vs.inf.ethz.ch/res/show.html?what=eco-data.

[191] Bund der Energieverbraucher – Der Kohlenklau. http://www.energieverbraucher.de/de/Der-Kohlenklau__1446.

[192] Bund der Energieverbraucher – Der Wattfrass. http://www.energieverbraucher.de/de/Der-Wattfrass__1437.

[193] Bundesministerium für Wirtschaft und Energie, Berlin. Erneuerbare Energien. http://bmwi.de/BMWi/Redaktion/Binaer/Energiedaten/energietraeger08-erneuerbare-energien,property=blob,bereich=bmwi2012,sprache=de,rwb=true.xls.

[194] Central Statistics Office of Ireland. Private households in permanent housing units by type of private accommodation, province county or city and census year. http://www.cso.ie/px/pxeirestat/Statire/SelectVarVal/Define.asp?maintable=CD511&PLanguage=0, 2011.

[195] Chih-Chung Chang and Chih-Jen Lin. LIBSVM – A library for support vector machines. http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[196] Department for Environment, Food and Rural Affairs (defra) – Household electricity study EV0702. http://randd.defra.gov.uk/Default.aspx?Menu=Menu&Module=More&Location=None&ProjectID=17359.

[197] Dojo Toolkit. https://dojotoolkit.org/.

[198] Echelon smart meter. http://en.wikipedia.org/wiki/Smart_meter#/media/File:Intelligenter_zaehler-_Smart_meter.jpg.

[199] edi@energy, OBIS-Kennzahlen-System, version 2.2a. http://www.edi-energy.de/files2%5COBIS-Kennzahlen-System%202.2a_20130401.pdf.

[200] Elster type R15 electricity meter. http://de.wikipedia.org/wiki/Intelligenter_Z%C3%A4hler#/media/File:Elster_Type_R15_electricity_meter.jpeg.

[201] ENERGY STAR rebate program. http://www.nolinrecc.com/energy-star-rebate-program.asp.

[202] ETH Zurich. Brutus cluster. http://en.wikipedia.org/wiki/Brutus_cluster.

[203] eurostat. Final energy consumption in households. http://ec.europa.eu/eurostat/web/products-datasets/-/t2020_rk200.

[204] eurostat. Number of private households by household composition, number of children and age of youngest child. http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=lfst_hhnhtych&lang=en.

[205] FAST FORWARD AG – FAST EnergyCam: The smart clip-on meter reader for home and business. http://fastforward.ag/eng/index.html#Product.

[206] Fluksometer v2. https://www.flukso.net/content/fluksometer-v2.

[207] German History in Documents and Images (GHDI) – The West German “ban on driving motor vehicles”, November 1973. http://germanhistorydocs.ghi-dc.org/sub_document.cfm?document_id=941&language=english.

[208] Nick Gillian. Gesture Recognition Toolkit (GRT) wiki. AdaBoost. http://www.nickgillian.com/wiki/pmwiki.php?n=GRT.AdaBoost.

[209] Juri Glass, Mathias Runge, and Nadim El Sayed. Implementation in C of the Smart Message Language (SML) protocol. https://github.com/dailab/libsml.

[210] Irish Social Science Data Archive. Commission for Energy Regulation (CER) electricity customer behaviour trial. Residential pre-trial survey questionnaire. http://www.ucd.ie/issda/static/documentation/cer/smartmeter/cer-residential-pre-trial-survey.pdf.

[211] Irish Social Science Data Archive. Commission for Energy Regulation (CER) smart metering project. http://www.ucd.ie/issda/data/commissionforenergyregulationcer/.

[212] Wilhelm Kleiminger and Daniel Pauli. Pylon smart metering framework in C. https://github.com/wkleiminger/pylon.

[213] Zico J. Kolter and Matthew J. Johnson. The reference energy disaggregation data set (redd). http://redd.csail.mit.edu.

[214] Charly Kühnast. Stromzähler auslesen. http://kuehnast.com/s9y/archives/460-Stromzaehler-auslesen.html.

[215] Laboratory of Computer and Information Science. SOM toolbox 2.0. http://www.cis.hut.fi/somtoolbox/.

[216] Landis+Gyr E750. http://www.landisgyr.com/webfoo/wp-content/uploads/2012/12/Landis+Gyr-E750-Brochure-English.pdf.

[217] Linksys WRT54GL wireless router. http://www.linksys.com/my/p/P-WRT54GL/.

[218] MathWorks. MATLAB. http://ch.mathworks.com/products/matlab/.

[219] MathWorks. Statistics and Machine Learning Toolbox. http://www.mathworks.de/products/statistics.

[220] Neural Energy. Italy – 27 million smart meters. http://www.neuralenergy.info/2009/07/italy.html.

[221] ONZO's Smart Energy Kit. http://www.thegreenitreview.com/2011/04/uk-company-onzos-smart-energy-kit-is.html.

[222] Openwrt. https://openwrt.org/.

[223] Oliver Parson. What even is supervised/unsupervised disaggregation? http://blog.oliverparson.co.uk/2015/05/what-even-is-supervisedunsupervised.html.

[224] Sven Petai. python-plugwise.
https://bitbucket.org/hadara/python-plugwise/wiki/Home.

[225] Plugwise. https://www.plugwise.com/.

[226] readwrite – The art, science and business of recommendation engines.
http://readwrite.com/2007/01/16/recommendation_engines.

[227] RF Code R130 dry contact active RFID tag.
http://www.codegate.co.uk/hardware/rf-code/r130-dry-contact-sensing-active-rfid-tag.

[228] Roving Networks, RN-134.
http://www.microchip.com/rn/products/RN_134_RN134_I_RM.html.

[229] S0-Recorder: Starter-Set EDL-Recorder mit Opto-Kopf OC3.
http://www.s0-recorder.com/shop/pi6/Starter-Set-EDL-Recorder-mit-Opto-kopf-OC3.html.

[230] SheevaPlugs.
http://www.globalscaletechnologies.com/t-sheevaplugs.aspx.

[231] timeanddate.com. Sunrise and sunset calculator.
http://www.timeanddate.com/worldclock/sunrise.html.

[232] United States Fuel Administration / Coles Phillips. Light consumes coal – Save light, save coal. http://www.loc.gov/pictures/item/2002712078/, 1917.

[233] U.S. Energy Information Administration. 2009 Residential Energy Consumption Survey (RECS). End-use consumption totals and averages, U.S. homes.
http://www.eia.gov/consumption/residential/data/2009/c&e/end-use/xls/CE3.1Site%20End-Use%20Consumption%20in%20US.xlsx, January 2013.

