On the Role of Smart Metering Data Analytics in the Energy Sector Digitization Process

Le Ray, Guillaume

Publication date:
2019

Document Version
Publisher's PDF, also known as Version of record

Technical University of Denmark

On the Role of Smart Metering Data Analytics in the Energy Sector Digitization Process

Guillaume Le Ray

Ph.D. Dissertation
Kgs. Lyngby, Denmark, April 2019

Center for Electric Power and Energy
Department of Electrical Engineering

There is nothing either good or bad, but
thinking makes it so.

— William Shakespeare, Hamlet

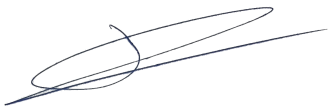
To Simone, Theodore
and Edith

Preface

This thesis is prepared at the Department of Electrical Engineering of the Technical University of Denmark in partial fulfillment of the requirements for acquiring the degree of Doctor of Philosophy in Engineering. The Ph.D. project was funded by the project EnergyLab Nordhavn (grant: EUDP 64015-0055).

This dissertation summarizes the work carried out by the author during his Ph.D. project. It started on 15 December 2015, and it was completed on 28 April 2019. During this period, he was hired by the Technical University of Denmark as a Ph.D. student at the Center for Electric Power and Energy (CEE).

The thesis consists of a summary report and six attached scientific papers, four have been peer-reviewed and published, whereas the two remaining ones are currently under review.



Guillaume Le Ray
28 April 2019

Acknowledgements

This thesis sums up the work of the last 3 years, 4 months and 13 days. What it does not tell is the human story behind it. Many persons contributed to this work one way or another. I cannot thank them all, and will therefore focus on the major contributors.

First of all, I would like to seize this opportunity to thank my Ph.D. supervisor Prof. Pierre Pinson. Pierre grew to be much more than a supervisor, to be a mentor and a role model. I grew up as better scientist but mostly a better man besides him thanks to the inspiration, motivation, support and kindness he provided me during all these years.

I would like to thank as well Prof. Ole Winther for his valuable advises that helped me to shape this work.

I would also like to thank Ph.D. Stephen Galsworthy and Quby's data science team for hosting me during my external stay and giving me the chance to see what is the future of metering data analytics.

I had the chance to have wonderful journey companions, current and former colleagues and friends at the ELMA group and at CEE who made the difficult times a bit less difficult. I especially thank Thibaut and Emerson for their friendship and the relaxing coffee and lunch breaks. In addition, I would like to thank Christos, Lejla and Lesia who started with me om this adventure. Thank you, Andrea, Andrea M., Andreas, Anna, Anubhav, Corey, Emil L., Emil B., Fabio, Jakob, Jalal, Lærke, Lucien, Morten, Pierre W., Spyro, Stefanos, Thanasis, Tiago Sousa, Tiago Soares, Trine, Tue, and Vladimir for being such great colleagues.

Last but not least, I would like to thank Simone for her dedication, support and unconditional love. I would like to thank Theodore and Edith for being such wonderful children. Finally I would like to thank my parents, Claudine and Michel for giving me the chance to be who I am today.

You all are part of this work!

Guillaume Le Ray

Kgs. Lyngby, Denmark, 2019

Table of Contents

List of Figures

List of Tables

Abstract

The electrical grid is facing profound structural changes due to the increased share of renewable energy generation dictated by the need to limit climate change impact and dependencies on imports. However renewable energy generation is variable, uncertain and not controllable, as it depends on the weather conditions and thus it is not dispatchable like conventional power generation. In the current grid operation framework, generation adjusts to cover the needs of demand and to keep the system balanced. Hence, a system integrating a large share of renewable energy sources requires other sources of flexibility among which demand flexibility that is controllable thanks to smart grid technologies. Smart grids and digitization of the energy sector through the deployments of smart meters open up to new business models for the utilities. Metering data analytics provides insights to the operators to make informed decisions in a close to real-time manner. Indeed, the analysis of streams of data as they are collected by smart meters can provide detailed information on the characteristics of the demand that can be used to improve the grid operation and planning.

This thesis addresses the crucial role of metering data analysis in the context of the digitization of the energy sector. The focus is on the implementation of methodologies for load analysis (load profiling) and load management (assets detection and management, Demand Response (DR) analysis) that can be implemented in the real-world and have a substantial positive impact on the grid operation and planning.

As a disruptive communication technology with large scale deployment, smart meters have raised privacy concerns. Beyond privacy problems, we first discuss the ethical implementation of smart grids and how the relationships between utilities and customers could secure investments in smart grid technologies. Indeed, the future of smart grids relies on the capacity to keep customers involved. Hence, beyond the technical aspects of smart grids, ethics should be the driver for decision making concerning interactions with customers.

Load profiling has been used to understand when and how electricity is consumed and to represent the loads with a limited number of typical load profiles summarizing the load behaviors. We propose in this work to take advantage of the stream nature of metering data to create a novel dynamic approach to load profiling using an online adaptive clustering algorithm. The online adaptive clustering algorithm was benchmarked against state-of-the-art clustering algorithm on a real-world dataset provided by Radius (Danish DSO) with more than 10 000 customers.

Among the new business models of utilities, assets detection and management for demand flexibility is probably the most important. Hence, the extraction of information about large appliances, that could be used to provide services to the grid is needed. We propose to use Non-Intrusive Load Monitoring (NILM) to decompose the aggregated consumption signal into individual appliances consumption signals. An unsupervised NILM algorithm is implemented and benchmarked against state-of-the-art NILM algorithm on the UK-Dale dataset. An application of the algorithm to the detection of Heat Pump (HP) is also presented.

The demand flexibility can be harnessed using DR through dynamic pricing. In a price-based DR framework there is significant variability of the individual response of participants. Hence, the analysis of the load to evaluate the responsiveness, quantify and characterize the response is fundamental to understand how the aggregated response is formed. The EcoGrid EU project in Denmark regroups 1900 participants into a real-world application of DR. The data collected during the EcoGrid EU project are used to test a responsiveness evaluation protocol as well as quantification and characterization of the response without the support of a baseline.

Resumé

El-nettet står overfor gennemgribende strukturelle ændringer grundet den øgede mængde af vedvarende energiproduktion, som er nødvendig for at begrænse klimaforandringerne samt reducere afhængigheden af import. Da produktion af vedvarende energi er variabel, usikker og svær at kontrollere pga. afhængighed af vejrforhold, kan den ikke styres påsamme måde som konventionel elproduktion. I den nuværende drift af el-nettet tilpasses produktionen til efterspørgslen for at holde systemet i balance. Et energisystem der skal kunne integrere store mængder af vedvarende energi kræver således andre kilder til fleksibilitet, bl.a. fleksibel efterspørgsel som nu kan styres takket være smart grid teknologier. Smart grids og digitalisering af energisektoren gennem brug af smarte målere åbner op for nye forretningsmodeller i forsyningssektoren. Metoder til analyse af målerdata giver operatørerne informationer der kan danne grundlag for at tage informerede beslutninger tæt pårealtid. Analyser af datastrømme, samtidig med at de bliver opsamlet i smarte målere, kan uden tvivl give detaljerede informationer om behovskarakteristika som kan bruges til at forbedre driften og planlægningen af el-nettet. Denne afhandling behandler den vigtige rolle analyse af målerdata spiller i forbindelse med digitalisering af energisektoren. Der fokuseres påimplementeringen af metoder til belastningsprofilering (load profiling) og styring af belastningen i el-nettet (assets detection and management, DR analyse) som kan anvendes i virkeligheden og have stor positiv betydning for driften og planlægningen af el-nettet. Som en disruptiv kommunikationsteknologi med storskala anvendelse har brugen af smarte målere givet anledning til bekymring omkring databeskyttelse og privatliv. Udover disse udfordringer, diskuterer vi først den etiske implementering af smart grids og dernæst hvordan forholdet mellem forsyningssektoren og kunderne vil kunne sikre investeringer i smart grid teknologier. Fremtidsudsigterne for brugen af smart grids afhænger således af evnen til at involvere kunderne. Udover de tekniske aspekter omkring brugen af smart grids bør etik således være drivkraften for beslutningstagning angående interaktioner med kunderne. Belastningsprofiler er blevet brugt til at forståhvornår og hvordan elektricitet forbruges samt til at repræsentere belastninger med et begrænset antal typiske belastningsprofiler som opsummerer belastningernes adfærd. Vi foreslår i denne afhandling at udnytte den kontinuerte strøm af målerdata til at anvende en ny dynamisk tilgang til at lave belastningsprofiler ved at benytte en online adaptiv clusteringalgoritme. Algoritmen er blevet bench-market op imod en state-of-the-art clusteringalgoritme pået virkeligt datasæt med mere end 10.000 kunder som er stillet til rådighed af Radius. Detektion og styring af fleksible aktivers efterspørgsel er sandsynligvis den vigtigste af de nye forretningsmodeller inden for forsyning. Derfor er der behov for at kunne udtrække information om store apparater der kan bruges til at yde service til el-nettet. Vi foreslår at benytte NILM til at dekomponere signalet for det aggregerede forbrug til signaler for de enkelte apparater. En ikke superviseret algoritme NILM implementeres og benchmarkes op imod en state-of-the-art NILM algoritme påUK-Dale datasættet. En anvendelse af algoritmen til detektion af varmepumper bliver ogsåpræsenteret. Fleksibiliteten i efterspørgslen kan udnyttes ved at bruge DR gennem dynamiske priser. Ved prisbaseret DR er der væsentlig forskellighed i det individuelle respons fra deltagerne. Derfor er analysen af belastningen, ved

først at evaluere følsomheden overfor prissignalet, dernæst kvantificere og karakterisere responset, afgørende for at forståhvordan det aggregerede respons er sammensat. EcoGrid EU projektet i Danmark regrupperer 1900 deltagere i projektet til en anvendelse af DR i virkelighedens verden. De data der blev opsamlet i EcoGrid EU projektet bliver brugt til at teste en protokol til evaluering af følsomhed overfor prissignalet samt til at kvantificere og karakterisere responset uden brug af en baseline.

Acronyms

AC Air Conditioner

ACC Accuracy

AMI Advance Metering Infrastructure

ANN Artificial Neural Network

CBA Cost Benefit Analysis

CBL Customer Base Line

CCA Curvilinear Component Analysis

CVA Canonical Variates Analysis

DFT Discrete Fourier Transform

DR Demand Response

DSO Distribution System Operator

DSM Demand Side Management

DTW Dynamic Time Warping

DWT Discrete Wavelet Transform

EC Estimated Accuracy

EH Electric Heating

EM Expectation Maximization

EV Electric Vehicle

FHMM Factorial Hidden Markov Model

FIR Finite Impulse Response

GDPR General Data Protection Regulation

GMM Gaussian Mixture Model

HAC Hierarchical Ascendant Clustering

HMM Hidden Markov Model

HP Heat Pump

ICA Independent Component Analysis

IC Independent Component

IoT Internet of Things

LSTM Long Short Term Memory

MP Matching Pursuit

NILM Non-Intrusive Load Monitoring

NMI Normalized Mutual Information

OMP Orthogonal Matching Pursuit

PCA Principal Component Analysis

PDF Probability Density Function

PPV Positive Predictive Value

PV Photovoltaics

RES Renewable Energy Sources

RMSE Root Mean Square Error

RNN Recurrent Neural Network

RTP Real Time Pricing

SAX Symbolic Aggregate Approximation

SOM Self-Organizing Map

SVM Support Vector Machine

TPR True Positive Rate

TSO Transmission System Operator

List of Publications

[Paper A] Le Ray, G., Pinson, P. (2019) The ethical smart grid: Enabling a fruitful and long-lasting relationship between utilities and customers. *Energy Policy*, Submitted.

[Paper B] Le Ray, G., Pinson, P. (2018). Online adaptive clustering algorithm for load profiling. *Sustainable Energy, Grids and Networks*, 17, 100181.

[Paper C] Winker, P., Le Ray, G., Pinson, P. (2019). Unsupervised energy disaggregation: from sparse signal approximation to community detection. *IEEE Transactions on Smart Grid*, Submitted.

[Paper D] Le Ray, G., Christensen, M. H., Pinson, P. (2019). Detection and characterization of domestic heat pumps. In *Proceedings of PowerTech Conference, 2019 IEEE Milano* (pp. 1-6). IEEE.

[Paper E] Le Ray, G., Larsen, E. M., Pinson, P. (2018). Evaluating price-based demand response in practice – with application to the EcoGrid EU experiment. *IEEE Transactions on Smart Grid*, 9(3), 2304-2313.

[Paper F] Le Ray, G., Pinson, P., Larsen, E. M. (2017). Data-driven demand response characterization and quantification. In *Proceedings of PowerTech Conference, 2017 IEEE Manchester* (pp. 1-6). IEEE.

CHAPTER 1

Introduction

1.1 The context of digitization in the energy sector

The emergence of the Internet, which was a major technological breakthrough, had an impact on almost every engineering field. Originally, digital data acquisition and transfer were limited to some activities like public administration, banking, and retail. Following the development of the Internet, the exchange of data increased at an exponential rate as households became connected to the network. Since the beginning of the 21st century, data collection has become part of the physical world with the democratization of GPS, smart cards for public transportation, smartphones and more generally Internet of Things (IoT) devices [1]. The development of connected devices and the physical infrastructures came alongside with the development of computational power allowing the analysis of large amounts of data in close to real-time. Utility services have also followed the digitization trend and deployed smart meters reporting consumption. The first utility company to do so was Enel (gas and electricity utility in Italy) with 30 million smart meters deployed between 2000 and 2005 to automate invoicing and detect fraud [2]. The rest of the Western world followed quickly and now most countries are in the process of deploying or are planning to deploy smart meters. Smart meters together with communication networks and data management systems form the Advance Metering Infrastructure (AMI), which collects the metering data and allows bi-directional information flows between operators and customers [3].

The modernization of electrical grids into smart grids aims at limiting dependencies on imports and reducing climate change impact by integrating an increased share of Renewable Energy Sources (RES) [4]. The deployment of smart grid technologies induce important structural changes. Electrical grids were originally designed to transport and distribute electricity from controllable (e.g., nuclear, hydro, gas or coal fired) power plants to customers so that the generation would meet the demand at any time to keep the grid balanced [5]. The integration of weather-dependent RES in the electrical grid increased uncertainty and variability on the generation side. A higher resolution of the metering data provides close to real-time data that can be used to adjust the generation to the demand in a more cost-efficient manner exploiting the flexibility of the grid. Utilities have many flexibility options to balance a system with large share of RES [6], (i) fast ramping generators, (ii) curtailment, (iii) storage, and (iv) demand flexibility. The solution lies in a combination of all these flexibility sources, among which demand flexibility is the most cost-efficient. Most of the smart meters have a two-way communication protocol, which allows for both sending metering data and receiving incentive signals. Hence, Demand Side Management (DSM), the action of modifying the aggregated consumption of customers, could be implemented on a large scale to exploit the potential flexibility of the demand and compensate RES inflexibility [7].

Smart meters, despite the name, are nothing more than digitally-connected meters. The grid becomes smart because of the analysis and exploitation of the information contained in metering data. Indeed, for utilities, the paradigm has drastically changed from a lack of data, i.e., one

data point per year per customer, to streams of data, i.e., one data point per hour (down to 15 minutes) per customer. The methods used to estimate the patterns of consumption based on the yearly consumption should now summarize them into a limited number of informative patterns. Moreover, additional information on the large and potentially flexible appliances can now be unveiled providing more accurate and detailed insights on load behaviors. The analysis of metering data can be divided into three branches [8], (i) load forecasting (e.g., individual and/or aggregated loads), (ii) load analysis (e.g., bad data detection, thief detection, and load profiling), and (iii) load management, (e.g., customer characterization, Demand Response (DR) program marketing, DR implementation). As a disruptive communication technology, smart meters raise some concerns. The implementation of smart grid technologies has shown that customers bear the risk as they are exposed to privacy breaching, while the benefits of smart meters are mostly on the utility side. The role of customers is changing, but the position of the customer in a smart grid organization is not. Smart grids should place customers at the center of the electrical grid. The need for metering data and analysis is not questioned, but the way they are used as well as the contributors are.

1.2 What can metering data be used for?

The objective of this thesis is to propose data analytics that take advantage of the quantity and the nature of metering data to provide insightful information to utilities to improve efficiency and sustainability of smart grids [9]. More precisely, it will focus on load analysis and load management analytics.

The growing amount of metering data has opened doors to analytics in the energy field. Moreover, the deregulation of the power industry has increased competition between energy providers. Innovative solutions in terms of tariffs or consumption insights are the main leverage to increase market share. Hence, analytics, which could be defined as the process of transforming data into insights, has a key role in this process [10]. A component of metering data measures human activities in relation to electricity. Hence, load behavior inherits partly the characteristics of human behavior consisting of habits (predictable) with well-scheduled activities, and changes (stochastic), which are difficult to predict. It becomes particularly true when considering important changes in the loads due to the acquisition of Electric Vehicle (EV), Photovoltaics (PV), Electric Heating (EH) and Air Conditioner (AC). Consequently, energy data analytics should require little (semi-supervised) or no (unsupervised) training to handle changing behaviors. Because of the ethical implications of data analytics, the role of the utilities in building a trust relationship with customers, which is an important part of smart grids success is also discussed.

Why should ethical smart grids be the end goal?

Authors in [11, 12] have underlined the need for more ethics and respect for privacy in the implementation of smart grids. Indeed, the long term development of smart grids relies on the active involvement of customers. Hence, customers are becoming key actors of smart grids. However, the deployment of smart meters and the metering data collection raises some questions on how customers are considered in the actual framework. The structural changes of the grid have modified the responsibilities of the actors, but the relationships between them have not evolved yet. The main issue raised by associations of customers is privacy breaching [13], but profound changes have to occur in the relationship between utilities and customers who are not merely consumers anymore. Before presenting the data analytics developed in this thesis, it is essential to

contextualize the use of metering data and understand under, which conditions it could have a positive impact on the smart grid and society.

How can load profiling be rethought by profiting of the stream nature of metering data?

In view of the growing metering data availability, more data-driven approaches are implemented. Load profiling using clustering has become state-of-the-art [14]. It is an improvement from simply categorizing customers by activities. However standard clustering approaches generate static typical load profiles, that do not take into account possible non-stationarity of the loads [15]. The algorithm has to be rerun to update typical load profiles in order to handle possible new trends in the load pattern. Considering the extensive size of the pool of customers to cluster, such an operation is time-consuming. Hence, looking closely at load patterns, three characteristics of their dynamic can be observed (i) stationarity, a load replicates over time; (ii) non-stationarity, a load changes pattern over time; and (iii) disruptive, a load follows a pattern never observed before. Our solution to answer this question and handle these characteristics is to implement an online adaptive clustering algorithm. The online part of the algorithm handles both stationary and non-stationary loads while the adaptive part allows the clustering to adjust to disruptive loads.

How much can we learn about appliances from metering data?

Utilities have approximated information, given by customers, about the appliances installed in each household. Beyond the standard domestic appliances, large power consumption and potentially flexible appliances (e.g., Heat Pumps (HPs), EH, ACs, EVs) are of interest to utilities [7]. Indeed, in the context of DSM these appliances are assets whose flexibility can be harnessed to provide services to the grid. Knowing precisely where assets are located on the grid topography allows to implement targeted recruitment of these customers into efficiency programs and manage assets optimally [16]. For example, the growing share of EVs could create important imbalances that could be prevented through the use of smart charging facilities [17] although the EVs have to be located first. We suggest developing a multi-purpose methodology based on Non-Intrusive Load Monitoring (NILM) approach to learn more about the load behaviors of assets. In contrast to state-of-the-art NILM methodologies, our approach is unsupervised, thus requires no training dataset with individual appliance consumption, which means that it can be applied to real-world dataset.

How can demand response be quantified and characterized in a data-driven manner?

Demand response is a complex framework that uses many control elements: smart home controllers, two-way communication smart meters, encryptions [18]. In theory, they all work perfectly together. In practice, troubleshooting has to be done and can result in a chaotic situation and be extremely time-consuming if the malfunctioning pieces of equipment are not identified. A test protocol is necessary to evaluate the responsiveness of equipment in each household, to detect potential malfunctions, and prioritize maintenance tasks. Furthermore, the amplitude of the response has to be evaluated to reward participants according to their contribution. Both quantification and characterization of the response have to be considered [19]. Indeed, quantification does not say anything about the dynamics of the response. The state-of-the-art evaluates the response using a baseline describing the load 'if' no incentives were received. Baselines are conceptually problematic as they rely on forecasting models that are trained using data with no DR, which are usually collected before the DR program starts. As the error of a model increases with the forecasting horizon, baselines present larger errors the longer the DR program lasts.

1.3 Scientific contributions

The energy sector digitization is part of the grid modernization process that was imposed by the increased share of RES [4]. The modernization of the grid not only modifies the grid structure but also the relationships between the actors. From a customer point of view, the goal of modernizing grids is not well understood and customers have been expressing distrust in the utilities in France, the Netherlands, the United Kingdom as the deployment of smart meters is occurring [20]. [Paper A] gives an overview of the context of smart meter deployments and metering data use to explain how legitimate concerns about the installation of smart meters and the uses of high-frequency metering data can engender distrust in the utilities. This work aims at presenting alternatives towards an ethical smart grid with more rewards for risk takers and better distribution of the benefits. Smart grids and deregulation of the power industry redistribute responsibilities between actors. Therefore, their roles should be modified; (i) customers are providing more information to the utilities through metering data and utilities are losing some control to aggregators, and (ii) customers are expected to provide services to the grid by changing their electricity consumption. Smart grids require active customers. It is not possible without a minimum of fairness, ethics, and respect for the role of each stakeholder. Smart meters are so far mostly used for billing purposes, grid maintenance and operation.

Load profiling, displaying the patterns of electricity consumption, provides insights for grid maintenance and operation. In the actual grid framework, the utilities had an approximate idea of the aggregated load profiles at different grid levels and how much electricity should be generated in each hour through the use of typical load profiles. Typical load profiles were implemented as the average daily pattern of each category of customers (e.g., grocery store, offices, demographics for domestics) over periods of time [14]. The typical load profiles are representative of categories only under the assumption of relative stationarity over the period and homogeneity in each category. Load profiling was constrained by the lack of data to obtain more accurate typical load profiles [21]. A large amount of metering data has changed the way typical load profiles are generated and today's state-of-the-art is data-driven through clustering approaches [22]. Indeed, clustering, summarizing a large set of loads through a limited number of cluster centers, is intuitively a suitable approach to generate typical load profiles from all metering data. Clustering tackles down the assumption of homogeneity of the categories as a category can be associated with many clusters representing subpopulations. However, clustering still relies on the assumption of stationarity, which is legitimately arguable as loads reflect human activities. We suggest to take advantage of the stream nature of metering data to cluster loads handle stationary, non-stationary, and disruptive load patterns. In [Paper B] an online and adaptive clustering algorithm is presented. It updates typical load profiles and adjusts the clustering to the data structure.

Clustering is the first step towards more advanced analytics on the demand side. Considering the amount of metering data collected and the challenges posed by smart grid operation and maintenance, much more can be done with the available data [23]. Metering data, and more specifically the information extracted from the data, has become the center of load management. The gap between metering data and machine learning was filled in the 1990s with the conceptualization of NILM, but it only became a new field of machine learning in the 2010s with the release of high-frequency consumption dataset and the increased computing power capacities [24, 25]. NILM consists in individualizing electricity consumption signal of appliances from the overall electricity consumption signal. Hence, information on the power consumption of individual appliances can

be obtained. The state-of-the-art approaches are variations of Hidden Markov Models (HMMs) and Artificial Neural Networks (ANNs),which both are supervised learning approaches requiring extensive training data and computational power [26]. [Paper C] presents an unsupervised NILM methodology, which is performing as well as state-of-the-art methodologies and shows stable performances with the degradation of the resolution. NILM applications have been limited to give customers feedback on the consumption of each appliance [27]. The unsupervised NILM methodology has been adapted to detect the presence and characterize the load behavior of targeted appliances (e.g., EVs, HPs, EH) to Distribution System Operators (DSOs). Such appliances can thereafter participate in a DR program to provide flexibility to the grid. Hence, they are strategic assets as they can either contribute to the unbalance or provide services to smart grids [16]. The approach is using a Bayesian framework to create a definition of the load behavior of a targeted appliance and detect its presence in households. An example using the EcoGrid EU dataset to the detection of domestic HPs is presented in [Paper D].

When customers owning assets like HPs are localized, they can be contacted to join a program like DR. The implementation of a DR framework is complex and relies on the communication between many devices (controllers, smart meters, central market) [28]. However, no work has been done on designing a protocol to identify households with faulty non-responsive equipment in order to do troubleshooting and maintain the DR capacity of the pool of participants. [Paper E] presents a protocol using a clinical trial approach to identify non-responsive households. Resources can then be allocated to solve non-responsiveness issues. In operation, the potential response is a portion of the flexible load capacity that depends on the initial conditions (i.e., indoor temperature, comfort settings) at the time of receiving a DR incentive. The usual way to quantify responses is with the support of a baseline, which is either a forecast of the participant consumption using an auto-regressive model trained on data collected before the DR framework starts or a representative set of participants, which are not receiving incentives [29, 30]. Nevertheless, baselines are biased either by the choice of model or the selection of participants to form the baseline. Hence, evaluating the response of participants without a baseline would reduce the bias. [Paper F] proposes a methodology to characterize and quantify the response of individual households in a real-time pricing DR framework without using a baseline. Characterizing and quantifying responses provide feedback to operators thus learning more about the behavior of the loads during DR event and can be used to fine-tune the incentives.

1.4 Thesis outline

This thesis is structured as follows. Part I is a report presenting the fundamental concepts that form the core of this work and summarizes the scientific contributions disseminated in the publications written during this Ph.D. project. Within this part, Chapter 2 corresponds to [Paper A] and gives perspective on the use of metering data and emphasizes the need for more ethics in the smart grid implementation. The methodologies that have been developed and their main results are reported in Chapter 3–5. In more detail, Chapter 3 focuses on load profiling and presents the implementation of an online adaptive clustering algorithm that accounts for highly variable loads. In Chapter 4 NILM is presented and an unsupervised NILM methodology using sparse signal approximation, clustering, and community detection is reported. Chapter 5 outlines the need for analysis of response in a DR framework and presents methodologies to evaluate the responsiveness of participants to characterize and quantify responses. The reporting closes with conclusions and

suggestions for future work in Chapter 6.

Part II contains all the publications that contribute to this thesis:

- **Paper A** is a manuscript submitted to *Energy Policy*. This paper provides a discussion on the need to build ethical smart grids, which would be the basis of a long-lasting trust relationship between utilities and customers. The context of today's information technology, privacy, and data protection is developed as it influences the perception of smart grid technologies. The focus is then on how bad practices regarding deployments of smart meters and metering data uses can jeopardize smart grid implementations and how ethical behavior can improve it.
- **Paper B** is a journal article published in *Sustainable Energy, Grids and Networks*. This paper presents an online and adaptive clustering algorithm to generate dynamic typical load profiles. This methodology is benchmarked against state-of-the-art clustering algorithms used in the load profiling literature. It is also implemented with district heating consumption data to illustrate its potential application beyond the electrical engineering field.
- **Paper C** is a manuscript submitted to *IEEE Transactions on Smart Grid*. This paper introduces an unsupervised NILM (i.e., energy disaggregation) algorithm using sequentially sparse signal approximation, clustering, and community detection to individualize appliances consumption. It compares its performance to the state-of-the-art NILM algorithm as well as the evolution of the performance with the degradation of the data resolution.
- **Paper D** is a peer-reviewed conference article in the *Proceeding of IEEE PowerTech Conference (2019)*. The paper is a use case application of the methodology described in [Paper C] to the detection of domestic HPs. It is a semi-supervised approach where first a prior definition of the HPs behavior is learned using a probabilistic Bayesian framework then the definition is used to detect HPs.
- **Paper E** is a journal article published in *IEEE Transactions on Smart Grid*. This paper proposes a methodology to evaluate the responsiveness of controllers installed in households in a DR framework. A test protocol has been developed and standard clinical trial statistics have been used to evaluate if the responses in the test group are significantly different from the one in the control group.
- **Paper F** is a peer-reviewed conference article in the *Proceeding of IEEE PowerTech Conference (2017)*. The paper introduces a methodology to evaluate quantitatively and qualitatively response of households in a DR framework without defining a baseline. A clustering based on the distance to the price signal is used to group participants to perform a standard finite impulse response and methodology using independent component analysis is implemented to characterize the responses.

CHAPTER 2

The Ethical Smart Grid: Enabling a Fruitful and Long-Lasting Relationship Between Utilities and Customers

This chapter proposes to discuss why the relationship between the utilities and the customers is fundamental in the future of smart grids. Beyond the technical aspects that ultimately shape the possible implementation of the smart grid, we want to emphasize its human aspect. The general context of data collection and privacy is indissociable from the customers' perception of smart meters, which is their entry point into the smart grid. The focus is then placed on the expected role and benefits of the different actors to highlight misalignments and bad-practices before proposing solutions to the implementation of more ethical smart grids. This chapter is [Paper A] as is.

2.1 Introduction

The European Union's (EU) energy policy is facing unprecedented challenges due to the increased dependencies on imports, scarce resources and the need to limit climate change [4]. Ambitious energy efficiency programs have been developed to tackle these challenges. Indeed, since 2009 and the 2020 Climate & Energy Package's road map to the 20/20/20 targets [31], EU has driven towards a *greener* energy sector to achieve energy efficiency, energy independence, and last but not least to reduce greenhouse gas emissions.

As most of the issues on power systems are observed at the distribution level, the program requires a modernization of the grid to foresee potential issues and to have a pervasive control to prevent them [5]. Information and communication technologies form the foundations of the smart grid pyramid (Figure 2.1), which support more advanced infrastructures. The Third Energy Package, adopted in 2009, strengthens the internal European market for gas and electricity by securing a competitive and sustainable supply of energy to the economy and the society [32]. To reach this goal, the EU has set the target of 80% of households equipped with smart meters by 2020 [33]. Smart meters are deployed to provide more transparency to consumers (billing, price, consumption), to improve awareness on energy consumption and empower the consumers to modify their energy behavior using metering data [32]. On the utility side, smart meters' data will help them to increase the efficiency and the reliability of the grid. In this paper, we define a utility as an entity that is given responsibility for the maintenance and operation of some infrastructure of public value and used for public service. As displayed in Figure 2.1, smart meters constitute the first fundamental application that involves customers. The perception of smart meters by customers will condition

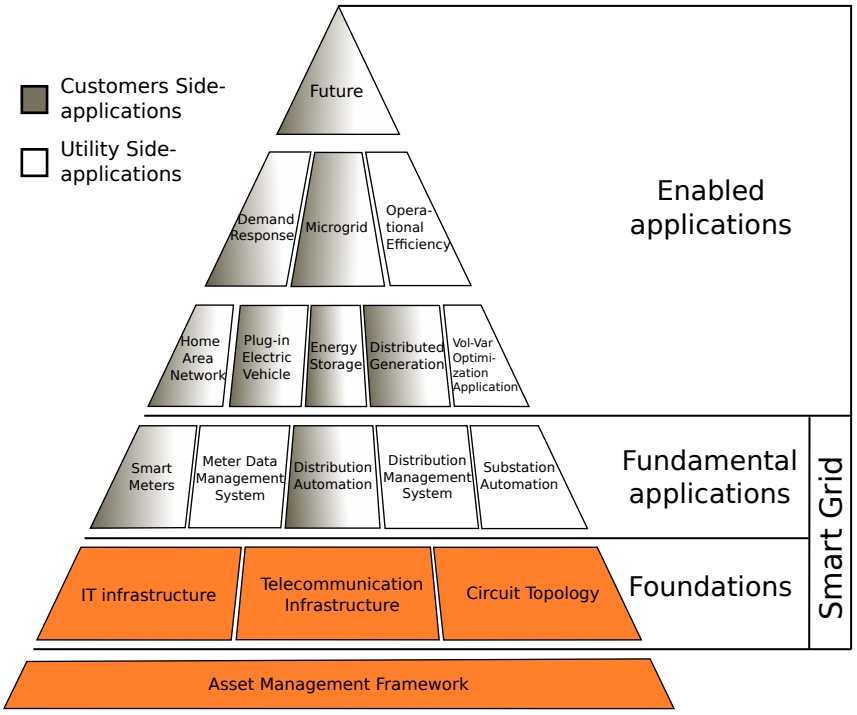


Figure 2.1: The Smart Grid pyramid (Source: [5]).

the future of smart grids in their capacity to transform customers into actors of the grid through the use of Demand Side Management (DSM) [20, 34, 35].

At the same time, concerns are raised about a possible backlash of domestic customers [36]. Indeed, the perspective of having smart meters reporting electricity consumption at high resolution in every home has engendered irrational fear (e.g., health issues and domestic accidents) and legitimate questions about the need of smart meters and their impact on privacy [12]. The concept of privacy has drastically evolved, from Aristotle making the distinction between the public sphere and the private sphere, to the creation of the right to privacy [37]. With the emergence of information technology, the legal framework has evolved to protect data and by extension data subjects. The legislation on data protection and privacy limits the legal use of data, but it also defines the limits of a gray area on how to handle customers and their data [38]. Even if the decisions made are in line with the existing legal framework, they can have a substantial (and potentially negative) impact on the future of smart metering and smart grid deployment plans by extension [39]. It is where ethics is fundamental. It is defined as *"A system of moral principles, which deals with what is good or bad for individuals and society. It is a collection of fundamental concepts and principles on an ideal human character that enables people to make decisions regarding what is right or wrong. Ethics is a code of conduct agreed and adopted by people in a society, which sets the norms of how a person should live and interact with other people."* [38].

Following this definition, an ethical and human perspective has to be placed on the smart grid technology, especially its extensions smart meters and the data they generate. The foundations of the smart grid are put in place today. Hence, it is essential to understand the perspective of the customers to build ethical smart grids to build a fruitful and long-lasting relationship between utilities and customers [39].

In this paper, we argue that beyond the technical and legal issues, ethics should be the driver for decision-making. The EU regulation defines a legal framework in which the utilities can implement the deployment and exploitation of smart meters with more or less focus on customers. However, the long term development of smart grid technology will change the status of consumers to prosumers and they should be treated as such: collaborators providing services to the grid. In an era where leakage of data is making newspapers' headlines regularly, this is the only reasonable way to build a sustainable relationship between domestic customers and utilities. Hence, a balance between the necessity of data to build the smart grid and the respect of the customers as partners has to be determined.

The following discussion is delimited by the legal and technical background of the roll-out and smart meters data utilization concerning privacy and data accessibility (Section 2.2). However, privacy is not only dependent on the legal and technical aspects, but also depends on what the citizens are willing to give. As privacy is the most critical issue raised with smart meters, it is crucial to understand what we mean today by privacy and what is really at stakes in this context when it is jeopardized (Section 2.3). In terms of ethics, the transition to smart grid redistributes the cards between utilities and consumers (prosumers) as the latter has a more prominent role now than before (Section 2.4). Customers are also exposed to giving away sensitive data to utilities. It has to be acknowledged and rewarded in an ethical manner (Section 2.5). The conclusions are gathered in Section 2.6 while opening up to broader perspectives.

2.2 Legal and technical background in relation to smart meters' data privacy

The legal framework of data protection and privacy has evolved, mainly due to the emergence of new technologies and new threats to privacy they create [20]. Here we aim at giving the background to both legal and technical aspects that are shaping data collection and use of data generated by smart meters. It scopes what is legally possible in Europe and how the technical setup decided by each Member State shapes the relationship between customers and utilities during the roll-out and after.

2.2.1 A compact historical review of the right to privacy in EU legislation

The origin of the *right to privacy* can be traced back to the Universal Declaration of the Human Rights (article 12) in 1948 [40]. It states that ' *No one shall be subjected to arbitrary interference with his privacy, family, home or correspondence, nor to attacks upon his honor and reputation. Everyone has the right to the protection of the law against such interference or attacks*'. It aims at protecting the right and interest of individuals rather than the data itself as data collection appeared to generate an unexpected impact on an individuals life. Soon after, the Council of Europe strengthened it in the European Convention on Human Rights [41].

The growth of information technology in the 1970s, especially in the public sector and in the banking industry, pushed the Committee of Ministers to the Member States to write 2 recommendations (Resolution 23 and 24) stating that every individual whatever his nationality or residence should have respect for his right to privacy with regards to automatic processing of personal data. These resolutions were received positively and the Council of Europe, which had an impact beyond Europe, as 46 countries ratified it [42]. It defines the concept of *personal data* as '*any information relating to an identified or identifiable individual ('data subject')*' and sets the foundation of data protection at an international level. The Convention aims to protect individuals against unjustified

collection, use, and dissemination of personal data. It then implicitly defines what will later be called *legitimate purpose*.

After years of negotiation between the Member States, the Data Protection Directive (Directive 95/46/EC) was adopted in order to harmonize the legal framework [43]. Some precisions were added to the definition of *personal data* about what *identifiable* meant; ' *an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity*'. It remains broad on purpose to extend its application to future information technologies. Despite being implemented on the same basic principles, it has generated different applications[1]. The Data Protection Directive has articulated around three points; (i) transparency: information on personal data being processed; (ii) legitimate purpose: specification, explicit and legitimate of the purposes of the data collection; and (iii) parsimony: adequacy to the purpose of the personal data collected.

Article 7 stipulates the lawful basis to process personal data:

(a) unambiguously consent; or

(b) processing is necessary for the performance of a contract; or

(c) processing is necessary for compliance with a legal obligation; or

(d) processing is necessary in order to protect vital interests; or

(e) processing is necessary for the performance of a task carried out in the public interest; or

(f) processing is necessary for the purposes of the legitimate interests pursued by the controller or by the third party.

In order to harmonize the Data Protection Directive among the EU Member States, the European Commission proposed the General Data Protection Regulation (GDPR) in 2012 [44]. It generalizes the basic principle of the Data Protection Directive and develops some further rules that apply to all data collected inside the EU by European or non-European organizations.

The main changes on the rights of the data subjects and responsibilities of controllers and processors concerning data protection and privacy of the data subject are:

- Explicit and provable consent (instead of unambiguous consent)(Article 7).
- transparency and modalities: The data controller should inform and communicate with the data subject in a *'concise, transparent, intelligible and easily accessible form, using clear and plain language'* (Article 12(1)). It should also facilitate the exercise of the data subject rights (Article 12(2)).
- Rectification and erasure: A person has the right to ask for his data to be erased (Article 17); to restrict the processing under certain condition (Article 18); to transfer personal data from one service to another (Data portability Article 20).
- Right to object to automated individual decision-making (Articles 21 and 22).

[1] As an EU Directive, it applies to all Member States, but each Member States transposes it in its national law

- Data protection by design and by default: The data protection and privacy should be included in the development of the service and the privacy settings should be set to a high level by default (Article 25).
- Communication of a personal data breach to the data subject (Article 34).

From the foundation of the right to privacy to the GDPR, the definition of privacy and data protection had to be updated according to the development of information technologies, which is going at a hectic pace. Nevertheless, the following discussion on smart meter data and their ethical use is bounded within the EU by this legal framework.

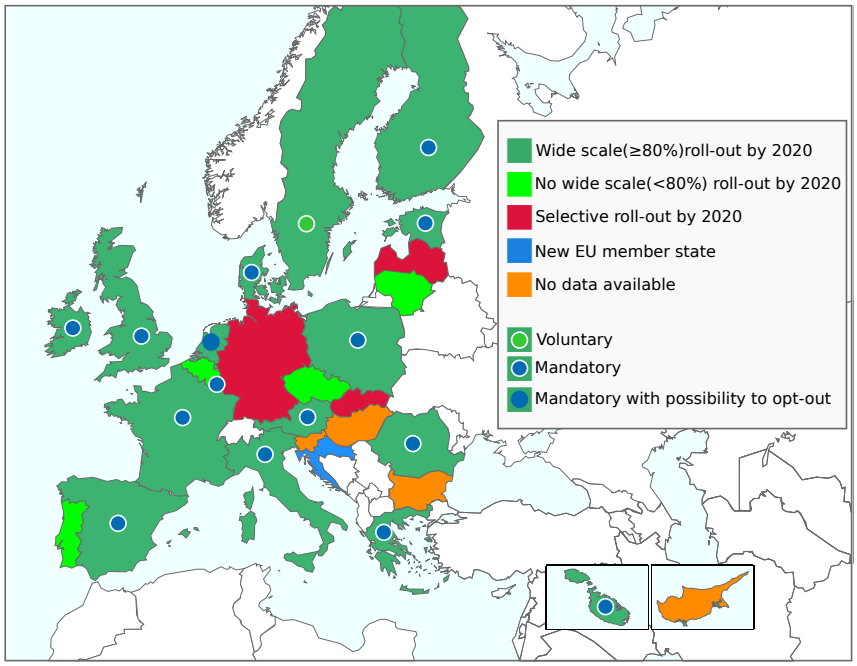


Figure 2.2: Map of the roll-out of smart meters in Europe [45].

2.2.2 A review of roll-outs and setups of smart metering in the EU

In the case of smart meters, the technological possibilities, as well as deployment strategies, are directly related to the problem of privacy and ethics. The scale of roll-out is decided based on a Cost Benefit Analysis (CBA), described in [32], which concludes whether the roll-out should be at least 80%, less or just selective. However, the roll-out strategy is entirely left to each EU Member State, which gives a broad diversity of setups and subsequently different data flows. Table 2.1 gives an overview of the roll-out status of the different Member States in 2014. The map in Figure 2.2 presents the roll-out scale as well as the recruitment strategy. Table 2.1 and Figure 2.2 give an overview of the diversity and the number of parameters to take into account in the roll-out of smart meters in EU. Temporal disparities are also observed; Italy and Sweden had already completed the deployment of smart meters before the adoption of the Directive 2012/27/EU. The Netherlands had planned an early deployment, but the roll-out, which was initially mandatory has been challenged by consumer protection organizations that sued the State to obtain the possibility to opt-out [13].

Smart metering has also changed the responsibilities of the Distribution System Operator (DSO) and Transmission System Operator (TSO) as they have to handle a large amount of data. Figure 2.3

Table 2.1: Overview of the roll-out (in 2014) in the EU. Source: [45]

Member State	roll-out scale	CBA[a]outcome	resolution	implementation/ownership	storage	Financing of the roll-out
Austria	95%	+	15 min	DSO	DSO	Metering & network tariffs
Belgium	<80%	-	NS[b]	DSO	DSO	Network tariffs
Bulgaria	TBA[c]	NA	NA	NA	NA	NA
Croatia[d]	NA[e]	NA	NA	NA	NA	NA
Cyprus	TBA	NA	NS	DSO	DSO	NA
Czech Republic	1%	-	NS	DSO	Central Hub	NA
Denmark	100%	+	15 min (hourly before 2011)	DSO	Central Hub	Network tariffs
Estonia	100%	+	hourly	DSO	Central Hub	Network tariffs
Finland	100%	+	1 hour (RT optional)	DSO	DSO	Network tariffs
France	95%	+	10-30 min	DSO/municipalities	DSO	NA
Germany	23%	-	15 min	DSO or meter operator	DSO or meter operator	NA
Greece	80%	+	NS	DSO	DSO	NA
Great Britain	99.5%	+	30 min (10s to customer)	Supplier	Central Hub	Funded by suppliers
Hungary	TBA	+	NS	NA	NA	NA
Ireland	100%	+	30 min (10s to customer)	DSO	DSO	Network tariffs
Italy	99%	NA	10 min	DSO	DSO	DSO & Network tariffs
Latvia	23%	-	NS	DSO	DSO	Network tariffs
Lithuania	<80%	-	NS	DSO	DSO	Network tariffs
Luxembourg	95%	+	NS	DSO	DSO	Network tariffs
Malta	100%	NA	NS	DSO	DSO	Network tariffs
Netherlands	100%	+	NS	DSO	DSO	Network tariffs
Poland	80%	+	NS	DSO	Central Hub	Network tariffs
Portugal	<80%	Inconclusive	15 min	DSO	DSO	DSO & Network tariffs
Romania	80%	+	NS	DSO	DSO	Network tariffs
Slovakia	23%	-	15 min	DSO	DSO/Central Hub	DSO & Network tariffs
Slovenia	TBA	NA	NS	NA	DSO	NA
Spain	100%	NA	NA	DSO	DSO	Network tariffs & SM rental
Sweden	100%	+	hourly	DSO	DSO	DSO &Network tariffs

[a] Cost-benefit analysis [b] Not Specified [c] To be announced [d] New Member State [e] missing information

is a schematic representation of the flow of data and actions between the different actors. The roles of the data controllers, data protection officer and supervisory authority are defined in the GDPR [44] and are taken in most cases by the DSO, TSO, or independent organism [46]. It could be considered in the context of smart metering as the perfect flow of data according to [47].

Some parameters, like the resolution of the data, the access to metering data and the implementation/ownership, have a direct impact on the setup and thus the data flow as shown in Figure 2.3 as well as the capacity of the customer to modify its consumption. The range of possibilities makes it difficult to standardize. However, most of the DSOs, as responsible authorities of the roll-out, (will) face the same ethical problems with their customers.

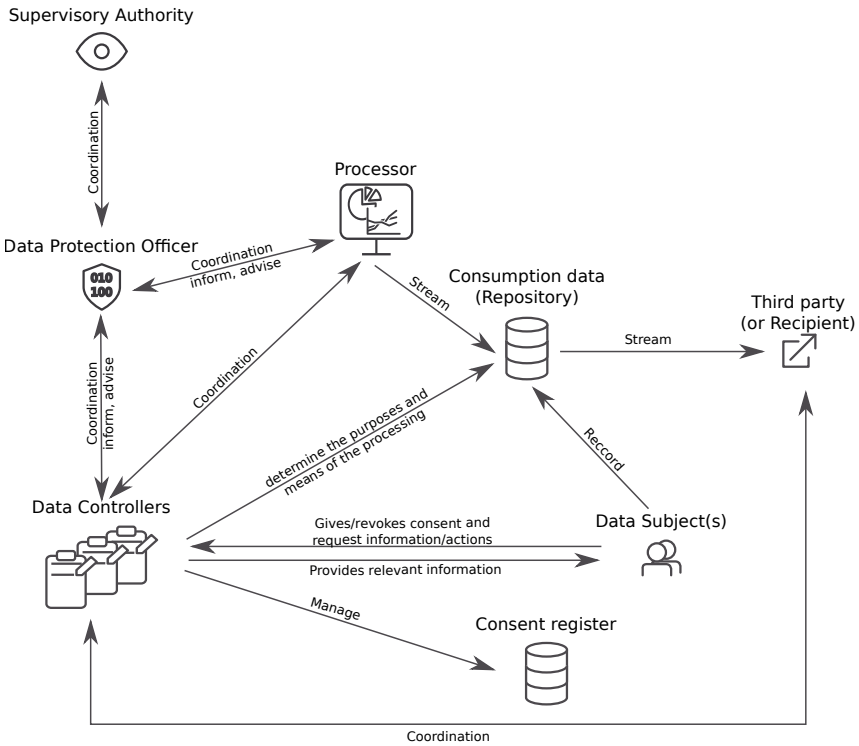


Figure 2.3: Interaction of actors and flow of smart meter data as described in the GDPR. Source: ([46]).

2.3 What privacy today?

Privacy is a generic word used to describe what we perceive as relating to private matters. Nevertheless, the definition of privacy is evolving. In this section, we give some examples revealing today's state of privacy and how much data we accept to give to obtain a service. A discussion is as well open on what is at stake when we talk about privacy breaching.

2.3.1 The state of privacy in the big data era

We have entered a new era called the *'Big data era'* [48]. Despite the term 'big', the root of big data pertains to (i) volume: the quantity of data being collected is growing exponentially [49]; (ii) velocity: The resolution at which data is being collected increases steadily; and (iii) variety:

The sources of data are getting more diverse. From the Data Protection Directive, data can be categorized into two types, personal data, which are protected by law and the non-personal data [50]. Hence, to go around restrictions on the use of personal data, the best way is to collect *more* non-personal data that can be combined to create a unique profile, defining an individual.

On the Internet, the most generic data collected concerns navigation information (i.e., browsing) and clicks. Cookies, saved on each computer, have been used to collect navigation information on users. Users' navigation information is then used to generate targeted advertisements. In Europe and until 2011, websites were not asking for consent on using cookies. In 2011, the so-called *'EU cookie legislation'*, Directive 2009/136/EC, detailing the use of cookies was added to Directive 2002/58/EC on digital user rights [51]. It stipulates that cookies ID are considered *personal data* from now on, and requires any website to ask for users' consent to retrieve information stored on cookies. Despite the efforts of the European Commission to regulate the exploitation of navigation information, new ways of collecting that information were already implemented. In order to optimize their visual aspect, websites collect information concerning the hardware (e.g., screen, computer) and the software (i.e., the browser type and version) with the genuine aim to give the best user experience. However, it can form a unique combination, which is called a *'browser fingerprint'* [52]. To be close to unique the fingerprint of a browser requires approximately 17 parameters. Thereby cookies are becoming obsolete and the online advertising business is still monetizing browsing information while avoiding legislation.

Google and Facebook emphasize concerns about data privacy as they have always been at the forefront of the data monetizing business models, providing services for free and monetizing data via advertisement. Thanks to the dimensions of their pool of users, they are self-sufficient in data to feed their advertisement algorithm. In 2017, Alphabet's (parent company of Google and Youtube) and Facebook's digital advertisement revenues combined represented a gigantic 191,8 Billion US dollars (respectively 123.5 B$ and 68.3 B$), which represents half of the global digital advertising revenue [53]. In itself, the use of data for targeted advertising is not much of a problem and can be considered as annoying when it is excessive. The problems come out of the methods used to maximize revenues.