[234] WDV-Molliné GmbH. Bluetooth Optokopf.
http://www.molline.de/produkte-loesungen/zubehoer-und-optionen/optokopf-bluetooth/.

[235] WeatherOnline. Weather information for Ireland. http://www.weatheronline.co.uk/weather/maps/current?LANG=en&CONT=euro&LAND=IE.

[236] Wharton Customer Analytics Initiative. Leveraging smartmeter data to increase customer engagement and reduce energy consumption.
http://wcai.wharton.upenn.edu/leveraging-smartmeter-data-

increase-customer-engagement-reduce-energy-consumption/, October 2012.

[237] The ZigBee Alliance. www.zigbee.org/.

Appendix A

Interviews with energy consultants

In the following, we present the interviews we performed with four energy consultants from different Swiss utilities. The goal of the interviews was to find out if and how the consultants would benefit from having knowledge about their customers' household characteristics.

We first provide the template we used for performing the interviews. Next, we print the questions and results of each of the interviews. Note that the template merely served as a guideline for the interview process by highlighting important aspects we wanted to discuss with the consultants. During the interview, we took handwritten notes and transferred the notes after the interviews were finished. This means that the interviews given in this section reflects the meaning of the original interviews, but both the questions and the answers are written in our own words.

We used the same template for each of the four interviews. The responses are written in italic underneath the question. Also, all the interviews were conducted in German and have been translated into English afterwards.

A.1 Interview 1

Date: June 2012
Interviewee: Energy consultant

1. Selection of customers

    a) What is the main reason for getting personal energy consulting? Who performs the initial contact? How are the customers selected?

    Mainly financial incentives (80%). Mostly families. Customers ask us for energy consulting, thus there is no selection.

    b) Who pays for personal energy consulting?

    The service is free for interested customers. Other utilities charge roughly CHF 75, which typically amortizes after 1 year.

    c) Which saving potential is required such that personal energy consulting pays off?

    Typically consumption decreases by 10%. Amortization depends on the bill—often it takes about one year.

    d) Why does the utility perform personal energy consulting?

    (1) Political mandate, (2) utilities are also interested in saving energy, since they can sell excess energy on the market, (3) happy customers.

2. Preparation and execution of a personal energy consulting session

    a) What is the procedure of a consulting session?

    Utilities have access to yearly consumption data, sometimes 6-month data (winter and summer). This way they can estimate if customers have an electric heating system or ventilation. During the consulting session, the consultant compares the meter with the previous dates:

    - *Main tariff (on-peak hours) high: This is an indicator for thermal consumers or wrong configurations (sometimes the meter is connected to the wrong apartment, or appliances or the boiler of other apartments are connected to the meter).*

- *Low tariff (off-peak hours) high: High standby consumption or high consumption during the weekend.*

b) How does personal energy consulting differ from providing generic energy saving recommendations?

Personalizing energy consulting is important due to the psychological effect when the own household is affected. Also, the recommendations actually apply in this scenario, it is not just a generic list.

c) What is the role of the customer's load curve for energy consulting?

Only main tariff and low tariff are known, no detailed load curve.

d) What is the role of the number of appliances and their usage?

It would be nice to know the "degree of technology" in the household (low, medium, high) in order to better assess the situation.

e) What is the role of the number of occupants and their "patterns" (*e.g.*, who is at home during the day, who is employed) for energy consulting?

Important information consists of:

- *Number of persons (*e.g.*, family, kids or not, women)—Women are for instance an indicator that much water is consumed in the household.*
- *Floor area?*
- *Boiler?*
- *Solar panels?*
- *Type of warm water heating?*

*Presence and absence would be a nice to have: People who are never at home but consume much energy should reduce their standby consumption. These types of information help to assess the situation in the household and find saving potentials and anomalies (*e.g.*, if the consumption is similar to the consumption of three-person households but only one person is living in the household). The same applies for the "degree of technology" in the household.*

f) Do the customers learn something new or do consultants show them their already-known vices?

*Often households lack the right assessment of their consumption (*e.g.*, lighting vs. microwave). Customers are often surprised about high saving potential of lighting.*

3. Saving potential

   a) What are typical saving measures? Which of those are related to heating, which are related to the electricity consumption?

   *Lighting (*i.e.*, installing energy-efficient light bulbs).*

   b) What are very specific / rare measures or recommendations?

   *Thermal consumers (*e.g.*, electric blankets), dehumidifiers that are running too long.*

   c) Are there types of households (*e.g.*, singles, families, employed, not employed) for whom energy-saving measures (and thus energy consulting) are particularly effective?

   Families are often requesting personal energy consulting. For what types of households measures are particularly effective is difficult to say.

   d) Are there houses (*e.g.*, old, young, valuable, not valuable) for which energy-saving measures (and thus energy consulting) are particularly effective?

   Old houses have typically a higher saving potential: Appliances are often old. Families with a new house often have money and saving is only a secondary goal.

   e) How much electricity saving potential does typically exist in a household?

   Most often 10% electricity savings are possible without large changes in consumption behavior. Large potential: Replacing light bulbs, reducing standby consumption.

   f) What measures to reduce electricity consumption are typically recommended?

   *Different things (*e.g.*, replacing light bulbs with energy-efficient light bulbs).*

   g) How can one save energy that is used for cooling (*e.g.*, refrigerator, freezer)?

   Refrigerators are often inefficient (broken seals or weak insulation) or used inefficiently (if people do not know different zones in the refrigerator or set it too high: Butter should be easily spreadable). Buying a new refrigerator is *worthwhile if the old one is more than 10 years old and has efficiency label F or G.*

4. Follow-up

   a) Are there partnerships to implement the measures?

      It is important to assist the customers, for instance by offering a shop that sells energy-efficient light bulbs. However, it is difficult to strike the balance between independent consulting and providing partnerships to support implementation.

   b) Is there supervision after the energy consulting session in order to monitor the implementation?

      Customers can rent different types of electricity measurement devices to identify energy guzzlers. In some cases we check the savings that have been realized after performing the consulting session. This is sometimes important to proof energy saving activities, for instance to the city or to the state.

5. Outlook

   a) What role will energy consulting play in the future considering the rise of smart metering and smart grids?

      A big role, if energy is getting more expensive. For the smart grid, automation will be more important than "manual" load shifting through behavior change.

   b) If one could categorize households—What would be relevant categories from the viewpoint of the utility (for energy consulting and beyond)?