Facebook generates a unique dataset, which appeals to psychometricians studying human behavior. The collection of *likes* from users can be used to generate precise psychological profiles like the *'Big five'* [54–56]. The Cambridge Analytica Scandal made citizens aware of how a breach into the security could contribute to private interests. Data from 100 000 of Facebook's users were initially collected with their consent for research. Despite rules and Non-Disclosure Agreements, access was given to Cambridge Analytica, which extended the data to 30 million users using interconnections between *friended* users. Data was after that not used for research, but to influence opinions through the targeted advertising algorithm of Facebook. The use of the data is thus not as questionable as the purpose. The exploitation of such a unique dataset for research purposes is valuable. However, the use of such a dataset for influencing opinion is a serious law infringement [57].

The *'privacy by default'* Article in the GDPR, has probably been designed based on the experience with Facebook's default privacy settings. Indeed, Facebook's privacy settings were left to a minimum level so that user's profiles could be *searchable*, and partly *visible* to all members, thus increasing traffic [58, 59]. From a user's point of view, they have to know (i) that access to their account is not restricted to 'friends', and (ii) that they should know how to restrict access [59]. The configuration as default to the lowest security settings is questionable from a user perspective, as

their personal information is not protected by default despite the existence of such parameters. Social networks benefit from data placed in them, but they benefit even more of connections created between user profiles (see [60] for more information) and generate their advertisement revenue from the traffic. Usually, users become aware of privacy issues, when terms and privacy policies have to be updated and some may modify their privacy settings, but the vast majority does not, as it is a non-trivial operation [59].

The use of smartphones and smart city application (e.g., public transportation card, traffic) adds a geographical dimension to information collected that anchors it in the physical world. Using mobility data from carriers' antennas it has been demonstrated that only four spatiotemporal points are needed to uniquely identify each carrier [61]. Using GPS data, the number of points decreases to collect unique patterns. Spatio-temporal data are highly sensitive personal data, information on where an individual is at any time can be used to intercept physical someone. They are personal data as they allow to uniquely identify a person from his data.

The Big Data era has induced in citizens a certain distrust as well as a necessity to stay connected, which create an ambivalent perception of technological products.

2.3.2 Privacy is not the problem anymore

Privacy comes from the Latin word *privatus*, which means 'withdraw from public life.' Indeed, the strict definition and application of privacy imply that each person should not in any way be uniquely identified using the collected data [40]. persoBut as the legal framework on data protection and privacy evolves with the emergence of new information technologies, the concept of privacy evolves as well. Privacy is usually guaranteed to data subjects by collecting data anonymously in the sense of *namelessness* (i.e., not identified by name, address, social security number). Examples have been given in Section 2.3.1 that shows that anonymized data can be used to uniquely identify individuals and thus questions the use of anonymity to protect privacy. Indeed, anonymity is used to collect data without naming the data subject, but keep them identifiable [52, 61]. It is crucial here to understand what is at stake in that context: names have no importance in themselves. However, identities, sets of information that define each person, are precious as well as sensitive. Personal data can be combined with other non-personal data to identify, contact or locate a *single* person. Discarding all this information, is a way to keep them *anonymous* (i.e., nameless), but still uniquely identifiable. Google has even created a word to describe these paradoxical IDs, *'anonymous identifier'*, which they use for targeted advertisement [1, 62].

A question arises then, how many data points are needed to uniquely identify users? Considering the profuseness of possibilities, only a few data points are needed to create a combination that uniquely identifies a user [52, 61]. The problem appears when the data collected can provide sufficient information to reach a person physically (e.g., through email, phone, address). In [62], the authors argue that the real value in anonymity is to prevent *reachability*, not to protect privacy. From the collected data, it should not be possible to communicate or reach data subjects physically. This concept is then much more meaningful and also reshapes the concept of privacy. It does not apply only to personal data, but also to non-personal data that could be used to reach a person. In [63], an algorithm is built to predict the social security number of American citizens based on their date and place of birth. They reach success rates from 7% to 61% in predicting the five first numbers (out of 9) using publicly available data depending on the period and state of birth. It proves that *any personal information can be sensitive information* when combined appropriately [63].

Respecting privacy is not respecting secrecy or granting control over personal information. It consists of respecting an appropriate flow of information. Nissenbaum calls it contextual integrity [47]; data (a type of information) collected in a specific context (e.g., finance, health, social norms) flow, following transmission principles (e.g., consent, buying, selling, confidentiality), between different actors (e.g., subject, sender, recipient) in an appropriate manner. Disruptive practices modifying the information flow are evaluated depending on how they move it from the ideal information flow. In other words, it evaluates the impact of disruptive flows on ethical values like fairness, justice, freedom, welfare or any other context-specific concepts.

Figure 2.3 could be a representation of the perfect flow of information in a smart metering context. In reality, from an ethical perspective, the relationships between utilities and consumers are far from perfect. The major problems are around data collection and the change of status of the actors.

2.4 Customers should be treated as collaborators

GDPR sets standards for data protection and privacy and the Third Energy Package gives guidelines and objectives for the roll-out and use of smart meters. However, in this new context of the smart grid, there are no clear guidelines or rules on how to work with customers. The main challenge is to keep a positive relationship between utilities and (domestic) customers in order to secure the investment and involve them as actors of the grid [34].

2.4.1 An ethical roll-out to improve acceptance of the customers

The roll-out scales as defined in the CBA (Figure 2.2) have been decided at the EU level and the DSOs are following the decision in deploying the smart meters. However, the decision to make the installation mandatory is raising concerns throughout Europe. As presented in Section 2.3 several privacy scandals have raised awareness about privacy issues and this should not be neglected. From a customer perspective, the roll-out of smart meters, especially when mandatory, is an intrusion to what is perceived as the last sanctuary of privacy, 'Home' [37]. Indeed, a meter is a foreign object in a household that inhabitants do not own (it is the DSO's property) and cannot remove/modify. The adoption of technology depends on the perception of customers [64]. The European Commission gave some guidelines on conducting the cost-benefit analysis where '*an assessment of the level of social resistance (or participation) to the project should be presented, including a description of means adapted to ensure social acceptance and their effectiveness*' [37]. The customers either put high expectation in smart meters (technophiles) and get disappointed or have realistic fears regarding privacy breaching and loss of control [65]. Hence, both situations lead to a negative perception of smart meters. To have a positive impact, the benefits of smart meters should be clearly stated and visible rapidly after installation to maximize customers' acceptance of the new technology.

The case of the Netherlands can be used as an example of what can go wrong when end-users are not appropriately considered in the smart metering framework [13]. Originally the roll-out was mandatory and refusing the installation was made punishable as an economic offense, with a fine of 17.000€ or imprisonment for a maximum of six months [66]. Besides privacy concerns transmitted to the Dutch Data Protection Authority on the use of high-resolution data, the utilities were not inclined to focus on a customer's inclusive solution to stimulate demand flexibility [13]. The Minister of Economic Affairs amended the Dutch Data Protection Authority's proposal by stipulating that the network operator could transfer hourly or 15-minute metering data to the

energy provider only if the customer gave his consent. To add up to the pile, the Dutch Consumer Union published a report stipulating that a mandatory roll-out of smart meter reporting 15-minute electricity information was an infringement of the right to privacy according to the article 8 of the European Convention on Human Right [41] and was thus not compatible with a democratic society. The problem was finally solved at the Senate by giving the right to customers to refuse to have a smart meter installed (opt-out). [66] considered that there are four factors for the rejection of the smart meter bill by the Senate (i) the high resolution of the data transferred up to the energy providers, (ii) the mandatory roll-out where resistance is sanctioned by high fines or even imprisonment, (iii) lack of explanation of the necessity of smart metering and by extension why the customers have to lose some privacy, and (iv) the combinations of different functionalities in one meter generating new risk and making the argumentation complex.

Research in social science on the topic of smart meters have also shown significant misalignment between the reality of the smart meters and customers' expectations. From a customer point of view, just the fact that a digitally connected meter is called 'smart' is inducing a wrong idea of what are its capabilities since it is not a smart home system [67]. This is a recipe for backlash. In [65], a behavioral study shows that most of the concerns and deceptions from the roll-out of smart meters could be solved in two ways(i) scale down the expectation of the customers in explaining clearly what the smart meters could do; and (ii) align the technology with the expectation by adding smart thermostats and in-home displays to visualize consumption in real-time. The smart grid framework requires that the customers know what their metering data is used for, even if it is technical, stakeholders have the responsibility of informing in a clear and understandable manner [34].

2.4.2 Evolution of the roles and relationships

The relationship between utilities and customers is ultimately changing due to potential consequences from two-way communication and deployment of decentralized generation [68]. Consumers become prosumers and provide service to the grid. They are actors of the grid and should be considered as collaborators in maintaining the stability of the grid. The European utilities have operated in the last decade important restructurations to cope with the opening of the energy market to concurrence. Indeed, utilites (e.g., EDF in France, ENEL in Italy, DONG in Denmark) used to be monopolistic and control the network from generation to distribution. As national companies, they had the trust of customers. Today the trust-relationship has to be rebuilt between customers and utilities to operate this transition and find a new balance between the actors of the grid. The goals of the relationship remain unclear to some extent as the benefits and the expectations are not aligned [20]. For example, the customers are expected to be more active, but smart meters alone do not provide this functionality, it requires an additional smart home device. Additionally, the contribution of customers to the stability and reliability of the grid should be highlighted as it can be used to develop new social norms concerning energy [20].

The development of aggregators could play an essential role in 'smoothing' the communication between the utilities and the customers as they would have fewer customers to handle. Indeed, beyond their technical role, they could act as representatives of the customers to utilities and have more weight in the decision making.

2.4.3 Smart meters to empower customers to become prosumers

The smart grid aims at transforming a centralized, utility-controlled network into a decentralized, consumer-interactive network allowed by high-resolution monitoring and two-way communication [68]. From a utility perspective, the need for metering data is almost mechanical. Indeed, a higher share of Renewable Energy Sources (RES) in the generation mix, as promoted by the EU, makes generation less adjustable to the demand. In order to compensate for the lack of flexibility on the generation side, the demand could be modulated according to some incentives (i.e., price, benefit) broadcast to the customers using a two-way communication [69].

DSM programs have been studied and implemented based on the idea of exploiting demand-side flexibility to reduce RES spillage [7]. DSM (including Demand Response (DR) frameworks as well as more complex pricing scheme) relies on a marginal dynamic price of generation [28]. Figure 2.4 gives an overview of the different price based solutions that can be used depending on resolutions of both price and metered data. To generate the corresponding bill, the energy providers need to know exactly how much power each customer has consumed during each time interval. Hence, the resolution of the metered data should then be higher than (or equal to) the one from the dynamic tariff. In such a framework, it is important that customers can access their electricity consumption and dynamic tariff to modify their energy behavior or to automate their white appliances (i.e., dishwasher, washing machine, electric heating) accordingly. High-resolution metering is then a way to make customers aware of their energy behavior so that they can shift their consumption from passive (consumers) to active (prosumers) who will provide services to the grid [70]. The incentive used to change the electricity consumption behavior of customers does not have to be financial; social norms are a powerful tool to change behaviors [71]. However, for such incentives to have a positive impact, the relationship between the utilities and the customer has to be positive as well [20]. A customer who manages his consumption closely should then be encouraged to get electricity cost reductions [11, 72].

In the context of smart grid, new business models and actors (aggregators) are relying on metering data to create portfolios and manage their assets [73]. However, it has been demonstrated that the success of such frameworks depends heavily on the magnitude of demand response triggered and subsequently active customers [74].

2.4.4 Control of access transferred to the data subject

The electricity metering data are stored on a data hub or the DSO's servers (Table 2.1). A consent register, as presented in Figure 2.3, can be created to record which third parties get access to the data. The consent register is here managed by the data controller (DSO), but it could have the form of the 'App' system developed for smartphones [46] where customers directly manage access grants. Hence, it will transfer responsibilities, risk assessment, and control to the data subject. It could then also generate the same problems as with 'Apps' on smartphones that are asking for access to data, which are not useful to the service provided. A third party can access data at high resolution (up to 1s depending on the model) wire or wireless to smart meters using a dedicated port. Again if there is no illegal intrusion to the household, it is assumed that customers should have control over what is connected to the port. The risk of abusive use of metering data by a third party is increased if the customers are not educated and made aware of how sensitive those data can be (as we can observe with smartphones). The risk is that information in the data (e.g.,

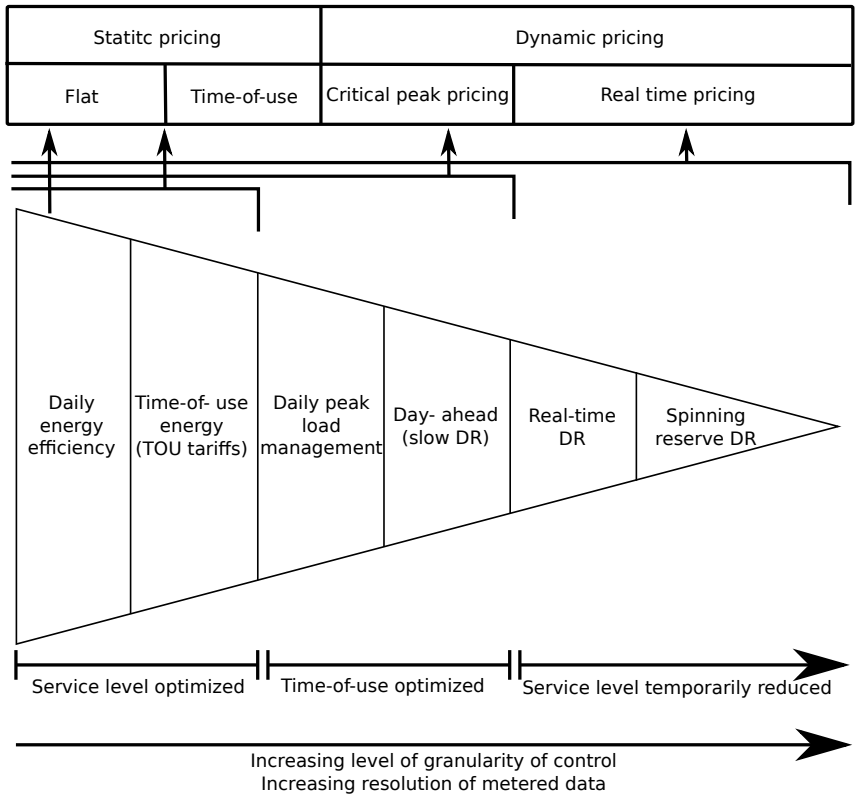


Figure 2.4: DSM service enabled in function of the resolution of the metering. Source: [18].

state of white appliances) is extracted and used by unsolicited third parties for sending targeted advertisement suggesting to replace an appliance [69].

The change from consumer to prosumer comes with changes in the distribution of the responsibilities and control between the actors of the grid. Indeed, a prosumer needs to control his/her consumption to act as such and provide services to the grid. The hierarchical structure of the actors of the grid, with consumers/prosumers at the bottom and utilities at the top, is actually, changing. The structure is flatter and prosumers are collaborators rather than consumers. Adjustment in terms of consideration has to occur.

2.5 Requirements in terms of ethics

Good practices can be implemented to make the roll-out and the use of metering data more ethical. It mainly consists of two global concept parsimony and equity. The need for data is acknowledged. Nevertheless, it should be parsimonious and come with a direct benefit to the data subject. Hence, both parties would be satisfied.

2.5.1 Data resolution in accordance to task to fulfill

Privacy concerns are often about the high resolution of metering data- [12]. The higher the resolution, the more precise the information (See Figure 2.5). By extension, concerns are also related to machine learning algorithm fed with such data like Non-Intrusive Load Monitoring (NILM) [75]. Beyond the potential information that can be extracted, the targeted use of the information is of higher interest. NILM applications, for example, are made in a specific context; it consists of

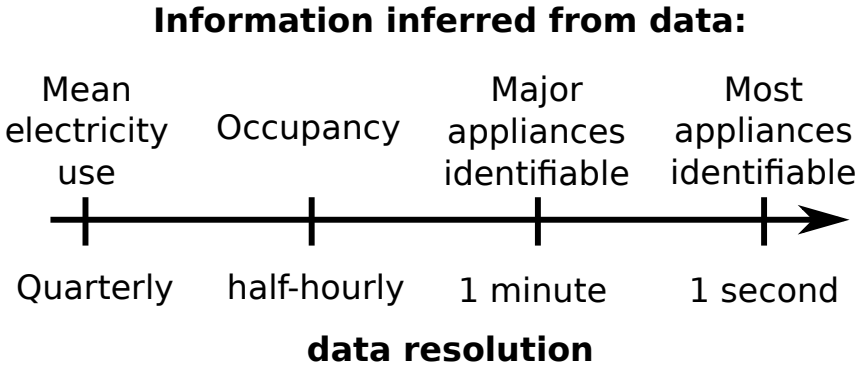


Figure 2.5: Representation of information that can be inferred from metering data in function of the resolution. Source:[12].

providing detailed information of individual appliances consumption to customers, who are also data subject, so that they can identify appliances with large unnecessary electricity consumption. This information should be provided to no one else. We could also imagine the use of NILM on data at a lower resolution to identify large and potentially flexible appliances. Hence, DSOs could use this information to propose flexibility contracts to customers; it would be beneficial for every party if the benefits are fairly distributed.

Different tasks can be completed using metering data, but they do not require the same level of information (data resolution). The data resolution should, therefore, be adjusted in accordance with the task to fulfill. In the same way that the purpose has to be legitimate, the resolution of the data has to be legitimate. For example, when billing customers under dynamic tariff, it does not improve anything to use electricity consumption at a higher resolution than the dynamic tariff. Hence, the resolution of the data should be chosen parsimoniously.

2.5.2 New risks require compensation

As with any connected device (e.g., computers, Internet of Things (IoT) devices), a risk of cyberattack exists. It can be organized by a foreign governmental agency, a malicious person or a malicious software [76]. The Russian attack on Ukrainian DSO Kyivoblenergo on December 23rd, 2015, is the first example of such an organized cyber attack used to temporarily shut down 30 substations of the distribution grid [77]. The grid is a strategic target and the use of a digital central control system makes them obvious targets for cyberattack. However, the attack did not target metering data, but the stability of the grid, which do not affect privacy in this specific case. Nevertheless, a cyber attack could also be conducted by customers on their own smart meter to steal electricity [72, 78], or by a malicious person on a specific customer to spy on him [12].

Hence, customers are exposed to new risks of privacy breaching in giving access to their metering data. The risk is considered and efforts on securing communication are made to limit it. However, the risk is not null, and it should be addressed with compensations/benefits [67]. As utilities transfer the cost of smart meters to the customers (Table 2.1), it would fair that customers get rewarded according to the amount of information transmitted to utilities [79].

We could imagine a voluntary basis system where customers could choose the resolution of the data provided and would have tariff/remunerations accordingly. The differences between the static tariff (i.e., for non-metered customers) and dynamic tariff (i.e., for metered customers) should then take into account the marginal cost of generation, but also a deduction for providing data for the forecast of the demand.

2.5.3 Balance of the benefits

Balancing the benefits resides in a trade-off between the loss of privacy and increased risk for the customer and the need for metering data for utilities [79]. We want to show that it is possible (even necessary) to do it ethically in order to be sustainable and avoid future backlash [36].

Today, the benefits of smart metering are going toward the utilities, which are saving the cost of employing meter readers, process invoice automatically and get more insights on the grid for fraud detection and maintenance [80]. As the meters are paid mostly through network tariffs (only Italy, Romania, Slovakia and Sweden are sharing costs between customers and DSO see Table 2.1), the tariff reductions due to potential savings will be, in a first time, shortened on the customer side. Real-time data has not shown that much interest and it seems fair to think that it will take some years before the technology becomes mature enough to deploy large scale DSM application. Hence, until dynamic tariffs are generalized to all EU Member States, some customers will pay for the technology without having any of the benefits. The access given to consult and analyze electricity consumption, as promoted by EU, would then have only little impact, as the customers could only reduce their consumption to reduce their electricity bill.

Moreover, customers undercharged because of malfunctioning electromagnetic meters, which is common because of the advanced age of electromagnetic meters, will observe an increase in their electricity bill due to increased metering accuracy [65]. A more accurate billing also means that it is easier for the energy providers to detect fraud in comparison to annual metering on electromagnetic meters. To give an idea of the cost of electricity theft, it is responsible in 2000 in the US of 0.5% to 3.5% losses of the annual growth revenue, which seems low, but still represents $10 billion, compensated by a higher price on the other customers [81]. With smart meters, the risk of undetected frauds decreases, which means that theoretically the cost of the fraud is reduced and can be translated into lower prices. Fraud is better monitored, but at the same time, the risk of electrocution in compromising smart meters (i.e., through software) is much lower than with electromagnetic meters and thus less repealing to possible thieves [72].

Smart meters are part of the Advance Metering Infrastructure (AMI), which forms the informational backbone of the smart grid and makes the grid smart. From a DSO perspective, AMI allows them to have precise information about the power flows at a distribution level beyond the substations. Again this value of metering is emphasized by the increase of variable RES in the generation mix and decentralized generation down to the distribution level [69]. This way it lowers the risks of outages, the DSOs can also anticipate the maintenance and solve problems faster as they do not need customers' calls to be aware of them.

Whatever the decision was taken to increase the capacity of RES and subsequently to limit climate change due to electricity generation, dynamic information on the demand side will be required to optimally use RES, invoice prosumers and balance the generation with the demand [11]. Hence, the roll-out of smart meter from an environment and grid management perspective is not negotiable, but the way the data is used and how the benefit of such infrastructure will be shared are still under discussion. If as in the case of the Netherlands [13], an opt-out is negotiated in most of the member states, customers will perform a cost-benefit analysis between the loss of privacy and the social, economic benefits generated [79]. The worst scenario could lead to the loss of an important part of metered households.

2.6 Conclusions and perspectives

Utilities are putting efforts into complying with the GDPR. GDPR however only protects the fundamental rights of the data subjects (i.e., customers). Nevertheless, DSOs must also respect the agenda of the Third Energy Package and deploy smart meters in due time. Hence, mandatory installation of smart meters appears to be the best solution to fulfill the task in time. This is not counting on the possible backlash of the customers. The Netherlands' case gives us a picture of what could happen in the EU during the next years [13].

Bad-practices during the roll-out (e.g., installing smart meters without informing customers) are giving a negative image of smart grid technology to the customers. The fact that it is required to improve the use of RES or that it is imposed by the EU are not arguments supporting the implementation of the smart grid as it is right now. On the contrary, all the technological developments and investments would be a vain attempt if customers do not adopt or at least accept the technology in the first place. Problems may not only arise during the roll-out, but also afterward as the relationship is dynamic and requires efforts on both sides to be fruitful.

It becomes then clear that insights from social science are necessary to understand how customers perceive 'smart meters' in *their* home. Indeed, from an engineering point of view, the customer as non-rational in its decision making is often seen as a problem. Hence, only the technico-economic aspects of the problem are used to evaluate a framework. The use of social sciences like consumer science, justice research, ethics and anthropology to large infrastructure projects involving citizens has shown a positive influence on their perception regarding a potentially non-appreciated project (i.e., construction of a dam, road).

In the general context of climate change, citizens have become aware of different levels of their responsibilities as well as how they can influence the outcome. Energy represents one of the fields were awareness is growing rapidly. It seems that customers are ready to become actors of the grid and support the development toward a greener electricity generation, but not at any cost.

CHAPTER 3

A Dynamic Approach to Load Profiling

Beyond the increasing amount of data collected, the deployment of smart meters has changed the nature of the metering data from data points to streams of data. Hence, load profiling, that used to be computed on yearly meter reading and information provided by customers, can now be data-driven. Furthermore, load profiling can take advantage of the stream nature of the metering data to provide more accurate information. This chapter discusses how load profiling approaches can profit from the stream nature of metering data to tackle the stationarity and homogeneity assumptions used with state-of-the-art data-driven load profiling approaches. This chapter uses the methodology developed and the results collected in [Paper B].

3.1 Typical load profiles generation and purpose

The implementation of 'greener' energy policies supporting the development and consumption of Renewable Energy Sources (RES) have motivated investments in Advance Metering Infrastructure (AMI) to modernize the grid (Third Energy Package in the European Union [82] and American Recovery and Reinvestment Act in the United States of America [83]). Such an infrastructure aims to update the status of the distribution grid more frequently in order to avoid or anticipate possible congestion problems due to RES intermittency. Indeed, intermittent RES increases uncertainty in matching the generation to the demand at any time, which can put the grid at risk of instability as variations in RES generation are not necessarily correlated to variations of consumption. Moreover the increasing deployment of Electric Vehicles (EVs), Heat Pumps (HPs), Photovoltaics (PV) in a non-homogeneous way over space requires observance at a finer granularity in time and space [84]. Solutions exist to optimize the use of RES (i.e., storage, Demand Side Management (DSM)), but they all rely on precise information on generation and demand statuses.

The need for information on when and how electricity is consumed has always existed at the Distribution System Operator (DSO) level to operate and maintain the grid. Historically, typical load profiles based on the categories of customers were the only insights available on the demand-side. The DSOs have been using them for, (i) electricity demand analysis, (ii) tariffs design, (iii) DSM, (iv) future load estimation, and (v) power system planning and operating. Before AMI, data was extremely scarce as electromagnetic meters were read once a year and only large customers were metered with an hourly resolution. The typical load profiles were generated by segmenting the metered customers into categories corresponding to their activities (e.g., offices, supermarket, streetlight, domestic) or demographics, building properties and white appliances for domestic customers. Hence, the objective was to generalize the information from a set of metered customers to the entire pool of customers in the most accurate manner. Hypotheses were formulated about

the stationarity of the load over time and homogeneity inside each category of customers. For the domestic customers, the Velander formula has been used to obtain the peak load and recreate load curves based on yearly consumption [84, 85]. Nevertheless, categorization is known to be inaccurate, as subpopulations exist in almost every category [14]. Grouping customers based on their consumption patterns is more cost-effective as the typical load profiles estimation is more accurate [21]. Deployment of smart meters has drastically changed the paradigm of load profiling: from estimating the demand using a few metered customers, to now summarizing the information of streams of metering data.

The problem of summarizing a large amount of data into a limited number of representative points is a well-known problem in data analysis, which has been studied extensively to produce clustering algorithms [86]. Hence, research on improving load profiling techniques using metering data explores what has been done in the field of data analysis, and clustering stands out as a well-suited approach to produce typical load profiles.

3.2 Clustering-based load profiling

The state-of-the-art in load profiling uses a classical clustering framework displayed in Figure 3.1. The framework is composed successively of a data preprocessing and feature extraction (optional), a clustering algorithm and a performance evaluation criterion to choose the best partition in K clusters [87].

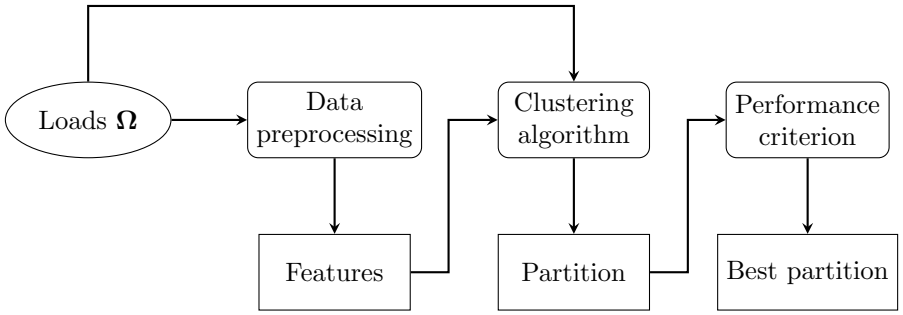


Figure 3.1: Classical data analysis clustering framework.

It is then applied on batches of historical data, $\mathbf{\Omega} = \{\mathbf{x}_1, \ldots, \mathbf{x}_n, \ldots, \mathbf{x}_N\}$ composed of N daily loads $\mathbf{x}$, to generate typical load profiles, which we call static as a complete rerun of the framework on a more recent batch of data is needed for the typical load profiles to be updated.

Table 3.1: The different types of clustering-based load profiling [15].

Number of clusters:	Static	Dynamic
Load static	Type 1	Type 2
Load dynamic	Type 3	Type 4

The load profiling methodologies can be classified as four types, as presented in Table 3.1, depending on whether it considers the time dependence of the load and whether the number of clusters K is changing over time. In this terminology, a static load profiling approach is a type 1 or type 3 depending on whether the preprocessing technique projects the load in a frequency domain or not. This is the first step towards data-driven load profiling.

In the following section, an overview of the methodologies used in load profiling is given, looking first at the data preprocessing then at clustering algorithms and finally at the performance criteria.

3.2.1 Data preprocessing

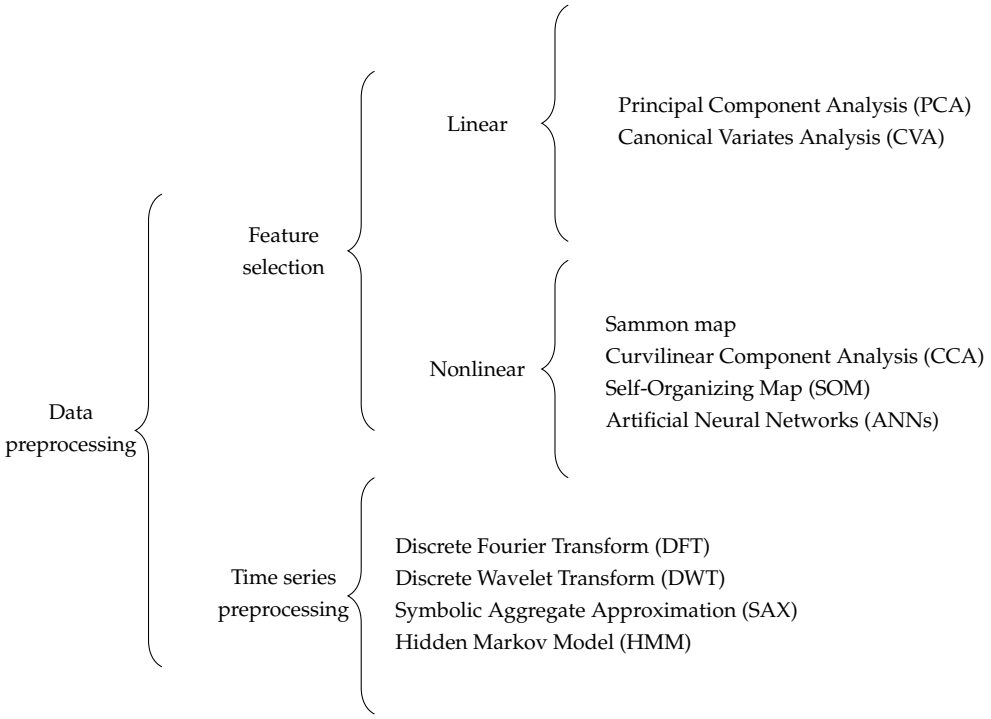


Figure 3.2: Classification of the preprocessing techniques used in the literature of cluster-based load profiling.

In clustering, preprocessing is often used in order to extract features containing non-redundant information derived from original variables (Figure 3.1). The clustering algorithm is then performed on the features, which concentrate the information contained in the variables. The preprocessing methods can be classified as presented in Figure 3.2 depending on their nature. In the following section, we give an overview of the feature extraction techniques and time series preprocessing encountered in the literature of load profiling.

Feature extraction

The principle of feature extraction is to remove redundancy among input variables, hence the number of dimensions T in space formed by the original variables is reduced to $D \ll T$ features in the final space. The projection can be a linear or non-linear transformation combining the original dimensions to form features. Hence, the features are uncorrelated and contain the same information as the original dimensions. The feature extraction speeds the clustering process up as fewer calculations are required.

In data analysis, the most well-known feature extraction technique is the Principal Component Analysis (PCA), which finds a set of orthogonal components representing the dimensions of highest variance in the dataset. Mathematically, it consists of finding the coordinates $\mathbf{P}$ of $\mathbf{\Omega}$ in a space so that the coordinates are ordered by decreasing variance. To do so, we need to find $\mathbf{W}$ the projection matrix with dimensions $N \times T$ so that

$$\mathbf{P} = \mathbf{\Omega W}, \tag{3.1}$$

that verifies

$$\mathbf{\Omega}^\top \mathbf{\Omega} = \mathbf{W}\mathbf{\Sigma}^2\mathbf{W}^\top, \tag{3.2}$$

where $\mathbf{\Sigma}^2$ is a diagonal matrix with the eigenvalues of $\mathbf{\Omega}^\top\mathbf{\Omega}$ on the diagonal. The eigenvalues of $\mathbf{\Omega}^\top\mathbf{\Omega}$, which is proportional to the covariance matrix of $\mathbf{\Omega}$, are thus conveniently ordered by decreasing order on the diagonal. $\mathbf{W}$ A subset of the components explaining a large amount of the covariance can be selected as $\mathbf{W}_D$ the D-first eigenvectors of $\mathbf{W}$ and forms a projection matrix of dimension $N \times D$. The coordinates $\mathbf{P}$ in the projected space of the loads in $\mathbf{\Omega}$

$$\mathbf{P}_D = \mathbf{\Omega}\mathbf{W}_D \tag{3.3}$$

with $\mathbf{P}$ matrix of dimensions $D \times N$. The matrix $\mathbf{P}_D$ can then be used as input to a clustering algorithm [22, 87].

$$\mathbf{\Omega}^\top\mathbf{\Omega} = \mathbf{W}(\mathbf{\Sigma}_B^\top\mathbf{\Sigma}_W)\mathbf{W}^\top \tag{3.4}$$

to obtain the eigenvectors of the product within groups and between group covariance matrices [22, 88]. Hence, the canonical variables are the dimensions that maximize the separation between groups while minimizing the separation within groups [89].

Unlike PCA and CVA, Sammon map performs a non-linear projection that preserves the original space topology [89]. To do so, the differences between the point-to-point distances in the original and projected spaces are measured [22]. The optimal mapping is the one that minimizes the difference according to the error

$$E = \frac{1}{\sum_{n=1}^{N-1}\sum_{m=n+1}^{M} d(\mathbf{x}_n, \mathbf{x}_m)} \sum_{n=1}^{N-1}\sum_{m=n+1}^{M} \frac{[d(\mathbf{x}_n, \mathbf{x}_m)d(\mathbf{y}_n, \mathbf{y}_m)]}{d(\mathbf{x}_n, \mathbf{x}_m)}, \tag{3.5}$$

where $d(\mathbf{x}_n, \mathbf{x}_n)$ and $d(\mathbf{y}_n, \mathbf{y}_m)$ are respectively the distance between two loads in the original and the projected spaces. It randomly initializes $\mathbf{y}_n$ and uses gradient descent to iteratively minimize the error E. It is more powerful than PCA as it can form nonlinear combinations of variables.

Artificial Neural Networks (ANNs) can also do feature extraction in a non-linear manner. Self-Organizing Maps (SOMs) are unsupervised ANNs whose goal is to find prototype vectors (i.e., weights) that represent the input data set while simultaneously realizing a continuous mapping from the input space to a lattice (i.e., grid). It can conserve the topology structuring the commonalities between loads in the initial data set [90]. Hence, a SOM is both a feature extraction technique and a clustering algorithm as it operates a projection from T-dimensional to a D-dimensional space with $0 < D \leq 2$. It forms a bi/unidimensional output space where the loads are grouped [91, 92]. Often the cluster formation is not apparent and a simple clustering process, K-means or Hierarchical Ascendant Clustering (HAC), is required to separate them. SOMs are considered a robust and computationally efficient approach. Example of its applications can be found in [90, 92, 93]. A Curvilinear Component Analysis (CCA) is a specific type of SOM where the lattice is a continuous space able to adjust perfectly to the shape of the subspace [22, 87]. In [94], a deep-learning auto-encoder is implemented to perform load profiling in real-time.

In the case of feature extraction, time steps of loads are decoupled and the time dependency is lost. Only the time step or a combination of time steps contributing to the discrimination of the space are kept, which can be an issue when patterns are similar, but slightly shifted in time.

Time series preprocessing

The problem of time dependency in feature extraction can be tackled with time series preprocessing. The most popular transformation is the projection from the time domain to the frequency domain. In the literature of load profiling, Discrete Fourier Transform (DFT) [92, 95] and Discrete Wavelet Transform (DWT) [96, 97] have been implemented as preprocessing to clustering. They respectively decompose the signal into harmonics and wavelets of different lengths [22]. The harmonics (or wavelets) can then be filtered to keep only the most representative harmonics. The main drawbacks of DFT are that, (i) it requires stationary behavior of a load, which can be reasonably considered non-stationary over a long period [98]; and (ii) as harmonics are filtered, the original gets degraded quickly. As DWT decomposes the signal locally into wavelets, it can be recomposed with accuracy after filtering[1].

Other time series preprocessing methods compress the information in time to reduce the length of the signals to process. Indeed, as the loads are the aggregation of consumption of different appliances or behaviors, patterns appear repetitively over time and can be either summarized by a letter as in Symbolic Aggregate Approximation (SAX) [99], or a state as in Hidden Markov Model (HMM) [100]. Hence, the features are not signals anymore, but sequences of index/symbol each referring to an element in a dictionary of patterns.

3.2.2 Clustering algorithms

There are plenty of clustering algorithms in the literature of data analysis. In the load profiling literature, only some that appear to be more suited for this task have been tested. Classification techniques have also been used in the literature, and are therefore included in this section. Figure 3.3 gives an overview of the type of the clustering and classification methodologies used in load profiling.

Clustering

All the unsupervised clustering techniques in the literature are distance-based clustering and can be divided into two groups; the Hierarchical Ascendant Clustering (HAC) and the partitioning algorithm, which is the K-means family algorithm and K-nearest neighbor. The metric used in clustering is often Euclidean, but it can be replaced by other distances [86].

The HAC takes a pairwise similarity square distance matrix of all the loads in the data. It starts with each load being a cluster and iteratively, (i) looks for the smallest value outside the diagonal to group the two closest clusters, and (ii) calculates their average typical load profile, until a single cluster is formed. A dendrogram, also called hierarchical tree, is then plotted to divide the partition into K clusters. Hierarchical clustering has been used in [87] to compare the performance of feature extraction techniques. HAC has also been used after SOM to refine the partitions into a lower number of clusters [90].

The K-means is also an iterative algorithm consisting of three steps, (i) an assignment step, where each load is assigned to the closest cluster center, (ii) an update step, where the cluster centers are recalculated as the average of the loads in the cluster, and (iii) a checking step, calculating the objective function to check for convergence (within-cluster sum of squares). The K-means algorithm starts with K cluster centers called centroids, randomly seeded among the loads $\mathbf{x_n}$ in

[1] DWT is used for this reason to compress music into MP3

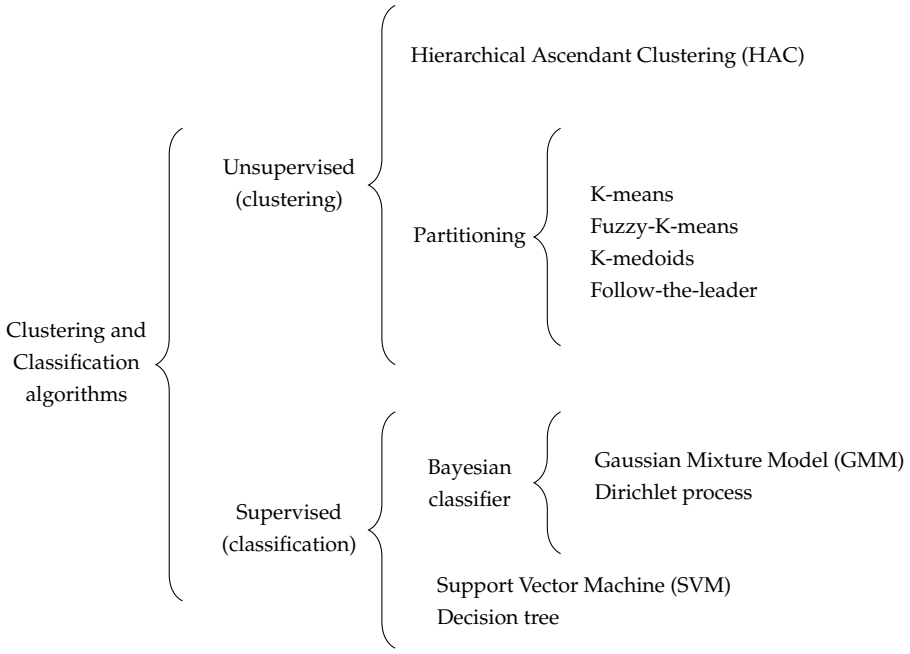


Figure 3.3: Classification of the clustering algorithms used in the literature of cluster-based load profiling.

$\mathbf{\Omega}$. The problem with random seeding, and more generally with K-means, is that it may converge towards a local optimum. After several repeated runs, the bias can be reduced, as the probability increases that the global optimum (lowest within-cluster sum of squares) is reached. The centroids can also be chosen more strategically. The K-means++ algorithm, for example, picks only the first center randomly and then finds the farthest loads iteratively from the selected centers [15, 101]. Hence, the risk of getting into a local optimum is reduced as the seeded centroids are picked at extremities of the cloud of points. Many variations of the K-means exist. Among the most popular are fuzzy-K-means [102] where the assignment is probabilistic, K-medoids [93] where the representative point of each cluster is the closest load in $\mathbf{\Omega}$ to the center, and more advanced variations [15, 103]. K-means has also been used after SOM to refine the partition into a lower number of clusters [90].

Follow-the-leader is an iterative clustering algorithm where a threshold σ, representing the largest distance from a point in a cluster to its center, is set instead of specifying K [91]. Hence, the first iteration of the clustering process sets K the number of clusters, the following steps refine the clustering, reassigning loads to the closest cluster center as in K-means [87]. The threshold σ is then tuned using a trial-and-error approach.

Classification

Classification algorithms have also been used for load profiling. Indeed, as the amount of metered households is constantly increasing, training a classifier on a representative set of loads collected previously to form groups (abusively called clusters) and then classify the remaining or new ones is an efficient approach if we assume that the load patterns are reproduced over time. As in the case of clustering, only a few classifiers have been tested on load profiling from the data analysis literature.

Among them is the Bayesian framework that models each class as a distribution and has the advantage of being inferential. It is a solution to the problem of updating the clusters to newly collected data in a computationally efficient manner. Gaussian Mixture Model (GMM), using multidimensional Gaussian distributions to represent the classes, has been demonstrated in [104, 105]. The number of classes K is defined *a priori* and a model, adjusting the parameters of the distributions, is fitted on the data using an Expectation Maximization (EM) algorithm that aims at optimizing the log-likelihood of the mixture model. Another Bayesian classifier observed in the literature is the Dirichlet process. The Dirichlet process is non-parametric and thus overcomes the disadvantages of GMM the number of classes does not have to be defined *a priori* [106]. It can theoretically fit an infinity of distributions. The distribution of each class is learned iteratively.

Support Vector Machine (SVM) is another branch of machine learning algorithms, which can be used for classification or regression [107]. The objective of SVM in the case of two classes would be to find a hyperplane that separates the data in two parts, maximizing the distance of the nearest points to the hyperplane on each side. The number of hyperplanes depends on how many classes have to be generated. In [108], a more advanced non-parametric version of SVM is proposed.

Decision trees regroup all the hierarchical models used to perform multiclass classification. It consists of nodes that split the data into two branches based on statistics of one or many features. The decision at each node can be deterministic or probabilistic. Decision tree application examples can be found in [107, 109, 110].

3.2.3 Clustering validity assessment

In classical clustering approach, the number of clusters K has to be empirically determined as it is a parameter of most clustering algorithms. To do so, the chosen algorithm is run many times $\{1, \ldots, q, \ldots, Q\}$ with different values of K_q in order to select the partition in K-clusters optimizing a criterion. Therefore this procedure is computationally costly as only one of the implementations is used afterward. Moreover, as for the preprocessing and clustering, many criteria exist, generating different results. The most known are the clustering dispersion indicator (CDI) [87, 91, 96], Davies Bouldin Index (DBI) [87, 90, 96, 103], Scatter Index (SI) [106], Silhouette Index [107], the ratio of the within-cluster sum of squares to between clusters sum of squares [96]. All of them have a different definition of what is a 'good' partition or a 'representative' set of cluster centers, which makes the choice of the 'best' partition complicated.

In the context of load profiling, the assumption is that the set of typical load profiles chosen to represent the pool of loads is representative in the sense that the error generated in assuming that a load corresponds to its assigned cluster center is minimal. Hence, we can choose a more general error estimator like the Root Mean Square Error (RMSE) calculated between the cluster centers and their assigned loads to estimate how much error is generated in assuming that a load can be accurately represented by its cluster center.

Beyond the methodologies presented in this section, some more advanced clustering approaches for load profiling have been proposed like ant colony clustering [111] and more complex multi-stage clustering composed of several clustering algorithm [96, 100, 112, 113]. The technological evolution of appliances and the penetration of new type of electrical appliances (e.g., EVs, HPs, batteries) are changing the loads on a time scale, which supposes the profiling to be rerun more often. In the classification presented in Table 3.1 all the methodologies referenced in this section are

type 1, which is problematic considering the non-stationarity nature of loads. Benítez *et al.* have implemented type 2 [15] and type 3 [114] clustering-based load profiling using variations of the K-means algorithm. In the following section, a type 4 clustering-based load profiling methodology is presented, which intends to rethink load profiling in a fully dynamic manner. Indeed, the clustering process is: (i) flexible, customers can change cluster; (ii) adaptive, the number of clusters can change according to data structure; (iii) online, the typical load profiles are dynamic and recursively updated; and (iv) it respects time dependency of load patterns.

3.3 An online adaptive clustering algorithm

From the analysis of load behavior, the implementation of an online adaptive clustering algorithm for load profiling has to accommodate the following three characteristics:

1. The loads clustered together may have the same behavior and thus have the same dynamics over time (stationary),
2. at times a load may disrupt from the cluster temporary or permanently to join another cluster (non-stationary),
3. behaviors never observed before may appear (disruption).

An overview of the algorithm presented is this section is given in Figure 3.4. The division of the section is represented by dashed rectangles; Section 3.3.2 (top rectangle in Figure 3.4) explains how the online adaptive clustering parameters are set; Section 3.3.3 (bottom rectangle in Figure 3.4) describes how the online part of the clustering is implemented; Section 3.3.4 (right inside bottom rectangle in Figure 3.4) presents how the online clustering can be adaptive.

3.3.1 Notations

In an online approach, loads are seen as elements (e.g., days, weeks) of a time series. We then write $\mathbf{\Omega}^t = \{\mathbf{x}_1^t, \ldots, \mathbf{x}_i^t, \ldots, \mathbf{x}_I^t\}$ where $\mathbf{\Omega}^t$ is a matrix containing loads $\mathbf{x}$ at time step t for customers $i \in \{1, \ldots, I\}$. Hence, loads are synchronized on the same time step t and the matrix $\mathbf{\Omega}^t$ can move in time.