      That would be interesting for the marketing department.

A.2 Interview 2

Date: July 2012
Interviewee: Energy consultant

1. Selection of customers

   a) How many customers are getting personal energy consulting?

      Our company has 400,000 *customers. I don't know how many of them get personal energy consulting.*

   b) What is the main reason for getting personal energy consulting? Who performs the initial contact? How are the customers selected?

      *Our customers visit the advisory center of our company. We do not visit customers at home. Some customers want concrete recommendations, others more generic information or they have questions with respect to recent developments (*e.g.*, nuclear phase-out, health concerns due to mercury). 90% of the recommendations are in the category of electricity usage and generation (*e.g.*, energy saving lamps, radioactivity, electromagnetic fields). The only consulting that takes place on site is energy coaching when houses are being refurbished.*

      Goals of energy consulting:

      - *Reduce costs without loss of comfort.*
      - *Image cultivation.*

      Target groups:

      - *Single-person households and couples without children.*
      - *Families not so much. They have other things to do than saving energy.*
      - *Old people (they often have time).*

      *In cooperatives, a responsible person (*e.g.*, a caretaker) often provides energy consulting for households. In general, there are few houses and many apartments (for geographical reasons). Building restorations are rarely performed for apartment houses because of the landlord/tenant dilemma. The main reasons for customers to ask for energy consulting are (1) questions with respect to energy efficiency, (2) questions about their energy bill, and (3) questions about recent trends and developments.*

c) Who pays for personal energy consulting?

   *It is free of charge, financed by the city. In addition to energy consulting, the city runs an energy efficiency fund to subsidize investments (*e.g.*, heat pumps and solar panels) and energy efficiency campaigns. Overall, 10% of our profit goes into energy consulting.*

d) Which saving potential is required such that personal energy consulting pays off?

   Performing energy consulting directly in households would be too expensive. It is an important question what efforts utilities should spend for private customers. We did a trial, in which we performed energy consulting in 100 households (2 hours per household plus coupons for energy saving lamps). The long-term savings are unclear.

e) Why does the utility perform personal energy consulting?

   Political mandate – we belong to the city. Also, demand must be reduced because energy production shrinks.

2. Preparation and execution of a personal energy consulting session

   a) What type of information about the customer does the utility before the consulting session?

      - *Address.*
      - *Yearly meter reading.*
      - *In theory: Does the customer have an electric boiler? If yes: How much power does it consume? In practice, this information is difficult to extract from the SAP system.*
      - *Signal of the ripple control.*
      - *Planned: CRM (*e.g.*, phone number, e-mail address) to improve customer retention (which is important due to the ongoing market liberalization).*

   b) What additional information would be helpful when selecting customers?

      Currently there is no customer selection.

   c) What additional information would be helpful to plan and execute the consulting session?

The optimum would be:

- *A pie chart that breaks down the consumption of a household.*
- *A form (which is already accessible online) filled out by the customer beforehand in order to get as much information as possible about the customer.*

This would be a great basis for the consulting session. The consulting session itself must be performed in a careful way: What recommendations are expected? Is it important to sensitize the occupants for the topic? Do they request energy consulting because they want to change a product?

Household occupancy is not so important at the moment.

d) What is the procedure of a consulting session?

A conversation in our customer center.

e) How does personal energy consulting differ from providing generic energy saving recommendations?

We already have a list of generic recommendations. In a personal session, we can better address the customer's personal needs.

f) What is the role of the customer's load curve for energy consulting?

None (for private customers).

g) What is the role of the number of appliances and their usage?

Energy saving recommendations are mostly related to the number and usage of household appliances.

h) What is the role of the number of occupants and their "patterns" (*e.g.*, who is at home during the day, who is employed) for energy consulting?

Occupancy barely plays a role for energy consulting. The number of occupants is important, because single-person households, couples, and retired persons are a popular target group for energy consulting compared to families.

i) Do the customers learn something new or do consultants show them their already-known vices?

Good question—It is mostly about sensitizing the customers.

3. Saving potential

   a) What measures have the biggest saving potential?

      Heat—Electrical water heating it the biggest consumer; saving warm water implies saving energy.

   b) What are typical saving measures? Which of those are related to heating, which are related to the electricity consumption?

      The biggest effect is caused by things that run frequently or for a long time. In addition, light and everything that produces heat: Dishwasher, washing machine, cooking, boiler.

      Building insulation reduces heating costs, however it implicitly leads to an increased use of air conditioning.

   c) Are there types of households (*e.g.*, singles, families, employed, not employed) for whom energy-saving measures (and thus energy consulting) are particularly effective?

      *Singles, couples without children, retirees. Families have little "room" (*i.e.*, time) to save energy.*

   d) How much electricity saving potential does typically exist in a household?

      30% without loss of comfort, excluding investment costs.

   e) What measures to reduce electricity consumption are typically recommended?

      - *Replacement of energy-efficient light bulbs.*
      - *Replacement of the fridge.*
      - *Boiler: Use less (*i.e.*, hot water).*

   f) How can one save heating costs?

      Heating costs are shrinking (per square meter). Reducing temperature by 1°C saves 6% heat energy.

   g) What is the role of water heating when performing energy consulting?

      A big role.

4. Follow-up

a) Are there partnerships to implement the measures?

- *Electrical scooters.*
- *Energy portal that provides energy-efficient devices.*
- *Energy performance certificate for buildings.*

5. Outlook

a) What role will energy consulting play in the future considering the rise of smart metering and smart grids?

Smart metering: Savings are low, ripple control is more effective, and there is no privacy problem. Consumption of individual consumers is not so important, because consumption only needs to be within a certain band. Randomness from households is evened out due to aggregation of several households. Smart grid: No major role for energy consulting.

b) If one could categorize households—What would be relevant categories from the viewpoint of the utility (for energy consulting and beyond)?

Currently the only existing classification consists of small*,* medium*, and* big *customers. Households are mostly small customers. A more detailed classification is only (potentially) relevant for marketing.*

A.3 Interview 3

Date: August 2012
Interviewees: Director energy consulting, director energy efficiency & smart metering

1. Selection of customers

   a) What is the main reason for getting personal energy consulting? Who performs the initial contact? How are the customers selected?

      The customer asks for it. We increasingly do commercials for it: Attachment to the bill, at fairs, and online. The main reason is a replacement of the electric heating or heat pump. Customers have many questions, for instance if investments are worth it or how they can get funding for solar energy. High electricity bills are no reason for energy consulting.

   b) Who pays for personal energy consulting?

      50/50 (utility/customer).

   c) Which saving potential is required such that personal energy consulting pays off?

      Saving potential is not important. More important are aspects that concern replacements of devices or infrastructure.

   d) Why does the utility perform personal energy consulting?

      No political mandate. Rather: Customer service.

2. Preparation and execution of a personal energy consulting session

   a) What type of information about the customer does the utility before the consulting session?

      - *Electricity consumption (semi-yearly).*
      - *Oil consumption.*
      - *Do they have electric heating?*

   b) What is the procedure of a consulting session?

      *A consulting session takes place on site. We evaluate their systems. Details about appliance usage are covered by additional material (*e.g.*, generic saving* *recommendations). In addition we inform customers about programs to replace light bulbs, refrigerators, or washing machines. Overall, the focus lies on space heating, water heating, and the warm water pump.*

c) How does personal energy consulting differ from providing generic energy saving recommendations?

Customers receive a 8-page report about their *house. This personalization is important. For "smaller" aspects, generic information is sufficient.*

d) What is the role of the customer's load curve for energy consulting?

One could recognize if a household has an old refrigerator or freezer. The optimum would be a pie chart that shows the consumption per category. This way, it also gets easier to compare households with each other. Easy would be: "Total consumption of your household compared to your neighbors". Interesting, but more difficult to achieve: "This device consumes this much compared to the rest of your household".

3. Saving potential

a) What measures have the biggest saving potential?

Warm water heating is always an important topic.

b) What are typical saving measures? Which of those are related to heating, which are related to the electricity consumption?

- *Reduce heat loss.*
- *Replace technical infrastructure (*e.g.*, heat pump).*
- *Behavioral aspects.*
  - *Optimize behavior (*i.e.*, reduction & load shifting).*
  - *Optimize appliance configuration.*

c) Are there types of households (*e.g.*, singles, families, employed, not employed) for whom energy-saving measures (and thus energy consulting) are particularly effective?

All types of customers are interested in energy consulting, people with houses as well as families a bit more than other groups. Families often do it for ecological reasons, retired people accept the fact that infrastructure must be *replaced. There is a pre-selection per phone. The size of the household plays no role.*

d) How much electricity saving potential does typically exist in a household?

15% without investments (7%–8% load reduction, the rest through load shifting).

e) What is the role of water heating when performing energy consulting?

An important role.

4. Follow-up

a) Are there long-term plans to implement energy-saving measures?

We do not check this.

b) Are there partnerships to implement the measures?

We have a shop where customers can buy energy saving lamps and standby killers. We also gave away (almost) for free 2500 *LED lamps, financed through a campaign named "subsidies per* kW*". Finally, we have a campaign that supports the replacement of refrigerators.*

c) Is there supervision after the energy consulting session in order to monitor the implementation?

No.

5. Outlook

a) Will personal energy consulting gain importance due to the plans of Switzerland to phase out nuclear energy?

Energy consulting for private customers costs 500 CHF *to* 1000 CHF *and is thus not cost-efficient (this will most likely be similar in the future). Its relevance will however increase. One can divide current measures into two groups:*

- *Sensitization for the masses through generic recommendations and different campaigns.*
- *High quality consulting.*

Currently, there is nothing in between.

A.4 Interview 4

Date: July 2012
Interviewee: Energy consultant

1. Selection of customers

   a) How many customers are getting personal energy consulting?

      300 customers per year.

   b) What is the main reason for getting personal energy consulting? Who performs the initial contact? How are the customers selected?

      Customers ask us for energy consulting, thus there is no selection. Customers want

      - *Confirmation (or disconfirmation) of an idea.*
      - *A second opinion as quality assurance (*e.g., *when buying a heat pump).*
      - *More and more consulting for funding opportunities: Putting together an product/investment packet, for instance when replacing walls or windows.*

   c) Who pays for personal energy consulting?

      *Customers pay a portion of the costs (*i.e., 200 CHF*). The consulting session typically lasts* 3 h *to* 4 h.

   d) Which saving potential is required such that personal energy consulting pays off?

      Consulting to save electricity is not worth it from a financial point of view, because electricity is too cheap.

   e) Why does the utility perform personal energy consulting?

      It is politically motivated.

2. Preparation and execution of a personal energy consulting session

   a) What type of information about the customer does the utility before the consulting session?

Almost nothing: Electricity consumption and energy consumed for space heating in the previous year. Until recently, customers received a device to measure the electricity consumption of individual appliances such as the refrigerator.

b) What additional information would be helpful to plan and execute the consulting session?

Numbers! (1) Total consumption, (2) relation between high tariff consumption and low tariff consumption, (3) information whether the consumption is high, medium, or low compared to others, (4) floor area of the household, (5) "degree of technology" in the household.

Numbers alone are not interesting, it is required to compare them with reference values.

c) What is the procedure of a consulting session?

*Only with pen & paper as well as a folder with information material. We do not use a laptop or other technical devices. First, we ask the customers if they have any specific questions. Next, we determine relevant numbers (*e.g.*, heat energy and electricity per square meter). We also go through the house and examine everything including roof, walls, floor, and windows. While going through the house we get an impression on the "degree of technology" in the household (just an impression, no detailed picture).*

d) How does personal energy consulting differ from providing generic energy saving recommendations?

*The most important thing is to show people where they stand compared to others. It is also important to verify that recommendations match expectations of the customers (*e.g.*, is it worthwile to install solar panels?). Based on the numbers we generate recommendations that are accepted by the customers. To this end, we need a good set of actual numbers and reference numbers, for instance with respect to the building (*e.g.*, floor area, with/without boiler).*

e) What is the role of the customer's load curve for energy consulting?

In particular the comparison between high tariff consumption and low tariff consumption is interesting. The load curve also tells us which part stems from consumers that are always and which part is caused by "activity" of the occupants. Having information about the load curve would be interesting, but *also requires interpretation. A week of consumption would not be sufficient, it should be measured over a longer time period.*

f) What is the role of the number of appliances and their usage?

Buying new appliances and investing in energy saving lamps is more worthwhile for commercial buildings than for households.

g) What is the role of the number of occupants and their "patterns" (*e.g.*, who is at home during the day, who is employed) for energy consulting?

Electricity consumption is almost independent of the number occupants. The base load is the most important indicator, more important then "presence".