We define the set of typical load profiles $\mathbf{\Upsilon}^t$ obtained after clustering $\mathbf{\Omega}^t$ formed by K^t vectors as $\mathbf{\Upsilon}^t = \{\mathbf{y}_1^t, \ldots, \mathbf{y}_k^t, \ldots, \mathbf{y}_{K^t}^t\}$, where K^t is the number of clusters at time step t. As K-means is used, the distances between the set of loads in $\mathbf{\Omega}^t$ and the typical load profiles $\mathbf{\Upsilon}^t$ at time step t are calculated and stored into a matrix

$$d(\mathbf{\Omega}^t, \mathbf{\Upsilon}^t) = \begin{bmatrix} d_{11}^t & \ldots & d_{1k}^t & \ldots & d_{1K^t}^t \\ \vdots & \ddots & & & \vdots \\ d_{i1}^t & & d_{ik}^t & & d_{iK^t}^t \\ \vdots & & & \ddots & \vdots \\ d_{I1}^t & \ldots & d_{Ik}^t & \ldots & d_{IK^t}^t \end{bmatrix}. \tag{3.6}$$

The vector of assignment at time step t, specifying to which typical load profile k each individual load $\mathbf{x}_i^t$ in $\mathbf{\Omega}^t$ is assigned to, is noted $\mathbf{A}^t = \{a_1^t, \ldots, a_I^t\}$ and generated using the operator,

$$a_i^t = \underset{k}{\operatorname{argmin}}(\mathbf{d}_i^t) \tag{3.7}$$

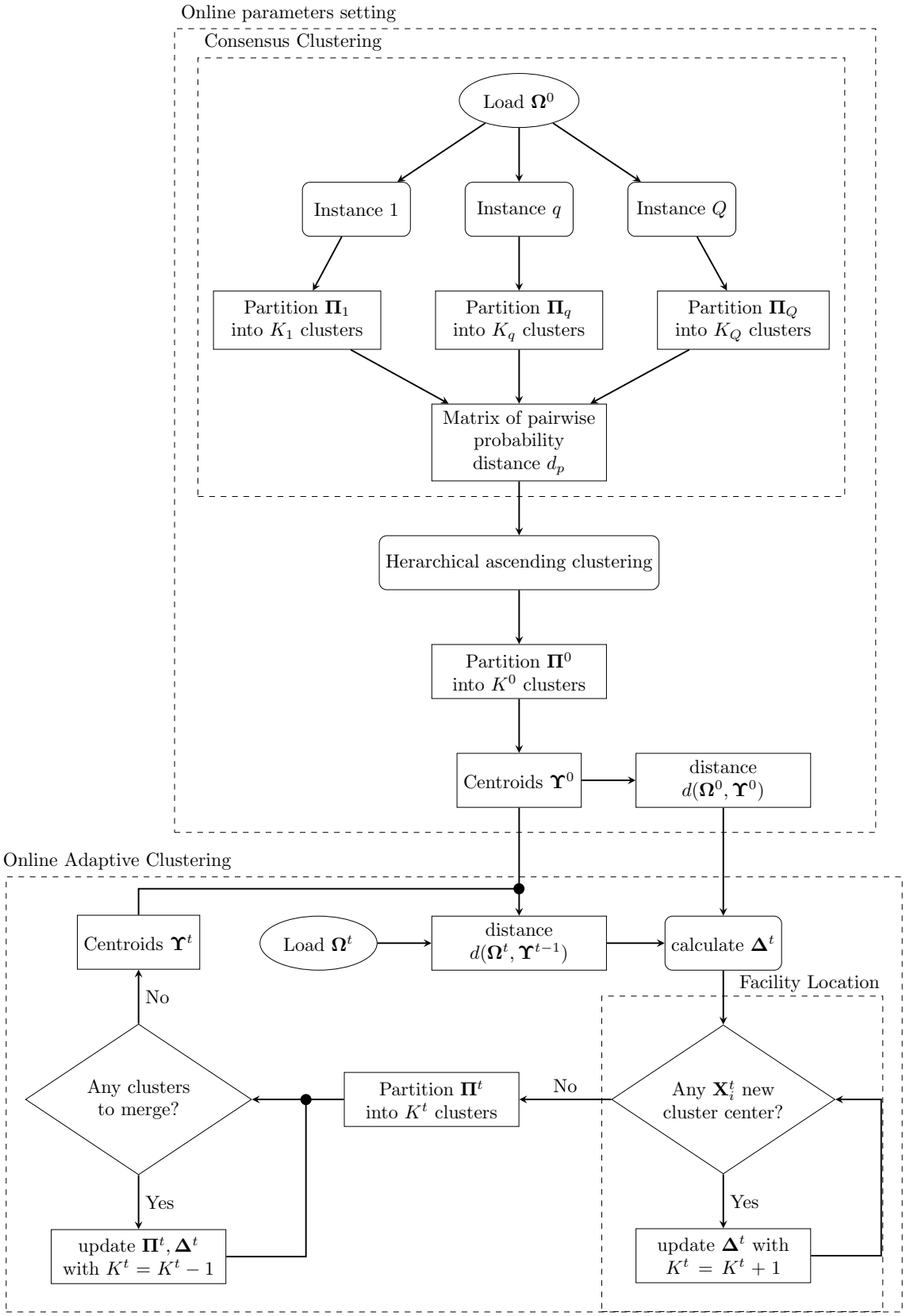


Figure 3.4: Overview of the online adaptive clustering algorithm.

on each line $\mathbf{d}_i^t$ of matrix $\mathbf{D}^t$ so that $1 \leq a_i^t \leq K^t$. The vector of assignment is used to generate a partition of the loads $\mathbf{\Pi}^t = \{\mathbf{P}_1^t, \ldots, \mathbf{P}_k^t, \ldots, \mathbf{P}_{K^t}^t\}$. A simple example with five loads $\{\mathbf{x}_1, \ldots, \mathbf{x}_5\}$ with a partition into two clusters could be, $\mathbf{\Pi} = \{\mathbf{P}_1 = \{\mathbf{x}_1, \mathbf{x}_3, \mathbf{x}_4\}, \mathbf{P}_2 = \{\mathbf{x}_2, \mathbf{x}_5\}\}$.

In the context of load profiling, patterns of consumption are the objects of interest. Hence, in order to compare patterns and not overall consumptions, the loads have to be 'aligned'. To do so, the loads $\mathbf{x}_i^t$ have to be normalized. A standard way, observed in the literature, is to divide each load by their reference power (i.e., maximum consumption over the period) to bind them to $[0, 1]$ [87].

3.3.2 A consensus clustering to set the online clustering parameters

Three parameters have to be defined to run the online adaptive clustering algorithm, (i) $\mathbf{\Upsilon}^0$, the first set of typical load profiles, (ii) K^0, the number of cluster centers at $t = 0$, and (iii) the distance matrix $d(\mathbf{\Omega}^0, \mathbf{\Upsilon}^0)$ at $t = 0$. Historical data are used offline to create the first partition $\mathbf{\Pi}^0$ and the listed parameters.

As explained in Section 3.2.3, the standard process of finding K is computationally costly and only a small part of the results computed is used afterwards. The alternative to this standard process is called consensus (or ensemble) clustering and consists in using all the partitions $\mathbf{\Pi}_q$ generated from instances q with different value of K_q to find a consensus partition $\mathbf{\Pi}^0$ into K^0 clusters with centroids $\mathbf{\Upsilon}^0$ [115]. The partitions $\{\mathbf{\Pi}_1, \ldots, \mathbf{\Pi}_q, \ldots, \mathbf{\Pi}_Q\}$ are combined into a probability distance $\mathbf{d}_p$

$$d_p(i, j) = 1 - \frac{\sum_{q=1}^{Q} K_q \delta(a_{qi}, a_{qj})}{\sum_{q=1}^{Q} K_q} \tag{3.8}$$

where δ is the co-occurrence matrix taking value 1 if i^{th} and j^{th} loads are assigned to the same cluster and 0 otherwise. The weighted sum of the co-occurrence matrices by the number of clusters K_q in each partition $\mathbf{\Pi}_q$ forms a probability that a pair of loads get assigned together.

The dendrogram of the HAC on the square matrix of distances $d_p(i, j)$ using Ward's criterion is then generated. The decision of how many clusters K^0 should be generated is then based on the expertise of the user in solving a specific problem. Subsequently, the initial typical load profiles, $\mathbf{\Upsilon}^0$ are calculated by the average of the loads in each cluster, and the initial distance matrix $d(\mathbf{\Omega}^0, \mathbf{\Upsilon}^0)$ is generated.

The benefits of consensus clustering are manifold, (i) all computed instances contribute to find K^0, and (ii) $\mathbf{\Pi}^0$ is a consensus between the different instances with less bias than each instance. We want to emphasize that the parameter-setting is the only input from the user to the algorithm. Hence, it is essential that the decision taken to set parameters is based on robust information as it will influence the evolution of the online adaptive clustering algorithm.

3.3.3 An online K-means clustering implementation

The parameters of the online adaptive clustering are set and we enter the online part of the algorithm corresponding to the bottom rectangle in dashed line in Figure 3.4. The iteration over time uses the core of the K-means algorithm to connect the time steps: (i) assignment of the loads in $\mathbf{\Omega}^t$ to the closest typical load profile in $\mathbf{\Upsilon}^{t-1}$, and (ii) update centroids by averaging loads from

time step t in each cluster. Hence, the calculation of $d(\mathbf{\Omega}^t, \mathbf{\Upsilon}^{t-1})$ connects consecutive time steps. The exponential smoothing of the distance matrix

$$\mathbf{\Delta}^t = \frac{\lambda^t d(\mathbf{\Omega}^0, \mathbf{\Upsilon}^0) + \sum_{\tau=1}^{t} \lambda^{t-\tau} d(\mathbf{\Omega}^\tau, \mathbf{\Upsilon}^{\tau-1})}{\sum_{\tau=0}^{t} \lambda^\tau} \tag{3.9}$$

is computed to transfer structural information from previous time steps. λ is the exponential smoothing coefficient and takes a value in $[0, 1]$, corresponding to how much information from the previous time step is transmitted to the next one. Hence, the assignment of the loads relies on the distance matrix from previous steps, which stabilizes the clustering. Implementation of the exponential smoothing handles the first characteristic, as loads are relatively stable and may evolve together if grouped in a same cluster. The typical load profiles are updated by averaging the loads, $\mathbf{x}_i^t \in \mathbf{P}_k^t$ from the latest loaded data $\mathbf{\Omega}^t$.

As mentioned in Section 3.2.1 metering data time series displays time dependency. To tackle this problem, the Euclidean distance used normally in K-means has been replaced by a Dynamic Time Warping (DTW) distance metric, which has proved to perform well to cluster time series [116, 117]. The distance metric calculated between any given load $\mathbf{x}$ and typical load profile $\mathbf{y}$ of the same length M is a dissimilarity index [118],

$$d(\mathbf{x}, \mathbf{y}) = \phi[\rho(\mathbf{x}, \mathbf{y})] d_{DTW}(\mathbf{x}, \mathbf{y}) \tag{3.10}$$

where the first order temporal correlation coefficient

$$\rho(\mathbf{x}, \mathbf{y}) = \frac{\sum_{m=1}^{M-1} (x_{m+1} - x_m)(y_{m+1} - y_m)}{\sqrt{\sum_{m=1}^{M-1} (x_{m+1} - x_m)^2} \sqrt{\sum_{m=1}^{M-1} (y_{m+1} - y_m)^2}}, \tag{3.11}$$

estimating the dynamic behaviors and the DTW distance

$$d_{DTW}(\mathbf{x}, \mathbf{y}) = \min_{r \in M} \left(\sum_{(i,j) \in \{1,\dots,M\}^2} |x_i - y_j| \right), \tag{3.12}$$

are balanced using the function ϕ

$$\phi(u) = \frac{2}{1 + \exp(u)}, \tag{3.13}$$

an adaptive tuning function [119]. DTW measures temporal similarities between two-time series by calculating the Euclidean distance between a point of a time series and all the points of the others to generate a distance matrix. A path called a warping path from the lower left corner to the upper right corner going through the lowest value is then defined through the distance matrix. The warping path can be constrained around the diagonal (i.e., a window around the diagonal of the distance matrix) by using the Paliwal window [120] to limit the shifting and speed up the algorithm as only the distances in the window have to be calculated.

3.3.4 Adaptivity: semi-online facility location

The last step toward a fully dynamic and flexible load profiling is to be able to handle the last characteristic, novelty, the apparition of unseen load behavior. In term of clustering, it means increasing the number of clusters (bottom right in Figure 3.4). To tackle this problem, we propose to use a probabilistic approach called facility location to adjust the number of clusters according to new unseen behaviors [121]. Facility location consists in setting a cost C_i^A proportional to the

distance d^t_{i,a_i} of each load $\mathbf{x}^t_i$ to the closest typical load profile $\mathbf{y}^{t-1}_{a_i}$ and a cost C^F to the generation of a new cluster center (i.e., facility) with $C^F \gg C^A_i$. C^A_i and C^F are combined to form a probability

$$p(\mathbf{y}^t_{k+1} = \mathbf{x}^t_i) = \min\left(\frac{C^A_i}{C^F}, 1\right), \tag{3.14}$$

that a load becomes its own cluster center. Hence, the farther a load is from the closest cluster center, the higher the cost of adding it to an existing cluster C^A_i and as a result the more probable that it becomes its own cluster center. Hence, facility cost limits clusters' range and should be empirically chosen depending on the clusters size and the number of clusters expected.

As the pattern of interest while doing load profiling is minimum a day, all the loads are transferred a the end of this period to the algorithm to form a block $\mathbf{\Omega}^t$, and not a load at a time [122]. Hence, the facility location is implemented in a semi-online manner as it evaluates the need for a new cluster center on each block $\mathbf{\Omega}^t$ and not a load at a time. A threshold γ_{min} of how many loads should exhibit a disruptive behavior to generate an extra typical load profile is set to reduce sensitivity. The process being probabilistic, 1000 Monte-Carlo simulations are run to obtain the distribution of the number of loads in $\mathbf{\Omega}^t$ above the threshold. The mode of the distribution is used to take the decision. If the threshold is reached, the farthest load from its closest typical load profile becomes its own cluster center. The distance matrix is then recalculated and checked for loads in $\mathbf{\Omega}^t$ that could be assigned to the new typical load profile. Finally $\mathbf{\Delta}^t$ is also updated with $K+1$ clusters.

The adaptive part also consists of reducing the number of clusters. Indeed, two cluster centers might converge, and the algorithm would ultimately merge them after many iterations. Meanwhile, it would create redundant typical load profiles. To avoid redundancy among the typical load profiles, a minimum distance d_{min} between two typical load profiles is defined. If $d_{min} > d(\mathbf{y}_u, \mathbf{y}_v)$ between two typical load profiles u and v, the two typical load profiles are considered similar and merged to form a single cluster using the average of their assigned loads as centroids. $\mathbf{\Delta}^t$ is then updated accordingly with $K-1$ clusters (bottom left in Figure 3.4).

3.3.5 Setting the parameters

The online adaptive clustering algorithm requires many parameters to be set empirically as the nature of the time series (e.g., electricity metering data, central heating district), the resolution and the preprocessing affect clustering. As mentioned previously the clustering can have a different objective depending on the application. Table 3.2 summarizes the parameters and which aspect of the clustering it tunes.

The influence of K^0 on the online clustering is mitigated by λ as the time increases thanks to the exponential forgetting. Beyond the first time steps, λ influences the amount of structural information transferred from one time step to another, a larger value of λ results in higher information persistence over time. C^F and d_{min} have opposite, but complementary actions as they respectively increase and decrease the number of typical load profiles. They control the radius of the clusters, a large C^F allows larger clusters thus a smaller K and a small d_{min} allows smaller clusters thus a larger K. Finally γ_{min} limits fluctuations of K over time.

3.4 Performance evaluation using simulated data

A synthetic dataset, where the number of underlying typical load profiles is controlled, is generated to evaluate the performance of the online adaptive clustering algorithm. We first describe how synthetic loads are generated, then the performance evaluation criterion and we finally present the results obtained.

Table 3.2: Influence of the parameters on the clustering process.

Parameter	Definition	Value	Influence
K^0	Initial number of cluster centers	$n \in \{2, \dots, N\}$	On the first iterations
λ	Exponential forgetting	$[0, 1]$	Smooths the clustering
C^F	Facility cost	Relative to C^A	Size (radius) of the clusters
d_{min}	Minimum distance between cluster centers	Relative to the expected number of clusters	Size (radius) of the clusters
γ_{min}	Number of disruptive loads needed to create a new cluster	$n \in \{2, \dots, 10\}$	Limits fluctuations of K

3.4.1 Data generation

As explained in Section 3.2.3, the performance evaluation of a clustering algorithm is dependent on the criterion chosen as the latent formation of typical load profiles cannot be unveiled. Hence, the only way to evaluate the true performance is to design a dataset that would simulate a realistic load behavior. The design of the synthetic data should prove that the online adaptive clustering algorithm can handle the three characteristics of a pool of loads: stationarity, non-stationarity, disruption, described in Section 3.3. The dataset consists in four slow changing (stationarity) and one disruptive typical load profiles (disruption) that produces 300 days of load data. They have been generated using nine typical load profiles with different behaviors at an hourly resolution sampled from ENTSO-E's data base [123]. The four slowly changing typical load profiles are generated as weighted averages

$$\mathbf{y}_k^t = \left(1 - \frac{t}{300}\right) \mathbf{y}_k^0 + \frac{t}{300} \mathbf{y}_k^{299} \tag{3.15}$$

of pairs $(\mathbf{y}_k^0, \mathbf{y}_k^{299})$ respectively the starting (Day 0) and ending (Day 299) profiles displayed by typical load profiles k as shown in Figure 3.5(a). The four typical load profiles $\boldsymbol{\Upsilon} = \{\mathbf{y}_1, \dots, \mathbf{y}_4\}$ are used to form $\mathbf{B^0} = \{b_1^0, \dots, b_{1000}^0\}$ with $1 \leq b_i^0 \leq 4$ random assignments of 1000 customers. Their daily typical load profile $\mathbf{y}_k^t$ is then individualized by adding multivariate Gaussian noise

$$\mathbf{x}_i^t = \mathbf{y}_k^t + \mathcal{N}(0, \boldsymbol{\Sigma}), \tag{3.16}$$

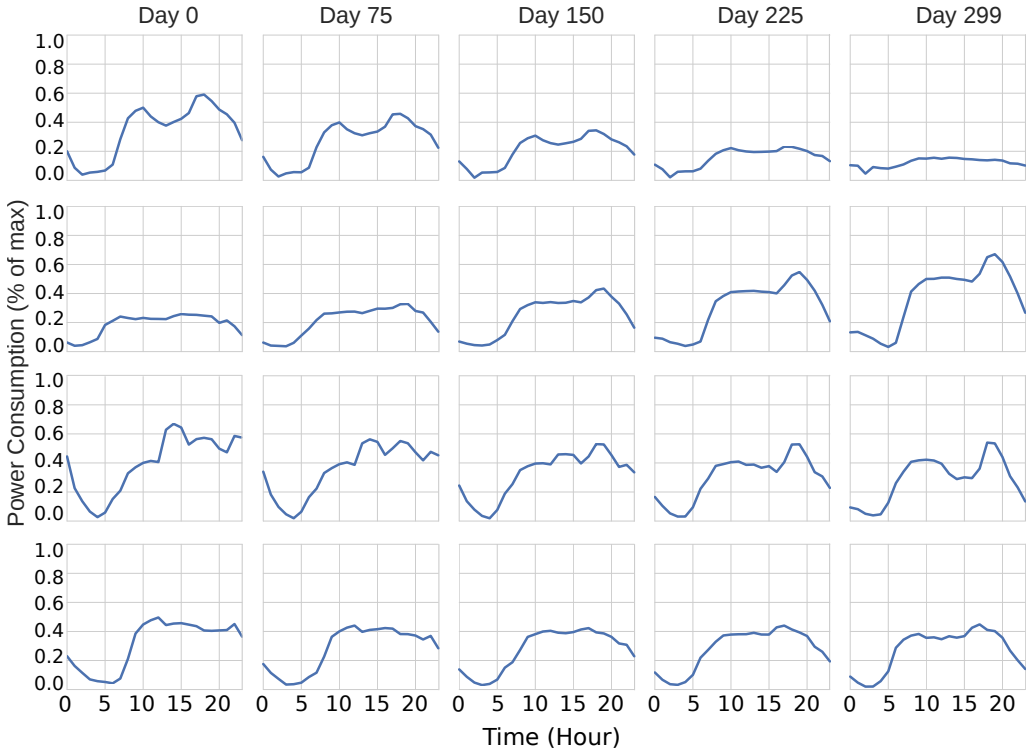
where the covariance matrix $\boldsymbol{\Sigma}$ is stochastically generated as

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2\rho_{11} & \dots & \sigma_1\sigma_l\rho_{1m} & \dots & \sigma_1\sigma_M\rho_{1M} \\ \vdots & \ddots & & & \vdots \\ \sigma_l\sigma_1\rho_{l1} & & \sigma_l^2\rho_{lm} & & \sigma_l\sigma_M\rho_{lM} \\ \vdots & & & \ddots & \vdots \\ \sigma_L\sigma_1\rho_{L1} & \dots & \sigma_L\sigma_m\rho_{Lm} & \dots & \sigma_L^2\rho_{LM} \end{bmatrix}, \tag{3.17}$$

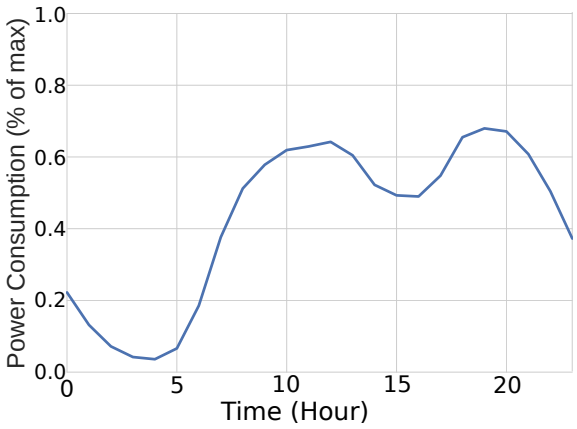
with $\sigma = \{\sigma_1, \dots, \sigma_{24}\}$ a normalized random vector of standard deviation and ρ a matrix of coefficient decreasing exponentially from the diagonal

$$\rho_{lm} = \exp\left(\frac{-|l - m|}{\tau}\right), \tag{3.18}$$

which adds some time shifting to the patterns. The daily load data were then normalized to $[0, 1]$ by dividing them by their reference power.



(a)



(b)

Figure 3.5: Synthetic data generation based of ENTSO-E load profiles. Morphing of the four typical load profiles (a) and the new load behavior appearing at day 200 (b).

To simulate the last characteristic (non-stationarity), 250 of the 1000 customers generated change from their slow transitioning typical load profile to the disruptive typical load profile (Figure 3.5(b)) at time step 100. The number of typical loads changes from four to five with $1 \leq b_i^{100} \leq 5$ and the number of clusters is expected to do the same and group the 250 customers. At time step 200, the fith typical load profile is remove and the 250 customers are return to their original typical load profiles with $1 \leq b_i^{200} \leq 4$. As the number of typical load profiles changes from five to four, the number of clusters is also expected to change from five to four.

3.4.2 Performance evaluation

We now have the ground truth $\mathbf{B}^t$ at each time step t to evaluate the performance. When the ground truth is known, standard classification evaluation can be applied. Here we opted for an information theory criterion called Normalized Mutual Information (NMI). For the two vector of assignment $(\mathbf{A}, \mathbf{B})$ it is

$$\mathrm{NMI}(\mathbf{A}, \mathbf{B}) = \frac{\mathrm{MI}(\mathbf{A}, \mathbf{B})}{\sqrt{\mathrm{H}(\mathbf{A})\mathrm{H}(\mathbf{B})}} \tag{3.19}$$

where H is the entropy

$$\mathrm{H}(\mathbf{A}) = \sum_{k=1}^{K^t} p(k) \log(p(k)), \tag{3.20}$$

defining the amount of disorder in assignment vector $\mathbf{A}$ with $K^t = \{4, 5\}$ depending on the time step. $p(k)$ is the probability that an assignment in $\mathbf{A}$ picked at random has value k. Cross-entropy or Mutual Information (MI)

$$\mathrm{MI}(\mathbf{A}, \mathbf{B}) = \sum_{g=1}^{G^t} \sum_{k=1}^{K^t} p(g, k) \log \left(\frac{p(g, k)}{p(g)p(k)} \right) \tag{3.21}$$

is basically the entropy of the joint probability between vectors of assignment $\mathbf{A} = \{a_1, \ldots, a_{1000}\}$ and $\mathbf{B} = \{b_1, \ldots, b_{1000}\}$. The joint probability $p(g, k)$ is the probability that a load is both from typical load profile g and is assigned to cluster k. It takes a value in $[0, 1]$ where 1 is a perfect match between $\mathbf{A}$ and $\mathbf{B}$.

3.4.3 Online clustering setup and results

This section aims at evaluating the online adaptive clustering algorithm. Hence, the consensus clustering is not performed on the synthetic data and the first clustering assignments $\mathbf{A}^0 is the ground truth \mathbf{B}^0$. The facility cost is fixed to $C^F = 100$, the number of points γ_{min} required to generate a new cluster is set to one, the exponential smoothing is $\lambda = 0.85$ and the minimum distance between cluster centers is $d_{min} = 0.07$.

Figure 3.6 presents in parallel the evolution of the NMI (solid green line), the number of clusters (blue dashed-dotted line) and the actual number of typical load profiles (red dashed line) over the test period.

First, looking only at the number of clusters and typical load profiles, we can observe that the online adaptive clustering algorithm is generating the correct number of clusters with only a slight delay of two days when the typical load profiles number is changing from five to four. Second, looking at the NMI the performance in term of classification is also high from day 0 to day 60 and from day 230 to day 299 as the NMI stays close to 1.0. Around day 60, the NMI decreases, which is probably a consequence of the information in partition Π^0 decaying with the exponential smoothing as well as a convergence between the typical load profiles (Figure 3.5(a)) resulting in some misclassifications. At day 100, a drop of the NMI to approximately 0.6 can be observed before it quickly increases to around 0.9 the next day and oscillates around 0.9 until day 200. At day 200, the NMI drops again down to 0.75 and comes quickly back to 0.95 the next day and increases steadily until 0.99 before it oscillates between 0.99 and 1.0. At the end of the test period, the NMI is back to 1.0, which means that $\mathbf{A}^{299} = \mathbf{B}^{299}$.

From the evaluation of the online clustering algorithm on synthetic data with known ground truth, we can be confident of its performance and test it on real-world datasets.

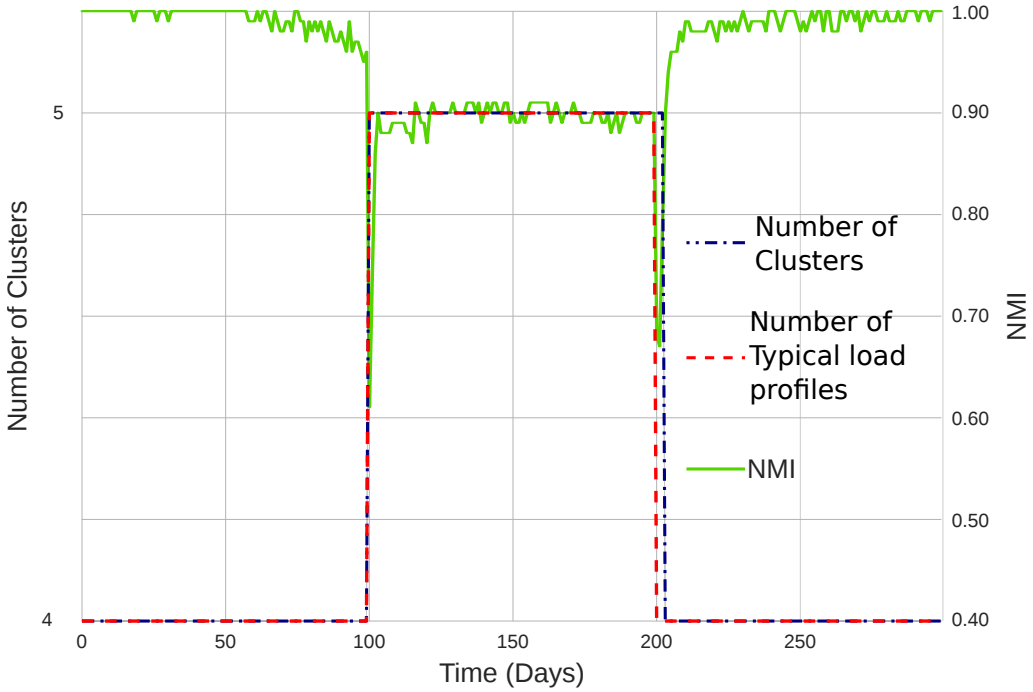


Figure 3.6: Mode of the NMI and the number of clusters over the test period.

3.5 Real-world data applications

In this section, the online adaptive clustering algorithm is tested on two real-world datasets. In Section 3.5.1 it is applied to electrical loads. In Section 3.5.2 we show that load profiling can be done with data provided by other sources than power systems by applying it to central district heating data.

3.5.1 Application on electricity metering data

Data description

Radius, the DSO in the Copenhagen area, provided a dataset that consists in a year ($T = 240$ days after removing missing data) of hourly power consumption data from $N = 13241$ customers. The customers metered are businesses, industries, and households with PV. The blocks $\mathbf{\Omega}^t$ consist of a day ($M = 24$) and the loads were preprocessed by dividing them by the peak over the period.

Implementation

The consensus clustering is implemented with a K-means algorithm using the dissimilarity index with d_{DTW}. The instances q are run with $K_q = \{10, \ldots, 100\}$ on day 0: 2015-01-12. The partition $\mathbf{\Pi}^0$ is obtained by cutting the dendrogram into ten clusters. The number of clusters intentionally underestimates the number of typical load profiles present in the dataset (approx. 20).

The online adaptive clustering algorithm starts with the following parameters: Clusters centers, $K = 10$; facility cost, $C^F = 950$; minimum number of customers to create a new cluster center, $\gamma_{min} = 5$; exponential smoothing coefficient, $\lambda = 0.85$ and the minimum distance between two cluster center $d_{min} = 0.13$. The online adaptive clustering is benchmarked against a standard

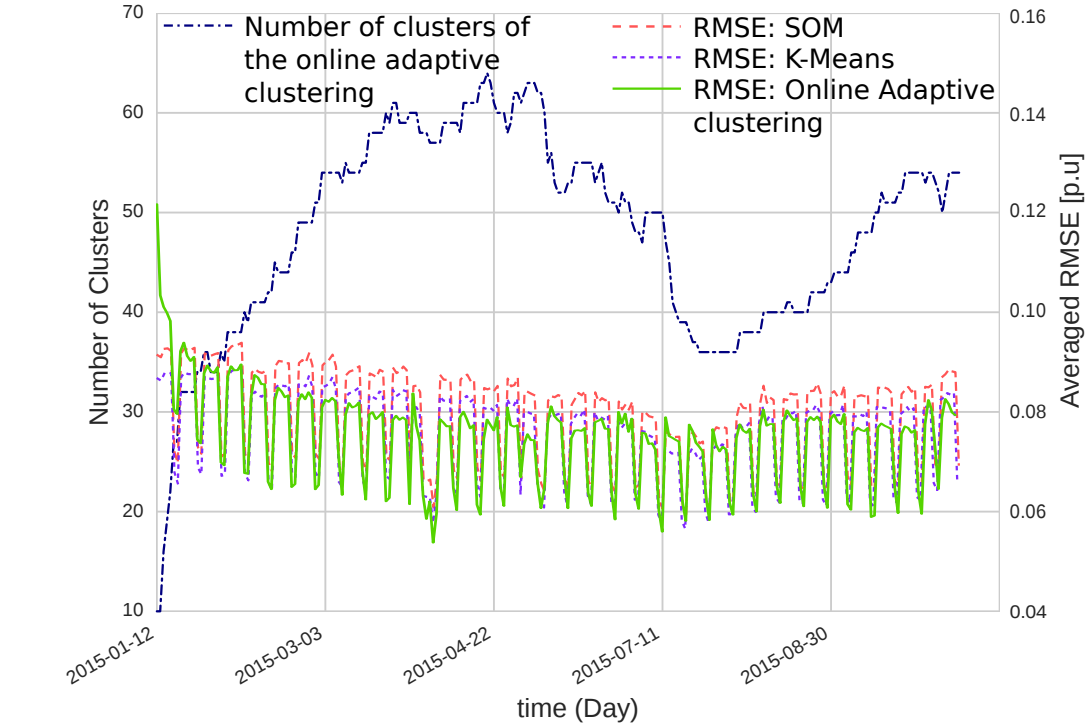


Figure 3.7: Number of clusters and Root Mean Square Error (RMSE) per unit of peak load, between daily loads and their assigned typical load profile over the test period for Benchmarks: SOM (7.9% RMSE overall score), K-means (7.5% RMSE overall score) and the online adaptive clustering algorithm (7.2% RMSE overall score).

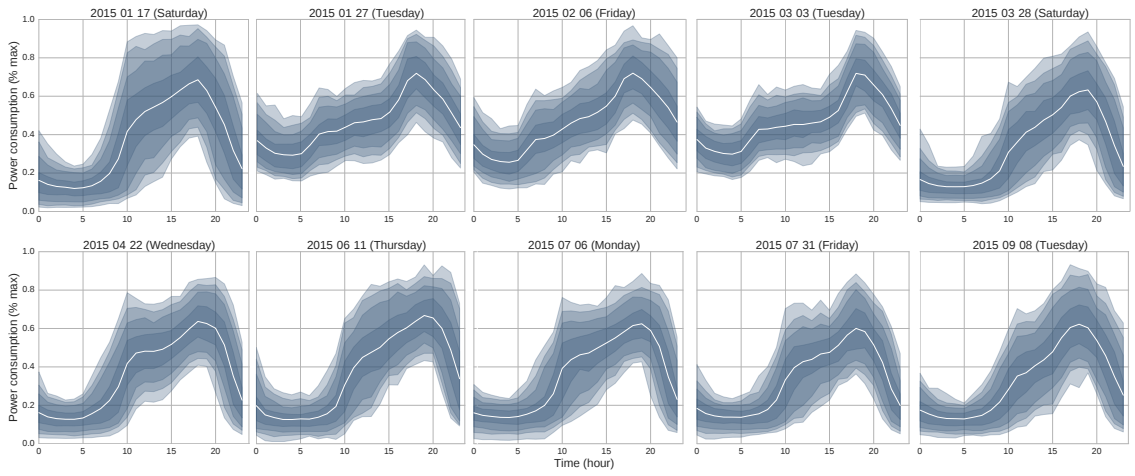


Figure 3.8: Evolution of cluster 15 along the test period.

K-means algorithm and a Self-Organizing Map (SOM) applied on the entire set of daily loads ($N \times T = 317784$) with $K = 50$ [87].

Results

The results of the computations are summarized in Figure 3.7 with the number of clusters and the averaged RMSE between daily loads and their assigned typical load profile for both the benchmarks and the online adaptive clustering algorithm. Looking at the number of clusters

generated by the online adaptive clustering algorithm, we observe first an increase up to 64 clusters around 2015-04-22, which later decreases to 36 during the summer and increases again to 54 in the fall. The weekly periodicity of the averaged RMSEs is explained by the activities of the customers in the study. Most of them are industries and businesses that are usually closed on weekends and generate more homogeneous patterns. After removing the 30 first days of convergence (2015-03-03), it appears that when the number of clusters is above 40, the online adaptive clustering beats the benchmarks consistently on the weekdays. Indeed, the facility location thresholds the error, which maintains the RMSE relatively stable between 0.08 (weekdays) and 0.06 (weekends). After 2015-03-03 the overall RMSE score of the online adaptive clustering algorithm is 7.2% against 7.5% and 7.9% respectively for the K-means and SOM. In terms of computational time, the consensus clustering (computed in parallel with 7 cores) took 15 minutes to be completed, and each iteration of the online adaptive clustering (on a single core) took approximately 1 minute 20 seconds (6 hours for 240 days) on a desktop computer equipped with an Intel Xeon CPU 3.50 GHz $\times$ 8 cores. As a comparison, the SOM was completed after 16 hours while the K-means took 20 hours.

These results highlight the benefits of having an online adaptive clustering process that estimates correctly the number of clusters in the data at time t and offers faster computation. The difference in terms of performance between the online adaptive clustering algorithm and the benchmarks would have probably been larger with a dataset displaying more non-stationary behavior like domestic loads involved in Demand Response (DR) programs.

An illustration of the evolution of a cluster over the period is given in Figure 3.8. The typical load profile shapes, as well as the bottom line, are changing over the period. The cluster consists mostly of restaurants, which are active daily from lunch to dinner time with a peak of activity at dinner time and along the year at high activity periods (Christmas holidays, Saturdays, and from April to August).

The application of the online adaptive clustering algorithm on electricity metering data has fulfilled expectations. Slow changing and fast-changing profiles are handled as expected and adjustment of the number of clusters to keep the same overall accuracy with a low computational cost at every iteration has proven to be more efficient than benchmarks.

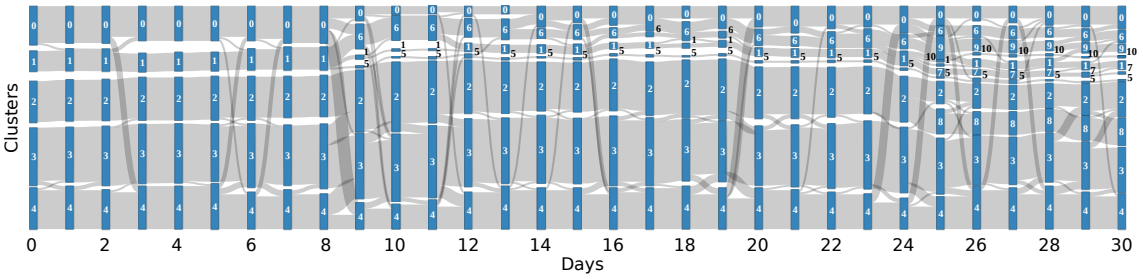


Figure 3.9: Sankey graph displaying the flow of buildings in the different clusters from one day to another during the test period.

3.5.2 Application on district heating

Data description

HOFOR, the central district heating operator in the Copenhagen area, provided a dataset that consists of heat consumption of 97 buildings over a month (31 days in Mach-April) at hourly

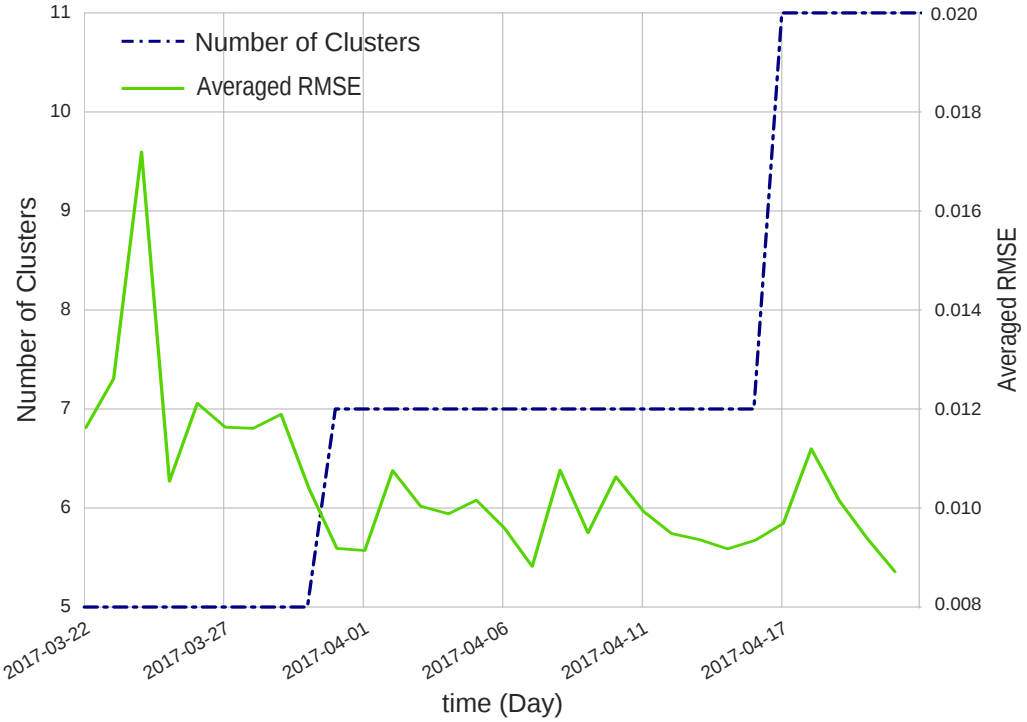


Figure 3.10: Number of clusters and Averaged RMSE between daily loads and their assigned typical load profile over the test period for the central district heating data.

resolution. A block $\mathbf{\Omega}^t$ groups a day of data ($M = 24$) and no preprocessing has been operated on the data as the consumptions lies within the same interval $[0, 1]$.

Implementation

The consensus clustering used the same modified version of the K-means algorithm that uses d_{DTW} as in Section 3.5.1 and instances q are run with $K_q = \{2, \ldots, 10\}$ on day 0. From the dendrogram of HAC, a partition $\mathbf{\Pi}^0$ into five clusters is generated.

The online adaptive clustering algorithm starts with $K = 5$ clusters, a facility cost $C^F = 7.5$, it needs only one building to create a new cluster center, the exponential smoothing is $\lambda = 0.85$ and the minimum distance between cluster centers is $d_{min} = 0.01$.

Results

An illustration of the evolution of the clusters' composition is given in (Figure 3.9) with a Sankey diagram of the 'flow' of customers between clusters and between time steps. The clustering is stable over time, besides small adjustments with some buildings changing cluster at each time step. At day 8-9, cluster 1 splits into cluster 1, 6 and 5. Cluster 5 is an outlier, isolated until the end of the period. Another splitting happens at day 24-25, cluster 1 splits into cluster 1, 10, 7. Cluster 9 takes most of the customers from cluster 1, but also some from cluster 0 and cluster 6. Figure 3.11 presents the evolution of the typical load profile formed by cluster 4. The shape changes slowly over time and confirm the stability observe in Figure 3.9.

The number of clusters, the RMSE between individual loads and their assigned typical load profile over the test period are presented in Figure 3.10. As seen in Section 3.5.1, the RMSE does not depend on the number of clusters. It first decreases and stabilizes around 0.01. We demonstrated

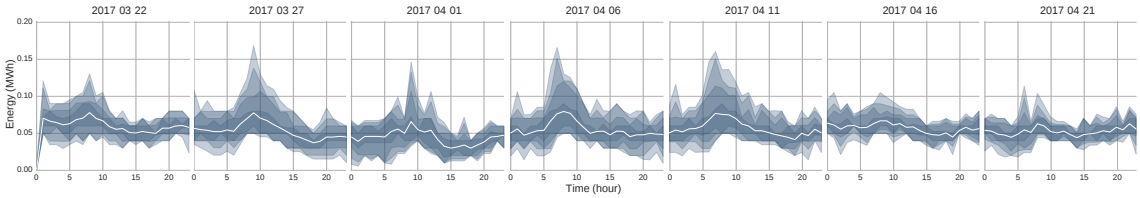


Figure 3.11: Evolution of cluster 4 along the test period.

here that the load profiling and more precisely our implementation of the online adaptive clustering for load profiling can provide insights beyond the power system field.

3.6 Conclusion

The online adaptive clustering algorithm is designed to take advantage of the stream nature of metering data. Its capacity to handle stationary or non-stationary loads as well as to adjust the number of clusters when previously unseen load patterns appear allows it to be run entirely autonomously and deliver typical load profiles with accuracy. Furthermore, it gives access to the typical load profiles shortly after data collection, and provides insights at a faster time scale than state-of-the-art load profiling that have to be rerun using the last data available to be updated. Hence, it provides the same functionalities than state-of-the-art load profiling with higher accuracy, but also new functionalities like demand flexibility assessment or anomaly (i.e., fraud, fault) detections. It is a first step towards fully dynamic load profiling as defined in Table 3.1 [15]. Radius is considering implementing a variant of the online adaptive clustering algorithm in their system. An implementation with mostly domestic loads has been done in collaboration with Quby (in the Netherlands). For domestic load implementation, the exponential smoothing is done with a days-of-the-week structure, as inhabitants' activities are usually based on a weekly schedule (i.e., a Monday will probably look like the previous Monday).

CHAPTER 4

Unsupervised Non-Intrusive Load Monitoring Methodology

Smart meters and the streams of data they generate give access to new information. Indeed, with high-frequency data collection (higher than 5-minute resolution) the load patterns of appliances are unveiled. This chapter discusses how much information about appliances behind the meter can be extracted from metering data. Non-Intrusive Load Monitoring (NILM) consists in separating individual appliances' power consumption signals from the aggregated power consumption signals. NILM state-of-the-art methodologies have shown promising results. However, they consist of supervised learning algorithms requiring data of a second or higher resolution and training dataset with individual appliance consumption signals. This is incompatible with real-world implementation. Hence, we propose an unsupervised NILM methodology and benchmark its performance against state-of-the-art methodologies as described in [Paper C]. We also present an application to the detection of Heat Pumps (HPs) as presented in [Paper D].

4.1 What is non-intrusive load monitoring?

The deployment of smart meters used to report power consumption to utilities and invoice customers generates a misalignment between what has been advertised and the real possibilities as explained extensively in Chapter 2. The data collected are not yet used to deliver the expected services despite the potential of improving energy efficiency. Machine learning algorithms, and precisely NILM, could play a key role in providing customers detailed information on their consumption without the deployment of additional hardware [23]. Moreover, what happens behind the meter could also be used to provide specific information (e.g., presence of HP, Electric Vehicle (EV), Electric Heating (EH)) to the utilities. Hart has described the concept of NILM in 1992, before smart meters deployment and before the computing capacities allowed such calculations [24].

The principle is relatively intuitive, since each of appliances ha a unique consumption signal that can be identified if observed individually (Figure 4.1b). Hence, as the overall consumption signal metered is the aggregation of all these individual consumption signals, it should be possible to separate them. The energy disaggregation problem can be formulated mathematically at time t as

$$x^t = \sum_{n=1}^{N} z_n^t + \varepsilon^t, \tag{4.1}$$

where x^t is the aggregated power consumption equal to the sum of N appliances power consumption signals z_n^t and ε^t are the residuals.

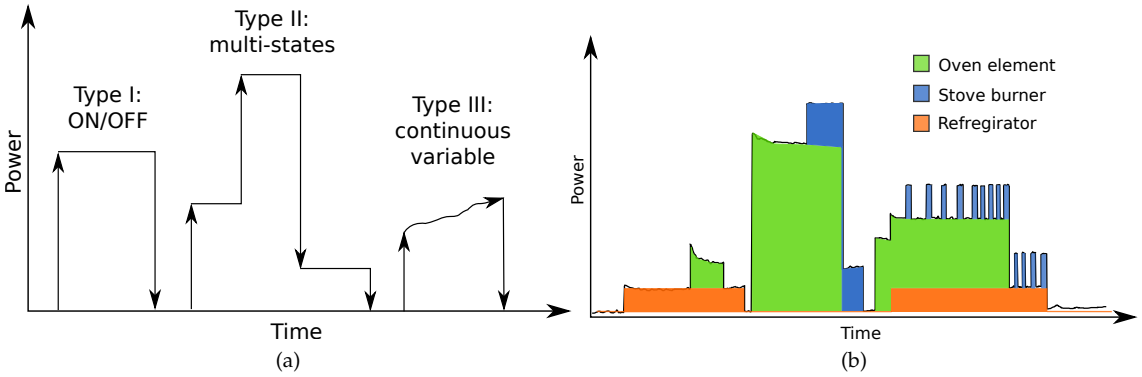


Figure 4.1: The different type of appliances observed in households (a) and an example of disaggregation of the overall signal into individual appliances (b).

Appliances can be categorized into four groups: (Figure 4.1a) (i) Type-I, simple two states, ON/OFF, appliances, e.g., kettle; (ii) type-II, multi-state appliances with a finite number of states, e.g., washing machine; (iii) type-III, continuous variable appliances whose power consumption has an uncountable number of states, e.g., dimmable light; and (iv) type-IV, appliances that are always active, also called vampire appliances, e.g., router [124]. Disaggregating even a small number of appliances is not a trivial task when considering the possible complexity of some appliances and their combination in the aggregated signal. Therefore a possible strategy would be looking for more data to discriminate appliances with close states more accurately. Real and reactive powers can be used to form a two-dimensional space, extra features like the harmonics on different signals, current and voltage can also be included in the model [125]. External features like time of the day, day of the year or weather information can also be used to obtain more information concerning the inhabitants' behavior [25].

The last distinction to make considering the NILM algorithm is the distinction between supervised and unsupervised learning algorithm. Supervised learning algorithms use individual load signal of each appliance to learn the characteristics of their states, power amplitude, operation time, steady or transient signature [126]. Using this collection of information, the states can be traced back to the overall consumption signal.

4.2 State-of-the-art methodologies

State-of-the-art methodologies are variations of Hidden Markov Models (HMMs) and Artificial Neural Networks (ANNs), which are supervised learning methodologies. The following section gives an overview of both methodologies and discusses their limitations.

4.2.1 Factorial hidden Markov models

HMMs rely on the idea that a system, called finite-state machine, generates observations $\mathbf{O}$ forming a time series $\{O^1, \ldots, O^T\}$, which is the materialization of hidden states $\mathbf{S} = \{S^1, \ldots, S^T\}$ (i.e., unobservable) following a Markov process (Figure 4.2b) [125]. A Markov process is a stochastic model that describes the change of states of a system probabilistically. For example, a single appliance of type-II with three states: (i) ON, (ii) OFF, and (iii) standby would generate the transition matrix $\mathbf{A}$, that can be illustrated by a chart as presented in Figure 4.2a. In addition

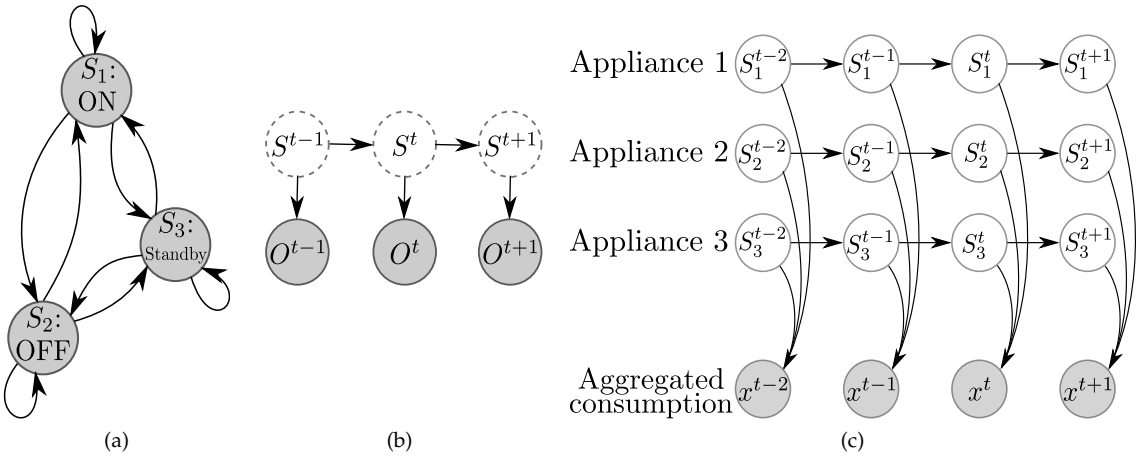


Figure 4.2: State transition model of a three-states appliance (a), representation of Markov chain (b), representation of a Markov chain with three appliances (c) [127].

to the arrows representing the transition from one state to another, probabilities are associated. The underlying set of states $\{S^1, \ldots, S^t, \ldots, S^T\}$ forms a Markov chain. An interesting property of Markov chain is that a state is independent of the past states [128]. In the case of NILM, the objective is to obtain a set of Markov processes (i.e., one per appliance) to disaggregate the overall consumption pattern. Factorial Hidden Markov Model (FHMM) is a generalization of HMM where the states of each appliance are factorized into a Markov process [129]. Figure 4.2c is a schematic representation of how the states $\{S_1^t, S_2^t, S_3^t\}$ of three appliances at each time step t contribute to the observed overall electricity consumption x^t.

The states and transition matrix are estimated using the Baum-Welch algorithm, a variation of the Expectation Maximization (EM) algorithm [127, 130]. The EM algorithm is doing iteration over two steps: (i) The E-step, using the parameter estimations to compute probability distributions over the hidden states, and (ii) The M-step, using the obtained probability distribution to maximize the expected log-likelihood of the parameters and adjust them accordingly [129]. The E-step is intractable due to the complexity of combining all the possible hidden states to generate the probability distribution.

With HMM, the classical algorithm used to obtain the chain sequence is the Viterbi algorithm [131]. From the observations, which are distorted information, the algorithm calculates the maximum likelihood estimates of the successive states of a finite-state machine. The complexity of the Viterbi algorithm is linear along a Markov chain. However, its complexity increases exponentially with the number of appliances considered in the model. This is because the probability of each transition must be evaluated at each time step. Since the variables correspond to each combination of all the appliances' states, the number of transitions is exponential to the number of appliances. Hence, the Viterbi algorithm suffers from the same problem as the E-step of the EM algorithm. It becomes intractable with a large number of appliances. In [132], a sparse Viterbi algorithm is implemented to reduce the number of states to consider and make it tractable.

In [127], a standard FHMM is implemented using the Baum-Welch algorithm to estimate the states and a Viterbi algorithm to obtain the Markov chains. In [133] variations of FHMM implementation are the additive FHMM, where each HMM emits an unobserved real-value, and the difference

FHMM, where the observation are the differences in aggregate power between consecutive time steps. An explicit duration HMM with differential observations to overcome that HMM cannot model the operation time of each state is proposed in [134]. In [128], a conditional FHMM integrating additional features, like the time of the day or factorial semi-Markov models to estimate operation time of appliances is presented. Both approaches are also combined into a conditional factorial semi-Markov model to the benefit of both characteristics. In [135], the difference FHMM proposed in [133] is incorporated into a graphical model. In [136], application of FHMM is done on a large campus building.

4.2.2 Artificial neural networks

ANNs have developed branches into different parts of supervised data analysis (i.e., regression, classification). As NILM can be assimilated to a complex classification task or a regression, ANNs intuitively became a good candidate to solve NILM problem. Depending on their architecture, ANNs can operate different tasks. ANNs mimic the way biological neural networks in the brain are organized, connected and activated. Neurons are connected by dendrites at the end of which are synapses. Inside neurons, the information is transported *via* electric impulse. At the synapse level, the electric impluse is converted into a chemical signal and its intensity is converted into concentration of neurotransmitter molecules. On the surface of the neuron receiving the signal, there are receptors to the neurotransmitter molecules. When the summation of all the receptors reaches a certain threshold, the receiver neuron is activated and emits an electric impulse to the dendrites.

ANNs basic structure is composed of: (i) an input layer, where the signal enters the network, (ii) one or many hidden layers, which form the depth of the network, (iii) the output layer generating the results (Figure 4.3a). In each layer, a certain number of neurons called units are laid. When designing ANN three aspects are important: (i) The architecture of the network (number of layers, number of units per layer), (ii) the type of activation functions to use in the different layers, and (iii) the training data. The activation function can be chosen according to the task to complete (Figure 4.3b). For example, linear activation functions are used for forecasting while logistic activation functions are preferred for classification. The combination of layers and units allows learning complex non-linear relationships between features. The training of ANNs consists of adjusting the weights of the network, represented with arrows in Figure 4.3a, so that the output of the network fits the training output data using a corresponding input data. To do so, training dataset, with both the input data and the corresponding output data, is needed. The input data enters the network and the information is passed through the network according to the weights. On the output layer, the error between the output generated by the ANN and the exact output data from the training set is calculated. The backpropagation algorithm is then used to distribute the responsibility of the error to the different weights and adjust their value accordingly [137].

The simplest ANN structure observed in the literature is the multilayer perceptron that consists of the input layer, several fully connected hidden layers, and the output layer [26, 138, 139]. Based on the multilayer perceptron, autoencoders are ANNs that take a signal as input, concentrate the information into a code layer (hidden layer with fewer units than the others) and decode it to match the output [26, 140]. Autoencoders are efficient at removing noise from a signal. In the context of NILM, the signal would be the consumption of a specific appliance, and the noise the consumption from the other appliances. The emergence of ANN architectures specialized in the

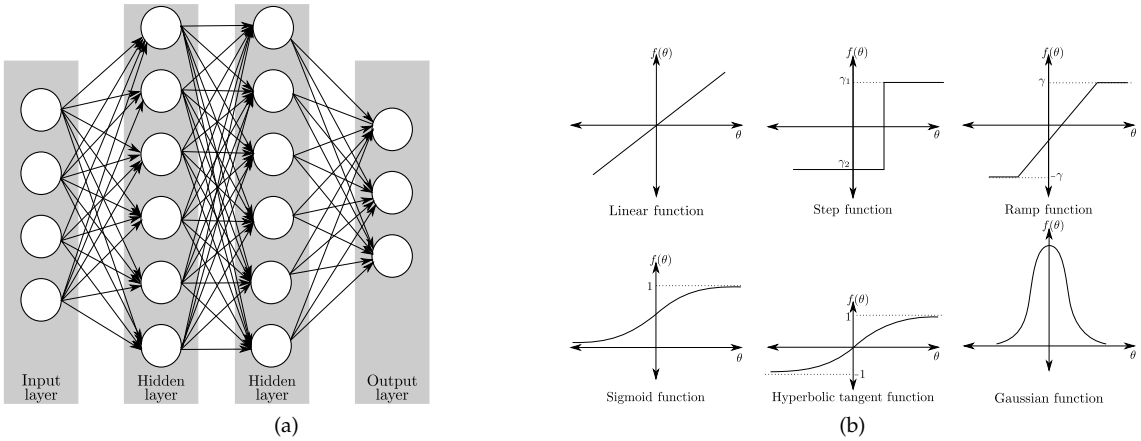


Figure 4.3: Example of structure of an ANN (multilayer perceptron) (a) and example of activation functions used with ANNs (b).

analysis of time series called Recurrent Neural Networks (RNNs) allowed to process online time series [141]. The most advanced ANN encountered in the literature, RNN with Long Short Term Memory (LSTM), is the most used architecture [26, 140, 142]. LSTM are specialized time series units that regulate RNNs learning gradients. Hence, it adapts the coefficients of the network to changing data.

The main problem of ANNs is that their construction and training is arbitrary [132]. Indeed, the design of ANNs is empirical and it is difficult to evaluate *a priori* what exact structure the network should take. Furthermore, the training requires labeled data, and the performance and generalization ability to unseen data of ANNs depend on the data observed during training. For NILM using state-of-the-art methodologies, the collection of high-resolution data for both individual appliances and aggregated consumption was the limiting factor [126, 143]. Dataset like REDD [25], UK DALE [144], Blued [145], AMPds [146] made training possible. The problem of NILM implementations is generally their lack of pragmatism towards what is possible to implement with real-world data and what information are relevant for customers or utilities [147]. The implementations focus on disaggregating the most accurately down to the smallest appliance, as a machine learning problem and not an engineering problem.

4.3 An unsupervised non-intrusive load monitoring methodology

In contrast to state-of-the-art NILM methodologies, we propose an unsupervised NILM methodology that combines, (i) a sparse signal approximation using an Orthogonal Matching Pursuit (OMP) algorithm, (ii) a Gaussian Mixture Model (GMM), and (iii) a community detection. They are materialized with dashed lined rectangles in Figure 4.4. The overall structure of the algorithm is presented in Figure 4.4.

4.3.1 Power signal sparse approximation

A power signal sparse approximation consists in approximating the aggregated power consumption using a set of functions φ. The observation of the load behavior and the categorization into four types of appliances shows that most appliances have non-transient states (i.e., type-I, type-II and type-IV

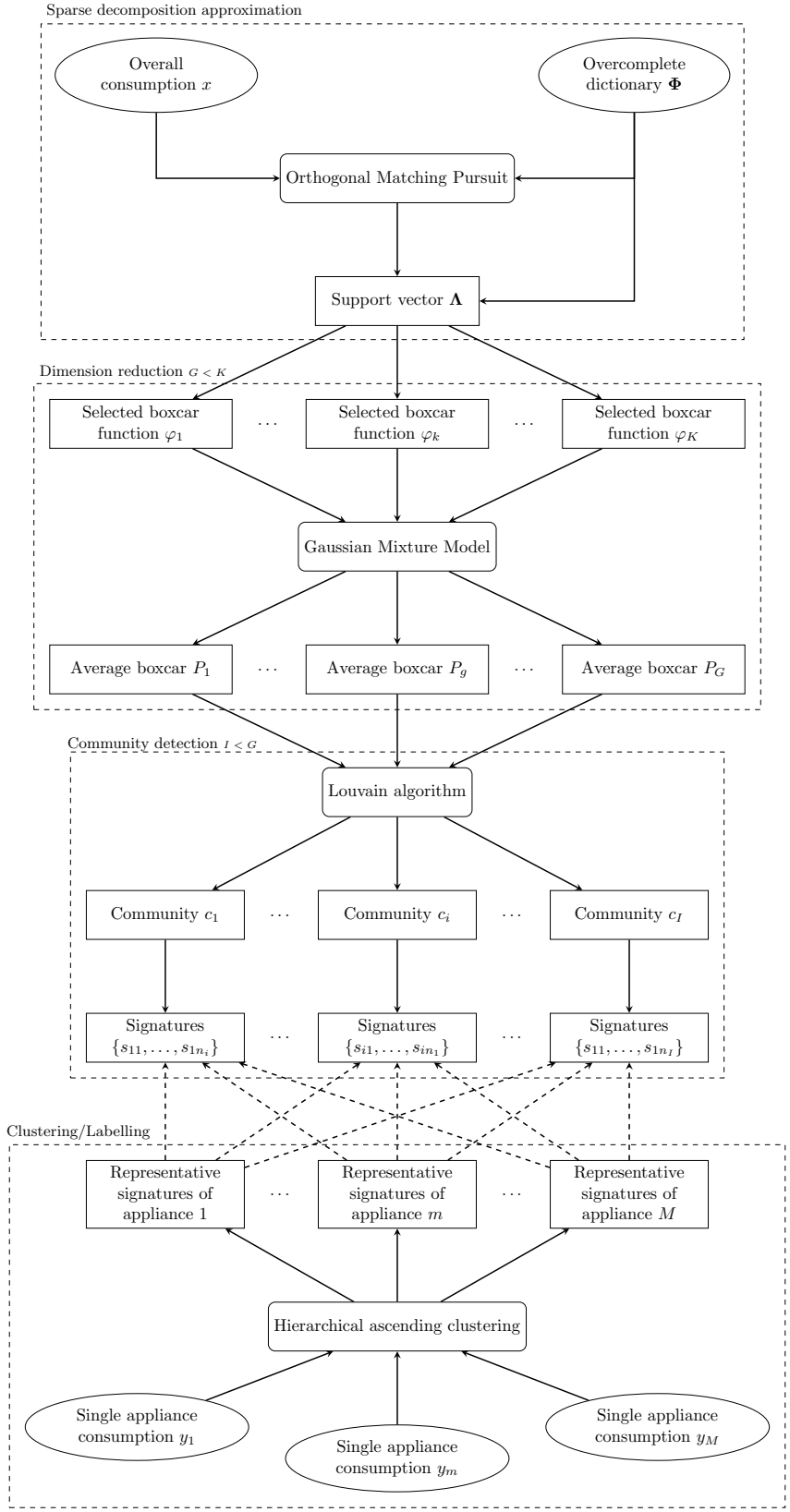


Figure 4.4: Overview of the OMP-GMM-community detection unsupervised NILM algorithm.

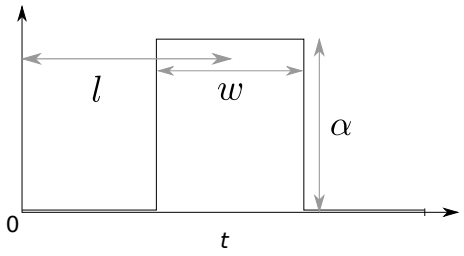


Figure 4.5: Representation of the general shape of boxcar functions.

appliances). Hence, load patterns can be approximated accurately using translation-invariant boxcar functions [148]. Each boxcar function takes the form

$$\varphi^t_{l,w} = \frac{1}{\sqrt{w}}\,\Pi^t_{l-w/2,l+w/2}, \tag{4.2}$$

where l is the boxcar translation and w the width and $\Pi^t_{a,b} = H(t-a) - H(t-b)$ with H a Heaviside step function. Figure 4.5 illustrates the general shape of a boxcar function and the different parameters involved in its design.

In the sparse approximation framework, (4.1) can then be written

$$x^t = \sum_{k=1}^{K} \alpha_k \varphi^t_k + \varepsilon^t \tag{4.3}$$

where α_k is the activation coefficient associated to function φ_k and takes value in $\mathbb{R}_{\geq 0}$. It can be written in a vectorial form as, $\mathbf{x} = \boldsymbol{\alpha}\,\boldsymbol{\Phi} + \boldsymbol{\epsilon}$ where $\mathbf{x}$ is the load vector for all time indices, $\boldsymbol{\Phi} = \{\varphi_1, \ldots, \varphi_k, \ldots, \varphi_K\}$ a dictionary of boxcar functions, $\boldsymbol{\alpha}$ is the vector of coefficients α_k and $\boldsymbol{\epsilon}$ a vector of residuals.

The dictionary $\boldsymbol{\Phi}$ is said to be overcomplete as all the possible boxcar functions are generated ahead of the approximation, but only a subset $J \ll K$ is used during the approximation [149–151]. Hence, $\boldsymbol{\alpha}$ is a sparse matrix as most of the coefficients are equal to zero. The non-zero entries in $\boldsymbol{\alpha}$ are gathered in a support vector Λ that restricts the dictionary $\boldsymbol{\Phi}$ to the subset $\boldsymbol{\Phi}_\Lambda$. The aim is to reach a degree of sparsity that eliminates transient states or internal states fluctuation. To do so, the changes in power consumption are detected using a fixed or adaptive threshold (i.e., minimum variation of power).

Direct approaches to sparse approximation, like combinatorial optimization, are complex and require important computing resources. Matching Pursuit (MP), an iterative process using a greedy algorithm, can be implemented instead [152]. MP finds iteratively the local optimal solution that makes an acceptable approximation to the global optimal solution. OMP outperforms MP by updating at every iteration all the activated coefficients simultaneously in generating the orthogonal projection of the selected functions [153]. Hence, it iteratively corrects the coefficients associated with each function. Nevertheless, OMP requires more computational resources as a consequence of the additional calculations.

The OMP algorithm proceed as follow: At iteration $j = 0$, the residuals r_0 are equal to $\mathbf{x}$, the coefficients in $\boldsymbol{\alpha}_0$ are set to 0 and subsequently Λ_0 is empty. At each iteration j, a single element k_j from $\boldsymbol{\Phi}$ that maximizes

$$k_j = \underset{k}{\operatorname{argmax}}\, ||r_{j-1}\varphi_k||_2, \tag{4.4}$$

is selected and corresponds to the widest boxcar function that can fits under the residuals at $j-1$. Subsequently, the sparse support vector includes the function k_j

$$\Lambda_j = \Lambda_{j-1} \cup k_j, \tag{4.5}$$

The coefficients $\boldsymbol{\alpha}_j$ are computed as least square estimates, i.e.,

$$\boldsymbol{\alpha}_j = \underset{\alpha}{\operatorname{argmin}} ||\mathbf{x} - \boldsymbol{\alpha}_{\Lambda_j} \boldsymbol{\Phi}_{\Lambda_j}||_2 = \boldsymbol{\Phi}_{\Lambda_j}^{\dagger} \mathbf{x}, \tag{4.6}$$

where $\boldsymbol{\Phi}_{\Lambda_j}^{\dagger}$ is the Moore-Penrose pseudo-inverse of $\boldsymbol{\Phi}_{\Lambda_j}$ [154]. The approximated signal $\widehat{\mathbf{x}}_j$ at iteration j is then calculated with $\widehat{\mathbf{x}}_j = \widehat{\mathbf{x}}_{j-1} + \alpha_{k_j} \varphi_{k_j}$ where α_{k_j} is the activation coefficient of φ_{k_j}. Finally, the residual is updated as $r_j = \mathbf{x} - \widehat{\mathbf{x}}_j$.

The output of the last iteration is the sparse support vector Λ_J restricting the overcomplete dictionary to $\boldsymbol{\Phi}_{\Lambda_J}$.

4.3.2 Dimension reduction

After sparse signal approximation, combinations of boxcar functions form the signature of an appliance. Over the repeated apparitions of the signature, the shapes of the boxcar functions used are slightly different. It can be because of the other appliance active at this moment or small variation in the signature. Nevertheless, these boxcar functions exhibit a strong cross-time dependency. However, a consequence of the overcomplete dictionary is that the boxcar functions selected by the OMP are mostly used only once. The community detection exploits the cross-time dependency of boxcar functions to generate the multi-state signatures. To identify boxcar functions that are close to each other as the same one, create redundancy and reinforce cross-time dependency between boxcar functions generated by the same appliance, boxcar functions are clustered. Hence, the clusters are used to form the communities. The clustering is implemented on the parameters $(\boldsymbol{\alpha}, \mathbf{w})$ that describe the shape (power amplitude, operation time) of boxcar functions when removing the translation l. The shapes of clusters in the 2D space $(\boldsymbol{\alpha}, \mathbf{w})$ are unknown *a priori*. A GMM generates clusters based on the fitting of Gaussian distribution on the points in the 2D space, which can be round or ellipsoidal depending on the structure of the covariance matrix [130]. Selected boxcar functions in Λ_J have coordinates $\zeta_k = (\alpha_k, w_k)$ and $\boldsymbol{\zeta} = \{\zeta_1, \ldots, \zeta_K\}$. A mixture of G clusters[1] is then defined as

$$p(\boldsymbol{\zeta}) = \sum_{g=1}^{G} \phi_g \, \mathcal{N}(\boldsymbol{\zeta} \mid \boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g), \tag{4.7}$$

where $\boldsymbol{\mu}_g$ is the mean, $\boldsymbol{\Sigma}_g$ is the covariance matrix and ϕ_g is the mixture coefficient of the g^{th} cluster. The total probability density function integrates to one with $\phi_g \geq 0$ and $\mathcal{N}(\boldsymbol{\zeta} \mid \boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g) \geq 0$. The mixture coefficients ϕ_g must satisfy

$$\sum_{g=1}^{G} \phi_g = 1 \quad \text{and} \quad 0 \leq \phi_g \leq 1. \tag{4.8}$$

Boxcar functions belonging to the same cluster are tagged as such and considered the same in the community detection. The performance of community detection depends on the number of clusters G. Hence, G is determined empirically with backward and forward fine-tuning between the community detection and the GMM.

[1] despite GMM being a classification approach, we will abusively use the term 'cluster' for simplicity.

4.3.3 Community detection

Graph theory analyzes relationships (connections) between objects in a network (graph). A graph $\Gamma(V, E)$ is a representation of pairwise relations between objects, where interactions are expressed with a set of vertices (i.e., boxcar functions) V and a set of edges (cross-temporal dependencies) $E = \{e(u, v) \ : \ u, v \in V\}$ [155]. Edges can be oriented in a direction or undirected (or bidirectional), thus each pair (u, v) results in $e(u, v) = e(v, u)$. In this work, the edges are undirected.

Community detection consists in forming strongly interconnected subsets from the graph [156]. Here, a community is a set of boxcar functions that have a strong cross-temporal dependency as they appear repeatedly close in time.

A weight sampled from a Gaussian distribution is assigned to each edge $e(u, v)$

$$\omega_{u,v} = \exp\left(\frac{-(t_u - t_v)^2}{2\sigma^2}\right), \tag{4.9}$$

where σ is the scaling parameter and t_u and t_v are respectively the position of the center of boxcar u and v. It measures the strength of the cross-temporal dependency between two boxcar functions u and v.

In practice, the community detection algorithm forms disjoint groups $C = \{c_1, \ldots, c_i, \ldots, c_I\}$ that cluster boxcar functions with strong cross-temporal dependencies. The core of the community detection process relies on the definition of the objective function that controls the aggregation into communities. The objective function determines the notion of community structure as groups of vertices with better internal connections than external [157]. The most known objective function is the modularity Q of the partition

$$Q = \frac{1}{2m} \sum_{c_i \in C} \left[\Sigma_{in}^{c_i} - \frac{\Sigma_{tot}^{c_i}}{4m^2}\right], \tag{4.10}$$

where m is the sum of all the weights in the network, $\Sigma_{in}^{c_i}$ is the sum of the weights from the internal edges of community c_i, and $\Sigma_{tot}^{c_i}$ the sum of the weights from the edges incident to a vertex in community c_i [158]. Q takes values in $[-1, 1]$ where a high value is defined as a high density of connections between the vertices in the same community and sparse or less dense connections between the vertices of the other communities.

Computing the modularity of the communities is an NP-complete optimization problem and thus impractical. The Louvain method is a heuristic approach, that uses an iterative greedy algorithm, inspired by Hierarchical Ascendant Clustering (HAC), which solves the problem in $O(n \log n)$ [156, 157]. As in the HAC, each vertex u starts as its own community and then the gain of taking u from its community and moving it into c_i of a neighbor vertex v is

$$\Delta Q_{u,c_i} = \frac{\sum_{v \in c_i} \omega_{u,v}}{2m} - \frac{\Sigma_{tot}^{c_i}\, \omega_u}{2m^2}, \tag{4.11}$$

where, ω_u is the sum of the weights of the edges incident to the vertex u, $\sum_{v \in c_i} \omega_{u,v}$ is the sum of the weights of the edges from the vertex u to vertices in community c_i (only one v at this stage). u is assigned to the community that maximized $\Delta Q_{u,c_i}$ and only if it is positive. This stage is repeated until stability of the partition is reached, i.e., when no individual move could improve the modularity [156]. The second step builds a new network from the communities obtained at the previous stage by summing the weights of the edges between vertices of the communities. Then the first step is reapplied on the network and so on until no improvement on the modularity is observed.

4.3.4 Labelling using clustering

A post-processing labeling of the signatures in each community is necessary as the disaggregation is unsupervised. An appliance can exhibit many signatures that correspond to the different programs/cycles of an appliance (i.e., cycles of a fridge or programs of a washing machine). To obtained patterns of typical behaviors, a clustering process is implemented on the activation events extracted from single appliance consumption signals. A training set of single appliance consumption signals, completely disjointed in time from the test set, is defined to extract activation events. The sparse approximation is computed on each activation event. The single activation events are aligned to all start at $t = 2$. As they have different durations, the standard Euclidean distance cannot be used in this clustering process. Dynamic Time Warping (DTW) is presented in Section 3.3, and it is used to measure the distance between time series that are shifted in time but is also convenient to measure the distance between time series of different length [159]. In this specific context, DTW allows us to compare time series, which have different lengths but still could have similar patterns. For each appliance, the pairwise DTW distance matrix of all the activation events is computed, then a HAC algorithm using Ward's criterion on the DTW distance matrix is computed. From the dendrogram of the HAC, the best partition into $3 \leq K \leq 6$ clusters is generated and the representative signatures are generated as averages over each cluster.

The labeling of the activation events from each community is done by calculating their DTW distances to the set of representative signatures of each appliance. The label of the closest representative signature is then assigned ot the activation event. The labeling is a mapping process between labels and the shape of the signature. Thus it needs *prior* information about how the signatures of the appliance of interest look like. In other words, labeling cannot be done in a completely unsupervised manner.

4.4 Performance evaluation

The algorithm performance is evaluated on the UK-DALE dataset [144] and benchmarked against the results obtained by [160] and [26] using respectively FHMM and RNN with LSTM on the same dataset. The dataset consists in power consumption data of one household over three months with five appliances: dishwasher, washing machine, kettle, microwave, and fridge, with a data resolution of 6 seconds (ten samples per minute). The details of the implementation are given in Section 4.4.1, Section 4.4.2 presents the performance metrics and the results are analyzed in Section 4.4.3. Afterward, we evaluate how the change in data granularity influences the performance of our algorithm. This is done by changing the resolution of the data from 6 seconds to 1 minute. State-of-the-art methodologies' performance is known to be highly dependent on the data resolution, which is not compatible with real-world implementation. The results are analyzed in Section 4.4.4.

4.4.1 Implementation

The dataset used to demonstrate the performance covers three months. In practice, it would take a large amount of memory and would be unpractical to generate the overcomplete dictionary on such an extensive dataset. Hence, the dataset is cut into batches of maximum 1000 time steps, and the corresponding overcomplete dictionary is generated and represents already 7.4 GB. The way batches are generated influences OMP performance as it performs the approximation over

each batch independently and the approximation of the signal is relative to the peak in each batch. Batches have different lengths and the consumption time series is cut in periods of lower consumption (less than 70 W), which are materialized with vertical gray lines in Figure 4.6. The batches length ranges from 30 seconds to 1 hour 20 minutes, with an average of 40 minutes. A low pass filter assimilating any consumption equal or lower to 5 W as null is implemented on all data.

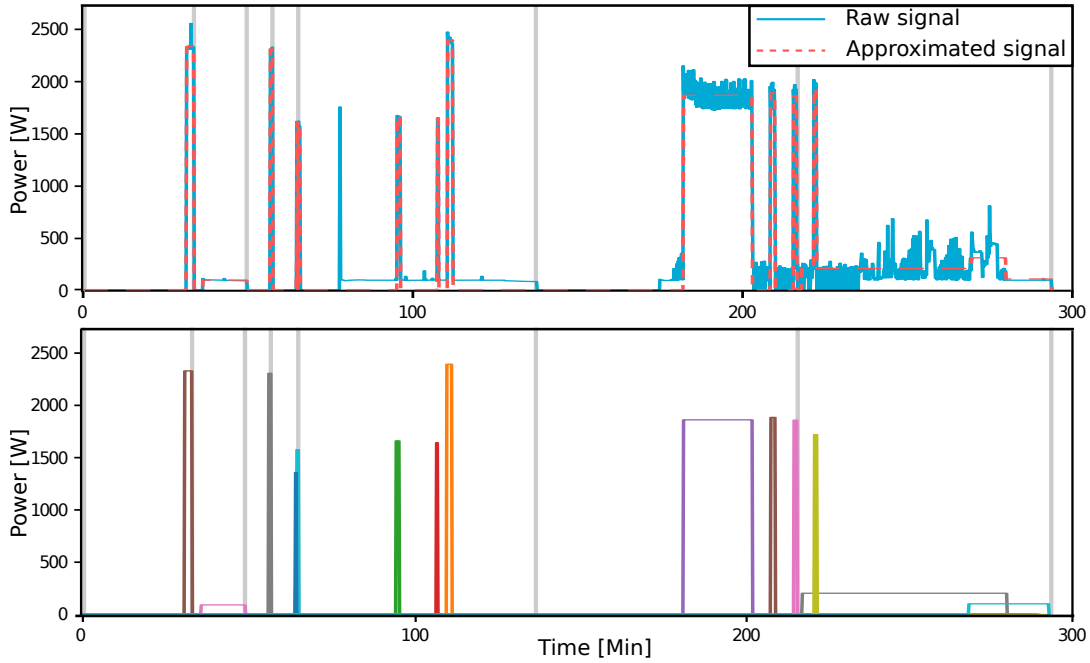


Figure 4.6: Top Figure: Raw consumption signal (blue) and OMP approximated signal (red dashed). Bottom Figure: detailed boxcar functions used to approximate the signal. The vertical gray lines correspond to the limits of the batches.

The generation of representative signatures is performed on the training set sampled from the same household and corresponds to a year of data. The clustering process is run on all the activation events collected from the individual appliances' consumption signals after sparse signal approximation. The HAC generates the following number of representative signatures; dishwasher: 5, washing machine: 6, microwave: 4, kettle: 2, fridge: 6 (Figure 4.7).

The OMP error tolerance is set empirically to 0.055 as it eliminates high-frequency variations and only keeps large power variations, materializing changes of states. An illustration of the approximation is given in Figure 4.6, where the blue solid line is the raw data and the red dashed line, is the OMP approximation. The resulting support vector Λ consists of 6660 unique boxcar functions. As each activation of type-I appliances (e.g., fridge, microwave, kettle) is approximated using a single boxcar function, their labeling is operated at this stage (Figure 4.7). For each type-I appliance, a threshold is empirically set to the maximum distance between a pattern from the representative signatures and boxcar functions in Λ. The type-I labeled boxcar functions are set aside for the remaining steps. The type-I appliances have a high-frequency activation and relatively short operation time, which means that they are likely to be repeatedly activated simultaneously with a type-II appliance (e.g., washing machine or dishwasher). Hence, it decreases the community detection performance as the generated signature is corrupted by unrelated boxcar functions.

The GMM generates 43 clusters of boxcar functions. In each cluster, the boxcar functions have

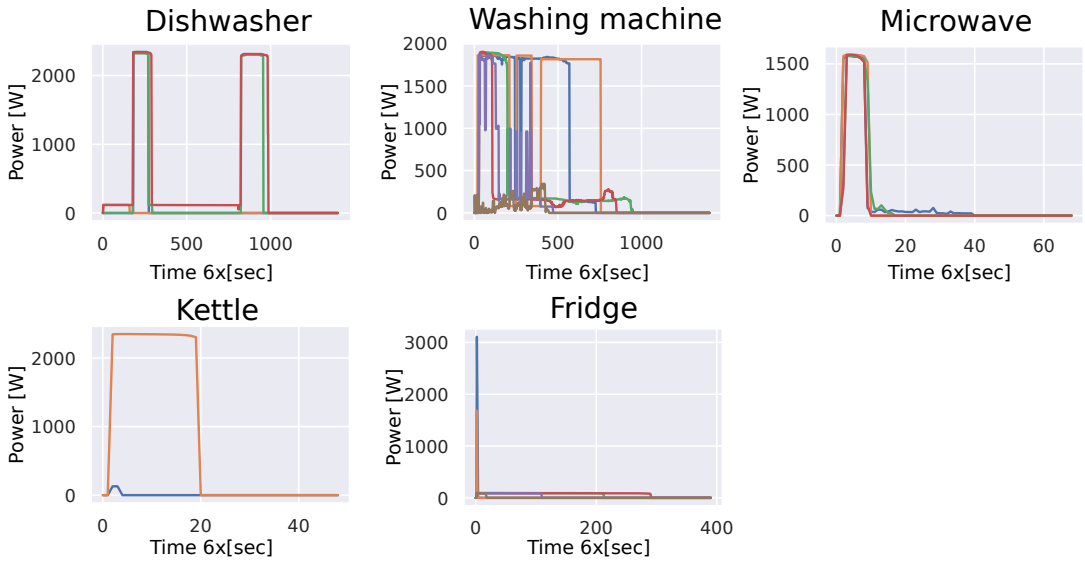


Figure 4.7: Representative signatures of each appliance.

approximately the same shape $(\boldsymbol{\alpha}, \mathbf{w})$ and are assumed to be generated by the same appliance. One cluster groups, more than 100 boxcar functions, and five clusters consist of a single boxcar function (Figure 4.8). The community detection takes the 43 clusters and looks for high cross-temporal

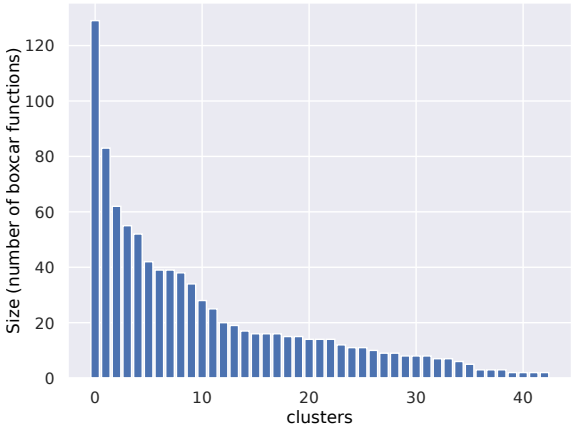


Figure 4.8: Histogram of the size of clusters generated by the GMM.

dependencies between them. The community detection has two parameters, the modularity threshold set to 1 and the scaling parameter for the weights set to 0.95. The community detection generates four communities. Activation events in each community are then labeled individually, by calculating their DTW distance to the representative signature in Figure 4.7.

4.4.2 Performance metrics

The performance of NILM algorithms is evaluated using both classification performance measures and the Estimated Accuracy (EC) [26, 161]. The classification performance measures compare the predictions to the ground truth. Each point falls into one of the four cells of the confusion matrix,

		Actual	
		Positive	Negative
Predicted	Positive	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)
		P	N

where True Positive/Negative (TP/TN) represents the number of times a disaggregated signal from a single appliance being ON/OFF is correctly assigned, False Positive (FP) represents the number of times a disaggregated signal from a single appliance is considered ON (consumption significantly larger than zero) but was actually OFF and False Negative (FN) is a disaggregated signal from a single appliance considered as OFF (consumption close to zero) but was ON. From the confusion matrix, the recall or True Positive Rate (TPR), the precision, also known as Positive Predictive Value (PPV), and the Accuracy (ACC) are respectively defined as

$$\text{TPR} = \frac{\text{TP}}{\text{TP + FN}}, \quad \text{PPV} = \frac{\text{TP}}{\text{TP + FP}}, \quad \text{ACC} = \frac{\text{TP+TN}}{\text{P + N}}.$$

From the PPV and the TPR, the F_1-score

$$F_1\text{-score} = 2\,\frac{\text{PPV TPR}}{\text{PPV + TPR}} \tag{4.12}$$

is calculated. In the NILM literature, the EC for each appliance n

$$\text{EC}^n = 1 - \frac{\sum_{t=1}^{T} |y_n^t - \hat{y}_n^t|}{2\sum_{t=1}^{T} y_n^t}, \tag{4.13}$$

where t is the time index and n is the appliance index, is used to estimate how accurately the consumption has been estimated [25, 26, 161–163]. The overall EC is given by

$$\text{EC} = 1 - \frac{\sum_{t=1}^{T}\sum_{n=1}^{N} |y_n^t - \hat{y}_n^t|}{2\sum_{t=1}^{T}\sum_{n=1}^{N} y_n^t}. \tag{4.14}$$

and tells how much of the overall signal is kept after summing up the disaggregated signals. All the introduced performance evaluation measures take values in $[0, 1]$ and are to be read the same way: the higher the value, the higher the performance.

4.4.3 Benchmark against state-of-the-art methodologies

The results of the implementation of our methodology are benchmarked against FHMM as implemented in [160] and RNN-LSTM as implemented in [26] (Table 4.1). Nevertheless, it is not a fair comparison to benchmark an unsupervised approach against supervised approaches. Hence, we do not expect to outperform them, but to reach similar performances. In the following section, LSTM is used for RNN-LSTM and OMP for the combination of OMP, GMM and community detection. In the *across appliances* column, only EC is calculated over all appliances, the other performance metrics are simply the average across all appliances.

For the performance across appliances, the OMP approach performs slightly better for the EC than the benchmarks. Regarding the classification metrics, OMP outperforms the benchmarks for all of them besides the TPR. For the TPR, the LSTM largely outperforms (0.85) OMP (0.55) and the

Table 4.1: Disaggregation performance of the OMP approach compared to FHMM [160] and RNN-LSTM [26].

Appliance	Accross Appliances			Dishwasher			Washing machine		
Method	OMP	FHMM	LSTM	OMP	FHMM	LSTM	OMP	FHMM	LSTM
EC	**0.93**	0.91	0.92	0.26	**0.91**	0.86	0.71	0.84	**0.88**
F_1-score	**0.54**	0.18	0.38	**0.26**	0.05	0.08	**0.61**	0.08	0.03
PPV	**0.62**	0.12	0.36	**0.24**	0.03	0.04	**0.73**	0.04	0.01
TPR	0.55	0.53	**0.85**	0.29	0.49	**0.87**	0.53	0.64	**0.73**
ACC	**0.96**	0.70	0.66	**0.96**	0.33	0.30	**0.96**	0.79	0.23

Appliance	Kettle			Microwave			Fridge		
Method	OMP	FHMM	LSTM	OMP	FHMM	LSTM	OMP	FHMM	LSTM
EC	0.56	0.92	**0.99**	0.67	0.84	**0.98**	0.81	0.94	**0.97**
F_1-score	0.47	0.19	**0.93**	**0.54**	0.01	0.13	**0.83**	0.55	0.74
PPV	0.33	0.14	**0.96**	**0.79**	0.01	0.07	**0.99**	0.40	0.72
TPR	0.82	0.29	**0.91**	0.41	0.34	**0.99**	0.71	**0.86**	0.77
ACC	0.99	0.99	**1.00**	**0.99**	0.91	0.98	**0.88**	0.50	0.81

FHMM (0.53). It means that the LSTM detects the activation events better (high TPR) but also has a higher false positive rate (lower PPV).

The detailed results of each appliance reveal large performance variability. The dishwasher appears to be difficult to detect accurately for all methodologies. The OMP algorithm detects it less often than FHMM and LSTM (lower TPR) but makes less false positives, considering it is ON when it is OFF (higher PPV). Hence, OMP has the highest F_1-score. For the washing machine, the EC of the OMP is closer to the benchmarks than for the dishwasher. The TPR of the OMP is slightly lower than FHMM and LSTM but its PPV is much higher (0.73 against 0.04 and 0.01), which again generates the highest F_1-score for OMP (0.61). For the kettle, the LSTM generates the individualized consumption signal almost perfectly as it has values close to one for all the performance metrics. OMP performs better than FHMM for the detection of the activation events (higher F_1-score, PPV, TPR). For the microwave, the detection of activation events (TPR) of OMP is higher than FHMM but lower that LSTM, which misses almost none (0.99). However, in terms of F_1-score the performance of OMP is again better as the PPV of the benchmarks are poor. For the fridge, all three methods perform well, FHMM has the highest TPR (0.86); OMP and LSTM perform similarly (respectively 0.71 and 0.77). The OMP has again a higher PPV, which generates a higher F_1-score for OMP compared to benchmarks.

The variability of performances from one appliance to another changes depending on the complexity of the load behavior. The OMP first performs a signal approximation into a square signal (Figure 4.6). Thus appliances with complex transient load behaviors obtain poorer performance (e.g., dishwasher). For the same reason appliances displaying similar average power consumptions and operation times are hard to separate. In general, the OMP approach performs well with classification metrics (especially with PPV) and the benchmarks have better EC as expected.

4.4.4 Evolution of the performance with the degradation of the resolution

The state-of-the-art approaches require high resolution (under 30 seconds) to identify signatures based on transient and internal state variations, which means that their performance drops quickly

with the degradation of the resolution [27]. OMP does not rely on transient states. It is then expected to perform accurately at a lower resolution. The OMP algorithm was implemented on the same dataset degraded to 1-minute resolution to assess the evolution of performance with the degradation of the resolution (Table 4.2).

Table 4.2: Comparison of the performance of the OMP approach with data at 6-second and 1-minute resolution.

Metrics	Across Appliances		Dishwasher		Washing machine		Kettle		Microwave		Fridge	
Resolution	6s	1min	6s	1min	6s	1min	6s	1min	6s	1min	6s	1min
EC	**0.93**	0.91	0.26	**0.57**	**0.71**	0.58	0.56	**0.73**	**0.67**	0.35	**0.81**	**0.81**
F_1-score	**0.54**	0.53	**0.26**	0.25	**0.61**	0.54	0.47	**0.60**	**0.54**	0.44	**0.83**	0.82
PPV	0.62	**0.78**	0.24	**1.00**	**0.73**	0.59	0.33	**0.89**	**0.79**	0.47	**0.99**	0.98
TPR	**0.55**	0.45	**0.29**	0.15	**0.53**	0.51	**0.82**	0.45	0.41	**0.42**	**0.71**	**0.71**
ACC	**0.96**	**0.96**	0.96	**0.98**	**0.96**	0.95	**0.99**	**0.99**	**1.00**	0.99	**0.88**	0.87

The performance across appliances at 6 seconds is better for all performance metrics besides PPV. Looking closely at the values, it appears that they are close for EC and F_1-score. The performance across appliances at 6-second and 1-minute resolution are similar.

Looking at the performance for each appliance, the 6-second resolution performs much better than the 1-minute resolution only for the washing machine and the microwave. For the dishwasher, the performance is poor for both resolutions, but the 1-minute resolution has a larger EC, which means that the individual signal is better recovered. For the kettle, the TPR is higher at 6 seconds but the PPV is lower than the 1-minute resolution. Hence, the 1-minute resolution has a better F_1-score. For the fridge, the performance of both resolutions is almost similar.

Between 6-second and 1-minute resolution, the number of data points is divided by ten, yet the change of resolution does not much deteriorate the performance of the OMP.

4.5 Application to heat pump detection

NILM applications are often pointed at for not focusing on real-world application and being more of a machine learning puzzle to solve rather than a solution to smart grid problems [147]. In the following section, we present an application of the OMP algorithm to the detection of HPs. Indeed, HPs as well as EH and Air Conditioners (ACs) could provide services to the grid if recruited in a Demand Response (DR) program for example [164].

4.5.1 Data presentation

The data used is a sample from the EcoGrid EU project, which was completed in 2015 [28]. More details about the EcoGrid EU framework are given in Section 5.1.3. A week of data from the 26$^{\text{th}}$ of December 2014 to the 31$^{\text{st}}$ of December 2014 that corresponds to the coldest period (i.e., where we know HPs are active) is used for the training and the testing. The resolution of the data is 5 minutes, which means that HPs activation events can be visually identified (from amplitude, operation time, frequency of activation) from the aggregated power consumption signal.

A part of the households in the dataset is known to be equipped with HP and labeled as such (284 in 1900 households of EcoGrid EU). 50 of them are randomly selected to be used during the training step to form a *prior*. The test set is unlabeled, which means that we do not know which and how many households are equipped with a HP. 65 of them are completely randomly selected from the households in EocoGrid EU and 10 households are randomly picked among the households labeled with a HP (to have a minimum number of 10 HPs) to form a test set of 75 households.

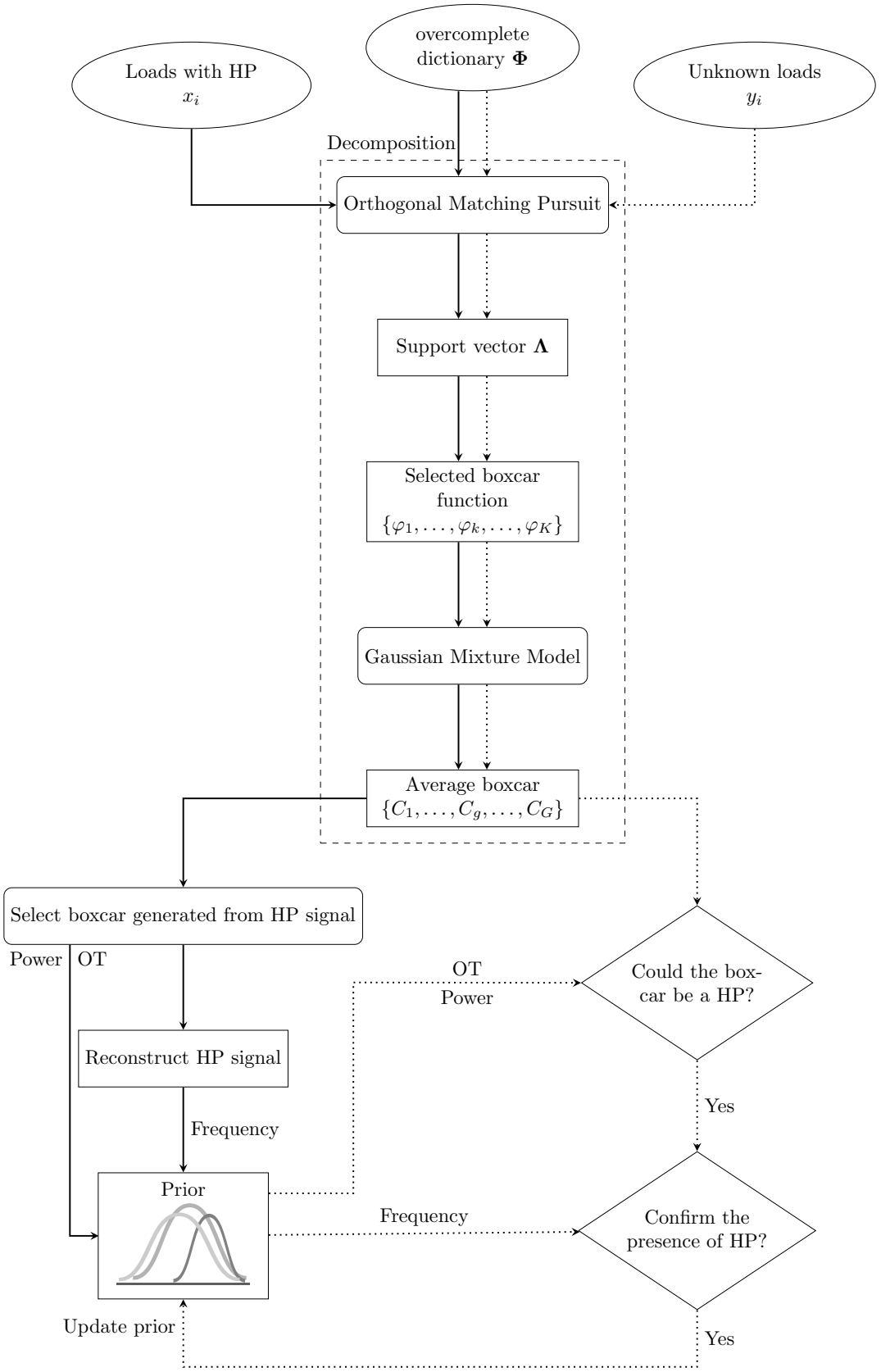


Figure 4.9: Overview of the HP detection algorithm.

4.5.2 A semi-supervised HP detection algorithm

The signal decomposition, dashed box in Figure 4.9, is the combination of the OMP and the GMM as presented in Section 4.3. HPs, ACs, EH, and EVs charging power consumption are type-I appliances (ON/OFF). Hence, using the power signal sparse approximation and a clustering is sufficient to isolate boxcar functions that are approximation of activation events of HPs.

The approach is semi-supervised, with a training phase, where a *prior* is learned from the training set (solid line in Figure 4.9); and a test phase, where HP among unlabeled households from the test set are detected using the prior, which is inferred after every iteration (dotted line in Figure 4.9).

Training phase: Prior definition

The training phase is materialized with a solid line and is placed on the left-hand side of Figure 4.9. During the training phase, the subset of households that are equipped with a HP is processed through signal sparse approximation and clustering. HP power consumption of each household is visually assessed from the approximate signal and the centroids of the clusters. Clusters with the corresponding power consumption are flagged as presented in the bottom graph of Figure 4.10. For each boxcar function from the flagged clusters, the power α (height) and the operation time w (width) are then collected. The power consumption signal of each HP is then reconstructed by summing up the flagged boxcar functions. The frequency of activation of HPs can be obtained by calculating the frequency of positive values over a rolling window of half a day.

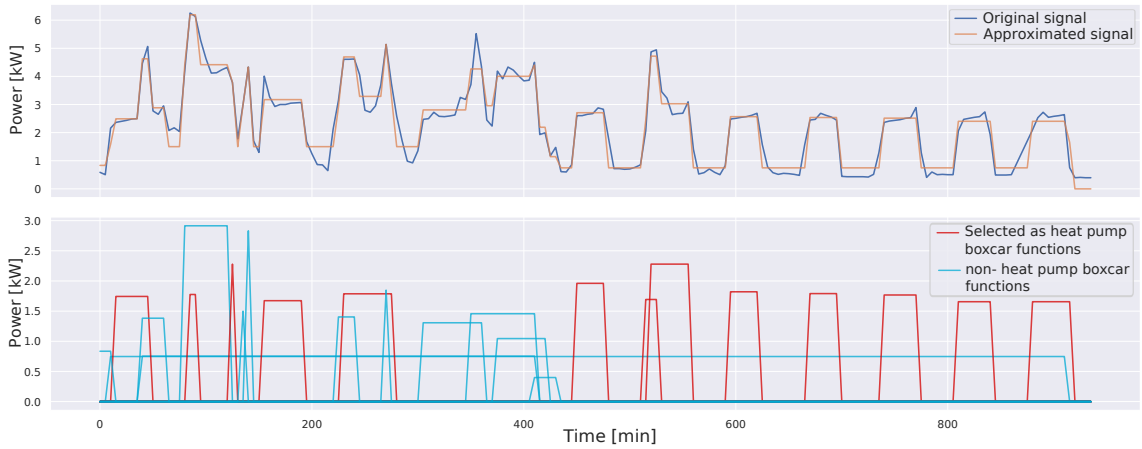
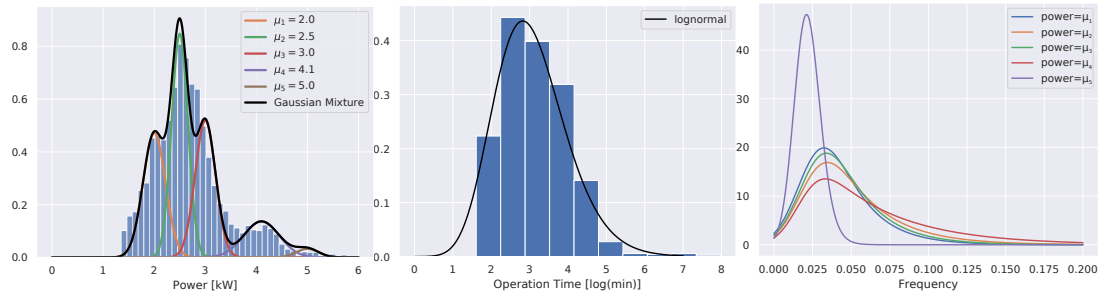


Figure 4.10: Raw and approximated power consumption (top) and corresponding boxcar functions of an household equipped with HP (bottom).