3. Saving potential

a) What measures have the biggest saving potential?

Biggest consumer: Warm water boiler.

b) Are there types of households (*e.g.*, singles, families, employed, not employed) for whom energy-saving measures (and thus energy consulting) are particularly effective?

*Everybody who has investment potential (*e.g.*, employed people). Retirees not so much, because they do not plan thus far in the future.*

- *Detached house: These households typically have money for investment, but they are not the primary target, because they are often "active" anyway.*
- *Multi party houses: High potential in particular in buildings from the 70s, 80s, or 90s. However, lots of consulting is required and the decision-making process is long.*
- *Very large buildings: Owners are often institutions, high threshold, perform investments only for financial benefits (*i.e.*, no ecological incentives).*
- *Commercial use of buildings: High potential, for instance in restaurants, which often still have ventilation from times in which people were allowed to smoke inside.*

c) How much electricity saving potential does typically exist in a household?

*Immediately: 5% to 10%. Through investments (*e.g.*, replacement of cooling appliances, heat pump, or electrical appliances): Maximum of* 1000 kWh *per household (*i.e.*, 20%–25%).*

d) What measures to reduce electricity consumption are typically recommended?

The low cost of electricity is the main barrier for investments: They are often not profitable. Therefore, we give recommendations how to replace things when they break. Industry is responsible to produce more energy-efficient appliances, politics are responsible to provide the right regulations.

e) How can one save heating costs?

Saving electricity is more difficult than saving heat energy: Cost of heating is decreasing anyway (per square meter). The problem is that the floor area is increasing. Using more and more glass is also a problem: In summer it gets hot, in winter it gets cool.

*10% savings are possible without investments (*e.g.*, by closing the windows and by configuring temperatures and boilers correctly).*

f) How can one save energy that is used for cooling (*e.g.*, refrigerator, freezer)?

Through the replacement of cooling appliances. We once offered to measure the consumption of appliances through a measuring device. However, it takes a very long time until replacement costs amortize.

4. Follow-up

a) Are there partnerships to implement the measures?

No.

b) Is there supervision after the energy consulting session in order to monitor the implementation?

No.

5. Outlook

a) Will personal energy consulting gain importance due to the plans of Switzerland to phase out nuclear energy?

Electricity costs should be 3x higher than they are today.

b) What role will energy consulting play in the future considering the rise of smart metering and smart grids?

Not directly related to the smart grid. Combined heat and power (CHP) will gain importance. When someone is cooking, for instance, the CHP could be switched on to heat the household and at the same time produce electricity for cooking.

c) What type of information about customers would be helpful when planning energy consulting sessions (*e.g.*, selection of customers, planning of the actual visits)?

Category of the dwelling:

- *Office? School? Restaurant?*
- *Detached house? (most often)*
- *Small apartment house?*
- *Large apartment house?*
- *Apartment?*

The age of the dwelling is important as well. Contacting those who are interested in energy consulting would be nice.

Appendix B

Features and characteristics used for the SOM analysis

This section shows in detail the features we used for the SOM analysis described in section 3.2.2 and the household characteristics we investigated. These are similar to the features and characteristics we ultimately implemented in our classification system, which are provided in section 3.2.1 and table 3.4, respectively. The contents of this section is mainly taken from [21].

Building and extending upon related work [38, 72, 155], we define features based on consumption values of individual days as well as aggregated over the entire week or over workdays and weekends separately. In particular, we identify 4 groups of features: (1) consumption figures, (2) ratios, (3) temporal properties, and (4) statistical properties.

Consumption figures correspond to simple aggregates of the actual consumption values of a household. For instance, the minimum or maximum consumption values of a day or the average consumption within a specific period (e.g., in the morning or during the night) are referred to as consumption figures. Ratios are quotients of average consumption values of different periods of a day. An example is the ratio between the average consumption in the morning and that during lunch-time. Temporal properties describe the time of the day in which certain events occur. Examples include the time where consumption reaches its daily maximum or the time of the day at which a given consumption threshold is exceeded for the first time. Finally, statistical properties allow to capture qualitative characteristics of the consumption curve. For instance, in order to determine how consumption profiles (of the same household) correlate to each other over subsequent days we compute the cross-correlation between these profiles. Table B.1 provides a list of all features we define and use in the context of this work. The table also shows the labels (on the right column), which we use to indicate the different features. The intervals *morning*, *noon*, *evening*, and *night* are defined as the time periods 6 a.m.–10 a.m., 10 a.m.–2 p.m., 6 p.m.–10 p.m., and 1 a.m.–5 a.m., respectively.

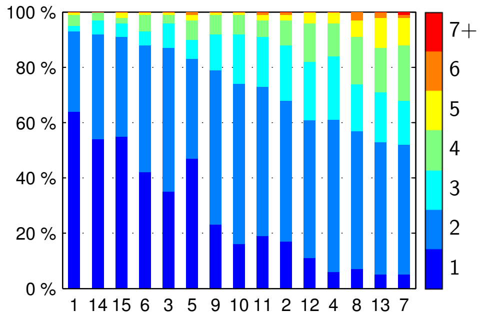
Table B.1: List of features used to build the input vectors of the self-organizing map. $\bar{P}$ denotes the 30-minute mean power samples provided by the data set.

(1) Consumption figures	
$\bar{P}$ (daily)	`c_day`
$\bar{P}$ (daily, weekdays only)	`c_weekday`
$\bar{P}$ (daily, weekend only)	`c_weekend`
$\bar{P}$ for (6 p.m.–10 p.m.)	`c_evening`
$\bar{P}$ for (6 a.m.–10 a.m.)	`c_morning`
$\bar{P}$ for (1 a.m.–5 a.m.)	`c_night`
$\bar{P}$ for (10 a.m.–2 p.m.)	`c_noon`
Maximum of $\bar{P}$	`c_max`
Minimum of $\bar{P}$	`c_min`
(2) Ratios	
Mean $\bar{P}$ over maximum $\bar{P}$	`r_mean/max`
Minimum $\bar{P}$ over mean $\bar{P}$	`r_min/mean`
`c_night` / `c_day`	`r_night/day`
`c_morning` / `c_noon`	`r_morning/noon`
`c_evening` / `c_noon`	`r_evening/noon`
(3) Temporal properties	
First time $\bar{P} > 1$kW	`t_above_1kw`
First time $\bar{P} > 2$kW	`t_above_2kw`
First time $\bar{P}$ reaches maximum	`t_daily_max`
Period for which $\bar{P} >$ mean	`t_above_mean`
(4) Statistical properties	
Variance	`s_variance`
$\sum(\lvert\bar{P}_t - \bar{P}_{t-1}\rvert)$ for all t	`s_diff`
Cross-correlation of subsequent days	`s_x-corr`
$\#\bar{P}$ with $(\bar{P}_t - \bar{P}_{t\pm1} > 0.2$ kW)	`s_num_peaks`