Probability Density Functions (PDFs) (Gaussian, Lognormal and Weibull) are then fitted on the empirical distribution of the power consumption, operation time and the frequency obtained from the training set. The best fits are presented in Figure 4.11. The PDF of the power consumption $\boldsymbol{\alpha}$ is actually a Gaussian mixture $p(\boldsymbol{\alpha}) = \sum_{p=1}^{P} \phi_p \, \mathcal{N}(\boldsymbol{\alpha} \mid \boldsymbol{\mu}_p, \boldsymbol{\sigma}_p^2)$ with average consumptions of each size of HP $\boldsymbol{\mu}_p = \{2.0, 2.6, 3.0, 4.1, 5.0\}$ (Figure 4.11a). The best PDF fit of the operation time is a lognormal distribution on the natural logarithm of the operation time (Figure 4.11b). For the frequency, a conditional PDF to each power consumption $\boldsymbol{\mu}_p$ is fitted. The best fits are lognormal PDF that adjust well to the long tail on the high end side (Figure 4.11c).



(a) (b) (c)

Figure 4.11: Priors generated during the training phase, power consumption (a), operation time (b) and frequency (c).

The set of three PDFs defines the prior of the load behavior of a HP. It can then be used to identify similar load behaviors in the overall electricity consumption of unlabeled households, unveiling the presence of a HP.

Test phase: Heat pump detection and characterization

The HP detection phase is materialized with a dotted line and is located on the right-hand side of Figure 4.9. Similarly to the learning phase, the overall consumption signals of households are processed through sparse signal approximation and GMM. Using the PDFs of the power consumption, the likelihood

$$\mathcal{L}(\mu_p \mid \alpha) = f(\alpha\,;\,\mu_p, \sigma_p^2) = \frac{1}{\sigma_p\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\alpha-\mu_p}{\sigma_p}\right)^2} \tag{4.15}$$

of each centroid power consumption α being sampled from a HP with average power consumption μ_p is calculated. The likelihood of the power consumption not being sampled from a HP that follows a lognormal distribution

$$\mathcal{L}(\mu_0 \mid \alpha) = f(\alpha\,;\,\mu_0, \sigma_0^2) = \frac{1}{\alpha}\frac{1}{\sigma_0\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\ln\alpha-\mu_0}{\sigma_0}\right)^2} \tag{4.16}$$

is calculated as well. The maximum likelihood decoding is

$$P[\alpha = \mu_p] = \frac{\mathcal{L}(\mu_p \mid \alpha)}{\sum_{p=0}^{P} \mathcal{L}(\mu_p \mid \alpha)} \tag{4.17}$$

for each value of μ_p. It returns a probability in $[0, 1]$ that a centroid with average power consumption α is sampled from the Gaussian distribution $\mathcal{N}(\boldsymbol{\mu}_p, \boldsymbol{\sigma}_p^2)$.

A counting of the centroids weighted by the size of the clusters is done per household to determine which μ_p is the most probable. The clusters with the selected μ_p as potential HP power consumption are then kept and the likelihood, the maximum likelihood decoding as well as the counting is then implemented on individual boxcar functions. The households with μ_0 as the most probable power consumption are discarded as they are considered not likely to be equipped with a domestic HP.

The same operation is done using the lognormal PDF on the natural logarithm of the operation time for the selected households. The complementary lognormal PDF of the natural logarithm of

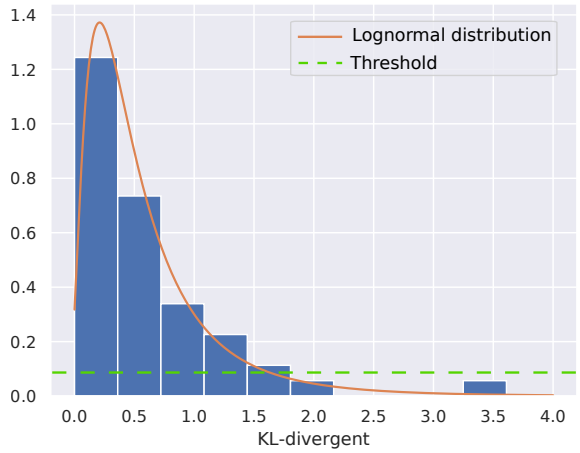


Figure 4.12: Distribution of the Kullback-Leibler divergence calculated between the frequency PDF fitted on individual households in the training set and the corresponding frequency PDF in Figure 4.11c. The orange line is the fitted lognormal PDF and the green dotted line is the threshold under which it is unlikely that a distribution of frequencies is sampled from a HP.

operation times of boxcar functions not generated by HPs is used to calculate the corresponding likelihood and subsequently do the decoding.

For each household, the set of boxcar functions that are likely to be generated by a HP are summed up to form the power consumption signal of the HP. It is then differentiated and the frequency of activation is calculated using a rolling window. As the frequency PDF is conditional to the power consumption of the HP μ_p, the corresponding PDF is used to calculate the likelihood. The decoding of the maximum likelihood cannot be done here as the likelihood of frequencies generated by other appliances cannot be calculated. The Kullback-Leibler divergence

$$D_{KL}(P \,||\, Q) = -\sum_i P(i) \ln\left(\frac{P(i)}{Q(i)}\right) \tag{4.18}$$

between the corresponding power consumption conditional PDFs of the frequency from the training phase, and each individual HP frequency PDF of the test set is calculated. It generates another lognormal PDF, Figure 4.12. A threshold of 0.1 is defined empirically under which it is unlikely that the frequency of activation events is generated by a HP.

The performance metrics used to evaluate the detection are the classification metrics : TPR, PPV and ACC as described in Section 4.4.2.

4.5.3 Results

The performances are evaluated for each step of the detection and the overall performance of the algorithm (Table 4.3). The results are presented with both the outlook of what we know *a priori* is the ground truth (left-hand side of Table 4.3) and after checking for unlabeled HPs detected and thus wrongly considered as false positive (right-hand side of Table 4.3). After visually checking for false positives and false negatives, three unlabeled households were displaying patterns that were identified as a HP and were correctly detected. No unlabeled HPs were classified as false negatives.

The first step of the detection has low PPV and a high TPR, which is expected as the range of possible average power consumption is wide and thus not discriminative. The two households

Table 4.3: Performance of the HP detection at each step, with or without correcting for false positives being actual HPs.

	A priori			Corrected		
	TPR	PPV	ACC	TPR	PPV	ACC
Power	0.80	0.11	0.15	0.84	0.16	0.19
Operation time	1.0	0.23	0.56	1.0	0.30	0.60
Frequency	0.88	0.32	0.21	0.91	0.45	0.35
Overall	0.70	0.31	0.75	0.77	0.46	0.80

labeled with a HP that are discarded at this stage are on the extremity of the range $[1.2, 6.0]$ that we consider possible for a domestic HP based on the PDFs in Figure 4.11a. The other households discarded at this step have maximum power consumption under 1.5kW, which is associated with a low probability of the presence of a HP. From an operation time perspective, the detection provides a better ACC of 0.56 associated with a higher PPV (0.23). No households labeled with a HP are discarded at this stage (TPR=1.0). The frequency step reduces the false positive rate (high PPV). Its ACC and TPR are a bit lower than the operation time, as one labeled household has been discarded since it displays a high frequency for a HP (probably an EH).

The overall performance of the detection algorithm with the *a priori* information is correct. After correction, all the performances for all metrics are increased (when not 1.0) by at least 0.4. The rate of false positives is still relatively high (low PPV), but in the context of HP detection it is preferred to have a high TPR and a low PPV so that most of the HP are kept. A visual examination of the false positives at the end of the process reveals that it consists of households equipped with large EH, which can have a loading behavior similar to HPs for power consumption and frequency. The operation times are in contrary usually short (5 minutes or less) but the operation time prior is not strict on the low side of the distribution and did not discard them (see Figure 4.11b).

4.6 Conclusion

The unsupervised disaggregation methodology presented in this chapter has shown similar performances to state-of-the-art supervised NILM methodologies. It is an application-oriented NILM methodology as it can be used with resolution data that are close to the utility standard. Indeed, the state-of-the-art NILM implementations aim at providing feedback to customers on individual appliance consumption to modify their energy behavior. They suffer from considerable computing costs. Thus, the potential power savings are low in comparison to the power used for computing. Hence, the implementation of an unsupervised NILM approach that uses standard metering data to detect appliances that can either contribute to the unbalance or provide services to the grid is of interest.

By definition, the algorithm used for HP detection is robust and can detect HPs even if other large appliances are present in the aggregated signal. As long as the characteristics of the appliances defined in the *prior* are not completely overlapping, it should be possible to individualize them. Similarly, auto-consumption using Photovoltaics (PV) or batteries should not affect the performance as long as sections of the signal (with a length of some hours) with low auto-consumption can be recovered. New business models rely on this type of approach to detect and manage flexible assets. The increase of EVs that could charge simultaneously and create important unbalances has emphasized the need for smart charging facilities. However, before installing smart charging facilities, standard domestic charging facilities have to be detected.

CHAPTER 5

Analysis of Metering Data in a Demand Response Framework

The modernization of the grid is a requirement for the implementation of 'greener' energy policies to increase the share of Renewable Energy Sources (RES). Indeed, with the increase of variable, uncertain and not controllable generation in the energy mix, flexibility has to be found somewhere else. Demand Response (DR) is a cost-effective approach that can be associated with other flexible solutions to balance the distribution system. DR is a decentralized source of flexibility, which requires metering and communication infrastructure. Hence, it can be a complex task to implement it. This chapter discusses how to evaluate responsiveness, then quantify and characterize the responses. As a DR implementation requires equipment for participating households, it is likely to fail, and identifying faulty equipment is a crucial task for troubleshooting as presented in [Paper E]. The response capacity is then shortened due to technical failures. The characteristics and the amplitude of the responses have to be evaluated as well to understand the dynamics of the responses. This is usually done by comparison to a baseline, which is an estimate of what would be the load without DR. Baselines being by definition biased, we propose a methodology, using the work in [Paper F], to evaluate the response without baseline.

5.1 Background of demand response and analysis of responses

5.1.1 Purpose of demand response

The integration of a higher share of RES in the energy mix, dictated by climate change and 'greener' energy policies, makes the generation more intermittent (i.e., weather dependent) and less adjustable to the demand. Hence, new ways are required to balance the demand to the generation. Utilities have many flexibility options to balance a system with large share of RES, (i) fast ramping generators, (ii) curtailment, (iii) storage, and (iv) flexibility of the demand [6]. All the options are developed and the solution lies in a combination of them. However, flexible demand remains the least costly source of flexibility. Demand Side Management (DSM), harnessing demand flexibility, enables carbon mitigation goals and reduction in generation capacity, which subsequently reduces investment in larger capacities [7]. The reliability of the grid is also increased as additional solutions can be used before reaching the point where an intentional brown-out happens leading to unknown consequences. Demand Response (DR) is the main tool of DSM with energy response (Figure 5.1). Under the term DR, all the frameworks that intend to modify load shapes through various incentives, especially reducing peak loads, are gathered. Indeed, the aggregated peak load is used to determine the generation capacity that has to be installed. Reduction/mitigation of the aggregated peak load is a strong leverage to postpone investments in

grid reinforcement [165]. In a nutshell, DSM is the policy that consists in modifying the demand consumption and it is evaluated on medium to long term, while DR is a tool to implement DSM and its evaluation is done on short term [166].

By taking a closer look to the DR branch in Figure 5.1, it can be noted that the output of a DR-event can be either dispatchable or not dispatchable, depending on whether an incentive results in an exact response or not. Non-dispatchable DR are economic incentives through time-sensitive pricing, which passes the marginal cost of generation onto end-customers, depending on the generation mix. Hence, highly carbonated generation, which has a high marginal cost, is penalized with a higher price to customers. The dispatchable DR consists mostly of tools used to maintain high reliability on the grid. DR, triggered at the Distribution System Operator (DSO) level, can be used for reliability by reducing the capacity of the aggregated load in disconnecting or reducing some individual loads [167]. The demand can also be used as a reserve (spinning or non-spinning) through DR based on operating reserve capacity requirements that are defined to reach applicable reliability standards [168, 169]. The last use of DR is for regulation in case of a change in grid frequency (i.e., frequency response) [170]. DR has emerged as a popular approach to provide services to the grid with its pros and cons [164].

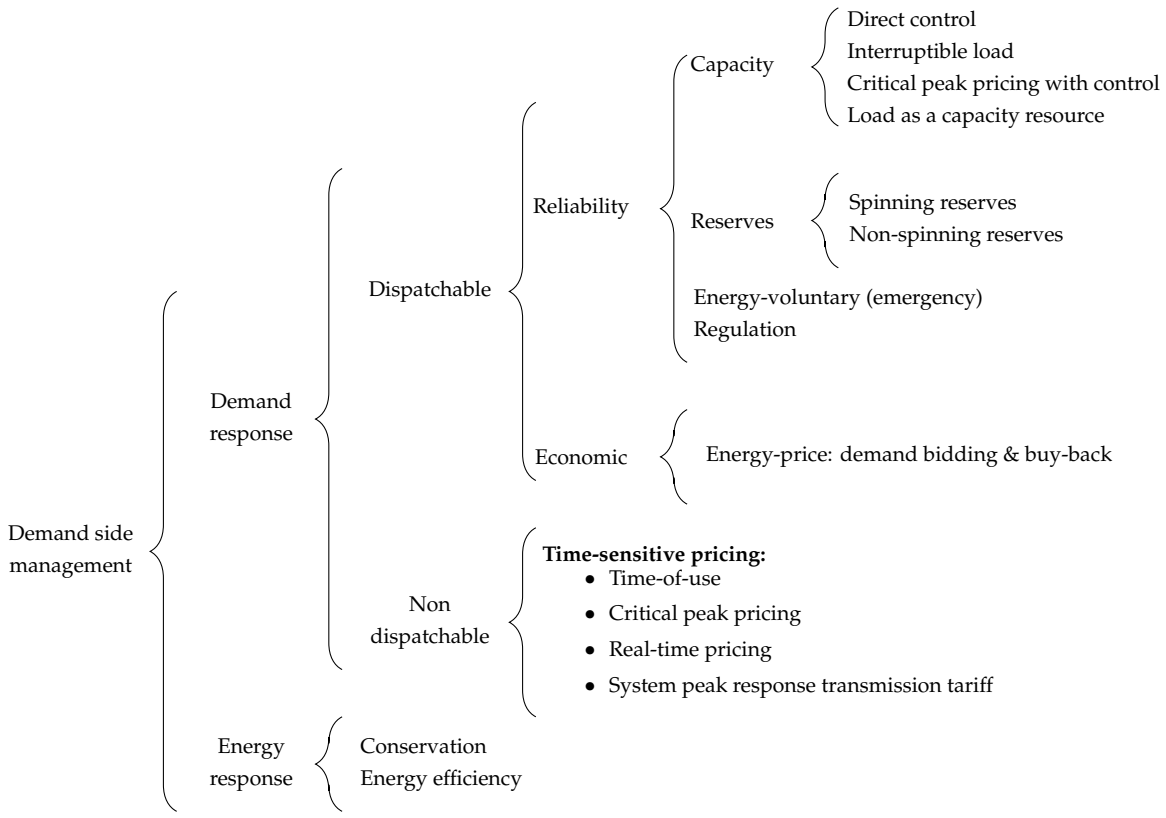


Figure 5.1: Relationship between DSM and DR [165, 168].

A multitude of DR frameworks exist, corresponding to the combinations of the targeted loads, the design, and the technical infrastructures [18]. DR can also be used to provide a variety of services. Thorough reviews of state-of-the-art DR deployment are given in [18, 35, 165, 171–173]. DR through time-sensitive pricing has emerged as a popular option due to its ability to respond quickly to changes in the generation mix (i.e., wind, solar) and the simplicity of financial settlements,

as customers are rewarded according to the provided flexibility [164]. In contrary to other DR frameworks, it allows customers to remain flexible and retain control [174]. The individual responses to price variation are seen as stochastic (i.e., many unknown parameters drive the response), but the response of the pool of participants remains statistically predictable. Hence, two major problems appear: (i) How can we check systematically that the equipment on the participant side responds? (ii) How can we dissociate the responsive part of the load from the non-responsive part? Besides mentions of the need to evaluate the responsiveness of controllers installed in each participant household, description of protocols and development of methodologies to identify non-responsive controllers remain lacking despite being crucial in practice for operating DR efficiently.

5.1.2 What is a baseline?

The challenge of assessing what the responsive part of a load is, is a well-studied topic in the DR literature [175–178]. The main tool used is baselines that estimate the load without DR incentives. Baselines are either formed by a representative households that are exposed to the same conditions as the participating households and generate a synchronous baseline, or metering data of the participants are collected before the DR program begins and their loads are forecast to generate a baseline. The main obstacle to baseline generation is that it requires metering data under non-DR conditions.

For forecast baseline, the standard model used is the auto-regressive integrated moving average (ARIMA) models with exponential smoothing [30]. It can also include geographical factors [179] and seasonal variations [180]. Baselines can also be generated using Artificial Neural Network (ANN) [181]. However, the black box nature of ANNs obscures understanding of the dynamics. The main benefit of ANN over linear modeling is the ability to capture non-linear relationships among external variables [30]. Nevertheless ANNs are not ultimately outperforming linear models [182, 183] and it is difficult to find the balance between overfitting and generalization of ANNs [181, 184]. Hence, if the DR has been running for some time, an assumption on the stationarity of the load signal has to be made to use a baseline modeled on data collected before DR implementation. However, as discussed in Chapter 3 it is a strong assumption.

The baseline can also consist of a set of households, which are statistically representative (i.e., have the same characteristics and the same weather condition) of the DR participants. Nevertheless, such a representative pool is difficult to obtain in practice. The obtained baseline can then be subtracted from the observed load signal under a DR event to analyze the response. The baseline serves two purposes, (i) it can be used in the market mechanism, and (ii) for the response evaluation. Hence, it would be an important achievement to get free from the baseline.

5.1.3 Empirical Framework: EcoGrid EU

The EcoGrid EU project is a large scale DR program implemented on the Danish island of Bornholm where RES represent a large share of the generation mix. The goal of EcoGrid EU was to exploit the thermal storage flexibility of residential participants in a large scale real-time market using a Real Time Pricing (RTP) [185]. The DR controls only heating elements, Heat Pumps (HPs) and Electric Heating (EH). Households participating in the program are equipped with automated devices provided by different manufacturers and subsequently controlled by different algorithms. Considering the indoor temperature and the comfort settings, households react to price at different

threshold depending on their thermal capacity margins. We assume that algorithms in controllers do not affect the amplitude of the response but the pace and the frequency of responses. The behavior of loads is stochastic. Each household may respond differently from one day to another, depending on the activity of the inhabitants.

The DR framework in EcoGrid EU is a hardware-in-the-loop market platform (Figure 5.2), which can be divided into four blocks: (i) **The forecasting module** estimates future load based on historical consumption, weather forecasts, and prices. (ii) **The day-ahead unit commitment** schedules dispatch according to spot prices and wind power forecast. (iii) **The EcoGrid market** finds 5-minute prices for the next hour. Its main objective is to maintain the system balance and maximize the social welfare by dispatching conventional balancing power and DR. (iv) **5-minute imbalance optimization** adjusts marginally the prices before committing the RTP.

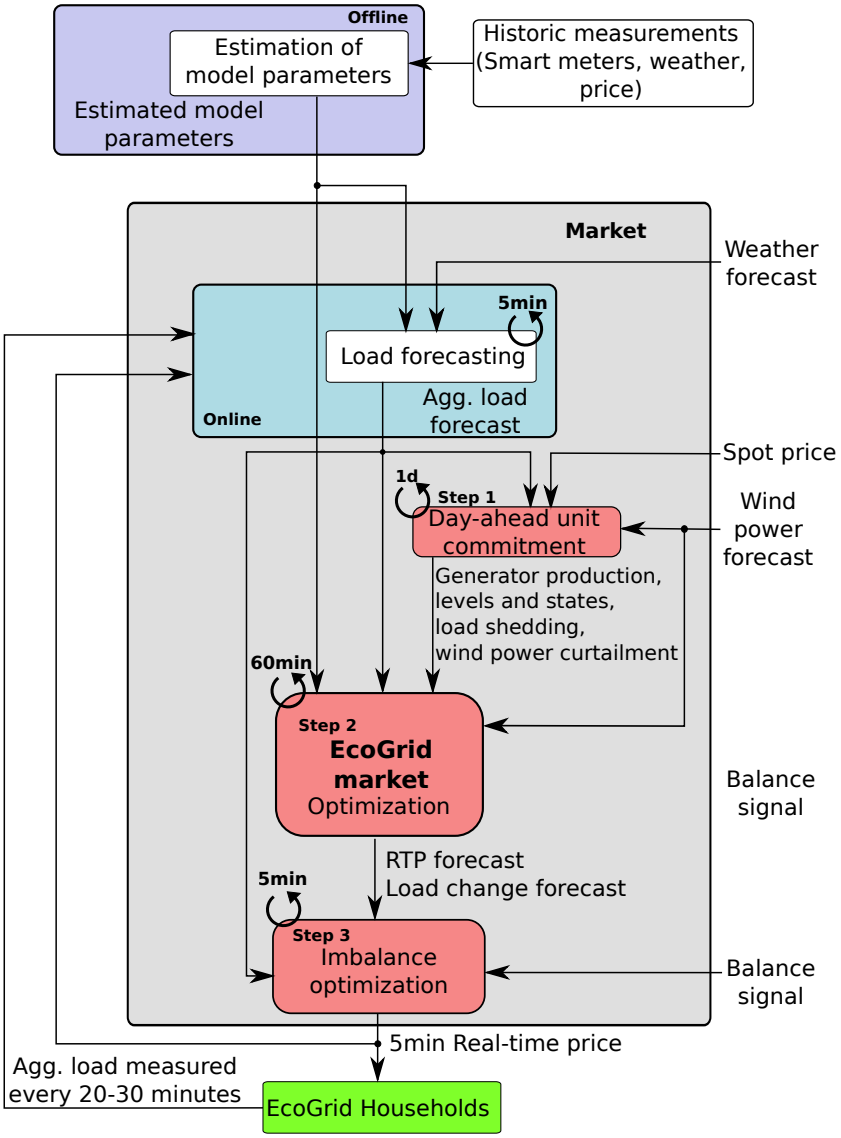


Figure 5.2: Overall representation of EcoGrid EU market and price generation.

The spot price is positively correlated with the consumption profile, i.e., a high demand engenders a higher spot price. As the spot price is a component of the RTP, the RTP generally follows the

daily load pattern. In practice, a responsive load is negatively correlated to RTP (i.e., consumption increases as the RTP decreases and *vice versa*). A full description of the market clearing mechanism that determines the RTP is given in [28].

Table 5.1: Test group descriptions.

Group	type of heating	Hardware	Software	type of DR	number
Group 1	mixed	none	none	none	253
Group 2	mixed	none	manual	none	455
Group 3	HP	manuf. 1	manuf. 1	auto.	195
Group 4	EH	manuf. 1	manuf. 1	auto.	322
Group 5	HP	manuf. 1	manuf. 2	auto.	84
Group 6	EH	manuf. 3	manuf. 3	auto.	398

Out of 1900 participating households, 1707 are kept in this work after filtering out households with more than 10% metering data missing. Four groups were fitted with automated equipment (groups 3–6 in Table 5.1). Group 1 was supposed to be the Customer Base Line (CBL), but appeared to be non-representative of the participant and thus could not be used as baseline [30]. Group 2 had no automation and participants were asked to change their consumption according to information provided *via* SMS. Controllers of groups 3–5 are from the same manufacturer, with either EH or HPs and a different algorithm for group 5. Group 6 is equipped with both hardware and software different from the other groups. The data were collected from September 2014 to February 2015 with an outdoor temperature between -10.6 °C and 18.0 °C. The resolution of the metering data is 5 minutes.

5.2 Responsiveness testing protocol

In the following section, we introduce a test-control protocol and the associated evaluation methodology to assess whether or not loads of households participating in the EcoGrid EU DR program are price-responsive. The test-control protocol is commonly used in clinical trials for the pharmaceutical industry to evaluate the efficiency of medicines or treatments. It consists of comparing the results from a 'control group', which does not receive the treatment to a treated 'test group'.

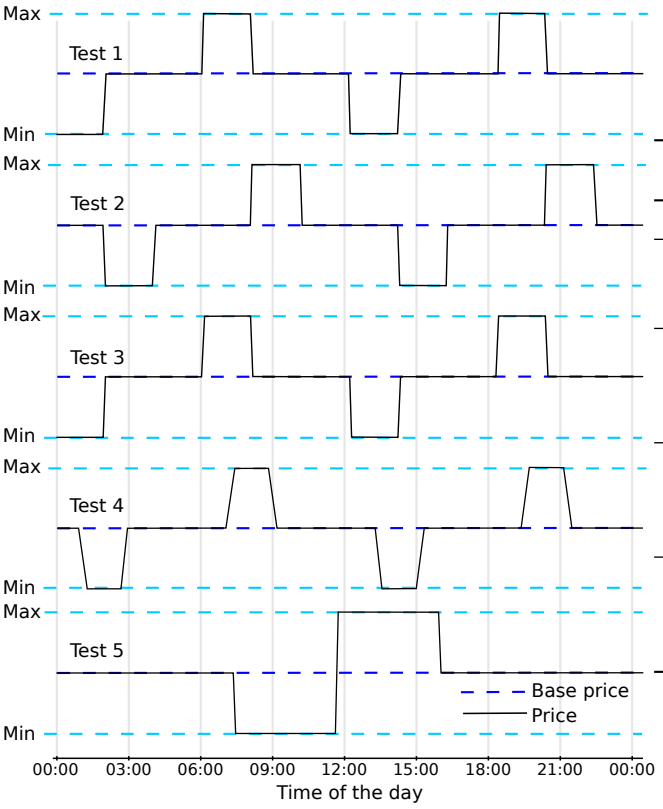
5.2.1 Experimental setup

Clinical trials were historically developed in the pharmaceutical industry [186]. It appeared to be impossible to conclude on the efficiency of a treatment based on tests on individual biological organisms. Hence, testing on populations reduced the potential bias produced by individuals on the overall assessment. The inherent uncertainty on the responsiveness (response to a treatment) resulting from the absence of homogeneity in the test group as well as in the reference group (e.g., behavior of the user, thermal comfort setup, energetic profile of the buildings), supports the idea that a clinical trial approach is relevant in this case.

Control group: Customer base line

The initial phase of the EcoGrid EU project was to recruit households by installing controllers. During the recruitment, the central operator disabled deliberately the automation of some households, while others were disabled because of technical problems. In the clinical test approach,

the CBL would constitute the reference group not exposed to the treatment 'price'. The CBL was created by averaging the power consumption signal of all disable households. The composition of both test and CBL groups varied from one test-case to the next as the technical problems evolved. Based on the observations of the CBL and non random appearance of the technical problems, it is concluded that the CBL is biased. Consequently, it influences the estimated confidence intervals and hypothesis tests performed.



(a)

Test period	Min/Max (€/MWh)	Base price (€/MWh)	Test signal
25/10/013	-53.7 / 148.0	47.2	Test 1
07/11/013	-53.7 / 148.0	47.2	Test 2
21/11/013	-61.3 / 140.3	39.4	Test 3
27/11/013	37.4 / 41.4	39.4	Test 3
06/12/2013	-134.5 / 134.5	0	Test 5
10/12/2013	-134.5 / 134.5	0	Test 5
11/12/2013	-134.5 / 134.5	0	Test 5
12/12/2013	-134.5 / 134.5	0	Test 5
20/01/2014	-134.5 / 134.5	0	Test 5
21/01/2014	-134.5 / 134.5	0	Test 5
22/01/2014	-134.5 / 134.5	0	Test 5
23/01/2014	-134.5 / 134.5	0	Test 5
08/03/2014	-61.3 / 140.3	39.4	Test 3
11/03/2014	37.4 / 41.4	39.4	Test 3
09/04/2014	-61.3 / 140.3	39.4	Test 4
13/04/2014	-61.3 / 140.3	39.4	Test 4

(b)

Figure 5.3: The different type of price signals (a) and price variations during the tests (b).

Treatments: Test cases

Test cases that stressed price-responsiveness in having extreme price variations were used as 'treatments' to observe and evaluate the responsiveness of controllers. As the power consumption should vary inversely to the RTP, a significant change in the power consumption was expected [187]. Test cases have the same duration of 24 hours and their shape and values are detailed in Table 5.3b and Figure 5.3a. Large step-down prices are strong incentives to consume electricity, especially when they are negative, as they generate benefits. In contrary large step up prices are designed to immediately decrease the consumption of HPs and EH to zero.

5.2.2 Load profiling

A load profiling of the loads in each test case (Figure 5.3b) and per group (Table 5.1) has been performed. Load profiling would reduce the number of load responses to analyze. Furthermore, clustering would allow us to identify outliers and discard them from the test group [19]. The causes

of variability of the responses are manifold. It can be, (i) technical problems, (ii) non-optimal setups (e.g., bad location of the sensors), or (iii) inhabitants interacting with the controllers (e.g., turning it off or changing comfort settings). Hence, we expect the responses to be diverse and changing in time as a load may be responsive during one test case and not responsive during the next one. The elements in the non-responsive clusters form a list of households to check for troubleshooting. The combination of 16 tests with large price variations leaves little doubt that a household where a large change of consumption opposite to the price variation is observed repeatedly, is responsive despite the bias on the CBL.

Section 3.2 gives an overview of the clustering algorithms used in the literature on load profiling. Depending on the *a priori* information, the number of clusters to generate can be known or unknown [188]. In this experiment, we have no prior information to determine K, the number of clusters to be generated; it has to be determined *a posteriori*. Hierarchical Ascendant Clustering (HAC) allows to determine K *a posteriori* from a dendrogram with only one instantiation.

Some preprocessing of the data is required to make the loads comparable for the HAC algorithm. Centering of the households' load is operated by removing the mean over the test period. Hence, after centering, all the loads have an average of zero over a test period. Their dynamics and response amplitude are conserved, thus allowing them to better compare the higher-order dynamics of the various households [189, 190].

Despite the possibility of having many typical possible responses for each test case, some synchronization in the responses is still expected as all households receive the same signal. In that context, the distance used for the HAC should account for covariances between the consumption signals. In our experimental framework, the space we have to explore has the dimension of the number of measurements performed in time. The resolution of the data is 5 minutes and a test case last 24 hours. Power consumption signals are time series and display temporal dependence. The Mahalanobis distance [191]

$$d_M(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i - \mathbf{x}_j)^\top \boldsymbol{\sigma}_{ij}^{-1} (\mathbf{x}_i - \mathbf{x}_j) \tag{5.1}$$

where $\boldsymbol{\sigma}_{ij}$ is the covariance matrix between the two load signals $\mathbf{x}_i$ and $\mathbf{x}_j$ respectively from household i and j, fulfills the aforementioned requirements. The covariance matrix $\boldsymbol{\sigma}_{ij}$ can be singular when the number of households I is smaller or about the same as the number of time steps T in the load signals [192]. To prevent singularity, $\boldsymbol{\sigma}_{ij}$ is replaced in (5.1) by a shrunk covariance matrix

$$\boldsymbol{\sigma}_{ij}^* = \lambda \Theta_{ij} + (1 - \lambda)\boldsymbol{\sigma}_{ij} \tag{5.2}$$

where Θ_{ij}, referred to as the target, is a diagonal matrix formed with the element on the main diagonal of the original covariance matrix $\boldsymbol{\sigma}_{ij}$ and λ is the shrinkage coefficient [192]. It is an efficient way to obtain a non-singular closest estimate of the covariance matrix $\boldsymbol{\sigma}_{ij}$. The shrinkage does a trade-off between a highly-structured matrix (Θ_{ij}) and a non-organized one ($\boldsymbol{\sigma}_{ij}$), while λ allows controlling the balance between the two [193]. We set

$$\lambda = \begin{cases} \lambda^*, & \text{if } \lambda^* \leqslant 1 \\ 1, & \text{otherwise} \end{cases} \quad \text{with,} \quad \lambda^* = \frac{\sum_{m \neq n} \widehat{\sigma}(c_{mn})}{\sum_{m \neq n} c_{mn}^2} \tag{5.3}$$

where c_{mn} are the components of the (sample) covariance matrix $\boldsymbol{\sigma}_{ij}$ and $\widehat{\sigma}(c_{mn})$ their estimated variance [194].

The HAC input is the pairwise Mahalanobis distance matrix $\mathbf{d_M}$ calculated between all pairs of loads (i, j). The clusters are then regrouped using Ward's method, also known as the minimum-variance method

$$E = \sum_{k=1}^{K} \sum_{i=1}^{I} d(\mathbf{x}_i, \mathbf{y}_k)^2 \tag{5.4}$$

where K is the number of clusters and $\mathbf{y}_k$ is the centroid of cluster K. It aims at minimizing the increase of the within-cluster sum of squared distances, E, at each iteration of the agglomerative process [86]. The total variance of a set of a clustered set of loads is the sum of the within-cluster variance plus the between-cluster centers variance [195]. As Ward's method minimizes the increase of within-cluster variance, it also maximizes the variance between cluster centers. The clusters are thereby the most homogeneous possible subgroups from the set of households at each iteration.

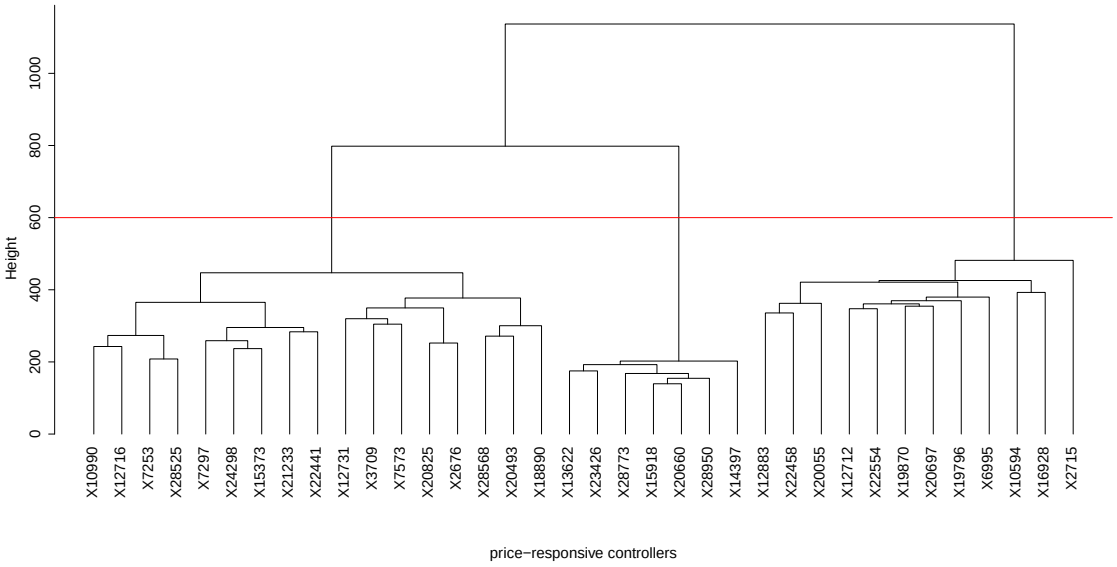


Figure 5.4: Dendrogram for the clustering of 35 households in one of the EcoGrid EU experiment. The red line indicates the cut made to obtain 3 clusters.

The output of the HAC is graphically represented in a dendrogram (Figure 5.4). The dendrogram is a basis to chose K, depending on the structure of the tree and the aim of the clustering. The decision is expert-based and relies on user interpretation of the structure of the tree and the knowledge on the problem to solve [196].

A partition into K clusters is obtained after cutting the dendrogram. The centroids can then be represented as the average of the loads in each cluster in parallel to the averaged load of the CBL and the RTP (Figure 5.5). From the graphical representation clusters with reactive adjustment to the price variations in comparison to CBL) can be sorted out from the non-responsive ones. Non-responsive clusters are then discarded at this stage. For example in Figure 5.5, the clusters 2 and 5 were discarded, as cluster 2 follows the CBL and cluster 5 has no day activity as the households are empty. The test group is now cleaned from non-responsive loads and in the remainder of the paper, we refer to the selected subgroups as test groups.

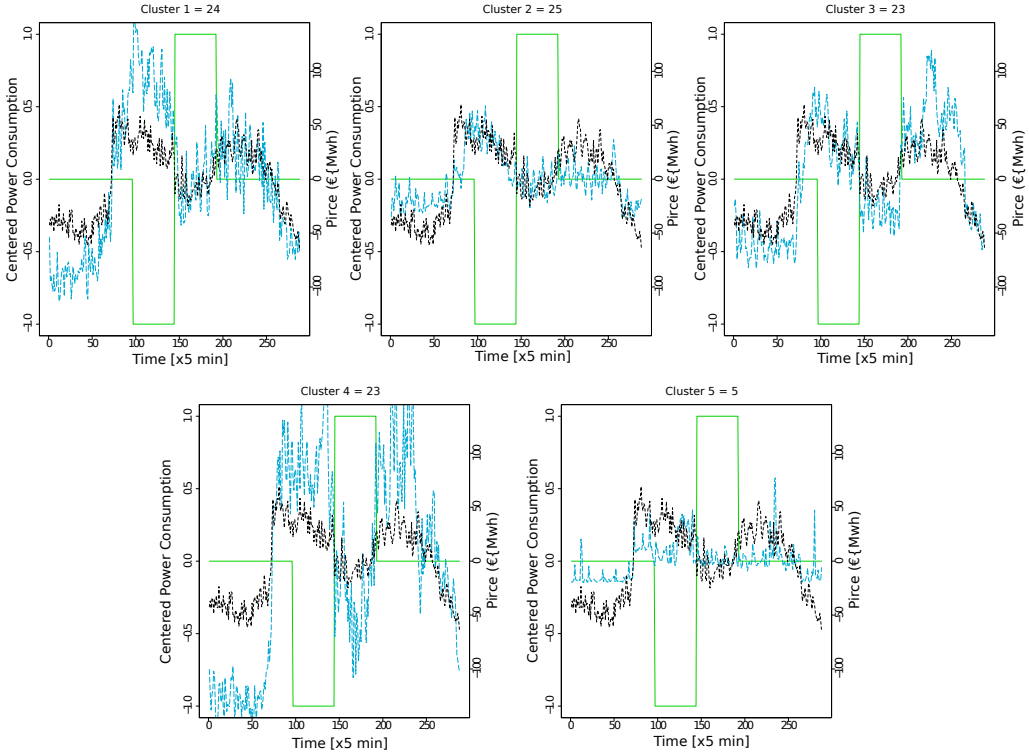


Figure 5.5: The averaged time series is calculated for each cluster in the test group and displayed as a colored dashed line. The black dashed line is the averaged time series of the CBL and the green line is the price.

5.2.3 A clinical trial test to determined responsiveness

By analogy to clinical trials in the pharmaceutical field, the CBL is our reference group, the price is our treatment, and the households receiving the price signals constitute the test group. The question we aim to answer is: *Do price variations induce significant opposite changes in power consumption?*

The clustering results can be analyzed, (i) visually by comparing the centroid of the selected clusters from each group to the CBL during the DR event, or (ii) statistically using hypothesis testing. They provide approximately the same information. The visual inspection is used to illustrate the responses, but also to get an impression of the range of possible behaviors. The statistical test is used for decision making.

Visual evaluation

The visual examination of a cluster centroid in parallel to the CBL and the RTP. The associated confidence intervals during a test case give clues to evaluate responsiveness (Figure 5.6).

Analysis of the distribution of the data shows that it does not follow a Gaussian distribution. Hence, a nonparametric approach (Non-Studentized pivotal method) using bootstrap sampling is used to generate confidence intervals [189]. From the average of 5000 samples, the 95% confidence intervals defined by 2.5% and 97.5% quantiles of the distributions are computed.

If the confidence intervals are not overlapping during the price variation, it can be inferred that the behavior of the test group is different from the CBL, for example from 7:05 to 8:05 in Figure 5.6. On

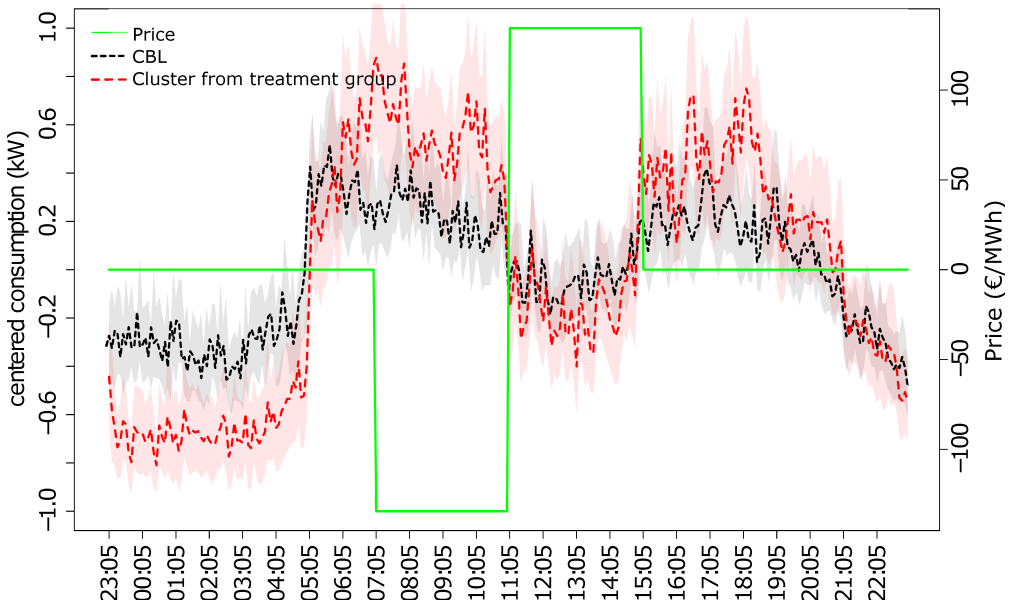


Figure 5.6: The average time series from the CBL and a cluster from the test group with their respective 95% confidence intervals generated from the bootstrap. The green line is the price.

the contrary, when the confidence intervals overlap or when an averaged signal lies within the confidence intervals of the other one, it cannot be concluded. The test and control groups are also different in period 23:05-05:05. As the controllers receive the day-ahead hourly price forecast (see Figure 5.2), they anticipate the price drop occurring at 07:05.

Hypothesis testing

Hypothesis testing is the statistical way to answer the question using the data collected during the test cases. EcoGrid EU's DR program goal is to displace a part of the load from periods with higher prices to periods with lower prices. The evaluation of whether or not the goal is achieved can be done on an economic basis using the RTP. Consequently, we could conclude whether a test group is price responsive or not. The average unit cost $\bar{C}_i$, for a test case with T time steps, is calculated as

$$\bar{C}_i = \frac{\sum_{t=1}^{T} x_i^t p^t}{\sum_{t=1}^{T} x_i^t} \tag{5.5}$$

where x_i^t is the consumption of electricity from household i at time t and p^t the price at time t.

As the variances of the average unit cost of the test and control groups are different and may have heavy tails, standard parametric tests cannot be used. A one-sided Mann-Whitney test (also known as the Wilcoxon rank sum test) is performed on the ranks of the average unit cost of the test and control groups. The hypotheses to test are

$$\begin{aligned} H_0 &: \mu_{test} \geqslant \mu_{CBL} \\ H_1 &: \mu_{test} < \mu_{CBL} \end{aligned} \tag{5.6}$$

where μ corresponds to the sum of ranks, H_1 is the one-sided tailed alternative hypothesis and H_0 is the null hypothesis. The null hypothesis means that either a controller is not responsive or that the response is not large enough to conclude that the average unit cost is lower for the test group.

If the H_0 is rejected, the alternative hypothesis is statistically confirmed. The Mann-Whitney test defines the statistic U with the following formula

$$U = \min\left(n_t n_c + \frac{n_t(n_t-1)}{2} - R_t, n_t n_c + \frac{n_c(n_c-1)}{2} - R_c\right) \tag{5.7}$$

where n_t, n_c are respectively the size of the test and the control groups and R_t, R_c are their respective sums of ranks. U follows a normal distribution and we can calculate the associated p-value as

$$P(U \geqslant U_{1-\alpha} | \mu_{test} \geqslant \mu_{CBL}) \tag{5.8}$$

The p-value can be interpreted as the probability of obtaining the result $\mu_{test} < \mu_{CBL}$ when the truth is $\mu_{test} \geqslant \mu_{CBL}$. The test is considered significant when the p-value is lower than the type-I error threshold α (set here to 0.05) [197].

5.2.4 Results

Clustering Results

The responsive clusters are visually selected by comparing the centroid of each cluster to the averaged CBL during the test cases (Figure 5.5). If the centroid of a cluster is flat (no activity), following the pattern of the averaged CBL (no response) or showing unexpected patterns (technical problems), it is discarded from the test group. When no cluster is price-responsive, only the aberrant ones will be removed.

Table 5.2: Number of households in various test cases: number in the test group (number deemed price-responsive after clustering) / number in the CBL.

Date	Manufacturer 1 EH	Manufacturer 1 HP	Manufacturer 2 EH
25/10/2013	68 (48) / 288	36 (24) / 197	88 (55) / 82
07/11/2013	65 (58) / 289	36 (33) / 197	88 (61) / 92
21/11/2013	67 (61) / 292	36 (36) / 200	87 (75) / 94
27/11/2013	66 (55) / 292	36 (34) / 201	89 (78) / 91
06/12/2013	—	—	115 (74) / 99
10/12/2013	—	—	103 (84) / 91
11/12/2013	—	—	100 (70) / 86
12/12/2013	—	—	106 (69) / 89
20/01/2014	—	—	230 (194) / 105
21/01/2014	—	—	223 (121) / 100
22/01/2014	—	—	230 (110) / 104
23/01/2014	—	—	229 (107) / 105
08/03/2014	30 (30) / 324	20 (17) / 236	237 (125) / 75
11/03/2014	38 (38) / 317	24 (18) / 229	232 (171) / 76
09/04/2014	101 (99) / 249	58 (43) / 188	249 (197) / 109
13/04/2014	38 (38) / 311	24 (22) / 222	269 (188) / 114

The range of the selection from the original data goes from 52% to 100% (Table 5.2). Hence, up to half (48%) of the controllers were in the test group, but did not visually appear to be price-responsive. The graphical representation of the centroids also summarizes the typical load response. For example, cluster 1 in Figure 5.5 are the controllers that respond to the first price variation, while cluster 3 are the ones that respond to the second price variation and cluster 4 gathers the ones that

respond to both variations. It also unveils the differences of strategies between the manufacturers (e.g., anticipate the price variation or rebound). Hence, we gain knowledge about the possible responses of controllers.

Results of the Clinical Trial Test Approach

The visual evaluation is summarized in Table 5.3, where light gray test cases have no cluster that displayed a responsive centroid and dark gray test cases have at least one responsive cluster.

Table 5.3: The color of the cell shows the results of the visual evaluation; gray is responsive, light gray is non-responsive. The figure is the p-value from the Mann-Whitney test. Significant test at $\alpha = 5\%$ are shown in bold and italic.

Date	Manufacturer 1 EH	Manufacturer 1 HP	Manufacturer 2 EH
25/10/2013	0.98	0.44	0.20
07/11/2013	0.39	***0.0013***	0.88
21/11/2013	0.95	0.20	0.09
27/11/2013	0.34	0.34	0.81
06/12/2013	—	—	0.13
10/12/2013	—	—	***0.00022***
11/12/2013	—	—	***0.0015***
12/12/2013	—	—	0.12
20/01/2014	—	—	0.99
21/01/2014	—	—	0.22
22/01/2014	—	—	0.21
23/01/2014	—	—	0.63
08/03/2014	0.54	0.59	***0.0068***
11/03/2014	0.86	0.80	0.28
09/04/2014	***0.0034***	***0.0096***	0.21
13/04/2014	0.96	0.12	***0.0014***

A progression over time of the number of responsive test cases is observed. During the roll-out phase of EcoGrid EU, controllers and other infrastructures were developed and improved, which explains the improvement of the DR as the heating period went on. Manufacturer 2 appears to be more often visually responsive.

Results of the hypothesis testing

The Mann-Whitney test results for the different test cases are providing statistical conclusions about the responsiveness (Table 5.3). The test cases where the test group has a significantly lower cost per unit are displayed in bold and italic. Only seven test cases display statistical significance among which four are from manufacturer 2. Another interesting result is that for the test occurring on 07/11/2013, loads of households equipped with controllers from manufacturer 1 are responsive when associated with a HP but not when associated with EH.

The comparison between the results from the visual examination and the Mann-Whitney test highlights the difference between price-responsiveness, which can be visually noticed but not statistically validated. Only test-cases that show an average unit cost significantly lower than the CBL have an economical impact.

The tendency observed with the visual evaluation is confirmed statistically, towards the end of the roll-out of the EcoGrid EU project. Price-responsiveness is validated for at least a subpopulation of

EH, HPs, and both manufacturers controllers. The improvement of the responsiveness over time is a direct consequence of the tuning performed on the controllers during the roll-out.

The CBL could be used, despite being biased, because the price variations were large and the variations of the electricity consumption were expected to be large as well. However, when the RTP signal is much smoother as it is in reality, it would not be possible to analyze the response using the CBL.

5.3 A baseline free response analysis

The RTP fluctuates and exhibits many local peaks if looking only at a single day. Controllers schedule the response to activate HPs and EH to obtain the lower cost and keep the indoor temperature within the range of the comfort zone, which means that controllers do not react to all price variations. Clustering is a good way to obtain the representative responses observed in the pool of participants. However, when more details are needed to characterize the response, the limits of the method are reached. We now analyze the response from EcoGrid EU DR with prices generated by the EcoGrid EU market (see Figure 5.2). The RTP now reflects the future generation mix (as part of the day-ahead price) and the forecast demand. Hence, the CBL cannot be used as it is not representative of the pool of participants. The behavior of a load can change from responsive to not responsive from one hour to another as the inhabitants can change comfort settings or the thermal capacity is saturated. In this section, we describe a methodology to characterize and quantify responses without baseline. The response delivered by each load daily is evaluated and summarized using the average and standard deviation per household rather than evaluating it over the entire heating season.

5.3.1 Response Categorization

The RTP is strongly correlated to the average daily pattern of the load. The more a load provides DR, the less it is correlated to the RTP. This is captured by a negative or low correlation

$$r_{p,x_i} = \frac{\sum_{t=1}^{T}(p^t - \overline{p})(x_i^t - \overline{x}_i)}{\sqrt{\sum_{t=1}^{T}(p^t - \overline{p})^2}\sqrt{\sum_{t=1}^{T}(x_i^t - \overline{x}_i)^2}} \tag{5.9}$$

between a load x_i of household i and RTP p. $\overline{x}_i$ and $\overline{p}$ are respectively the average over t of a load and the RTP. When a load is reactive to a price variation, the change in power consumption is opposite to the price, and the correlation between the load and RTP is then negative. The correlation-based distance

$$d_{cor}(p, x_i) = \sqrt{\left(\frac{1 - r_{p,x_i}}{1 + r_{p,x_i}}\right)} \tag{5.10}$$

between the individual loads x_i and the RTP p of a single day, measures the distance of a load to the price [198]. Note that an average correlation of 0 over a day corresponds to a d_{cor} of 1 and the more negative the correlation is, the larger the d_{cor}, so the farther from the price, the larger the response on average.

The d_{cor} is calculated for each individual daily load over the heating season. The mean and standard deviation is thereafter calculated per household i to cluster them according to their average distance from the RTP and the stability of the responses during the season. We used the K-means algorithm with $k = 5$ and it was run with 25 different sets of starting points to keep the best partition.

5.3.2 Response quantification

Finite Impulse Responses (FIRs) are used to measure the response of each cluster to the price [30]. FIRs are obtained from a general linear model of x^t the observed load,

$$x^t = \boldsymbol{\lambda}^{t\top}\theta_\lambda + \boldsymbol{z}^{t\top}\theta_z + \boldsymbol{\chi}^{t\top}\theta_\chi + \varepsilon^t \tag{5.11}$$

where $\boldsymbol{\lambda}^{t\top}$ is a vector including forecast, real-time and historic electricity prices (e.g., day-ahead, hour-ahead and RTP); $\boldsymbol{z}^{t\top}$ is a vector of exogenous variables (e.g., solar radiance, wind speed, exterior temperature, Fourier series of the daily independent base load); $\boldsymbol{\chi}^{t\top}$ is a vector of interactions of some of the precedent variables, θ_λ, θ_z, θ_χ are their respective coefficients and ε^t is the normally distributed error.

Concretely, the FIRs is the vector of relative real-time price coefficients noted θ_λ in (5.11). The relative price is the difference between the RTP and the hour-ahead price. A Lasso penalization is used to fit the model with the smallest number of coefficients [199].

5.3.3 Response characterization

First we consider each load as a linear combination of k statistically independent consumption patterns s_k. The vector of loads $\mathbf{x} = \{x_1, \ldots, x_i, \ldots, x_I\}$ can then be decomposed as

$$\mathbf{x} = \sum_{k=1}^{K} \mathbf{a}_k s_k = \mathbf{A}\mathbf{s}, \tag{5.12}$$

where $\mathbf{a}_k$ is a vector of coefficient of length I and $\mathbf{A}$ a matrix with dimensions $I \times K$. The components $\mathbf{s}$ are load behaviors (thermostat setting, price response, daily variation) and are called Independent Components (ICs) [200]. The method used to obtain the ICs is called Independent Component Analysis (ICA) and was first developed for signal processing to solve blind source separation, like the *cocktail party problem*.

In practice the ICA algorithms estimate $\mathbf{A}$ and use its inverse matrix of $\mathbf{W}$

$$\mathbf{s} = \mathbf{W}\mathbf{x}, \tag{5.13}$$

which corresponds to weights of loads $\mathbf{x}$ to obtain the ICs in $\mathbf{s}$. ICA assume that the components in $\mathbf{s}$ should be statistically *independent*, thus they must have *non-gaussian* distributions. The ICA estimation is based on measure of non-gaussianity (e.g., kurtosis, negentropy).

Before running an ICA, signals must be centered and whitened using for example Principal Component Analysis (PCA). Thus all signals have a zero mean and unit variance and the input variables are orthogonal.

The *fastICA* algorithm, first initiates randomly $\boldsymbol{w}$ by sampling from Gaussian distribution $\mathcal{N}(0,1)$ (Figure 5.7) [200]. A fixed point iteration process then finds the maximum non-gaussianity by approximating the negentropy using its approximative Newton iteration

$$w_i^+ = E\{\boldsymbol{x}\, g(\boldsymbol{w}_k^\top \boldsymbol{x})\} - E\{g'(\boldsymbol{w}_k^\top \boldsymbol{x})\}\boldsymbol{w}_k \tag{5.14}$$

where E is the expectation, the function g and g' are respectively the derivative and second derivative of the non-quadratic function G,

$$G(u) = -\exp(\frac{-u^2}{2}) \tag{5.15}$$

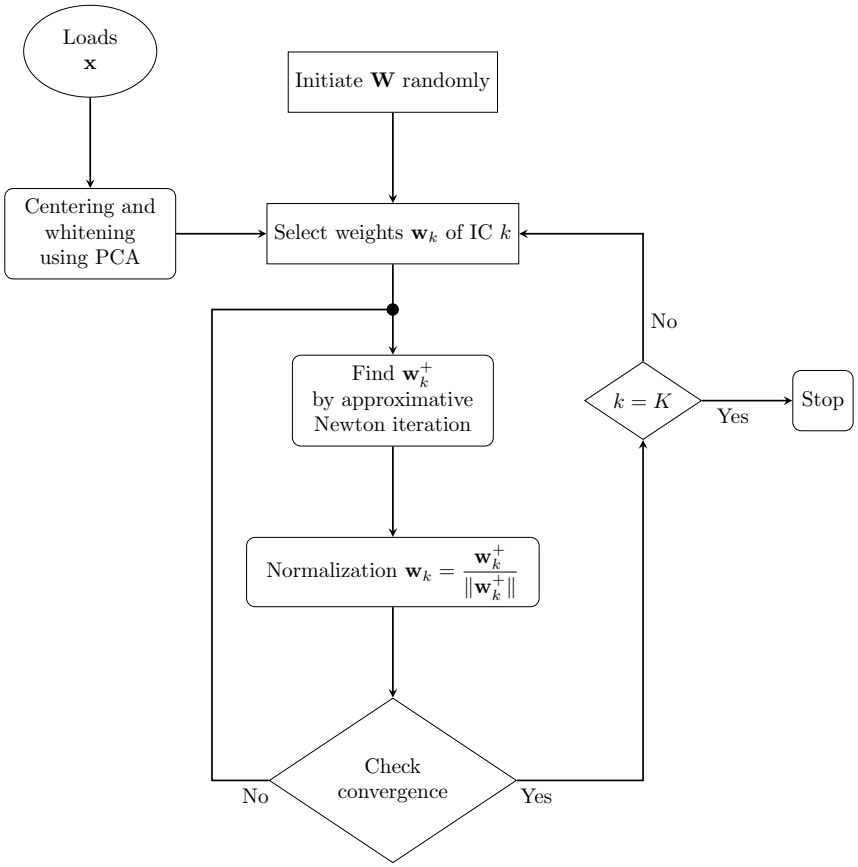


Figure 5.7: Representation of the FastICA algorithm.

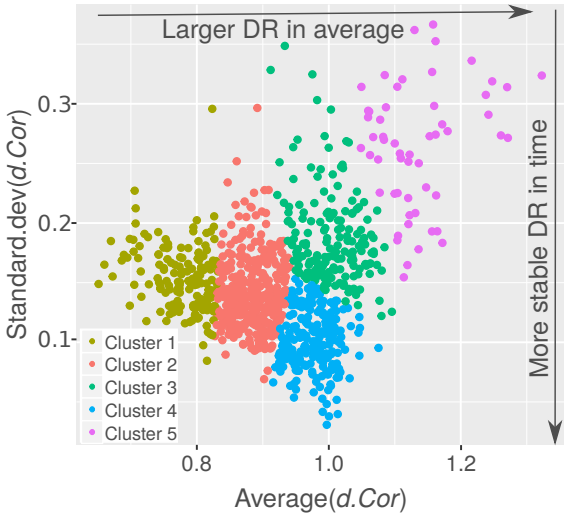


Figure 5.8: Clusters generated with K-means (k=5) on the mean and standard deviation of the d_{cor} distance.

5.3.4 Results

Response categorization results

The K-means implementation divides the pool of participants into five clusters in the 2D space formed by the average and the standard deviation of d_{cor} calculated between each daily load and

the corresponding daily RTP (Figure 5.8). The standard deviation is mostly used to discriminate between houses with the same average participation but different variability over the heating season.

The distribution of the groups presented in Table 5.1 among the clusters reveals that besides cluster 5, which is mostly composed of group 6 households, the manufacturer or the type of heating does not influence the average response (Figure 5.9a). Indeed, cluster 5, which is the farthest from the RTP on average, is mostly composed of households from the group 6 and 4, which are both equipped with EH but it could result from other factors as group 4 and 6 are also present in other clusters (e.g., cluster 3). Besides cluster 5, the distribution of the groups among the clusters supports the hypothesis that the group and subsequently the technologies (i.e., controllers or heating systems) are not the main driver of the average response of the load as the groups are well distributed in the different clusters.

The distribution of the dwelling types in each cluster shows that the three clusters with the highest response to RTP on average present also the largest shares of holiday houses (Figure 5.9b). The fact that holiday houses present the largest d_{cor} from the price has could be explained by the fact that (i) their base load does not follow daily patterns of standard load, and (ii) the absence of inhabitant gives the controller full control on planning the heating over the day.

Figure 5.9: Split of the groups (a) and the type of dwelling (b) into the K-means clusters.

Response quantification results

The response quantification has been done using FIR on each cluster (Figure 5.10). The results are given in percentage of the maximum load. The price variation occurs at $t = 0$, so the response is observed from 0 to 175 minutes. Hence, a drop in consumption is displayed with a negative value.

Cluster 4 exhibits the smallest response, peaking at 10% of the load 35 minutes after the price change, followed by a stiff and large rebound (21%) where the energy recovery is larger than the initial response. A small increase in consumption can be observed when the price variation occurs; it may be a time-delay caused by minimum run-times since this coincides with the slow DR delivery. Cluster 1 and 2 display the second smallest DR volume, at 13%, with a long-lasting

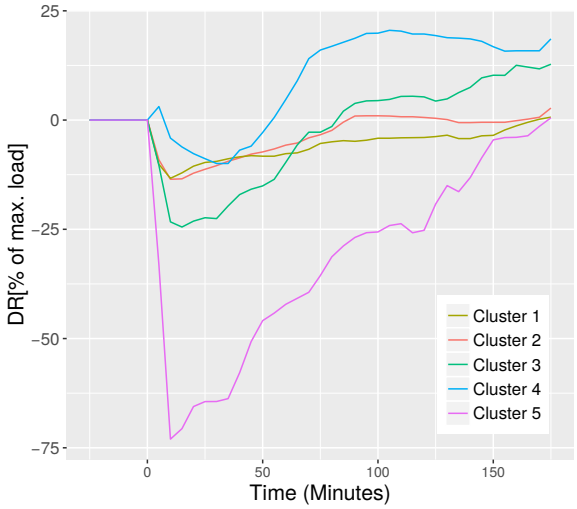


Figure 5.10: FIRs of the clusters obtained based on their average response and variability in the response.

response (85 minutes for cluster 2 and 175 minutes for cluster 1) without a rebound. Cluster 3 delivers 24% DR, with a maximum delivery after 15 minutes. It exhibits a slow but larger rebound after 85 minutes, which increases until the end of the FIR.

Cluster 5 is the most homogeneous cluster with mostly houses from group 6. It is also the cluster delivering by far the largest response. Its response amplitude reaches 72% of its maximum load, with a delay of just 10 minutes and does not have a rebound after the DR event. These results confirm that; (i) The responses are diverse, (ii) quantifying is not enough (rebound), and (iii) the controller type and the heating type are not the main driver of the response.

Response characterization results

The ICA has been implemented on data collected Tuesday, January 6th, 2015. After whitening the loads using PCA, the ten first principal components, collecting 72% of the total variance, are kept for the ICA, which means that ten ICs are generated. ICA projects loads into a space of dimension $K = 10$ where the ICs are the base. It is important to note that in contrary to PCA, ICA does not order ICs when generating the base. Hence, they cannot be ranked according to a specific metric. The ICs do not have a preferred orientation (positive or negative), which means that some times the taking -IC signal is easier to interpret.

Table 5.4: Distribution of the groups above the 95% quantile and under the 5% quantile of the density of weights for each IC.

	5% quantile										95% quantile									
IC #	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10
Group 3	33	5	**10**	5	14	4	0	5	4	5	12	9	10	7	19	6	15	11	8	31
Group 4	12	21	**29**	7	16	21	0	10	**27**	23	20	27	10	**26**	15	17	17	20	**27**	9
Group 5	2	4	4	1	7	3	14	7	2	4	4	1	0	0	7	6	7	6	1	3
Group 6	3	20	7	**37**	13	22	36	28	**17**	18	14	13	**30**	17	9	21	11	13	**14**	7

To interpret how the ICs have been generated, the extreme 5% quantiles (above 95% and under 5%) of the weight density distribution, which corresponds to the most contributive loads, are analyzed

Table 5.5: Distribution of the clusters above the 95% quantile and under the 5% quantile of the density of weights for each IC.

	5% quantile										95% quantile									
IC #	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10
Cluster 1	4	9	**16**	1	8	7	23	1	**27**	8	16	13	0	**23**	9	10	1	26	**13**	6
Cluster 2	23	26	**26**	8	30	19	20	11	**17**	20	17	24	10	**16**	21	19	24	22	**28**	19
Cluster 3	10	5	3	**24**	3	15	6	18	2	12	4	3	**14**	7	7	11	13	2	3	6
Cluster 4	11	7	5	3	6	4	0	13	4	9	13	8	**8**	3	12	3	7	0	5	19
Cluster 5	2	3	0	**14**	3	5	1	7	0	1	0	2	**18**	1	1	7	5	0	1	0

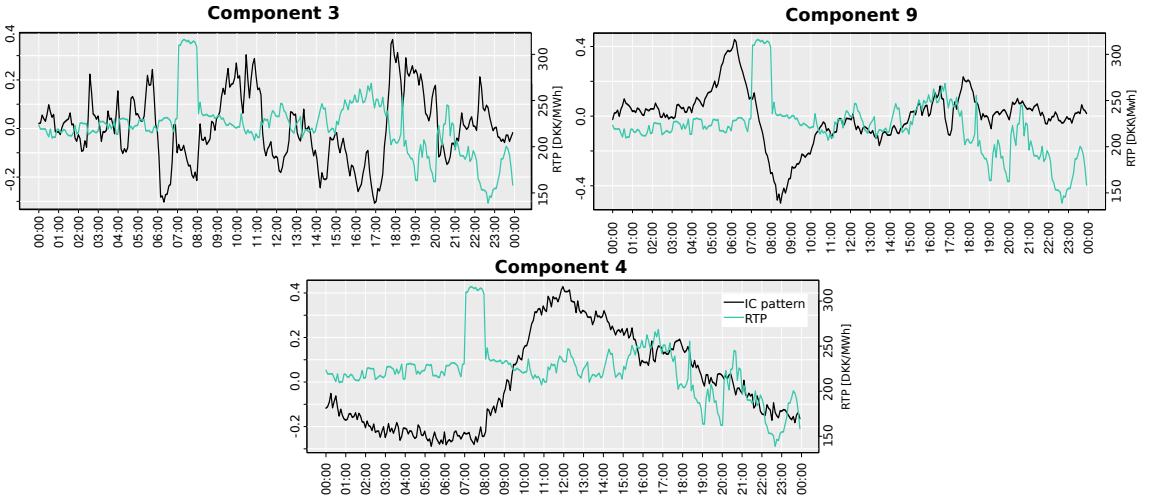


Figure 5.11: Representation of the ICs number 3, 4 and 9 with the RTP in the same period (06-01-2015)

for each IC to evaluate which groups (Table 5.4) and which clusters (Table 5.5) contribute highly (positively or negatively) to each IC. From Table 5.4, Figure 5.11 and our expertise of the different equipment, the ICs numbers 3, 4 and 9 are construed. The ICs illustrations in Figure 5.11 are the positive orientation of the ICs.

IC 3 responds to every RTP variation, which means that the loads with a large positive weight (above the 95% quantile) for IC 3 responds to every variation of the RTP. Inversely, if the weight is negative (under 5% quantile), they follow the RTP. From Table 5.4, groups 3, and 4 are the ones following the variation of the RTP, while group 6 is the one responding inversely to every variation of the RTP. For the clusters, the most negative weights are observed for clusters 1 and 2. The highest positive are from clusters 3, 4, and 5, which confirms their response observed with the FIRs.

IC 4 displays a pattern, which is not correlated to the RTP. With a positive weight, it displays an increase in consumption from 08:00 to 12:00. When the weight is negative, it is the opposite, the consumption drops from 08:00 to 12:00. This pattern seems to materialize the presence of inhabitants during the day when the weight is positive or an economy setting when negative. This hypothesis can be confirmed by the large negative weights of clusters 3 and 5, which are the ones with the largest share of holiday houses and the large positive weight of cluster 1 and 2, mostly composed of inhabited houses (Table 5.5). It also mitigates the idea that the type of housing is a factor facilitating the response to a small variation of price, which corresponds to IC 3.

IC 9 displays a good example of anticipation (positive) of the price variation at 07:00 or a rebound (negative). It anticipates the RTP peak by heating up and using the thermal inertia of the house to

stop heating during the peak. It shows how the controllers use the day-ahead price or hour-ahead price to schedule the heat by displacing some energy before or after the RTP peak. Groups 6 and 4 are both divided in approximately the same proportion between large positive and large negative weights on IC 9, which tells that the choice of anticipating the peak or having a rebound is not linked to a specific controller but may be dictated by the presence of EH as both groups are equipped with them (Table 5.4). Clusters 1 and 2 combined are in the same numbers under 5% and above 95%. This is due to the composition of these clusters, which have a majority of households from groups 4 and 6, which presents the same cleavage into half rebound before and half rebound after.

The ICs 3, 4 and 9 thoroughly analyzed here illustrate three different load behavior, (i) response to RTP (IC 3), (ii) specific settings of the thermostat (IC 4), and (iii) power consumption scheduling (IC 9).

5.4 Conclusion

DSM evaluates the amount of energy displaced in time from the aggregated consumption of customers as a service to the grid. Hence, when operating DR, the number of active participants is a crucial aspect that influences the capacity of the response and subsequently the DSM impact. The detection of non-responsive households using simple statistical tests is efficient to manage to pool the participants. The quantification of the responses is commonly done to evaluate the response. However, assessing the characteristics of the response, like obtaining the characteristics of a generator or a battery, is fundamental to understand how the aggregated response could look like and to deliver a more finely-tuned response.

We could imagine implementing a DR where different prices are sent to households in order to obtain the expected aggregated load. It assumes stationarity of the time series, which are known to be non-stationary when looking at single consumption time series, but it is a fair assumption when looking at the aggregated consumption time series. In the continuity of this idea, a unit commitment model with different prices per subgroup of households could be implemented to evaluate the benefit of broadcasting different prices.

The ICA algorithm could be extended to an online (i.e., sliding window) ICA to take advantage of the stream nature of the metering data and go around the computing limitation of ICA. A sliding window PCA with component tracking would solve the problem. A step-wise model selection of the ICs could also be implemented to select the ICs that respond to the RTP.

CHAPTER 6

Conclusions and Future Work

In the present thesis, we addressed the four main research questions related to metering data analytics to provide feedback to utilities. The first research question gives the context of energy digitization from a societal point of view and evaluates under which conditions this work can have a positive impact on society. It discusses the necessity of building ethical smart grids that are customer-centric and do not see customers as consumers but as collaborators of the grid operations. It also emphasizes the need for smart grids, metering data and analytics to reduce the impact of power systems on climate change. The remaining questions focus on proposing and evaluating data analytics methodologies that take advantage of the stream nature of metering data. Load analysis, and more precisely load profiling, is explored and a dynamic approach to load profiling is implemented to address the second question. Load management with detection and characterization of assets as well as Demand Response (DR) quantification and characterization is also developed. The detection and characterization of assets address the third research question. To answer it, a Non-Intrusive Load Monitoring (NILM) approach is proposed and an application to the detection of Heat Pumps (HPs) is tested on real-world data. The last research question concerns DR data analysis. First, a protocol to detect non-responsive participants is described and tested. Then a methodology is proposed to quantify and characterize the response without the support of a baseline.

6.1 Overview of contributions

A number of original contributions to the state-of-the-art were presented in this thesis related to metering data analytics to provide insights on the demand side to the utilities. A non-technical introduction defends the need for ethics in the development of analytics-based capabilities in smart grids. The technical contributions focused on electric load analysis and load management.

In the context of our modern societies, information technologies are ubiquitous and they are legitimately raising concerns among citizens. It becomes necessary to understands under which conditions the smart grid can be a viable concept from a customer point of view. It appears that there is an important misalignment between how smart grids are promoted to the public and how these are developed and implemented. First, there is a phase shifting of the benefit of smart meters deployments. The utilities benefit today of smart meters, with more accurate billing and data while customers are still waiting for the savings and the better control on their electricity consumption as the use of raw metering data (when available) remains limited. Nevertheless, customers are still expected to provide flexibility to the grid in the near future. The smart meters are crystallizing the lack of trust the customers have in the utilities as a consequence to bad-practices in terms of communication and customer service. Climate change awareness has raised since the COP21 and the Paris Agreements in 2015. It tends to prove that a fringe of the population is ready to actively change the way they use energy and is just waiting for opportunities. Ethical smart grids, where the customers and utilities collaborate and share the benefit of the analytics and the smartness of

the grid, should be the end goal. All the technology and engineering work would be pointless if the customers do not adopt it.

An online adaptive clustering algorithm was proposed and evaluated, in order to exploit streams of metering data and generate efficiently typical load profiles. It is an improvement of the state-of-the-art load profiling as it handles (i) stationary (ii) non-stationary, and (iii) disruptive loads. Furthermore, as it is online each iteration (i.e., time step) relies on the typical load profiles computed at the previous time step, which makes each iteration inexpensive to compute (less than 2 minutes for more than 10 000 loads on a standard desktop computer). Last but not least, the adaptive part of the algorithm adjusts the number of clusters to the underlying number of typical load profiles present in the dataset (according to the parameter set). One of the benefits is that the maximum distance from a cluster center is thresholded. Another benefit is that it isolates outliers instead of forcing to assign them to a typical load profile. Hence, it could also be used for anomaly detection or any application that track the daily patterns over time.

The detection and characterization of potentially flexible appliances like Electric Vehicle (EV), HP, Electric Heating (EH), Air Conditioner (AC) using available metering data constitute one of the new business models utilities/aggregators have to implement. Indeed, before harnessing demand flexibility, assets have to be listed and recruited to form a portfolio. NILM is probably the most studied analytics approach to obtain individual appliances consumption from the aggregated metering data. The methodology proposed, combining Orthogonal Matching Pursuit (OMP), Gaussian Mixture Model (GMM) and community detection, performs as well as the supervised state-of-the-art NILM methodologies while it is unsupervised. Moreover, its performance is marginally affected by the degradation of the resolution of the data. Hence, it fulfills the requirements for real-world implementation. Finally, an application of the methodology to the detection and characterization of domestic HPs is demonstrated using the EcoGrid EU data. The method is based on a Bayesian framework, which means that the load characteristics of HPs (e.g., power amplitude, operation times, and frequency of an activation) can be inferred with new HPs identified.

The assets can then be recruited into a DR program that can harness their potential flexibility. Price based-DR that passes the marginal cost of generation over to the customer is the most popular as the customers are rewarded or penalized according to their contribution to the grid balance status. The two-way communication protocol of smart meters associated with a controller in each household makes such a framework possible. However, it remains a complex system that relies on many pieces of equipment. In the real-world failures happen and have to be detected quickly to maintain the response capacity as high as possible. Hence, the responsiveness protocol using clinical trial approach is a fast and straightforward way to evaluate responsiveness and detect malfunctioning pieces of equipment. When responsive, the response is usually assessed in comparison to a baseline, which is either modeled using data collected before the DR program starts, or a Customer Base Line (CBL) generated by a representative set of households that are not included in the DR program. Both approaches present a bias, (i) the model relies on the parameter selected, (ii) statistical representativity of the CBL is almost impossible to reach or at least impossible to test. Hence, developing a methodology to both quantify and characterize responses without a baseline is moving the state-of-the-art forward. Instead of considering a baseline, our approach evaluates the response in relation to the Real Time Pricing (RTP), which we know is as the exact incentive signal. Hence, we assume to obtain a less biased evaluation of the responses.

The contributions of this thesis cover a wide range of applications of metering data analytics in the field of load analysis and load management in the context of digitization of the energy sector. Precisely we focused on exploiting the stream nature of the data to identify patterns (load profiling, detection of assets) and separating consumption signals (NILM, Independent Component Analysis (ICA)). The increase computing capacities and improved data acquisition will make large scale implementation of such methodologies possible. Quby is an example of such an application with its product Toon. Both utilities and customers need more comprehensive information on their consumption. Metering data analytics as a specific field is still in his infancy and many developments have to be expected in the near future.

6.2 Perspective for future research

Based on the ideas investigated and the results obtained, we identified several paths for future research to be explored. The development of machine learning approaches tailored to current and future needs in the energy field will provide better analytics for utilities but also to customers, as well as emerging actors of the energy system.

The online adaptive clustering algorithm we proposed opens the door to new types of analytics for profiling, like the analysis of time series formed by the assignments time series. With a tracking of the clusters over time, the variability, the frequency at which a customer change cluster and the number of cluster visited, yield a broader perspective on consumption characteristics (like the Symbolic Aggregate Approximation (SAX) described in Section 3.2.1). The online adaptive clustering algorithm could also be combined with the characterization and quantification of response approach to analyze large scale DR implementations and provide feedback to the operators on the last day responses. In a broader context, the methodology and resulting algorithm are generic: they have the potential to be used in many fields where profiling based on data streams is needed.

The main caveats of supervised learning approaches are known: they require large amounts of data, they are non-adaptive to unseen data during training, and they are potentially computationally expensive. Unsupervised approaches, in turn, have a major default: they require long periods for tuning as parameters are set empirically. Hence, the best solution is probably in-between. A Bayesian framework like the HP detection presented in Section 4.5 takes the best of both worlds. The *prior* can be weak and the inference of data tunes it, or it can be defined based on the *prior* (expert) knowledge gathered for the problem at hand. Indeed, it is rare to face a situation where no *prior* knowledge exists. Beyond the case of inference on one application, transfer learning, that consists in using the knowledge gathered from a previous problem as input to solving a related, but different problem, can be operated in a Bayesian framework. For example, the distributions learned for the detection of HPs are inferred on EcoGrid EU data. They can then be used to do HP detection in other projects or regions.

A growing number of analytics approaches will be designed and implemented to provide insights on the demand. In the meanwhile, the need for data will grow and data will increasingly become valuable. By this time, utilities will be interested in buying and selling data. Hence, it will be time to explore the design of market structures to exchange data or information about the loads without disclosing sensitive information on customers. Today's largest data application are data brokers, processing Internet browsing data for the purpose of online targeted advertising. They also exchange data following the law on privacy and data protection. The digitization of the

energy sector is just starting and the foundations are laid down today. The monopoly of utilities has limited data exchange, the liberalization of the energy market is reshaping utilities and a deregulation of the energy data will most likely come next. It is then vital to think about how to exchange data without exposing customers.

A trade-off between privacy and the use of data is being considered as part of the deployment of smart grids. It is legitimate to collect and use data from customers with the aim of improving social welfare. It must be said that today's use of data is somehow limited and that impact on society is moderate, while the privacy is mostly challenged by the collection of data rather than its analysis. However, the fast development of analytics will modify the balance due to more invasive analysis of metering data (NILM being the first in the line). Hence, future analytics implementations have to keep this balance, the ethical optimal balance being that the analytics limits the privacy breaching while yielding benefits to both individual customers and society.

Bibliography

[1] Rob Kitchin. The ethics of smart cities and urban science. *Phil. Trans. R. Soc. A*, 374(2083):20160115, 2016.

[2] Ahmad Faruqui, Dan Harris, and Ryan Hledik. Unlocking the 53 billion Euro savings from smart meters in the EU: How increasing the adoption of dynamic tariffs could make or break the EU's smart grid investment. *Energy Policy*, 38(10):6222–6231, 2010.

[3] 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, 2012.

[4] European Parliament. Directive 2012/27/EU. *Official Journal of the European Union*, L315/1(October):1–56, 2012.

[5] Hassan Farhangi. The path of the smart grid. *IEEE power and energy magazine*, 8(1), 2010.

[6] Paul Denholm, Erik Ela, Brendan Kirby, and Michael Milligan. Role of energy storage with renewable electricity generation. Technical report, National Renewable Energy Lab.(NREL), Golden, CO (United States), 2010.

[7] Goran Strbac. Demand side management: Benefits and challenges. *Energy Policy*, 36(12):4419–4426, dec 2008.

[8] Yi Wang, Qixin Chen, Tao Hong, and Chongqing Kang. Review of smart meter data analytics: Applications, methodologies, and challenges. *IEEE Transactions on Smart Grid*, 2018.

[9] Xiufeng Liu and Per Sieverts Nielsen. Streamlining smart meter data analytics. In *Proceedings of the 10th Conference on Sustainable Development of Energy, Water and Environment Systems*. International Centre for Sustainable Development of Energy, Water and Environment Systems, 2015.

[10] Damminda Alahakoon and Xinghuo Yu. Advanced analytics for harnessing the power of smart meter big data. In *Proceedings of the 2013 IEEE International Workshop on Inteligent Energy Systems (IWIES)*, pages 40–45. IEEE, 2013.

[11] Alexandra B. Klass and Elizabeth J. Wilson. Remaking energy: The critical role of energy consumption data. *Cal. L. Rev.*, 104:1095, 2016.

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

[13] Robin Hoenkamp, George B. Huitema, and Adrienne J. C. de Moor-van Vugt. Neglected consumer: The case of the smart meter rollout in the Netherlands. *Renewable Energy L. & Policy Rev.*, page 269, 2011.

[14] Yi Wang, Qixin Chen, Chongqing Kang, Mingming Zhang, Ke Wang, and Yun Zhao. Load profiling and its application to demand response: A review. *Tsinghua Science and Technology*, 20(2):117–129, 2015.

[15] Ignacio Benítez, Alfredo Quijano, José Luis Díez, and Ignacio Delgado. Dynamic clustering segmentation applied to load profiles of energy consumption from Spanish customers. *International Journal of Electrical Power and Energy Systems*, 55:437–448, 2014.

[16] Hongliang Fei, Younghun Kim, Sambit Sahu, Milind Naphade, Sanjay K Mamidipalli, and John Hutchinson. Heat pump detection from coarse grained smart meter data with positive and unlabeled learning. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1330–1338. ACM, 2013.

[17] Yijia Cao, Shengwei Tang, Canbing Li, Peng Zhang, Yi Tan, Zhikun Zhang, and Junxiong Li. An optimized EV charging model considering TOU price and SOC curve. *IEEE Transactions on Smart Grid*, 3(1):388–393, 2012.

[18] Pierluigi Siano. Demand response and smart grids – A survey. *Renewable and Sustainable Energy Reviews*, 30:461–478, 2014.

[19] Philip Bartholomew, Wayne Callender, Cheryl Hindes, Clifford Grimm, Kathy Johnson, Mary Straub, Dave Williams, Mark Williamson, Debbie Hayes, Wilbur Johnson, et al. Demand response measurement & verification. *AEIC Load Research Committee*, 2009.

[20] Christine Horne, Brice Darras, Elyse Bean, Anurag Srivastava, and Scott Frickel. Privacy, technology, and norms: The case of smart meters. *Social Science Research*, 2015.

[21] Niels Andersen, Heiko Vester, and Tommy Heine Bentsen. Grid planning by integrate customer meters. In *Proceedings of the 24th International Conference on Electricity Distribution*, pages 1–5. CIRED, 2017.

[22] Ioannis P. Panapakidis, Minas C. Alexiadis, and Grigoris K. Papagiannis. Enhancing the clustering process in the category model load profiling. *IET Generation, Transmission Distribution*, 9(7):655–665, 2015.

[23] Carrie K. 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, 2013.

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

[25] Zico J. Kolter and Matthew J. Johnson. REDD: A public data set for energy disaggregation research. In *Proceedings of the Workshop on Data Mining Applications in Sustainability (SIGKDD), San Diego, CA*, volume 25, pages 59–62. Citeseer, 2011.

[26] Jack Kelly and William Knottenbelt. Neural nilm: Deep neural networks applied to energy disaggregation. In *Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments*, pages 55–64. ACM, 2015.

[27] Anthony Faustine, Nerey Henry Mvungi, Shubi Kaijage, and Kisangiri Michael. A Survey on Non-Intrusive Load Monitoring Methodies and Techniques for Energy Disaggregation Problem. *arXiv*, (March), 2017.

[28] Yi Ding, Salvador Pineda, Preben Nyeng, Jacob Østergaard, Emil M. Larsen, and Qiuwei Wu. Real-time market concept architecture for EcoGrid EU - A prototype for European smart grids. *IEEE Transactions on Smart Grid*, 4(4):2006–2016, 2013.

[29] Johanna L. Mathieu, Phillip N. Price, Sila Kiliccote, and Mary Ann Piette. Quantifying changes in building electricity use, with application to demand response. *IEEE Transactions on Smart Grid*, 2(3):507–518, 2011.

[30] Emil M. Larsen, Pierre Pinson, Fabian Leimgruber, and Florian Judex. Demand response evaluation and forecasting - methods and results from the ecogrid eu experiment. *Sustainable Energy, Grids and Networks*, 10:75–83, 2017.

[31] European Parliament. Directive 2009/29/EC. *Official Journal of the European Union*, 140:63–87, 2009.

[32] European Commission. Energy efficiency plan 2011. Technical report, European Commission, 2011.

[33] European Union. Directive of 2009/72/EC. *Official Journal of the European Union*, L211(August):L 211/55 – L 211/93, 2009.

[34] Raquel Bertoldo, Marc Poumadère, and Luis Carlos Rodrigues. When meters start to talk: The public's encounter with smart meters in France. *Energy Research and Social Science*, 2015.

[35] Vincenzo Giordano, Flavia Gangale, Gianluca Fulli, Manuel Sánchez Jiménez, Ijeoma Onyeji, Alexandru Colta, Ioulia Papaioannou, Anna Mengolini, Corina Alecu, Tauno Ojala, et al. Smart grid projects in europe. *JRC Ref Rep Sy*, 8, 2011.

[36] Gregg P. Zachary. Saving smart meters from a backlash. *IEEE Spectrum*, 8(48):8, 2011.

[37] Vagelis Papakonstantinou and Dariusz Kloza. Legal protection of personal data in smart grid and smart metering systems from the European perspective. In *Smart Grid Security*, chapter 2, pages 41–129. London: Springer, 2015.

[38] Spyros G. Tzafestas. Ethics and law in the internet of things world. *Smart Cities*, 1(1):98–120, 2018.

[39] Maya Jegen and Xavier D Philion. Power and smart meters: A political perspective on the social acceptance of energy projects. *Canadian Public Administration*, 60(1):68–88, 2017.

[40] United Nations. United Nations universal declaration of Human rights 1948. *Office of the Hight Comissioner for Human Rights*, page 11, 1949.

[41] European Court of Human Rights. European convention on Human rights. *Convention for the Protection of Human Rights and Fundamental Freedoms*, page 30, 1950.

[42] of Europe Council. Convention for the protection of individuals with regard to automatic processing of personal data, 1981.

[43] European Parliament. Directive 95/46/EC. Technical report, European Parliament, 1995.

[44] European Union. Regulation 2016/679. *Official Journal of the European Union*, 2001, 2016.

[45] European Commission. Benchmarking smart metering deployment in the EU-27 with a focus on electricity. *European Commission*, pages 1–10, 2014.

[46] Smart Grids Task Force Expert Group 1- Standards and Interoperability. My energy data. Technical report, European Commission, 2016.

[47] Helen Nissenbaum. A contextual approach to privacy online. *Daedalus*, 140(4):32–48, 2011.

[48] Irving Wladawsky-Berger. The big data era is here, feb 2015.

[49] OECD. Exploring data-driven innovation as a new source of growth: Mapping the policy issues raised by "Big Data". *OECD Digital Economy Papers*, no. 222, 2013.

[50] European Commission. What is personal data?, 2018.

[51] European Parliament. Directive 2009/136/EC. *Official Journal of the European Union*, 337:11–36, 2009.

[52] Pierre Laperdrix, Walter Rudametkin, and Benoit Baudry. Beauty and the beast: Diverting modern web browsers to build unique browser fingerprints. In *Proceedings of the 37th IEEE Symposium on Security and Privacy (S&P 2016)*, pages 878–894, 2016.

[53] Rona Molla. Google leads the world in digital and mobile ad revenue, 2018.

[54] M. Kosinski, D. Stillwell, and T. Graepel. Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences*, 110(15):5802–5805, 2013.

[55] Robert R. McCrae and Oliver P. John. An introduction to the five-factor model and its applications. *Journal of personality*, 60(2):175–215, 1992.

[56] Samuel D. Gosling, Adam A. Augustine, Simine Vazire, Nicholas Holtzman, and Sam Gaddis. Manifestations of personality in online social networks: Self-reported Facebook-related behaviors and observable profile nformation. *Cyberpsychology, Behavior, and Social Networking*, 14(9):483–488, 2011.

[57] Michal Kosinski, Sandra C. Matz, Samuel D. Gosling, Vesselin Popov, and David Stillwell. Facebook as a research tool for the social sciences: Opportunities, challenges, ethical considerations, and practical guidelines. *American Psychologist*, 70(6):543–556, 2015.

[58] Ralph Gross and Alessandro Acquisti. Information revelation and privacy in online social networks. In *Proceedings of the 2005 ACM workshop on Privacy in the electronic society*, page 80, 2005.

[59] Yabing Liu, Krishna P. Gummadi, Balachander Krishnamurthy, and Alan Mislove. Analyzing facebook privacy settings: User expectations vs. reality. In *Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference*, pages 61–70. ACM, 2011.

[60] David W. Mcdonald and Mark S. Ackerman. Expertise recommender: A flexible recommendation system and architecture. In *Proceedings of the ACM conference on Computer supported cooperative work*, pages 231–240, 2000.

[61] Yves-Alexandre De Montjoye, César A. Hidalgo, Michel Verleysen, and Vincent D. Blondel. Unique in the crowd: The privacy bounds of human mobility. *Scientific reports*, 3, 2013.

[62] Solon Barocas and Helen Nissenbaum. Big data's end run around anonymity and consent. *Privacy, big data, and the public good: Frameworks for engagement*, 1:44–75, 2014.

[63] Alessandro Acquisti and Ralph Gross. Predicting social security numbers from public data. *Proceedings of the National Academy of Sciences of the United States of America*, 106(27):10975–10980, 2009.

[64] Pedro Ponce, Kenneth Polasko, and Arturo Molina. End user perceptions toward smart grid technology: Acceptance, adoption, risks, and trust. *Renewable and Sustainable Energy Reviews*, 60:587–598, 2016.

[65] Tamar Krishnamurti, Daniel Schwartz, Alexander Davis, Baruch Fischhoff, Wändi Bruine de Bruin, Lester Lave, and Jack Wang. Preparing for smart grid technologies: A behavioral decision research approach to understanding consumer expectations about smart meters. *Energy Policy*, 41:790–797, 2012.

[66] Serge Gutwirth, Ronald Leenes, Paul De Hert, and Yves Poullet. *European data protection: Coming of age*. Springer Science & Business Media, 2012.

[67] Charlie Wilson, Tom Hargreaves, and Richard Hauxwell-Baldwin. Benefits and risks of smart home technologies. *Energy Policy*, 2017.

[68] Himanshu Khurana, Mark Hadley, Ning Lu, and Deborah A. Frincke. Smart-grid security issues. *IEEE Security & Privacy*, 8(1), 2010.

[69] Soren Finster and Ingmar Baumgart. Privacy-aware smart metering: A survey. *IEEE Communications Surveys & Tutorials*, 16(3):1732–1745, 2014.

[70] Gianfranco Chicco. Customer behaviour and data analytics. In *Proceedings of the 2016 International Conference and Exposition on Electrical and Power Engineering, EPE 2016*, pages 771–779, 2016.

[71] Hunt Allcott. Social norms and energy conservation. *Journal of public Economics*, 95(9-10):1082–1095, 2011.

[72] Patrick McDaniel and Stephen McLaughlin. Security and privacy challenges in the smart grid. *IEEE Security & Privacy*, 7(3), 2009.

[73] Daniel E. M. Bondy, Kai Heussen, Oliver Gehrke, and Anders Thavlov. A functional reference architecture for aggregators. In *Proceedings of the IEEE International Conference on Emerging Technologies and Factory Automation, ETFA*, volume 2015-Octob, pages 1–7, 2015.

[74] Guido Pepermans. Valuing smart meters. *Energy Economics*, 2014.

[75] Christoph Klemenjak and Peter Goldsborough. Non-intrusive load monitoring: A review and outlook. *archiv*, 2016.

[76] Rainer Knyrim and Gerald Trieb. Smart metering under eu data protection law. *International Data Privacy Law*, 1(2):121–128, 2011.

[77] Robert M. Lee, Michael J. Assante, and Tim Conway. Analysis of the cyber attack on the ukrainian power grid: Defense use case. *SANS Institute & the Electricity Information Sharing and Analysis Center*, 2016.

[78] Ilhami Colak, Seref Sagiroglu, Gianluca Fulli, Mehmet Yesilbudak, and Catalin-Felix Covrig. A survey on the critical issues in smart grid technologies. *Renewable and Sustainable Energy Reviews*, 54:396–405, 2016.

[79] Mary J. Culnan and Robert J. Bies. Consumer privacy: Balancing economic and justice considerations. *Journal of social issues*, 59(2):323–342, 2003.

[80] Zheng Hu, Jin-ho Kim, Jianhui Wang, and John Byrne. Review of dynamic pricing programs in the US and Europe: Status quo and policy recommendations. *Renewable and Sustainable Energy Reviews*, 42:743–751, 2015.

[81] Thomas B. Smith. Electricity theft: A comparative analysis. *Energy Policy*, 32(18):2067–2076, 2004.

[82] EU. 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, 2009.

[83] Recovery Act. American recovery and reinvestment act of 2009. *Public Law*, pages 111–5, 2009.

[84] Joerg Dickert and Peter Schegner. Residential load models for network planning purposes. In *Proceedings of the 2010 Modern Electric Power Systems*, pages 1–6. IEEE, 2010.

[85] Sten Velander. Methods of operational analysis applied to distribution of electric power. *Journal of Teknisk Tidskrift*, 82:293–299, 1952.

[86] Rui Xu and Donald C. Wunsch II. *Clustering*. Wiley-IEEE Press, 2008.

[87] Gianfranco Chicco, Roberto Napoli, and Federico Piglione. Comparisons among clustering techniques for electricity customer classification. *IEEE Transactions on Power Systems*, 21(2):933–940, 2006.

[88] Xiaoli Li, Chris P Bowers, and Thorsten Schnier. Classification of energy consumption in buildings with outlier detection. *IEEE Transactions on Industrial Electronics*, 57(11):3639–3644, 2010.

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

[90] Teemu Räsänen, Dimitrios Voukantsis, Harri Niska, Kostas Karatzas, and Mikko Kolehmainen. Data-based method for creating electricity use load profiles using large amount of customer-specific hourly measured electricity use data. *Applied Energy*, 87(11):3538–3545, 2010.

[91] Gianfranco Chicco, Roberto Napoli, Federico Piglione, Petru Postolache, Mircea Scutariu, and Cornel Toader. Load pattern-based classification of electricity customers. *IEEE Transactions on Power Systems*, 19(2):1232–1239, 2004.

[92] Sergio V. Verdú, Mario O. García, Carolina Senabre, Antonio G. Marín, and Francisco J. G. 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, 2006.

[93] Fintan McLoughlin, Aidan Duffy, and Michael Conlon. A clustering approach to domestic electricity load profile characterisation using smart metering data. *Applied Energy*, 141:190–199, 2015.

[94] Ervin D. Varga, Sándor F. Beretka, Christian Noce, and Gianluca Sapienza. Robust real-time load profile encoding and classification framework for efficient power systems operation. *IEEE Transactions on Power Systems*, 30(4):1897–1904, 2015.

[95] Shiyin Zhong and Kwa-Sur Tam. Hierarchical classification of load profiles based on their characteristic attributes in frequency domain. *IEEE Transactions on Power Systems*, 30(5):2434–2441, 2015.

[96] Kevin Mets, Frederick Depuydt, and Chris Develder. Two-stage load pattern clustering using fast wavelet transformation. *IEEE Transactions on Smart Grid*, 7(5):2250–2259, 2016.

[97] Yong Xiao, Jinfeng Yang, Huakun Que, Mark Junjie Li, and Qin Gao. Application of wavelet-based clustering approach to load profiling on ami measurements. In *Proceedings of the 2014 China International Conference on Electricity Distribution (CICED)*, pages 1537–1540. IEEE, 2014.

[98] Anestis Antoniadis, Xavier Brossat, Jairo Cugliari, and Jean-Michel Poggi. Clustering functional data using wavelets. *International Journal of Wavelets, Multiresolution and Information Processing*, 11(01):1350003, 2013.

[99] Antonio Notaristefano, Gianfranco Chicco, and Federico Piglione. Data size reduction with symbolic aggregate approximation for electrical load pattern grouping. *IET Generation, Transmission & Distribution*, 7(2):108–117, 2013.

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

[101] David Arthur and Sergei Vassilvitskii. K-means++: The advantages of careful seeding. In *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 1027–1035. Society for Industrial and Applied Mathematics, 2007.

[102] Yanlong Wang, Li Li, and Qinmin Yang. Application of clustering technique to electricity customer classification for load forecasting. In *Proceedings of the 2015 IEEE International Conference on Information and Automation, ICIA 2015 - In conjunction with 2015 IEEE International Conference on Automation and Logistics*, pages 1425–1430. IEEE, 2015.

[103] Sérgio Ramos, João M. Duarte, F. Jorge Duarte, and Zita Vale. A data-mining-based methodology to support MV electricity customers' characterization. *Energy and Buildings*, 91:16–25, 2015.

[104] Stephen Haben, Colin Singleton, and Peter Grindrod. Analysis and clustering of residential customers energy behavioral demand using smart meter data. *IEEE Transactions on Smart Grid*, 7(1):136–144, 2016.

[105] Bruce Stephen, Antti J Mutanen, Stuart Galloway, Graeme Burt, and Pertti Järventausta. Enhanced load profiling for residential network customers. *IEEE Transactions on Power Delivery*, 29(1):88–96, 2014.

[106] Ramon Granell, Colin J. Axon, and David C. H. Wallom. Clustering disaggregated load profiles using a Dirichlet process mixture model. *Energy Conversion and Management*, 92:507–516, 2015.

[107] Joaquim L. Viegas, Susana M. Vieira, João M. C. Sousa, R. Melicio, and Victor M. F. Mendes. Electricity demand profile prediction based on household characteristics. In *Proceedings of the International Conference on the European Energy Market, EEM*, volume 2015-Augus, pages 1–5. IEEE, 2015.

[108] Gianfranco Chicco and Irinel-Sorin Ilie. Support vector clustering of electrical load pattern data. *IEEE Transactions on Power Systems*, 24(3):1619–1628, 2009.

[109] Barney D. Pitt and Daniel S. Kirschen. Application of data mining techniques to load profiling. In *Proceedings of the 21st International Conference on Power Industry Computer Applications. Connecting Utilities. PICA 99. To the Millennium and Beyond (Cat. No. 99CH36351)*, pages 131–136. IEEE, 1999.

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

[111] Gianfranco Chicco, Octavian Marcel Ionel, and Radu Porumb. Formation of load pattern clusters exploiting ant colony clustering principles. In *Proceedings of the IEEE EuroCon 2013*, pages 1460–1467, 2013.

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

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

[114] Ignacio Benítez, José-Luis Díez, Alfredo Quijano, and Ignacio Delgado. Dynamic clustering of residential electricity consumption time series data based on Hausdorff distance. *Electric Power Systems Research*, 140:517–526, 2016.

[115] Thomas Grotkjær, Ole Winther, Birgitte Regenberg, Jens Nielsen, and Lars K. Hansen. Robust multi-scale clustering of large DNA microarray datasets with the consensus algorithm. *Bioinformatics*, 22(1):58–67, 2006.