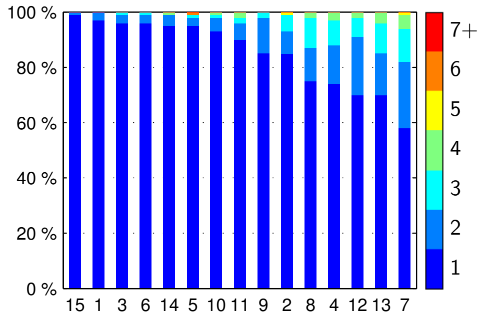
Table B.2 shows the complete list of characteristics that we extracted from the questionnaires in order to investigate to what extent they can be inferred from the electricity consumption traces. The distribution of each property over the SOM is shown in figure B.1. Figure 3.9 shows the corresponding cluster map. The plots are taken from [153].

Table B.2: Household characteristics we extracted from the questionnaires accompanying the CER data set and investigated in the SOM analysis.

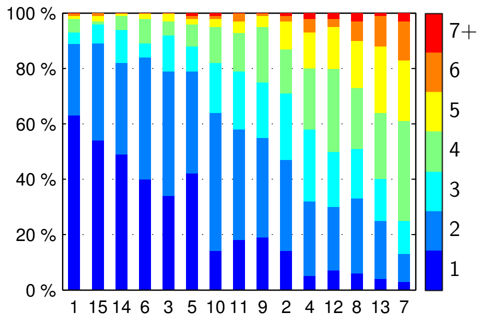
(1) Characteristics related to the occupants of the household	
Number of adults	`#adults`
Number of children	`#children`
Number of adults and children	`#persons`
Adults at home during day	`#adults@home`
Children at home during day	`#children@home`
Persons at home during day	`#pers@home`
Age of chief income earner	`age`
Employment of chief income earner	`employment`
Social class of chief income earner	`soc_class`
(2) Characteristics related to the dwelling	
Type of dwelling	`type_house`
Relationship to property	`own_house`
Age of building	`age_house`
Floor area	`floor_area`
Number of bedrooms	`#bedrooms`
(3) Characteristics related to the appliances in the household	
Type of cooking facilities	`type_cook`
Type of water heating	`type_water`
Type of space heating	`type_heat`
Number of appliances	`#appliances`
Entertainment	`#entertainment`
Percentage of energy saving lamps	`lighting`
Timed space heating system	`timed_heat`
Timed water heating system	`timed_water`



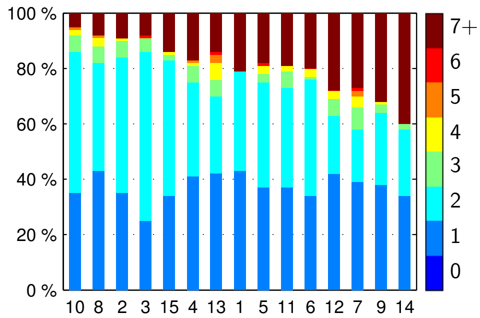
(a) #adults.



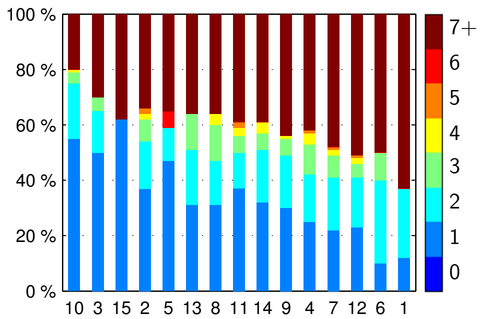
(b) #children.



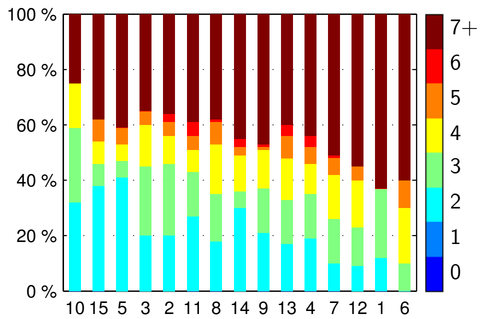
(c) #persons.



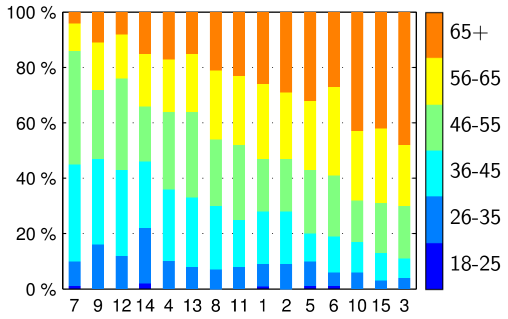
(d) #adults@home.



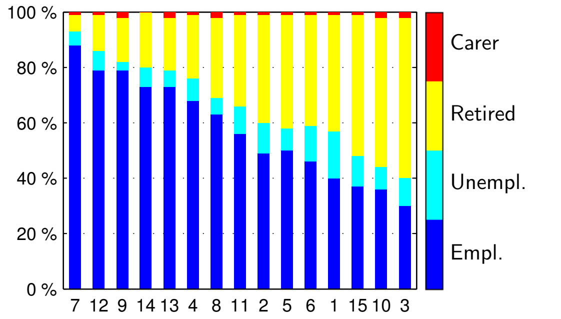
(e) #children@home.



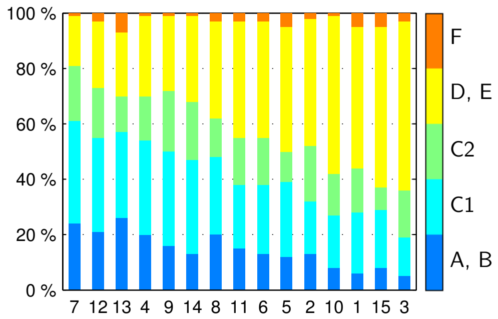
(f) #persons@home.