[116] T. Warren Liao. Clustering of time series data - A survey. *Pattern Recognition*, 38(11):1857–1874, 2005.

[117] John Paparrizos and Luis Gravano. K-shape: Efficient and accurate clustering of time series. In *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data*, pages 1855–1870. ACM, 2015.

[118] Ahlame D. Chouakria and Panduranga N. Nagabhushan. Adaptive dissimilarity index for measuring time series proximity. *Advances in Data Analysis and Classification*, 1(1):5–21, 2007.

[119] Alon Efrat, Quanfu Fan, and Suresh Venkatasubramanian. Curve matching, time warping, and light fields: New algorithms for computing similarity between curves. *Journal of Mathematical Imaging and Vision*, 27(3):203–216, 2007.

[120] Kuldip K. Paliwal, Anant Agarwal, and Sarvajit S. Sinha. A modification over sakoe and chiba's dynamic time warping algorithm for isolated word recognition. *Signal Processing*, 4(4):329–333, 1982.

[121] Adam Meyerson. Online facility location. In *Proceedings of the 2001 IEEE International Conference on Cluster Computing*, pages 426–431. IEEE, 2001.

[122] Edo Liberty, Ram Sriharsha, and Maxim Sviridenko. An algorithm for online k-means clustering. In *Proceedings of the Eighteenth Workshop on Algorithm Engineering and Experiments (ALENEX)*, pages 81–89. SIAM, 2016.

[123] ENTSO-E. Consumption data hourly load value of a specific country for a specific day March 27th 2013, 2013.

[124] 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, 2012.

[125] Michael Zeifman and Kurt Roth. Nonintrusive appliance load monitoring: Review and outlook. *IEEE transactions on Consumer Electronics*, 57(1):76–84, 2011.

[126] Nur Farahin Esa, Md Pauzi Abdullah, and Mohammad Yusri Hassan. A review disaggregation method in non-intrusive appliance load monitoring. *Renewable and Sustainable Energy Reviews*, 66:163–173, 2016.

[127] Ahmed Zoha, Alexander Gluhak, Michele Nati, and Muhammad Ali Imran. Low-power appliance monitoring using factorial hidden markov models. In *Proceedings of the 2013 IEEE Eighth International Conference on Intelligent Sensors, Sensor Networks and Information Processing*, pages 527–532. IEEE, 2013.

[128] Hyungsul Kim, Manish Marwah, Martin Arlitt, Geoff Lyon, and Jiawei Han. Unsupervised disaggregation of low frequency power measurements. In *Proceedings of the 2011 SIAM international conference on data mining*, pages 747–758. SIAM, 2011.

[129] Zoubin Ghahramani and Michael I. Jordan. Factorial hidden markov models. In *Advances in Neural Information Processing Systems*, pages 472–478, 1996.

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

[131] David G. Forney. The viterbi algorithm. *Proceedings of the IEEE*, 61(3):268–278, 1973.

[132] Stephen Makonin, Fred Popowich, Ivan V Bajić, Bob Gill, and Lyn Bartram. Exploiting HMM sparsity to perform online real-time nonintrusive load monitoring. *IEEE Transactions on Smart Grid*, 7(6):2575–2585, 2016.

[133] Zico J. Kolter and Tommi Jaakkola. Approximate inference in additive factorial HMMs with application to energy disaggregation. In *Artificial Intelligence and Statistics*, pages 1472–1482, 2012.

[134] Zhenyu Guo, Z Jane Wang, and Ali Kashani. Home appliance load modeling from aggregated smart meter data. *IEEE Transactions on power systems*, 30(1):254–262, 2015.

[135] Oliver Parson, Siddhartha Ghosh, Mark Weal, and Alex Rogers. Non-intrusive load monitoring using prior models of general appliance types. In *Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence*, 2012.

[136] Nipun Batra, Oliver Parson, Mario Berges, Amarjeet Singh, and Alex Rogers. A comparison of non-intrusive load monitoring methods for commercial and residential buildings. *arXiv preprint arXiv:1408.6595*, 2014.

[137] Yann LeCun, Bernhard Boser, John S. Denker, Donnie Henderson, Richard E. Howard, Wayne Hubbard, and Lawrence D. Jackel. Backpropagation applied to handwritten zip code recognition. *Neural computation*, 1(4):541–551, 1989.

[138] Henning Lange and Mario Bergés. The neural energy decoder: Energy disaggregation by combining binary subcomponents. In *Proceedings of the NILM-2016 workshop.*, 2016.

[139] Chaoyun Zhang, Mingjun Zhong, Zongzuo Wang, Nigel Goddard, and Charles Sutton. Sequence-to-point learning with neural networks for non-intrusive load monitoring. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[140] Wan He and Ying Chai. An empirical study on energy disaggregation via deep learning. In *Proceedings of the 2016 2nd International Conference on Artificial Intelligence and Industrial Engineering (AIIE 2016)*. Atlantis Press, 2016.

[141] Alex Graves. *Supervised sequence labelling with recurrent neural networks*. Springer, 2012.

[142] Lukas Mauch and Bin Yang. A new approach for supervised power disaggregation by using a deep recurrent lstm network. In *Proceedings of the 2015 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pages 63–67. IEEE, 2015.

[143] Nipun Batra, Amarjeet Singh, and Kamin Whitehouse. If you measure it, can you improve it? exploring the value of energy disaggregation. In *Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments*, pages 191–200. ACM, 2015.

[144] Jack Kelly and William Knottenbelt. UK-DALE: A dataset recording UK domestic appliance-level electricity demand and whole-house demand. *ArXiv*, 59, 2014.

[145] Adrian Filip. Blued: A fully labeled public dataset for event-based nonintrusive load monitoring research. In *Proceedings of 2nd Workshop on Data Mining Applications in Sustainability (SustKDD)*, page 2012, 2011.

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

[147] Sean Barker, Sandeep Kalra, David Irwin, and Prashant Shenoy. Nilm redux: The case for emphasizing applications over accuracy. In *Proceedings of the NILM-2014 workshop*. Citeseer, 2014.

[148] Simon Arberet and Andreas Hutter. Non-intrusive load curve disaggregation using sparse decomposition with a translation-invariant boxcar dictionary. In *Proceedings of the Innovative Smart Grid Technologies Conference Europe (ISGT-Europe), 2014 IEEE PES*, pages 1–6. IEEE, 2014.

[149] Michal Aharon, Michael Elad, and Alfred Bruckstein. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on Signal Processing*, 54(11):4311–4322, 2006.

[150] Laura Rebollo-Neira and Zhiqiang Xu. Sparse signal representation by adaptive non-uniform b-spline dictionaries on a compact interval. *Signal processing*, 90(7):2308–2313, 2010.

[151] Jin Xu, Haibo He, and Hong Man. Active dictionary learning in sparse representation based classification. *arXiv*, pages 1–7, 2014.

[152] Laurent Daudet. Audio sparse decompositions in parallel. *IEEE Signal Processing Magazine*, 27(2):90–96, 2010.

[153] Yagyensh C. Pati, Ramin Rezaiifar, and Perinkulam S. Krishnaprasad. Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition. In *Proceedings of 27th Asilomar conference on signals, systems and computers*, pages 40–44. IEEE, 1993.

[154] Rémi Gribonval, Holger Rauhut, Karin Schnass, and Pierre Vandergheynst. Atoms of all channels, unite! average case analysis of multi-channel sparse recovery using greedy algorithms. *Journal of Fourier analysis and Applications*, 14(5-6):655–687, 2008.

[155] Santo Fortunato. Quality functions in community detection. In *Proceedings of the SPIE Fourth International Symposium on Fluctuations and Noise, Noise and Stochastics in Complex Systems and Finance*, volume 6601, page 660108. International Society for Optics and Photonics, 2007.

[156] Vincent D. Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):P10008, oct 2008.

[157] Xinyu Que, Fabio Checconi, Fabrizio Petrini, and John A. Gunnels. Scalable community detection with the louvain algorithm. In *Proceedings of the Parallel and Distributed Processing Symposium (IPDPS), 2015 IEEE International*, pages 28–37. IEEE, 2015.

[158] Mark E. J. Newman. Analysis of weighted networks. *Physical review E*, 70(5):056131, 2004.

[159] Saeed Aghabozorgi, Ali Seyed Shirkhorshidi, and Teh Ying Wah. Time-series clustering–a decade review. *Information Systems*, 53:16–38, 2015.

[160] 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 *Proceedings of the 5th international conference on Future energy systems*, pages 265–276. ACM, 2014.

[161] Stephen Makonin and Fred Popowich. Non-intrusive load monitoring (NILM) performance evaluation: A unified approach for accuracy reporting. *Energy Efficiency*, 8(4):809–814, jul 2015.

[162] Nilson Henao, Kodjo Agbossou, Sousso Kelouwani, Yves Dubé, and Michaël Fournier. Approach in nonintrusive type i load monitoring using subtractive clustering. *IEEE Transactions on Smart Grid*, 8(2):812–821, 2017.

[163] Bochao Zhao, Lina Stankovic, and Vladimir Stankovic. On a training-less solution for non-intrusive appliance load monitoring using graph signal processing. *IEEE Access*, 4:1784–1799, 2016.

[164] Niamh O'Connell, Pierre Pinson, Henrik Madsen, and Mark O'Malley. Benefits and Challenges of Demand Response: A Critical Review. *Renewable and Sustainable Energy Reviews*, 39:686–699, 2013.

[165] Murthy V.S.K. Balijepalli, Vedanta Pradhan, Shrikrishna A. Khaparde, and RM Shereef. Review of demand response under smart grid paradigm. In *Proceedings of the ISGT2011-India*, pages 236–243. IEEE, 2011.

[166] Peter Palensky and Dietmar Dietrich. Demand side management: Demand response, intelligent energy systems, and smart loads. *IEEE transactions on industrial informatics*, 7(3):381–388, 2011.

[167] Ryan Hledik and Ahmad Faruqui. Valuing demand response: International best practices, case studies, and applications. Brattle group website, 2015.

[168] Mark G. Lauby, Jessica J. Bian, and James Powell. 2011 north american demand response availability. In *Proceedings of the 2013 IEEE Power & Energy Society General Meeting*, pages 1–5. IEEE, 2013.

[169] A Yousefi, TT Nguyen, H Zareipour, and OP Malik. Congestion management using demand response and facts devices. *International Journal of Electrical Power & Energy Systems*, 37(1):78–85, 2012.

[170] Zhao Xu, Jacob Ostergaard, and Mikael Togeby. Demand as frequency controlled reserve. *IEEE Transactions on power systems*, 26(3):1062–1071, 2011.

[171] P Jazayeri, A Schellenberg, WD Rosehart, JADJ Doudna, SAWS Widergren, D Lawrence, JAMJ Mickey, and SAJS Jones. A survey of load control programs for price and system stability. *IEEE Transactions on Power Systems*, 20(3):1504–1509, 2005.

[172] Jacopo Torriti, Mohamed G. Hassan, and Matthew Leach. Demand response experience in Europe: Policies, programmes and implementation. *Energy*, 35(4):1575–1583, April 2010.

[173] Jamshid Aghaei and Mohammad-Iman Alizadeh. Demand response in smart electricity grids equipped with renewable energy sources: A review. *Renewable and Sustainable Energy Reviews*, 18:64–72, 2013.

[174] Andrew J Roscoe and G Ault. Supporting high penetrations of renewable generation via implementation of real-time electricity pricing and demand response. *IET Renewable Power Generation*, 4(4):369–382, 2010.

[175] Johanna L. Mathieu, Duncan S. Callaway, and Sila Kiliccote. Examining uncertainty in demand response baseline models and variability in automated responses to dynamic pricing. In *Proceedings of the IEEE Conference on Decision and Control and European Control Conference*, pages 4332–4339. IEEE, December 2011.

[176] Olivier Corradi, Henning Ochsenfeld, Henrik Madsen, and Pierre Pinson. Controlling electricity consumption by forecasting its response to varying prices. *IEEE Transactions on Power Systems*, 28(1):421–429, 2012.

[177] Miriam L Goldberg and G Kennedy Agnew. Protocol development for demand response calculation–findings and recommendations, 2003.

[178] Katie Coughlin, Mary Ann Piette, Charles Goldman, and Sila Kiliccote. Statistical analysis of baseline load models for non-residential buildings. *Energy and Buildings*, 41(4):374–381, 2009.

[179] Zhong Fan, Parag Kulkarni, Sedat Gormus, Costas Efthymiou, Georgios Kalogridis, Mahesh Sooriyabandara, Ziming Zhu, Sangarapillai Lambotharan, and Woon Hau Chin. Smart grid communications: Overview of research challenges, solutions, and standardization activities. *IEEE Communications Surveys and Tutorials*, 15:21–38, 2013.

[180] James W Taylor. Triple seasonal methods for short-term electricity demand forecasting. *European Journal of Operational Research*, 204(1):139–152, 2010.

[181] Henrique Steinherz Hippert, Carlos Eduardo Pedreira, and Reinaldo Castro Souza. Neural networks for short-term load forecasting: A review and evaluation. *IEEE Transactions on power systems*, 16(1):44–55, 2001.

[182] Tao Hong. Short term electric load forecasting. 2010.

[183] Ahsan Raza Khan, Anzar Mahmood, Awais Safdar, Zafar A Khan, and Naveed Ahmed Khan. Load forecasting, dynamic pricing and DSM in smart grid: A review. *Renewable and Sustainable Energy Reviews*, 54:1311–1322, 2016.

[184] Chirag Deb, Fan Zhang, Junjing Yang, Siew Eang Lee, and Kwok Wei Shah. A review on time series forecasting techniques for building energy consumption. *Renewable and Sustainable Energy Reviews*, 74:902–924, 2017.

[185] Per Lund, Preben Nyeng, Rune N. Grandal, Stig H. Sørensen, and Others. Deliverable 6.7 Overall evaluation and conclusion integration of DER and DR. Technical report, EcoGrid EU Final Report, 2016.

[186] Simon B. Harvey. The Harvard Medical School Guide to Men's Health. *Publishers Weekly*, 249(31), 2002.

[187] Niels E. Helstrup, Per Lund, Maja F. Bendtsen, Dieter Gantenbein, and Andreas Arendt. Deliverable 6.3 - System operation and monitoring: Large-scale smart grids demonstration of real time market-based integration of DER and DR. Technical report, EcoGrid EU Internal Report, 2013.

[188] Leonard Kaufman and Peter J. Rousseeuw. Finding groups in data: an introduction to cluster analysis. *Journal of the American Statistical Association*, 86(415):830–832, 1991.

[189] Bradley Efron. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science*, 1(1):54–75, 1986.

[190] Robert A. Yaffee. Introduction to time series analysis and forecasting with applications of SAS and SPSS. *International Journal of Forecasting*, 17(2):301–302, 2001.

[191] Prasanta C. Mahalanobis. On the generalized distance in statistics. *Proceedings of the National Institute of Sciences*, pages 49–55, 1936.

[192] Zoltán Prekopcsák and Daniel Lemire. Time series classification by class-specific mahalanobis distance measures. *Advances in Data Analysis and Classification*, 6(3):185–200, 2012.

[193] Olivier Ledoit and Michael Wolf. Honey, I shrunk the sample covariance matrix. *UPF economics and business working paper*, (691), 2003.

[194] Juliane Schaefer and Korbinian Strimmer. A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Statistical Applications In Genetics and Molecular Biology*, 4(1), 2005.

[195] Jean-Pierre Benzécri. *Analyse des données, Tome 1, La Taxinomie*. Dunod, 1 edition, 1976.

[196] Charles Romesburg. *Cluster analysis for researchers*. Lulu. com, 2004.

[197] Rand R. Wilcox. *Applying contemporary statistical techniques*. Elsevier, 2003.

[198] Pablo Montero and José Vilar. TSclust: An R package for time series clustering. *JSS Journal of Statistical Software*, 62(1):1–43, 2014.

[199] Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An introduction to statistical learning*, volume 112. Springer, 2013.

[200] Aapo Hyvärinen and Erkki Oja. Independent component analysis: algorithms and applications. *Neural networks*, 13(4-5):411–430, 2000.

Collection of Relevant Publications

[Paper A] Le Ray, G., Pinson, P. (2019) The ethical smart grid: Enabling a fruitful and long-lasting relationship between utilities and customers. *Energy Policy*, Submitted.

[Paper B] Le Ray, G., Pinson, P. (2018). Online adaptive clustering algorithm for load profiling. *Sustainable Energy, Grids and Networks*, 17, 100181.

[Paper C] Winker, P., Le Ray, G., Pinson, P. (2019). Unsupervised energy disaggregation: from sparse signal approximation to community detection. *IEEE Transactions on Smart Grid*, Submitted.

[Paper D] Le Ray, G., Christensen, M. H., Pinson, P. (2019). Detection and characterization of domestic heat pumps. In *Proceedings of PowerTech Conference, 2019 IEEE Milano* (pp. 1-6). IEEE.

[Paper E] Le Ray, G., Larsen, E. M., Pinson, P. (2018). Evaluating price-based demand response in practice – with application to the EcoGrid EU experiment. *IEEE Transactions on Smart Grid*, 9(3), 2304-2313.

[Paper F] Le Ray, G., Pinson, P., Larsen, E. M. (2017). Data-driven demand response characterization and quantification. In *Proceedings of PowerTech Conference, 2017 IEEE Manchester* (pp. 1-6). IEEE.

[Paper A] The ethical smart grid: Enabling a fruitful and long-lasting relationship between utilities and customers.

The ethical smart grid: Enabling a fruitful and long-lasting relationship between utilities and customers

G. Le Ray , P. Pinson

Centre for Electric Power and Energy, Technical University of Denmark, Kgs. Lyngby, Denmark

Abstract

The European Union is implementing ambitious programs to tackle energy efficiency, energy independence and climate change challenges. To reach the 20/20/20 targets, the EU aims at modernizing power grids to make them 'smart' by collecting close to real-time data and subsequently operate these grids more optimally. One of the smart grids' purposes is to integrate a growing share of renewable energy sources while efficiently accommodating their variability and limited predictability through actuation of consumer flexibility. To do so, 80% of the customers throughout Europe will be equipped with smart meters by 2020. On the one hand, the success of energy programs relies on customer involvement in altering their energy consumption through the use of analytics and incentive-based demand side management. On the other hand, the utilities are arguably showing limited ethics in the way smart meters are rolled-out and on the use of collected data. Indeed, beyond legal binds and technical obstacles, the deployment of smart meters and the use of collected data is unclear in terms of how to deal with customers. We argue that ethics should be an import driver for decision-making to guarantee the sustainability of smart grid programs which relies on the active participation of customers.

Keywords: Big Data, Privacy, Smart meter, Smart grid, Ethics

1. Introduction

The European Union's (EU) energy policy is facing unprecedented challenges due to increased dependencies on imports, scarce resources and the need to limit climate change (European Parliament, 2012). Ambitious energy efficiency programs have been developed to tackle these challenges. Indeed, since 2009 and the 2020 Climate & Energy Package's road map to the 20/20/20 targets (European Parliament, 2009b), EU has driven towards a *greener* energy sector to achieve energy efficiency, energy independence and last but not least to reduce greenhouse gas emissions.

As most of the issues on power systems are observed at the distribution level, the program requires a modernization of the grid to foresee potential issues and to have a pervasive control to prevent them (Farhangi, 2010). Information and communication technologies forms

Email address: [gleray,ppin]@elektro.dtu.dk (G. Le Ray , P. Pinson)

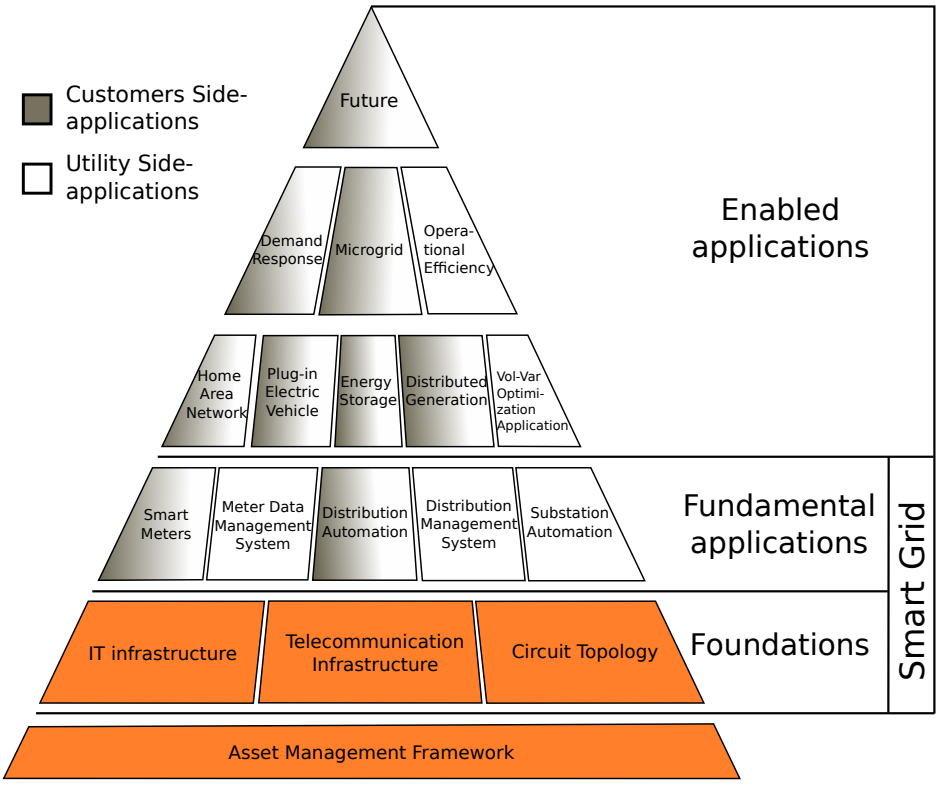


Figure 1: The Smart Grid pyramid (Source: Farhangi (2010))

the foundations of the smart grid pyramid (Figure 1) which support more advanced infrastructures. The Third Energy Package, adopted in 2009, strengthens the internal European market for gas and electricity by securing a competitive and sustainable supply of energy to the economy and the society (European Commission, 2011). To reach this goal, the EU has set the target of 80% of households equipped with smart meters by 2020 (European Union, 2009). Smart meters are deployed to provide more transparency to consumers (billing, price, consumption), to improve awareness on energy consumption and empower the consumers to modify their energy behavior using metering data (European Commission, 2011). On the utility side, smart meters' data will help them to increase the efficiency and the reliability of the grid. In this paper we define a utility as an entity that is given responsibility for the maintenance and operation of some infrastructure of public value and to be used for a public service. As displayed in Figure 1, smart meters constitute the first fundamental application that involves customers. The perception of smart meters by customers will condition the future of smart grids in their capacity to transform customers into actors of the grid through the use of demand side management (Bertoldo et al., 2015, Horne et al., 2015, Giordano et al., 2011).

At the same time, concerns are raised about a possible backlash of domestic customers (Zachary, 2011). Indeed the perspective of having smart meters reporting electricity consumption at high resolution in every home has engendered irrational fear (e.g. health issues and domes-

tic accidents) and legitimate questions about the need of smart meters and their impact on privacy (McKenna et al., 2012). The concept of privacy has drastically evolved, from Aristotle making the distinction between public sphere and private sphere, to the creation of the right to privacy (Papakonstantinou and Kloza, 2015). With the emergence of information technology, the legal framework has evolved to protect data and by extension data subjects. The legislation on data protection and privacy limits the legal use of data but it also defines the limits of a gray area on how to handle customers and their data (Tzafestas, 2018). Even if the decisions made are in line with the existing legal framework, they can have a substantial (and potentially negative) impact on the future of smart metering and smart grid deployment plans by extension (Jegen and Philion, 2017). It is where ethics is fundamental, it is defined as: *"A system of moral principles which deals with what is good or bad for individuals and the society. It is a collection of fundamental concepts and principles on an ideal human character that enable people to make decisions regarding what is right or wrong. Ethics is a code of conduct agreed and adopted by people in a society, which sets the norms of how a person should live and interact with other people."* (Tzafestas, 2018).

Following this definition, an ethical and human perspective has to be placed on the smart grid technology, especially its extensions smart meters and the data they generate. The foundations of the smart grid are put in place today. Hence it is important to understand the perspective of the customers to build an ethical smart grid to build a fruitful and long-lasting relationship between utilities and customers (Jegen and Philion, 2017).

In this paper, we argue that beyond the technical and legal issues, ethics should be the driver for decision-making. The EU regulation defines a legal framework in which the utilities can implement the deployment and exploitation of smart meters with more or less focus on customers. However, the long term development of smart grid technology will change the status of consumers to prosumers and they should be treated as such: collaborators providing services to the grid. In an era where leakage of data is making newspapers' headlines on a regular basis, this is the only reasonable way to build an sustainable relationship between domestic customers and utilities. Hence a balance between the necessity of data to build the smart grid and the respect of the customers as partners has to be determined.

The following discussion is delimited by the legal and technical background of the roll-out and smart meters data utilization in relation to privacy and data accessibility (Section 2). However privacy is not only dependant on the legal and technical aspects but also depends on what the citizens are willing to give. As privacy is the most important issue raised with smart meters, it is crucial to understand what we mean today by privacy and what is really at stakes in this context when it is jeopardized (Section 3). In terms of ethics, the transition to smart grid redistributes the cards between utilities and consumers (prosumers) as the latter have a more important role now than before (Section 4). Customers are also exposed in giving away sensitive data to utilities and it has to be acknowledge and rewarded in an ethical manner (Section 5). The conclusions are gathered in Section 6 while opening up to broader perspectives.

2. Legal and technical background in relation to smart meters' data privacy

The legal framework of data protection and privacy has evolved over time, mainly due to the emergence of new technologies and new threats to privacy they create (Horne et al., 2015). Here we aim at giving the background to both legal and technical aspects that are shaping data collection and use of data generated by smart meters. It scopes what is legally possible in Europe and how the technical setup decided by each Member State shapes the relationship between customers and utilities during the roll-out and after.

2.1. A compact historic review of the right to privacy in EU legislation

The origin of the *right to privacy* can be traced back to the Universal Declaration of the Human Rights (article 12) in 1948 (United Nations, 1949). It states that ' *No one shall be subjected to arbitrary interference with his privacy, family, home or correspondence, nor to attacks upon his honour and reputation. Everyone has the right to the protection of the law against such interference or attacks*'. It aims at protecting the right and interest of individuals rather than the data itself as data collection appeared to generate unexpected impact on individuals life. Soon after, the Council of Europe strengthened it in the European Convention on Human Rights (European Court of Human Rights, 1950).

The growth of information technology in the 1970's, especially in the public sector and in the banking industry, pushed the Committee of Ministers to the Member States to write 2 recommendations (Resolution 23 and 24) stating that every individual whatever his nationality or residence should have respect for his right to privacy with regards to automatic processing of personal data. These resolutions were received positively and the Council of Europe implemented them in the Data Protection Convention which had impact beyond Europe, as 46 countries ratified it (Council of Europe, 1985). It defines the concept of *personal data* as '*any information relating to an identified or identifiable individual ('data subject')*' and sets the foundation of data protection at an international level. The aim of the Convention is to protect individuals against unjustified collection, use and dissemination of personal data. It then implicitly defines what will later be called *legitimate purpose*.

After years of negotiation between the Member States, the Data Protection Directive (Directive 95/46/EC) was adopted in order to harmonize the legal framework (European Parliament, 1995). Some precisions were added to the definition of *personal data* about what *identifiable* meant; ' *an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity*'. It remains broad on purpose to extend its application to future information technologies. Despite being implemented on the same basic principles, it has generated different applications[1]. The Data Protection Directive is articulated around three points; i) transparency: information on personal data being processed; ii) legitimate purpose: specification, explicit and legitimate of

[1] As an EU Directive, it is applicable to all Member States but each Member States transposes it in its national law

the purposes of the data collection; iii) parsimony: adequacy to the purpose of the personal data collected.

Article 7 stipulates the lawful basis to process personal data:

(a) unambiguously consent; or

(b) processing is necessary for the performance of a contract; or

(c) processing is necessary for compliance with a legal obligation; or

(d) processing is necessary in order to protect vital interests; or

(e) processing is necessary for the performance of a task carried out in the public interest; or

(f) processing is necessary for the purposes of the legitimate interests pursued by the controller or by the third party.

In order to harmonize the Data Protection Directive among the EU Member States, the European Commission proposed the General Data Protection Regulation (GDPR) in 2012 (European Union, 2016). It generalizes the basic principle of the Data Protection Directive and develops some further rules that are applicable to all data collected inside the EU by European or non-European organizations.

The main changes on the rights of the data subjects and responsibilities of controllers and processors in relation to data protection and privacy of the data subject are:

- Explicit and provable consent (instead of unambiguous consent)(Article 7).
- Transparency and modalities: The data controller should inform and communicate with the data subject in a *'concise, transparent, intelligible and easily accessible form, using clear and plain language'* (Article 12(1)). It should also facilitate the exercise of the data subject rights (Article 12(2)).
- Rectification and erasure: A person has the right: to ask for his data to be erased (Article 17); to restrict the processing under certain condition (Article 18); to transfer personal data from one service to another (Data portability Article 20).
- Right to object to automated individual decision-making (Articles 21 and 22).
- Data Protection by design and by default: The data protection and privacy should be included in the development of the service and the privacy settings should be set to a high level by default (Article 25).
- Communication of a personal data breach to the data subject (Article 34).

From the foundation of the right to privacy to the GDPR, the definition of privacy and data protection had to be updated according to the development of information technologies which is going at a hectic pace. Nevertheless, the following discussion on smart meter data and their ethical use is bounded within the EU by this legal framework.

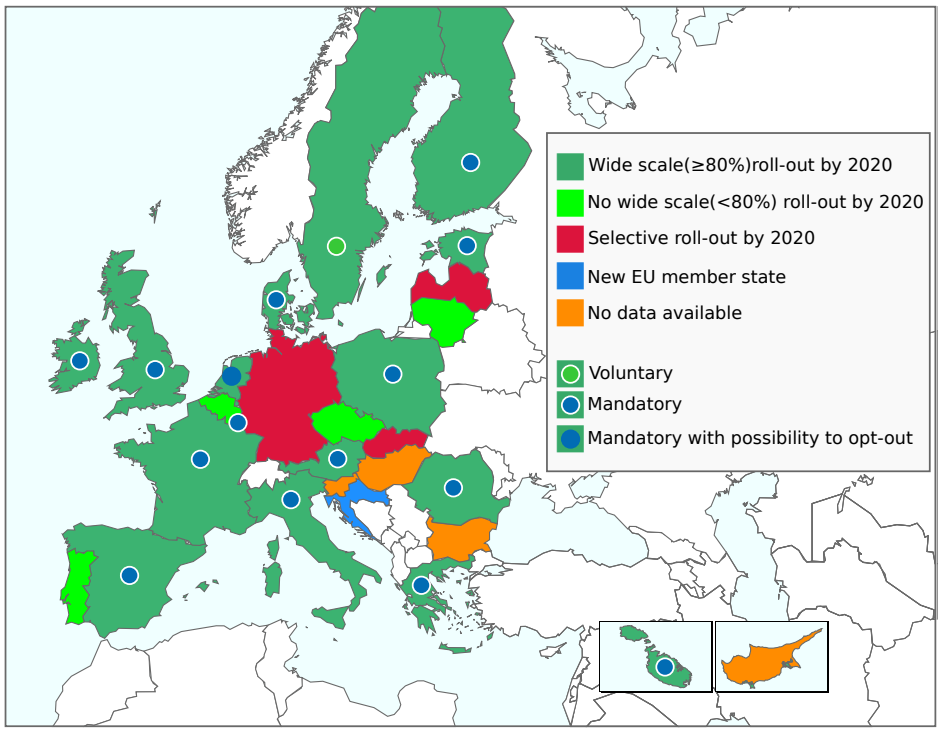


Figure 2: Map of the roll-out of Smart meters in Europe (European Commission, 2014).

2.2. A review of roll-outs and setups of smart metering in the EU

In the case of smart meters, the technological possibilities as well as deployment strategies are directly related to the problem of privacy and ethics . The scale of roll-out is decided based on a cost-benefit analysis (CBA), described in (European Commission, 2011), which concludes wether the roll-out should be at least 80%, less or just selective. However the roll-out strategy is left completely to each EU Member State which gives a large diversity of setups and subsequently different data flows. Table 1 gives an overview of the roll-out status of the different Member States in 2014. The map in Figure 2 presents the roll-out scale as well as the recruitment strategy. Table 1 and Figure 2 give an overview of the diversity and the number of parameters to take into account in the roll-out of smart meters in EU. Temporal disparities are also observed; Italy and Sweden had already completed the deployment of smart meters before the adoption of the directive 2012/27/EU. The Netherlands had planned an early deployment but the roll-out which was originally mandatory has been challenged by consumer protection organizations that sued the State to obtain the possibility to opt-out (Hoenkamp et al., 2011).

Smart metering has also changed the responsibilities of the DSO and TSO as they have to handle large amount of data. Figure 3 is a schematic representation of the flow of data and actions between the different actors. The role of the data controllers, data protection officer and supervisory authority are defined in the GDPR (European Union, 2016) and are taken in most cases by the DSO, TSO or independent organism (Smart Grids Task Force Expert Group 1- Standards and Interoperability, 2016). It could be considered in the con-

Table 1: Overview of the roll-out (in 2014) in EU. Source: European Commission (2014)

Member State	roll-out scale	CBA[a] outcome	resolution	implementation/ownership	storage	Financing of the roll-out
Austria	95%	+	15 min	DSO	DSO	Metering & network tariffs
Belgium	<80%	-	NS[b]	DSO	DSO	Network tariffs
Bulgaria	TBA[c]	NA	NA	NA	NA	NA
Croatia[d]	NA[e]	NA	NA	NA	NA	NA
Cyprus	TBA	NA	NS	DSO	DSO	NA
Czech Republic	1%	-	NS	DSO	Central Hub	NA
Denmark	100%	+	15 min (hourly before 2011)	DSO	Central Hub	Network tariffs
Estonia	100%	+	hourly	DSO	Central Hub	Network tariffs
Finland	100%	+	1 hour (RT optional)	DSO	DSO	Network tariffs
France	95%	+	10-30 min	DSO/municipalities	DSO	NA
Germany	23%	-	15 min	DSO or meter operator	DSO or meter operator	NA
Greece	80%	+	NS	DSO	DSO	NA
Great Britain	99.5%	+	30 min (10s to customer)	Supplier	Central Hub	Funded by suppliers
Hungary	TBA	+	NS	NA	NA	NA
Ireland	100%	+	30 min (10s to customer)	DSO	DSO	Network tariffs
Italy	99%	NA	10 min	DSO	DSO	DSO & Network tariffs
Latvia	23%	-	NS	DSO	DSO	Network tariffs
Lithuania	<80%	-	NS	DSO	DSO	Network tariffs
Luxembourg	95%	+	NS	DSO	DSO	Network tariffs
Malta	100%	NA	NS	DSO	DSO	Network tariffs
Netherlands	100%	+	NS	DSO	DSO	Network tariffs
Poland	80%	+	NS	DSO	Central Hub	Network tariffs
Portugal	<80%	Inconclusive	15 min	DSO	DSO	DSO & Network tariffs
Romania	80%	+	NS	DSO	DSO	Network tariffs
Slovakia	23%	-	15 min	DSO	DSO/Central Hub	DSO & Network tariffs
Slovenia	TBA	NA	NS	NA	DSO	NA
Spain	100%	NA	NA	DSO	DSO	Network tariffs & SM rental
Sweden	100%	+	hourly	DSO	DSO	DSO &Network tariffs

[a] Cost-benefit analysis [b] Not Specified [c] To be announced [d] New Member State [e] missing information

text of smart metering as the perfect flow of data according to Nissenbaum (2011).

Some parameters, like the resolution of the data, the access to metering data and the implementation/ownership, have a direct impact on the setup and thus the data flow as shown in Figure 3 as well as the capacity of the customer to modify its consumption. The range of possibilities makes it difficult to standardize, however most of the DSOs, as responsible authorities of the roll-out, (will) face the same ethical problems with their customers.

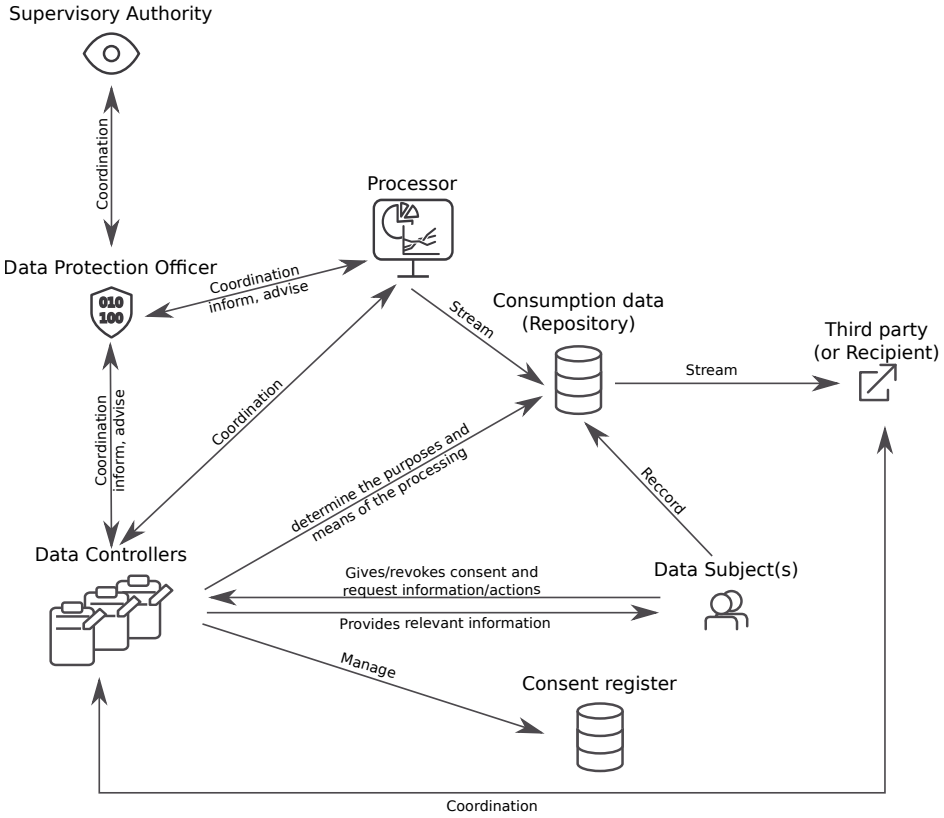


Figure 3: Interaction of actors and flow of smart meter data as described in the GDPR. Source: (Smart Grids Task Force Expert Group 1- Standards and Interoperability (2016)).

3. What privacy today?

Privacy is a generic word used to describe what we perceive as relating to private matters. Nevertheless the definition of privacy is evolving over time. In this section we give some examples revealing today's state of privacy and how much data we accept to give to obtain a service. A discussion is as well open on what is at stake when we talk about privacy breaching.

3.1. The state of privacy in the Big Data era

We have entered a new era called the *'Big Data era'* (Wladawsky-Berger, 2015). Despite the term 'big', the root of Big Data pertains to *(i)* volume: the quantity of data being

collected is growing exponentially (OECD, 2013); *(ii)* velocity: The resolution at which data is being collected increases steadily; and *(iii)* variety: The sources of data are getting more diverse. From the Data Protection Directive, data can be categorized into 2 types, personal data, which are protected by law and the non-personal data (European Commission, 2018). Hence to go around restrictions on the use of personal data, the best way is to collect *more* non-personal data that can be combined to create a unique profile, defining an individual.

On the Internet, the most generic data collected concerns navigation information (i.e. browsing) and clicks. Cookies, saved on each computer, have been used to collect navigation information on users. Users' navigation information are then used to generate target advertisement. In Europe and until 2011, website were not asking for consent on using cookies. In 2011, the so called *'EU cookie legislation'*, Directive 2009/136/EC, detailing the use of cookies was added to Directive 2002/58/EC on digital user rights (European Parliament, 2009a). It stipulates that cookie ID are considered *personal data* from now on, and requires any website to ask for users' consent to retrieve information stored on cookies. Despite the efforts of the European Commission to regulate the exploitation of navigation information, new ways of collecting those information were already implemented. In order to optimize their visual aspect, websites collect information concerning the hardware (e.g. screen, computer) and the software (i.e. the browser type and version) with the genuine aim to give the best user experience. However it can form a unique combination which is called a *'browser fingerprint'* (Laperdrix et al., 2016). To be close to unique the fingerprint of a browser requires approximately 17 parameters. Thereby cookies are becoming obsolete and the online advertisement business is still monetizing browsing information while avoiding legislation.

Google and Facebook emphasize concerns about data privacy as they have always been at the forefront of the data monetizing business models, providing services for free and monetizing data via advertisement. Thanks to the dimensions of their pool of users, they are self-sufficient in data to feed their advertisement algorithm. In 2017, Alphabet's (parent company of Google and Youtube) and Facebook's digital advertisement revenues combined represented a gigantic 191,8 Billion US dollars (respectively 123.5 B$ and 68.3 B$) which represents half of the global digital advertising revenue (Molla, 2018). In itself only, the use of data for targeted advertisement is not much of a problem and can be considered as annoying when it is excessive. The problems comes out of the methods used to maximize revenues.

Facebook generates a unique dataset which appeals psychometricians studying human behavior. The collection of *likes* from users can be used to generate really precise psychological profiles like the *'Big five'* (Kosinski et al., 2013, McCrae and John, 1992, Gosling et al., 2011). The Cambridge Analytica Scandal made citizens aware of how a breach into the security could contribute to private interests. Data from 100 000 of Facebook's users were originally collected with their consent for research. Despite rules and Non-Disclosure Agreements, access was given to Cambridge Analytica which extended the data to 30 millions users using interconnections between *friended* users. Data was thereafter not used for research but to influence opinions through the targeted advertising algorithm of Facebook. The use of the data is thus not as questionable as the purpose. The exploitation of such a unique dataset for research purposes is valuable, however the use of such a dataset for

influencing opinion is a serious law infringement (Kosinski et al., 2015).

The *'privacy by default'* Article in the GDPR, has probably been designed based on the experience with Facebook's default privacy settings. Indeed, Facebook's privacy settings were left to minimum level so that user's profiles could be *searchable*, and partly *visible* to all members, thus increasing traffic (Gross et al., 2005, Liu et al., 2011). From a user's point of view, they have to know i) that access to their account is not restricted to 'friends' ii) that they should know how to restrict access (Liu et al., 2011). The configuration as default to the lightest security settings is questionable from a user perspective, as their personal information are not protected by default despite existence of such parameters. Social networks benefit from data placed in them but they benefit even more of connections created between user profiles (see (Mcdonald and Ackerman, 2000) for more information) and generate their advertisement revenue from the traffic. Usually users become aware of privacy issues, when terms and privacy policies have to be updated and some may modify their privacy settings but the large majority does not, as it is a non trivial operation (Liu et al., 2011).

The use of smart phones and smart city application (e.g. public transportation card, traffic) adds a geographical dimension to information collected that anchors it in the physical world. Using mobility data from carriers' antennas it has been demonstrated that only 4 spatio-temporal points are needed to uniquely identify each carrier (De Montjoye et al., 2013). Using GPS data, the number of points decreases to collect unique patterns. Spatio-temporal data are highly sensitive personal data, information on where an individual is at any time can be used to intercept physical someone. They are personal data as they allow to uniquely identify a person from his data.

The Big Data era has induced in citizens a certain distrust as well as a necessity to stay connected, which create a ambivalent perception of technological products.

3.2. Privacy is not the problem anymore

Privacy comes from the Latin word *privatus* which means 'withdraw from public life'. Indeed the strict definition and application of privacy implies that each individual should not in any way be uniquely identifiable using the collected data (United Nations, 1949). But as the legal framework on data protection and privacy evolves with the emergence of new information technologies, the concept of privacy evolves as well. Privacy is usually guaranteed to data subjects by collecting data anonymously in the sense of *namelessness* (i.e. not identified by name, address, social security number). Examples have been given in Section 3.1 that shows that anonymized data can actually be used to uniquely identify individuals and thus questions the use of anonymity to protect privacy. Indeed, anonymity is used to collect data without naming the data subject but keep them identifiable (Laperdrix et al., 2016, De Montjoye et al., 2013). It is important here to understand what is at stake in that context: names have no importance in themselves. But identities,sets of information which define each individual, are extremely valuable as well as sensitive. Personal data can be combined with other non-personal data to identify, contact or locate a *single* person. Discarding all these information, is actually a way to keep them *anonymous* (i.e. nameless)

but still uniquely identifiable. Google has even created a word to describe these paradoxical IDs, *'anonymous identifier'*, which they use for targeted advertisement (Kitchin, 2016, Barocas and Nissenbaum, 2014).

A question arises then, how many data points are needed to uniquely identify users? Considering the profuseness of possibilities, only a few data points are needed to create a combination that uniquely identifies a user (Laperdrix et al., 2016, De Montjoye et al., 2013). The problem appears when the data collected can provide sufficient information to reach a person physically (e.g. through email, phone, address). Barocas and Nissenbaum (2014) argue that the real value in anonymity is to prevent *reachability*, not to protect privacy. From collected data it should not be possible to communicate or reach physically data subjects. This concept is then much more meaningful and reshapes also the concept of privacy. It does not apply only to personal data but also to non-personal data that could be used to reach a person. In (Acquisti and Gross, 2009) an algorithm is built to predict social security numbers of American citizen based on their date and place of birth. They reach success rates from 7% to 61% in predicting the 5 first numbers (out of 9) using publicly available data depending on the period and state of birth. It proves that *any personal information can be sensitive information* when combined appropriately (Acquisti and Gross, 2009).

Respecting privacy is not respecting secrecy or granting control over personal information. It consists of respecting an appropriate flow of information. Nissenbaum (2011) calls it contextual integrity; data (type of information) collected in a certain context (e.g. finance, health, social norms) flow, following transmission principles (e.g. consent, buying, selling, confidentiality), between different actors (e.g. subject, sender, recipient) in an appropriate manner. Disruptive practices modifying the information flow are evaluated depending on how they move it from the ideal information flow. In other words, it evaluates the impact of disruptive flows on ethical values like fairness, justice, freedom, welfare or any other context specific concepts.

Figure 3 could be a representation of the perfect flow of information in a smart metering context. In the reality, from an ethical perspective, the relationships between utilities and consumers are far from perfect. The major problems are around data collection and the change of status of the actors.

4. Customers should be treated as collaborators

GDPR sets standards for data protection and privacy and the Third Energy Package gives guidelines and objectives for the roll-out and use of smart meters. However in this new context of smart grid there are no clear guidelines or rules on how to work with customers. The main challenge is to keep a positive relationship between utilities and (domestic) customers in order to secure the investment and involve them as actors of the grid (Bertoldo et al., 2015).

4.1. An ethical roll-out to improve acceptance of the customers

The roll-out scales as defined in the CBA (Figure 2) have been decided at the EU level and the DSOs are following the decision in deploying the smart meters. However the decision

to make the installation mandatory is raising concerns throughout Europe. As presented in Section 3 several privacy scandals have raised awareness about privacy issues and this should not be neglected. From a customer perspective, the roll-out of smart meters, especially when mandatory, is an intrusion to what is perceived as the last sanctuary of privacy, 'Home' (Papakonstantinou and Kloza, 2015). Indeed a meter is a foreign object in a household that inhabitants do not own (it is the DSO's property) and cannot remove/modify. The adoption of the technology depends on the perception of the customers (Ponce et al., 2016). The European Commission gave some guidelines on conducing the cost-benefit analysis where '*an assessment of the level of social resistance (or participation) to the project should be presented, including a description of means adopted to ensure social acceptance and their effectiveness*' (Papakonstantinou and Kloza, 2015). The customers are either putting high expectation in smart meters (technophiles) and get disappointed, or they have realistic fears regarding privacy breaching and loss of control (Krishnamurti et al., 2012). Hence both situations lead to a negative perception of smart meters. To have a positive impact, the benefits of smart meters should be clearly stated and visible rapidly after installation to maximize customers' acceptance of the new technology.

The case of the Netherlands can be used as an example of what can go wrong when end-users are not considered properly in the smart metering framework (Hoenkamp et al., 2011). Originally the roll-out was mandatory and refusing the installation was made punishable as an economic offense, with a fine of 17.000€ or imprisonment for a maximum of six months (Gutwirth et al., 2013). Beside privacy concerns transmitted to the Dutch Data Protection Authority on the use of high resolution data, the utilities were not inclined to focus on a customers inclusive solution to stimulate demand flexibility (Hoenkamp et al., 2011). The Minister of Economic Affairs amended the Dutch Data Protection Authority's proposal by stipulating that the network operator could transfer hourly or 15-minute metering data to the energy provider only if the customer gave his consent. To add up to the pile, the Dutch Consumer Union published a report stipulating that a mandatory roll-out of smart meter reporting 15-minute electricity information was an infringement of the right to privacy according to the article 8 of the European Convention on Human Right (European Court of Human Rights, 1950) and was thus not compatible with a democratic society. The problem was finally solved at the Senate by giving the right to customers to refuse having a smart meter installed (opt-out). Gutwirth et al. (2013) considered that there are four factors for the rejection of the smart meter bill by the Senate 1) the high resolution of the data transferred up to the energy providers, 2) the mandatory roll-out where resistance is sanctioned by high fines or even imprisonment, 3) lack of explanation of the necessity of smart metering and by extension why the customers have to loose some privacy 4) the combinations of different functionalities in one meter generating new risk and making the argumentation complex.