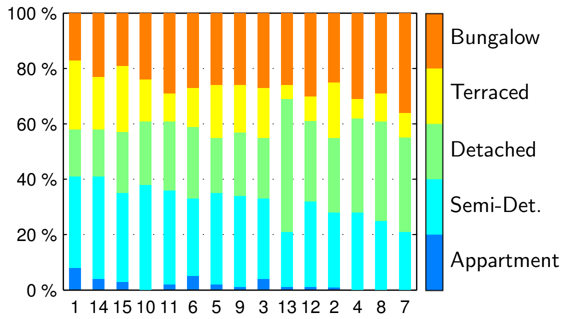
(g) age.



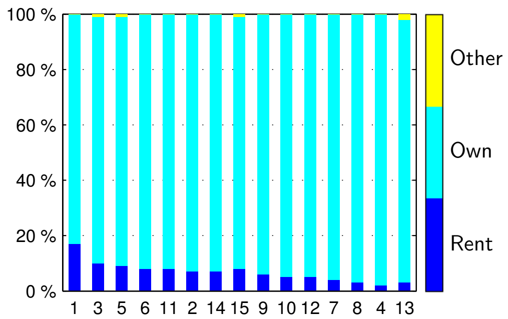
(h) employment.



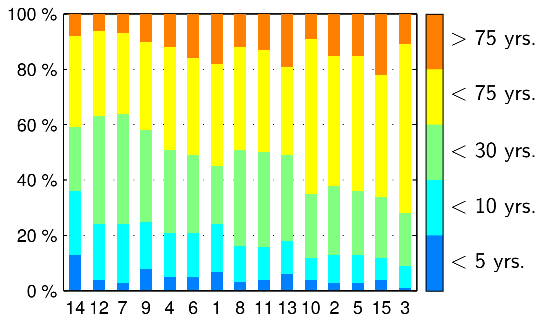
(i) soc_class.



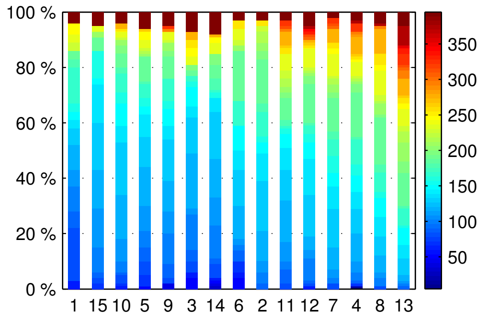
(j) type_house.



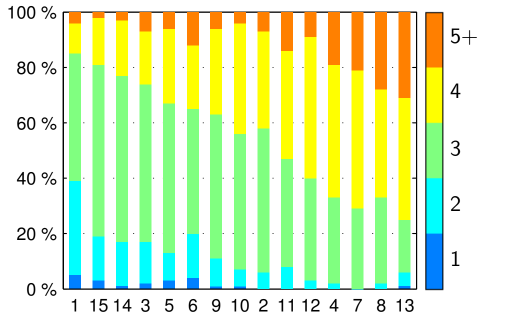
(k) own_house.



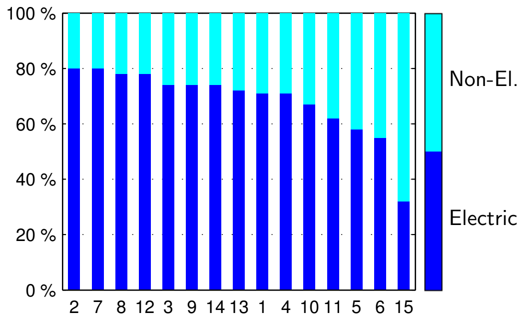
(l) age_house.



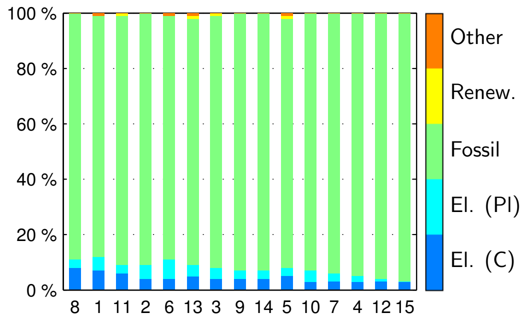
(m) floor_area.



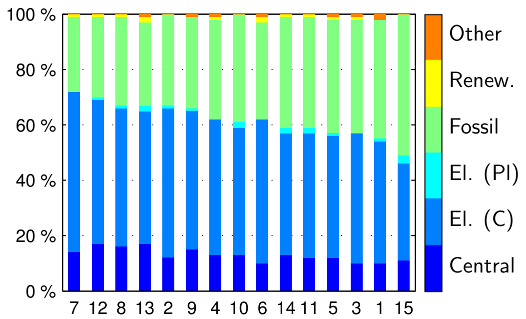
(n) #bedrooms.



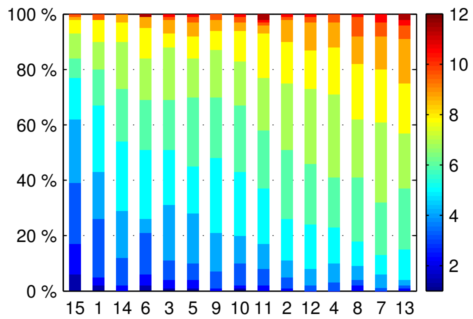
(o) type_cook.



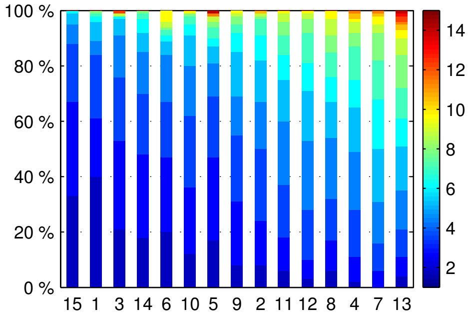
(p) type_heat.



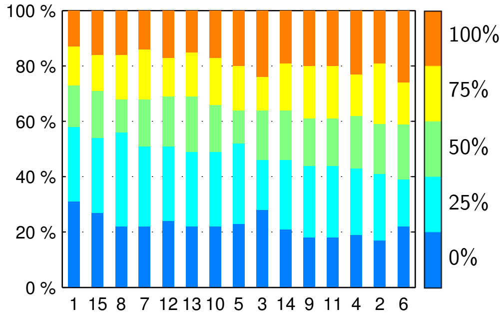
(q) type_water.



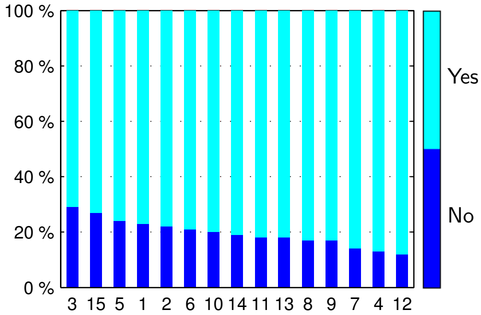
(r) #appliances.



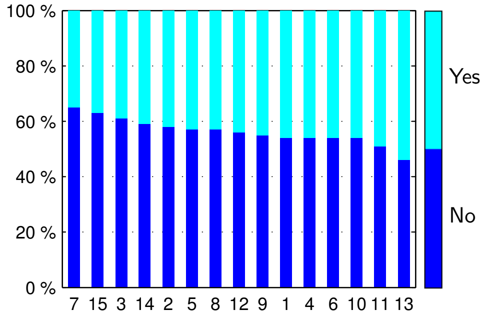
(s) #entertainment.



(t) lighting.



(u) timed_heat.



(v) timed_water.

Figure B.1: Distribution of household characteristics over the different clusters.

Appendix C

Detailed appliance information for household 5

Table C.1: Overview of the electrical appliances installed in household 5 of the ECO data set.

Appliance	Power consumption	Control	Comments
Fountain (inside): Pump and light	Connected to a plug	Pump: 7x24h Light: 7 p.m. – 10 p.m.	Plug connected to radio (since 2 October 2012)
TV (Entertainment)	Connected to a plug	TV over Ecoman	Satellite system always connected
Pond	Pump: 300 W Light: 70 W	8 a.m. – 8 p.m. Sunset – 8 p.m. & 6 a.m. – sunrise	Ripple control signal
Waterbed	2 x 400 W	Thermostat-controlled from 10 p.m. – 7 a.m.	
Bathroom (Heating, hairdryer, ventilation)	Total: 1000 W. Standby: 1 W or 9 W (with heating)	Thermostat-controlled when used	
Warm water heating	6000 W or solar collector heat pump: 95 W	Low tariff, thermostat-controlled	Was switched off during the study. Heat from solar collector
1st heating system: Sun	Solar collector pump: 95 W. Circulating pump: 5 W to 22 W	Thermostat	
2nd heating system: Wood	Control and ventilation: 120 W, Circulating pump: 80 W, Circulating pump: 5 W to 22 W, Boiler charging pump: 95 W	Thermostat	
3rd heating system: Electric	Water storage (2000 L): 18,000 W	Thermostat	Only used during absence of multiple days
Washing machine & dishwasher		Typically used during low tariff	Washing machine: Warm water connection during the summer
Telephone	3 devices	7 x 24h	

Gerät	Leistung (W)	Steuerung	Bemerkung
Zimmerbrunnen Pumpe Beleuchtung	Wird als Einzelverbraucher gemessen	7x24h 19:00 - 22:00	Seit 2.10.12 ist ein DAB-Radio gemessen.
Fernseher (Entertainment)	Wird als Einzelverbraucher gemessen	TV über Ecoman	SAT Anlage ist immer am Netz
Gartenteich, "Wasserfall"	Pumpe: 300W Beleuchtung: 70W	08:00 - 20:00 Dämmerung - 20:00 06:00 - Dämmerung	Signal Rundsteuerung Energie Thun
Wasserbett	2x 400W	Thermostatsteuerung gesperrt von 22:00 - 07:00	
Dusch-WC Heizung, Fön, Geruchsabsaugung	Total: 1000W Standby: 1/9 (mit Boilerheizung) W	Bei Benutzung Automat. Thermostatgesteuert	
Heizung /Warmwasser			
Warmwasser	6000 W oder: Pumpe Sonnenkollektor: 95W	In NT-Zeit Thermostat	War ausgeschaltet, Wärme von Sonnenkollektor, wenn Speicher voll wird rückgekühlt
1.Heizsystem (Sonne)	Pumpe Sonnenkollektor: 95W Umwälzpumpe: 5-22W	Thermostat Leistungsregelung	
2.Heizsystem (Holz)	Steuerung , Ventilator:120W Umwälzpumpe: 80W Umwälzpumpe: 5-22W Boilerladepumpe: 95W	Abgastemp. geregelt Thermostat Leistungsregelung Thermostat	
3.Heizsystem (elektrisch)	Wasserspeicher 2000l: 18`000W	Thermostat	Nur in Betrieb bei mehrtägiger Abwesenheit
Waschen /Abwaschen		In der Regel während NT	WM besitzt einen Warmwasseranschluss, welcher im Sommer in Betrieb ist
Fixtelefonie schnurlos	3 Geräte	7x24h	

Figure C.1: Overview of the electrical appliances installed in household 5 (original version, in German).

Curriculum Vitae: Christian Beckel

Personal Data

Date of Birth	25 April 1983
Place of Birth	Ludwigsburg, Germany
Citizenship	German

Education

2011 – 2015	PhD student, Dept. of Computer Science, ETH Zurich, Switzerland Supervision: Prof. Dr. Friedemann Mattern
2013	Visiting researcher, Stanford University, USA
2002 – 2008	Studies of Computer Science, Eberhard Karls University Tübingen, Germany. Diplom (M.Sc. equiv.) in CS (major) and BA (minor)
2005 – 2006	Graduate exchange student, University of Oregon, USA
2001	Abitur (high-school degree), Bietigheim-Bissingen, Germany

Employment

2011 – 2015	Research assistant, ETH Zurich, Switzerland
2011 – 2015	Senior researcher, Bits to Energy Lab (ETH Zurich & University of St. Gallen), Zurich, Switzerland
2008 – 2011	Development engineer, Robert Bosch Ltd. Corporate Research & Advance Engineering, Germany
2007	Software engineer (intern), IBM Germany Research & Development
2004 – 2005	Student research assistant, Eberhard Karls University Tübingen
2001 – 2002	Civilian service, kindergarten for mentally handicapped children

Awards & honors

2011 – 2014	PhD scholarship, sponsored by Bosch through the “Hans L. Merkle Foundation for Excellence in Science and Research”
2005 – 2006	Fulbright scholarship, University of Oregon, USA
2005 – 2006	Scholarship of the Landesstiftung Baden-Württemberg
1989	Skipped first grade in elementary school