Research in social science on the topic of smart meters have also shown large misalignment between the reality of the smart meters and customers' expectations. From a customer point of view, just the fact that a digital connected meter is called 'smart' is actually inducing a wrong idea of what are its capabilities, since it is not a smart home system (Wilson et al., 2017). This is a recipe for backlash. In Krishnamurti et al. (2012), a behavioral study shows that most of the concerns and deceptions from the roll-out of smart meters could be

solved in two ways 1) scale down the expectation of the customers in explaining clearly what the smart meters could do; 2) align the technology with the expectation by adding smart thermostats and in-home displays to visualize consumption in real-time. The smart grid framework requires that the customers know what their metering data is used for, even if it is technical, stakeholders have the responsibility of informing in a clear and understandable manner (Bertoldo et al., 2015).

4.2. Evolution of the roles and relationships

The relationship between utilities and customers is ultimately changing due to potential consequences from two-way communication and deployment of decentralized generation (Khurana et al., 2010). Consumers become prosumers and provide service to the grid, they are actors of the grid and should be considered as collaborators in maintaining stability of the grid. The European utilities have operated in the last decade important restructurations to cope with the opening of the energy market to concurrence. Indeed they (e.g. EDF in France, ENEL in Italy, DONG in Denmark) used to be monopolistic and control the network from generation to distribution. As national companies they had the trust of customers. Today the trust-relationship has to be rebuilt between customers and utilities to operate this transition and find a new balance between the actors of the grid. The goals of the relationship remain unclear to some extend as the benefits and the expectations are not aligned (Horne et al., 2015). For example, the customers are expected to be more active but smart meters alone do not provide this functionality, it requires an additionally smart home device. Additionally the contribution of customers to the stability and reliability of the grid should be highlighted as it can be used to develop new social norms in relation to energy (Horne et al., 2015).

The development of aggregators could play an important role in 'smoothing' the communication between the utilities and the customers as they would have less customers to handle. Indeed beyond their technical role they could act as representatives of the customers to utilities and have more weight in the decision making.

4.3. Smart meters to empower customers to become prosumers

The smart grid aims at transforming a centralized, utility-controlled network into a decentralized, consumer-interactive network allowed by high-resolution monitoring and two-way communication (Khurana et al., 2010). From a utility perspective, the need of metering data is almost mechanical. Indeed, a higher share of RES in the generation mix, as promoted by EU, makes generation less adjustable to the demand. In order to compensate the lack of flexibility on the generation side, the demand could be modulated according to some incentives (i.e. price, benefit) broadcast to the customers using a two-way communication (Finster and Baumgart, 2014).

Demand side management (DSM) programs have been studied and implemented based on the idea of exploiting demand-side flexibility to reduce RES spillage (Strbac, 2008). DSM (including DR frameworks as well as more complex pricing scheme) relies on a marginal dynamic price of generation (Ding et al., 2013). Figure 4 gives an overview of the different price based solutions that can be used depending on resolutions of both price and metered

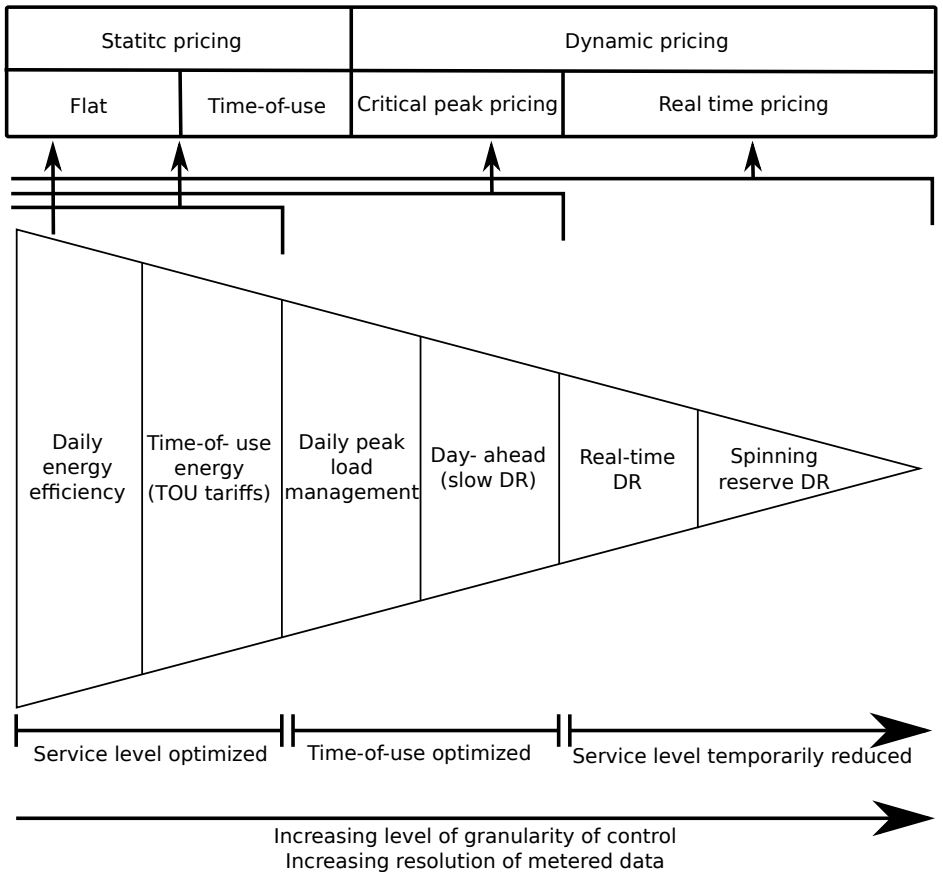


Figure 4: DSM service enabled in function of the resolution of the metering. Source: Siano (2014).

data. To generate the corresponding bill, the energy providers need to know exactly how much power each customer has consumed during each time interval. Hence the resolution of the metered data should then be higher than (or equal to) the one from the dynamic tariff. In such a framework, it is important that customers can access their electricity consumption and dynamic tariff to modify their energy behavior or to automate their white appliances (i.e. dishwasher, washing machine, electric heating) accordingly. High resolution metering is then a way to make customers aware of their energy behavior so that they can shift their consumption from passive (consumers) to active (prosumers) who will provide services to the grid (Chicco, 2016). The incentive used to change the electricity consumption behavior of customers does not have to be financial, social norms are a powerful tool to change behaviors (Allcott, 2011). However, for such incentives to have a positive impact, the relationship between the utilities and the customer has to be positive as well (Horne et al., 2015). A customer who manages correctly his consumption, should then be encouraged in getting electricity cost reductions (McDaniel and McLaughlin, 2009, Klass and Wilson, 2016).

In the context of smart grid, new business models and actors (aggregators) are relying on metering data to create portfolios and manage their assets (Bondy et al., 2015). How-

ever, it has been demonstrated that the success of such frameworks depends heavily on the magnitude of demand response triggered and subsequently active customers (Pepermans, 2014).

4.4. Control of access transferred to the data subject

The electricity metering data are stored on a datahub or on the DSO servers (Table 1). A consent register, as presented in Figure 3, can be created to record which third parties get access to the data. The consent register is here managed by the data controller (DSO) but it could have the form of the 'App' system developed for smartphones (Smart Grids Task Force Expert Group 1- Standards and Interoperability, 2016) where customers directly manage access grants. Hence it will transfer responsibilities, risk assessment and control to the data subject. It could then also generate the same problems as with 'Apps' on smartphones that are asking for access to data which are not useful to the service provided. A third party can access data at high resolution (up to 1s depending on the model) wire or wireless to smart meters using a dedicated port. Again if there is no illegal intrusion to the household, it is assumed that customers should have control over what is connected to the port. The risk of an abusive use of metering data by a third party is increased if the customers are not educated and made aware of how sensitive those data can be (as we can observe with smartphones). The risk being that information in the data, state of white appliances for example, are extracted and used by an unsolicited third parties for sending targeted advertisement suggesting to replace an appliance (Finster and Baumgart, 2014).

The change from consumer to prosumer comes with changes in the distribution of the responsibilities and control between the actors of the grid. Indeed a prosumer needs to control his/her consumption to act as such and provide services to the grid. The hierarchical structure of the actors of the grid, with consumers/prosumers at the bottom and utilities at the top, is actually, changing. The structure is flatter and prosumers are collaborators rather than consumers. Adjustment in terms of consideration have to occur.

5. Requirements in terms of ethics

Good practices can be implemented to make the roll-out and the use of metering data more ethical. It mainly consists of two global concept parsimony and equity. The need of data is acknowledged, nevertheless it should be parsimonious and come with a direct benefit to the data subject. Hence both parties would be satisfied.

5.1. Data resolution in accordance to task to fulfill

Privacy concerns are often about the high resolution of metering data- (McKenna et al., 2012). Obviously the higher the resolution the more precise the information (See Figure 5). By extension, concerns are also related to machine learning algorithm fed with such data like Non-Intrusive Load Monitoring (NILM) (Klemenjak and Goldsborough, 2016). Beyond the potential information that can be extracted, the targeted use of the information is of higher interest. NILM applications, for example, are made in a certain context; it consists of providing detailed information of individual appliances consumption to customers, who

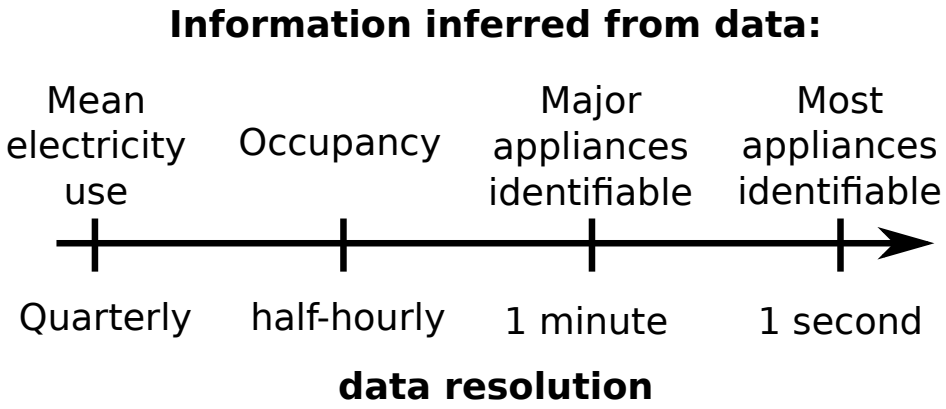


Figure 5: Representation of information that can be inferred from metering data in function of the resolution. Source:McKenna et al. (2012).

are also data subject, so that they can identify appliances with large unnecessary electricity consumption. This information should be provided to no one else. We could also imagine the use of NILM on data at lower resolution to identify large and potentially flexible appliances. Hence DSOs could use this information to propose flexibility contracts to customers; it would be beneficial for every party if the benefits are fairly distributed.

Different tasks can be completed using metering data but they do not require the same level of information (data resolution). The data resolution should therefore be adjusted in accordance to the task to fulfill. In the same way that the purpose has to be legitimate, the resolution of the data has to be legitimate. For example when billing customers under dynamic tariff, it does not improve anything to use electricity consumption at a higher resolution that the dynamic tariff. Hence the resolution of the data should be chosen parsimoniously.

5.2. New risks require compensation

As with any connected device (e.g. computers, Internet of Things devices), a risk of cyberattack exists. It can be organized by a foreign governmental agency, a malicious person or a malicious software (Knyrim and Trieb, 2011). The Russian attack on Ukrainian DSO Kyivoblenergo on December 23, 2015 is the first example of such an organized cyberattack used to temporarily shut down 30 substations of the distribution grid (Lee et al., 2016). The grid is a strategic target and the use of digital central control system makes them obvious targets for cyberattack. However, the attack did not target metering data but the stability of the grid which in this specific case do not affect privacy. Nevertheless, a cyberattack could also be conducted by customers on their own smart meter to steal electricity (McDaniel and McLaughlin, 2009, Colak et al., 2016), or by a malicious person on a specific customer to spy on him (McKenna et al., 2012).

Hence customers are exposed to new risks of privacy breaching in giving access to their metering data. The risk is considered and efforts on securing communication are made to limit it. However, the risk is not null, and it should be addressed with compensations/benefits (Wilson et al., 2017). As utilities transfer the cost of smart meters to the

customers (Table 1), it would fair that customers get rewarded according to the amount of information transmitted to utilities (Culnan and Bies, 2003).

We could imagine a voluntary basis system where customers could chose the resolution of the data provided and would have tariff/remunerations accordingly. The differences between the static tariff (i.e. for non-metered customers) and dynamic tariff (i.e. for metered customers) should then take into account the marginal cost of generation but also a deduction for providing data for the forecast of the demand.

5.3. Balance of the benefits

Balancing the benefits resides in a trade-off between the loss of privacy and increased risk for customer and the need of metering data for utilities (Culnan and Bies, 2003). We want to show that it is possible (even necessary) to do it ethically in order to be sustainable and avoid a future backlash (Zachary, 2011).

Today the benefits of smart metering are going toward the utilities which are saving cost of employing meter readers, process invoice automatically and get more insights on the grid for fraud detection and maintenance (Hu et al., 2015). As the meters are payed mostly through network tariffs (only Italy, Romania, Slovakia and Sweden are sharing costs between customers and DSO see Table 1), the tariff reductions due to potential savings will be, in a first time, shortened on the customer side. Real-time data has not shown that much interest and it seems fair to think that it will take some years before the technology becomes mature enough to deploy large scale DSM application. Hence until dynamic tariffs are generalized to all EU Member States, some customers will pay for the technology without having any of the benefits. The access given to consult and analyze electricity consumption, as promoted by EU, would then have only little impact, as the customers could only reduce their consumption to reduce their electricity bill.

Moreover, customers undercharged because of malfunctioning electromagnetic meters, which is common because of the advanced age of electromagnetic meters, will observe an increase of their electricity bill due to increased metering accuracy (Krishnamurti et al., 2012). A more precise billing means also that it is easier for the energy providers to detect fraud in comparison to annual metering on electromagnetic meters. To give an idea of the cost of electricity theft, it is responsible in 2000 in the US of 0.5% to 3.5% losses of the annual growth revenue which seems low but still represents $10 billions, compensated by higher price on the other customers (Smith, 2004). With smart meters, the risk of undetected frauds decreases which means that theoretically the cost of the fraud is reduced and can be translated into lower prices. Fraud is better monitored, but at the same time, the risk of electrocution in compromising smart meters (i.e. through software) is much lower than with electromagnetic meters and thus less repealing to possible thieves (McDaniel and McLaughlin, 2009).

Smart meters are part of the Advanced Metering Infrastructure (AMI) which forms the informational backbone of the smart grid and makes the grid smart. From a DSOs perspective, AMI allows them to have precise information about the power flows at a distribution level beyond the substations. Again this value of metering is emphasized by the increase of variable RES in the generation mix and decentralized generation down to the distribution

level (Finster and Baumgart, 2014). This way it lower the risks of outages, the DSOs can also anticipate the maintenance and solve problems faster as they do not need customers' calls to be aware of them.

Whatever the decision taken to increase the capacity of RES and subsequently to limit climate change due to electricity generation, dynamic information on the demand side will be required to optimally use RES, invoice prosumers and balance the generation with the demand (Klass and Wilson, 2016). Hence the roll-out of smart meter from a environment and grid management perspective is not negotiable, but the way the data is used and how the benefit of such infrastructure will be shared are still under discussion. If as in the case of the Netherlands (Hoenkamp et al., 2011), an opt-out is negotiated in most of the member states, customers will perform a cost-benefit analysis between the loss of privacy and the social, economical benefits generated (Culnan and Bies, 2003). The worst scenario could lead to the loss of an important part of the metered households.

6. Conclusions and perspectives

Utilities are putting efforts in complying with the GDPR. GDPR however only protects the basic rights of the data subjects (i.e. customers). Nevertheless DSOs must also respect the agenda of the Third Energy Package and deploy smart meters in due time. Hence mandatory installation of smart meters appears to be a good solution. This is not counting on possible backlash of the customers. The Netherlands' case gives us a picture of what could happen in EU during the next years (Hoenkamp et al., 2011).

Bad practices during the roll-out (e.g. installing smart meters without informing customers) are giving a negative image of smart grid technology to the customers. The fact that it is required to improve the use of RES or that it is imposed by EU are not arguments supporting the implementation of smart grid as it is right now. On the contrary, all the technological development and investment would be a vain attempt if customers do not adopt or at least accept the technology in a first time. Problems may not rise during the roll-out but also afterwards as the relationship is dynamic and requires efforts on both side to be fruitful.

It becomes then clear that insights from social science are necessary to understand how customers are perceiving 'smart meters' in *their* home. Indeed, from a engineer point of view, customer as non-rational in its decision making is often seen as a problem. Hence only the technico-economic aspects of the problem are used to evaluate a framework. The use of social science like consumer science, justice research, ethics and anthropology to large infrastructure projects involving citizens has shown positive influence on their perception regarding potentially non-appreciated project (i.e. construction of a dam, road).

In the general context of climate change, citizens have become aware at different levels of their responsibilities as well as how they can influence the outcome. Energy and electricity by extension, represents one of the fields were awareness is growing rapidly. It seems that customers are ready to become actors of the grid and support the development toward a greener electricity generation but not at any cost.

Acknowledgement

The authors thank the EUDP for funding through the EnergyLab Nordhavn project (EUDP 64015-0055).

References

Acquisti, A., Gross, R., 2009. Predicting Social Security numbers from public data. Proceedings of the National Academy of Sciences of the United States of America 106 (27), 10975–10980.

Allcott, H., 2011. Social norms and energy conservation. Journal of public Economics 95 (9-10), 1082–1095.

Barocas, S., Nissenbaum, H., 2014. Big data's end run around anonymity and consent. Cambridge University Press, NY.

Bertoldo, R., Poumadère, M., Rodrigues, L. C., 2015. When meters start to talk: The public's encounter with smart meters in France. Energy Research and Social Science.

Bondy, D. E. M., Heussen, K., Gehrke, O., Thavlov, A., 2015. A Functional Reference Architecture for Aggregators. In: IEEE International Conference on Emerging Technologies and Factory Automation, ETFA. Vol. 2015-Octob. pp. 1–7.

Chicco, G., 2016. Customer behaviour and data analytics. In: Proceedings of the 2016 International Conference and Exposition on Electrical and Power Engineering, EPE 2016. pp. 771–779.

Colak, I., Sagiroglu, S., Fulli, G., Yesilbudak, M., Covrig, C.-F., 2016. A survey on the critical issues in smart grid technologies. Renewable and Sustainable Energy Reviews 54, 396–405.

Council of Europe, 1985. Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data.

Culnan, M. J., Bies, R. J., 2003. Consumer privacy: Balancing economic and justice considerations. Journal of social issues 59 (2), 323–342.

De Montjoye, Y.-A., Hidalgo, C. A., Verleysen, M., Blondel, V. D., 2013. Unique in the crowd: The privacy bounds of human mobility. Scientific reports 3.

Ding, Y., Pineda, S., Nyeng, P., Østergaard, J., Larsen, E. M., Wu, Q., 2013. Real-time market concept architecture for EcoGrid EU - A prototype for European smart grids. IEEE Transactions on Smart Grid 4 (4), 2006–2016.

European Commission, 2011. Energy Efficiency Plan 2011. Tech. rep., European Commission.

European Commission, 2014. Benchmarking smart metering deployment in the EU-27 with a focus on electricity. European Commission, 1–10.

European Commission, 2018. What is personal data? URL https://ec.europa.eu/info/law/law-topic/data-protection/reform/what-personal-data{_}en

European Court of Human Rights, 1950. European Convention on Human Rights. Convention for the Protection of Human Rights and Fundamental Freedoms, 30.

European Parliament, 1995. Directive 95/46/EC. Tech. rep., European Parliament.

European Parliament, 2009a. Directive 2009/136/EC. Official Journal of the European Union 337, 11–36.

European Parliament, 2009b. Directive 2009/29/EC. Official Journal of the European Union 140, 63–87.

European Parliament, 2012. Directive 2012/27/EU. Official Journal of the European Union L315/1 (October), 1–56.

European Union, 2009. Directive of 2009/72/EC. Official Journal of the European Union L211 (August), L 211/55 – L 211/93.

European Union, 2016. Regulation 2016/679. Official Journal of the European Union 2001.

Farhangi, H., 2010. The path of the smart grid. IEEE power and energy magazine 8 (1).

Finster, S., Baumgart, I., 2014. Privacy-aware smart metering: A survey. IEEE Communications Surveys & Tutorials 16 (3), 1732–1745.

Giordano, V., Gangale, F., Fulli, G., Jiménez, M. S., Onyeji, I., Colta, A., Papaioannou, I., Mengolini, A., Alecu, C., Ojala, T., et al., 2011. Smart grid projects in europe. JRC Ref Rep Sy 8.

Gosling, S. D., Augustine, A. A., Vazire, S., Holtzman, N., Gaddis, S., 2011. Manifestations of Personality in Online Social Networks: Self-Reported Facebook-Related Behaviors and Observable Profile Information. Cyberpsychology, Behavior, and Social Networking 14 (9), 483–488.

Gross, R., Acquisti, A., Heinz III, H., 2005. Information revelation and privacy in online social networks. In: Proceedings of the 2005 ACM workshop on Privacy in the electronic society. p. 80.

Gutwirth, S., Leenes, R., De Hert, P., Poullet, Y., 2013. European data protection: Coming of age. In: European Data Protection: Coming of Age. Springer Science & Business Media, pp. 1–440.

Hoenkamp, R., Huitema, G. B., de Moor-van Vugt, A. J. C., 2011. Neglected consumer: The case of the smart meter rollout in the netherlands, the. Renewable Energy L. & Policy Rev., 269.

Horne, C., Darras, B., Bean, E., Srivastava, A., Frickel, S., 2015. Privacy, technology, and norms: The case of Smart Meters. Social Science Research.

Hu, Z., Kim, J.-h., Wang, J., Byrne, J., 2015. Review of dynamic pricing programs in the US and Europe: Status quo and policy recommendations. Renewable and Sustainable Energy Reviews 42, 743–751.

Jegen, M., Philion, X. D., 2017. Power and smart meters: A political perspective on the social acceptance of energy projects. Canadian Public Administration 60 (1), 68–88.

Khurana, H., Hadley, M., Lu, N., Frincke, D. A., 2010. Smart-grid security issues. IEEE Security & Privacy 8 (1).

Kitchin, R., 2016. The ethics of smart cities and urban science. Phil. Trans. R. Soc. A 374 (2083), 20160115.

Klass, A. B., Wilson, E. J., 2016. Remaking Energy: The Critical Role of Energy Consumption Data. Cal. L. Rev. 104, 1095.

Klemenjak, C., Goldsborough, P., 2016. Non-Intrusive Load Monitoring: A Review and Outlook. archiv.
URL http://arxiv.org/abs/1610.01191

Knyrim, R., Trieb, G., 2011. Smart metering under EU data protection law. Access 1 (2), 121–128.

Kosinski, M., Matz, S. C., Gosling, S. D., Popov, V., Stillwell, D., 2015. Facebook as a research tool for the social sciences: Opportunities, challenges, ethical considerations, and practical guidelines. American Psychologist 70 (6), 543–556.

Kosinski, M., Stillwell, D., Graepel, T., 2013. Private traits and attributes are predictable from digital records of human behavior. Proceedings of the National Academy of Sciences 110 (15), 5802–5805.

Krishnamurti, T., Schwartz, D., Davis, A., Fischhoff, B., de Bruin, W. B., Lave, L., Wang, J., 2012. Preparing for smart grid technologies: A behavioral decision research approach to understanding consumer expectations about smart meters. Energy Policy 41, 790–797.

Laperdrix, P., Rudametkin, W., Baudry, B., 2016. Beauty and the Beast: Diverting modern web browsers to build unique browser fingerprints. In: 37th IEEE Symposium on Security and Privacy (S&P 2016). pp. 878–894.

Lee, R. M., Assante, M. J., Conway, T., 2016. Analysis of the cyber attack on the Ukrainian power grid. SANS Industrial Control Systems, 23.
URL https://ics.sans.org/media/E-ISAC{_}SANS{_}Ukraine{_}DUC{_}5.pdf

Liu, Y., Gummadi, K. P., Krishnamurthy, B., Mislove, A., 2011. Analyzing facebook privacy settings: user expectations vs. reality. In: Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference. ACM, pp. 61–70.

McCrae, R. R., John, O. P., 1992. An introduction to the five-factor model and its applications. Journal of personality 60 (2), 175–215.

McDaniel, P., McLaughlin, S., 2009. Security and privacy challenges in the smart grid. IEEE Security & Privacy 7 (3).

Mcdonald, D. W., Ackerman, M. S., 2000. Expertise recommender: a flexible recommendation system and architecture. In: Proceedings of the ACM conference on Computer supported cooperative work. pp. 231–240.

McKenna, E., Richardson, I., Thomson, M., 2012. Smart meter data: Balancing consumer privacy concerns with legitimate applications. Energy Policy 41, 807–814.

Molla, R., 2018. Google leads the world in digital and mobile ad revenue.
URL hhttps://www.recode.net/2017/7/24/16020330/google-digital-mobile-ad-revenue-world-leader-faceb

Nissenbaum, H., 2011. A Contextual Approach to Privacy Online. Daedalus 140 (4), 32–48.
OECD, 2013. Exploring Data-Driven Innovation as a New Source of Growth: Mapping the Policy Issues Raised by "Big Data". OECD Digital Economy Papers no. 222.
Papakonstantinou, V., Kloza, D., 2015. Legal Protection of Personal Data in Smart Grid and Smart Metering Systems from the European Perspective. In: Smart Grid Security. London: Springer, Ch. 2, pp. 41–129. URL http://link.springer.com/10.1007/978-1-4471-6663-4
Pepermans, G., 2014. Valuing smart meters. Energy Economics.
Ponce, P., Polasko, K., Molina, A., 2016. End user perceptions toward smart grid technology: Acceptance, adoption, risks, and trust. Renewable and Sustainable Energy Reviews 60, 587–598.
Siano, P., 2014. Demand response and smart grids – A survey. Renewable and Sustainable Energy Reviews 30, 461–478.
Smart Grids Task Force Expert Group 1- Standards and Interoperability, 2016. My Energy Data. Tech. rep., European Commission.
Smith, T. B., 2004. Electricity theft: A comparative analysis. Energy Policy 32 (18), 2067–2076.
Strbac, G., dec 2008. Demand side management: Benefits and challenges. Energy Policy 36 (12), 4419–4426.
Tzafestas, S. G., 2018. Ethics and law in the internet of things world. Smart Cities 1 (1), 98–120.
United Nations, 1949. United Nations Universal Declaration of Human Rights 1948. Office of the Hight Comissioner for Human Rights, 11.
Wilson, C., Hargreaves, T., Hauxwell-Baldwin, R., 2017. Benefits and risks of smart home technologies. Energy Policy.
Wladawsky-Berger, I., feb 2015. The Big Data Era Is Here.
Zachary, G., 2011. Saving smart meters from a backlash. IEEE Spectrum 8 (48), 8.

[Paper B] Online Adaptive Clustering Algorithm for Load Profiling.

Sustainable Energy, Grids and Networks 17 (2019) 100181

Contents lists available at ScienceDirect

Sustainable Energy, Grids and Networks

journal homepage: www.elsevier.com/locate/segan

Online adaptive clustering algorithm for load profiling

G. Le Ray *, P. Pinson

Centre for Electric Power and Energy, Technical University of Denmark, Kgs. Lyngby, Denmark

ARTICLE INFO

Article history:
Received 17 April 2018
Received in revised form 9 August 2018
Accepted 6 December 2018
Available online 12 December 2018

Keywords:
Clustering
Load profiling
Smart grid
Time-series analysis

ABSTRACT

With the large-scale deployment of smart metering, energy sector is facing 'Big Data' related challenges. While metered customers generate streams of data, load profiling methods are not taking advantage of this structure. Indeed, insights on the demand are traditionally provided by static typical load profiles. Renewable energy sources generate intermittency in the production and subsequently uncertainty in aligning the generation to the demand at any time. This work proposes a new view on load profiling that takes benefit of the stream structure of the data, an adaptive and recursive clustering method that generates typical load profiles updated to newly collected data. The online adaptive clustering algorithm is based on an online K-means approach using a dynamic time warping based distance associated with a facility location to adjust the number of typical load profiles. The performance of the algorithm is evaluated on a synthetic dataset and applications are presented on real-world dataset from both electricity and central district heating.

1. Introduction

The energy sector is following the trends of Big Data and growing interest is placed in collecting and analyzing energy data. The development of Advanced Metering Infrastructure (AMI), and Information and Communication Technologies (ICT), thanks to public investments (Third Energy Package in Europe [1] and the American Recovery and Reinvestment Act in the United States of America [2]), constitutes the informational backbone of the energy sector. These investments were motivated by the implementation of 'greener' energy policies supporting the use of larger amount of Renewable Energy Sources (RES). RES are known to introduce intermittency in the production, hence, if no actions are taken, the larger the share of RES in the generation, the farther it can be from the consumption. It is then a necessity to have information about the demand status at a resolution that allows operators to take actions to minimize the use of conventional energy sources.

Typical load profiles, that describe the consumption of group of customers, are used to provide information to the utilities about the demand status. Load profiles have been mainly used in the electricity sector, but other energy fields (e.g. central district heating, gas) can benefit from its development. The Electricity sector is also the main leverage in transitioning to more sustainable energy production/consumption providing the highest share of RES. As production has to meet the consumption at any time, RES intermittency has to be compensated on the consumption side. Solutions exist to optimize the use of RES (i.e. storage, demand side management) but they all rely on precise information on generation and demand statuses.

Typical load profiles were traditionally generated for the electricity sector by segmenting customers by activities . With only a few customers metered, demographics or customers categories available, the goal was to estimate load profiles of non-metered customers, using assumptions to extrapolate hourly load profiles from yearly consumption. It was known to be inaccurate as subpopulations exist in almost each category [3]. Deployment of AMI has drastically changed the paradigm of load profiling from estimating the demand using a few metered customers to summarizing the information contained in a large pool of metered customers.

State-of-the-art load profiling methods use clustering algorithms combined with dimension reduction techniques on batches of historical data and generate static typical load profiles which suppose that loads are repeated over years taking into account seasonality and other temporal periodicity of the data [3–7]. This is a first step toward data-driven load profiling. Nevertheless, technological evolution of white appliances and the penetration of new type of electrical appliances (e.g. electric vehicle, heat pumps, batteries) are actually changing the loads on a time scale which requires the profiling to be rerun more often. A recursive approach to load profiling would be a computationally inexpensive way to do so. The problem of temporal dependence, inherent to time series, has been tackled by using Wavelet transform [8] or Fourier transform [9] that can be seen as dimension reduction techniques. However, both the number of clusters and the assignment of customers to a cluster remain static and the clustering has to be rerun to be updated with newly collected data. Bayesian framework is by essence inferential and is a solution to the problem of

* Corresponding author.
E-mail address: gleray@elektro.dtu.dk (G. Le Ray).

https://doi.org/10.1016/j.segan.2018.100181
2352-4677/

Table 1
The different type of clustering-based load profiling [12].

Number of clusters:	Static	Dynamic
Load static	Type 1	Type 2
Load dynamic	Type 3	Type 4

updating the clusters to newly collected data. Example of Gaussian mixture models [10] and Dirichlet process [11] can be found in the literature but none of them challenges the temporal static structure of the clustering.

Benítez et al. have defined four types of clustering-based load profiling (Table 1) depending on whether loads are considered dynamic or static and whether the number of clusters K is evolving or not. In this categorization, the work in [5–11] are Type 1. In their latest works, Benítez et al. have implemented type 2 [12] and type 3 [13] clustering-based load profiling. In this paper, we present a Type 4 clustering-based load profiling methodology. The clustering process is (i) flexible, customers can change cluster; (ii) adaptive, the number of clusters can change according to data structure; (iii) online, the typical load profiles are dynamic and recursively updated and (iv) it respects time dependency of load patterns.

The remainder of this paper is organized as follow: Section 2 presents the notations and the preprocessing of the data, the methodology is introduced in Section 3, the performance of the online adaptive clustering is evaluated on a synthetic dataset in Section 4. Real-world data applications are presented in Section 5 and the work is ultimately concluded with an outlook in Section 6.

2. Preliminaries: Notations and data preprocessing

The notations used to present the online adaptive clustering algorithm in the following sections are first introduced. Smart meters can record consumption at high frequency (up to 1 s), however data are broadcast into batches at regular intervals during a day to minimize the communication costs. In this framework, metering data are then not collected online but as blocks of data $\mathbf{\Omega}^t$ of fixed length (e.g. days, weeks). Each block $\mathbf{\Omega}^t$, $t \in \{0, \dots, T\}$ consists into a set of vectors

$$\mathbf{\Omega}^t = \{\mathbf{X}_1^t, \dots, \mathbf{X}_i^t, \dots, \mathbf{X}_I^t\}, \tag{1}$$

where each vector $\mathbf{X}_i^t$ is a load from meter i at time step t with all the same length (e.g. 24 for blocks of a day with hourly resolution). The set of typical load profiles $\mathbf{\Upsilon}^t$ generated after clustering $\mathbf{\Omega}^t$ at step t is formed by K^t vectors

$$\mathbf{\Upsilon}^t = \{\mathbf{Y}_1^t, \dots, \mathbf{Y}_k^t, \dots, \mathbf{Y}_{K^t}^t\}, \tag{2}$$

where K^t is the number of clusters at time step t. The clustering algorithm used in this work is distance-based. The distances between the set of loads $\mathbf{\Omega}^t$ and the typical load profiles $\mathbf{\Upsilon}^{t-1}$ at previous time step $t-1$ are calculated and stored into a matrix

$$\mathbf{D}^t = d(\mathbf{\Omega}^t, \mathbf{\Upsilon}^{t-1}) = \begin{bmatrix} d_{11}^t & \dots & d_{1k}^t & \dots & d_{1K^t}^t \\ \vdots & \ddots & & & \vdots \\ d_{i1}^t & & d_{ik}^t & & d_{iK^t}^t \\ \vdots & & & \ddots & \vdots \\ d_{I1}^t & \dots & d_{Ik}^t & \dots & d_{IK^t}^t \end{bmatrix}. \tag{3}$$

The distance between a load $\mathbf{X}_i^t$ and a typical load profile $\mathbf{Y}_k^{t-1}$ is noted d_{ik}^t. A vector of labels $\mathbf{A}^t = [A_1^t, \dots, A_I^t]$ specifying to which typical load profile k each individual load $\mathbf{X}_i^t$ in $\mathbf{\Omega}^t$ is assigned to is generated using the operator,

$$\forall i \in [1, I], A_i^t = \underset{k}{\operatorname{argmin}}\,(\mathbf{d}_i^t) \tag{4}$$

on each line $\mathbf{d}_i^t$ of matrix $\mathbf{D}^t$. The result of assigning each load to a typical load profile is a partition of the loads $\mathbf{\Pi}^t = \{\mathbf{P}_1^t, \dots, \mathbf{P}_k^t, \dots, \mathbf{P}_{K^t}^t\}$. As an example, we define a set of five loads $\{\mathbf{X}_1, \dots, \mathbf{X}_5\}$, and a possible partition into two clusters could be, $\mathbf{\Pi} = \{\mathbf{P}_1 = \{\mathbf{X}_1, \mathbf{X}_3, \mathbf{X}_4\}, \mathbf{P}_2 = \{\mathbf{X}_2, \mathbf{X}_5\}\}$.

If raw loads are clustered, the main information used to cluster would be the average consumption, the amplitude of the peak and the exact peak location. To make the loads comparable, the loads $\mathbf{X}_i^t$ have to be normalized. In this work, we have opted to divide each load by their reference power (i.e. maximum consumption over the period) so that they are bounded to [0, 1] [5].

In the remainder of this paper, the t index will be omitted to simplify the notation when objects from the same time step are used.

3. Proposal clustering algorithm

The splitting of this section into subsections is materialized by dotted rectangles in Fig. 1. Section 3.1 explains the setting of the online clustering algorithm parameters. Section 3.2 presents the online clustering that consists into an iterative process based on the K-means algorithm that connects time steps. Section 3.3 presents the facility location approach, used to evaluate if extra cluster centers should be created.

3.1. Consensus clustering: Online parameters setting

The online adaptive clustering algorithm requires three parameters: $\mathbf{\Upsilon}^0$, the first set of typical load profiles; K^0, the number of cluster centers at $t = 0$ and $d(\mathbf{\Omega}^0, \mathbf{\Upsilon}^0)$, the distance matrix at $t = 0$. To set these parameters, available historical data can be used offline to create the first partition $\mathbf{\Pi}^0$.

Determining the optimal number of clusters K without *prior* information is a non-trivial and computationally expensive problem. The classical way to determine it is by running several instances, $q \in \{1, \dots, q, \dots, Q\}$, of a clustering algorithm with different values of K_q to select K^0 that minimizes a criterion [14]. Thereafter only the partition generated with the optimal K^0 is kept, the other instances being used only to determine K^0. In this work, we opted for consensus (or ensemble) clustering to determine K^0 and generate a robust partition [15]. Consensus clustering consists into running in parallel several instances q of clustering algorithm(s) (possibly different) with different values of K_q that return q partitions $\mathbf{\Pi}_q$ of historical data $\mathbf{\Omega}^0$. Every instance contributes to determine K^0, the partition $\mathbf{\Pi}^0$ and the cluster centroids $\mathbf{\Upsilon}^0$. From the partitions $\{\mathbf{\Pi}_1, \dots, \mathbf{\Pi}_q, \dots, \mathbf{\Pi}_Q\}$ a probability distance $\mathbf{D}_p$ is calculated,

$$\mathbf{D}_p(i, l) = 1 - \frac{\sum_{q=1}^{Q} K_q \delta(\mathbf{A}_{qi}, \mathbf{A}_{ql})}{\sum_{q=1}^{Q} K_q}, \tag{5}$$

$$\text{with } \delta(a, b) = \begin{cases} 1, & \text{if } a = b \\ 0, & \text{if } a \neq b \end{cases} \tag{6}$$

δ the co-occurrence matrix specifies if two points are in the same cluster in each instance and creates a linkage between each pair of loads in $\mathbf{\Omega}^0$. The sum of the co-occurrence matrices is weighted by the number of clusters K_q set in each instance. The benefits of consensus clustering are multiple; all instances are used to build the probability distance matrix; the final partition is a consensus between the different instances so it reduces the bias of each instance; it gets freed from a potential bias due to the choice of initial set of cluster centers.

Using the dendrogram of the hierarchical ascending clustering implemented on the probability distance matrix with the Ward criterion, the user determines the number of clusters K^0 . The expertise of the user in solving a specific problem is necessary to

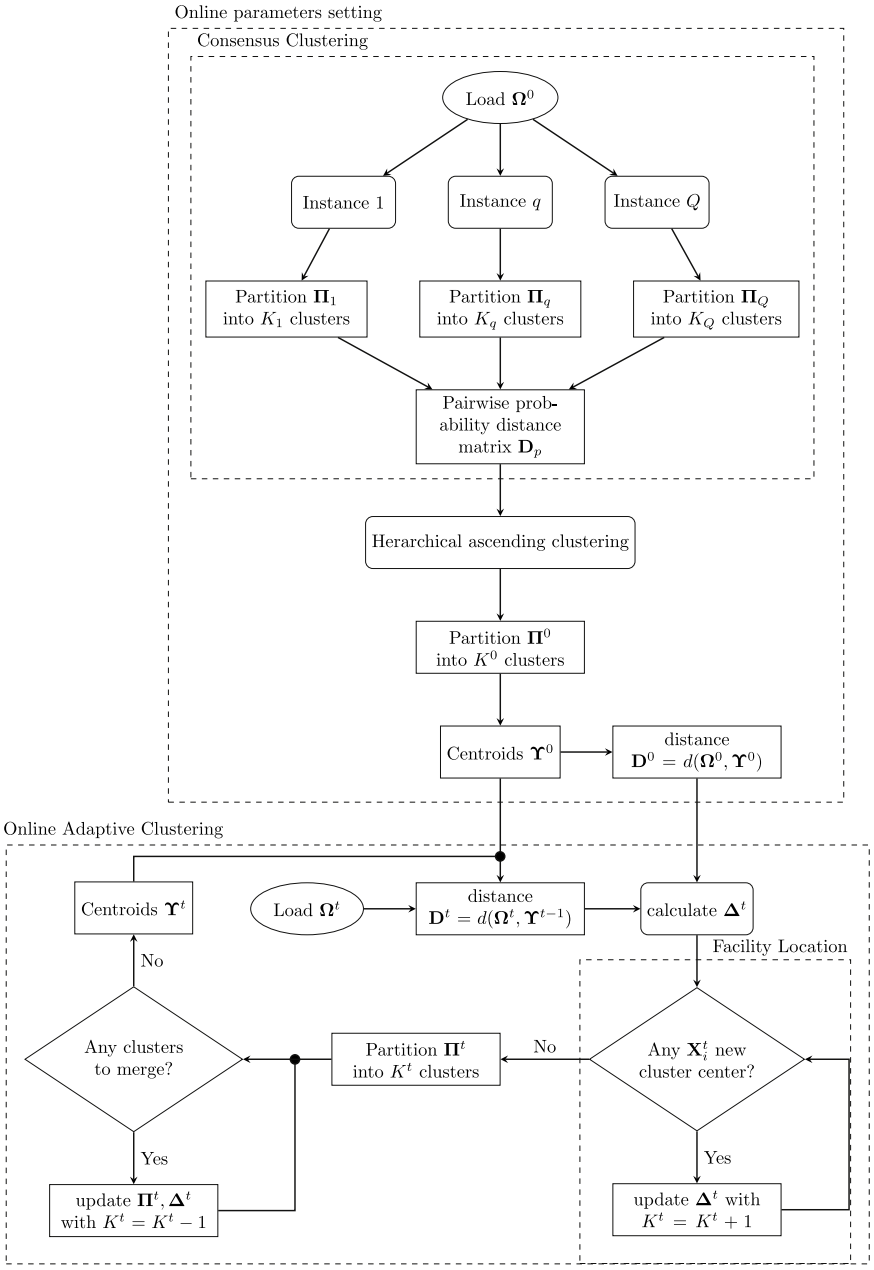


Fig. 1. Overview of the online adaptive clustering algorithm.

determine the optimal number of clusters K^0. From the K^0 clusters, the initial typical load profiles, $\mathbf{\Upsilon}^0$ are calculated by the average of the loads in each cluster, and the initial distance matrix $\mathbf{D}^0 = d(\mathbf{\Omega}^0, \mathbf{\Upsilon}^0)$ of each loads in $\mathbf{\Omega}^0$ to the typical load profiles $\mathbf{\Upsilon}^0$ is generated.

3.2. Online clustering

The online clustering algorithm takes as input the results of the consensus clustering, the typical load profiles $\mathbf{\Upsilon}^0$ and $\mathbf{D}^0$ the matrix of distances between the loads in $\mathbf{\Omega}^0$ and $\mathbf{\Upsilon}^0$ (Fig. 1). The time iterative process of the online clustering uses the core of K-means algorithm with:

- An assign step, where loads in $\mathbf{\Omega}^t$ are assigned to the closest centroid in $\mathbf{\Upsilon}^{t-1}$,
- Update centroids by averaging loads in each cluster.

It differs from K-means algorithm in using exponential smoothing to transfer structural information from previous time steps during the calculation of the distance matrix,

$$\mathbf{\Delta}^t = \frac{\sum_{\tau=0}^{t} \lambda^{t-\tau} \mathbf{D}^\tau}{\sum_{\tau=0}^{t} \lambda^\tau} \tag{7}$$

λ is the exponential smoothing coefficient and takes a value in [0, 1], which corresponds to how much of the previous time step information is transmitted to the next one. The assignment of the loads in $\mathbf{\Omega}$ is based on matrix $\mathbf{\Delta}$. Implementation of the exponential smoothing relies on the assumption that loads are relatively stable and those grouped in a same cluster may exhibit the same dynamic over time and thus remain together.

The centroids are updated by averaging the loads $\mathbf{X}_i^t \in \mathbf{P}_k^t$. Hence the most recent data are used in the calculation of the typical load profiles which provides up-to-date typical load profiles.

Consumption data display time-dependency. The clustering process presented in this work is distance-based, so the distance

definition chosen should respect time dependency and create clusters based on pattern. Several distance measures using correlation coefficients, Euclidean distance and Dynamic Time Warping (DTW) have been used in clustering time series [16,17]. To tackle this problem, we suggest to use a dissimilarity index [18],

$$d(\mathbf{X}, \mathbf{Y}) = \phi[\rho(\mathbf{X}, \mathbf{Y})] d_{DTW}(\mathbf{X}, \mathbf{Y}) \tag{8}$$

that balance a first order temporal correlation coefficient

$$\rho(\mathbf{X}, \mathbf{Y}) = \frac{\sum_{m=1}^{M-1} (x_{m+1} - x_m)(y_{m+1} - y_m)}{\sqrt{\sum_{m=1}^{M-1} (x_{m+1} - x_m)^2} \sqrt{\sum_{m=1}^{M-1} (y_{m+1} - y_m)^2}}, \tag{9}$$

estimating the dynamic behaviors and the DTW distance

$$d_{DTW}(\mathbf{X}, \mathbf{Y}) = \min_{r \in M} \left(\sum_{(i,j) \in \{1,\dots,M\}^2} |x_i - y_j| \right), \tag{10}$$

for any given load $\mathbf{X}$ and typical load profile $\mathbf{Y}$ of same length M using the function ϕ

$$\phi(u) = \frac{2}{1 + \exp(u)}, \tag{11}$$

an adaptive tuning function to balance automatically the distance d_{DTW} according to the temporal correlation coefficient ρ [19]. DTW has been broadly used in time series analysis and pattern recognition, it measures temporal similarities between time series. It calculates the Euclidean distance between a point of a time series and all the points of the other time series to create a distance matrix between each pair of points. It then finds the shortest way r from the lower left corner to the upper right corner which is called a warping path. We used the Paliwal window [20] to limit the shifting to only a few time steps (i.e. a window around the diagonal of the distance matrix) depending on the data resolution and to speed up the calculation.

3.3. Adaptivity: semi-online facility location

So far the algorithm is only online, but not yet adaptive. In this section, we present a probabilistic approach called facility location to adjust the number of clusters according to new unseen data [21]. When clustering loads, two antagonist processes have to be handled:

- Stability: loads changing shape simultaneously (e.g. seasonality) should stay in the same cluster,
- Novelty: disruptive load behavior should generate a new cluster center.

Facility location tackles the second point. A cost C_i^a is set to the assignment of each load, which is proportional to the distance d_{i,A_i} of each load $\mathbf{X}_i^t$ to the closest typical load profile $\mathbf{Y}_{A_i}^{t-1}$ and a cost C^f to the creation of a new facility (i.e. cluster center) with $C^f \gg C_i^a$. C_i^a and C^f are combined to form a probability

$$p(\mathbf{Y}_{k+1} = \mathbf{X}_i) = \min\left(\frac{C_i^a}{C^f}, 1\right), \tag{12}$$

that a load becomes a new cluster center. Hence the larger d_{i,A_i}, the larger the marginal cost C_i^a of adding a new load to existing clusters, the higher the probability that it becomes its own cluster center. The facility cost is regulating the clusters' size and can be empirically chosen according to the clusters size wanted.

As the newly generated loads arrive simultaneously in a block $\mathbf{\Omega}$, facility location is implemented in a semi-online way and evaluates the creation of new cluster centers on $\mathbf{\Omega}$ instead of a load at a time as in the completely online setup [22]. A threshold γ_{min} is defined to set how many loads should exhibit a disruptive behavior to generate an extra typical load profile, thus sensitivity to outliers is reduced. Monte-Carlo simulations are run 1000 times to obtain the distribution of the number of loads above the threshold, and the mode of the distribution is used to evaluate if the threshold is reached or not. Hence it makes the algorithm consistent. If the threshold is reached, the load which is the farthest from its closest centroid is used as extra cluster center and the distance matrix $\mathbf{D}$ is then recalculated to check if other loads in $\mathbf{\Omega}$ should be assigned to the new typical load profile. Thereafter $\mathbf{\Delta}$ is also updated with $K + 1$ clusters (Fig. 1).

If two cluster centers are converging, the algorithm would ultimately merge them, but it would take many iterations. To avoid redundant typical load profiles, a minimum threshold $d_{min} < d(\mathbf{Y}_m, \mathbf{Y}_n)$ between two cluster centers m and n is defined. If a pair of centroids gets under d_{min}, they are considered similar and are merged to form a single cluster with the average of their assigned loads as centroid. Thereafter $\mathbf{\Delta}$ is also updated with $K - 1$ clusters (Fig. 1). Hence the algorithm reacts faster to decreasing number of typical load profiles over time.

3.4. Setting the parameters

The online adaptive clustering algorithm requires setting parameters empirically as the nature of the data (e.g. electric load data, central heating district data), the resolution, the preprocessing affect the clustering process. Moreover the objective of the clustering can differ from one application to another, which is why the setting of the parameters is left to the user. Table 2 gives guidelines on the action of each parameter on the clustering process.

Setting K^0 has influence only on the few first iterations (depending on λ) as the initial structure will fade out progressively thanks to the exponential forgetting. λ is influencing how much structural information is transferred from one iteration to another, the larger λ the more information transmitted. C^f and d_{min} are parameters controlling the adaptivity and more precisely controlling the radius of the clusters. A large value of C^f will allow larger clusters and thus a smaller K. A small value of d_{min} will allow the creation of smaller clusters and thus a larger K. γ_{min} can be used to limit fluctuation of K.

4. Online adaptive clustering performance evaluation using simulated data

The performance of the online adaptive clustering algorithm is first evaluated on a synthetic dataset where the true typical load profile for each consumer is known. The following section describes how synthetic loads were generated, how the clustering process was evaluated and presents the results.

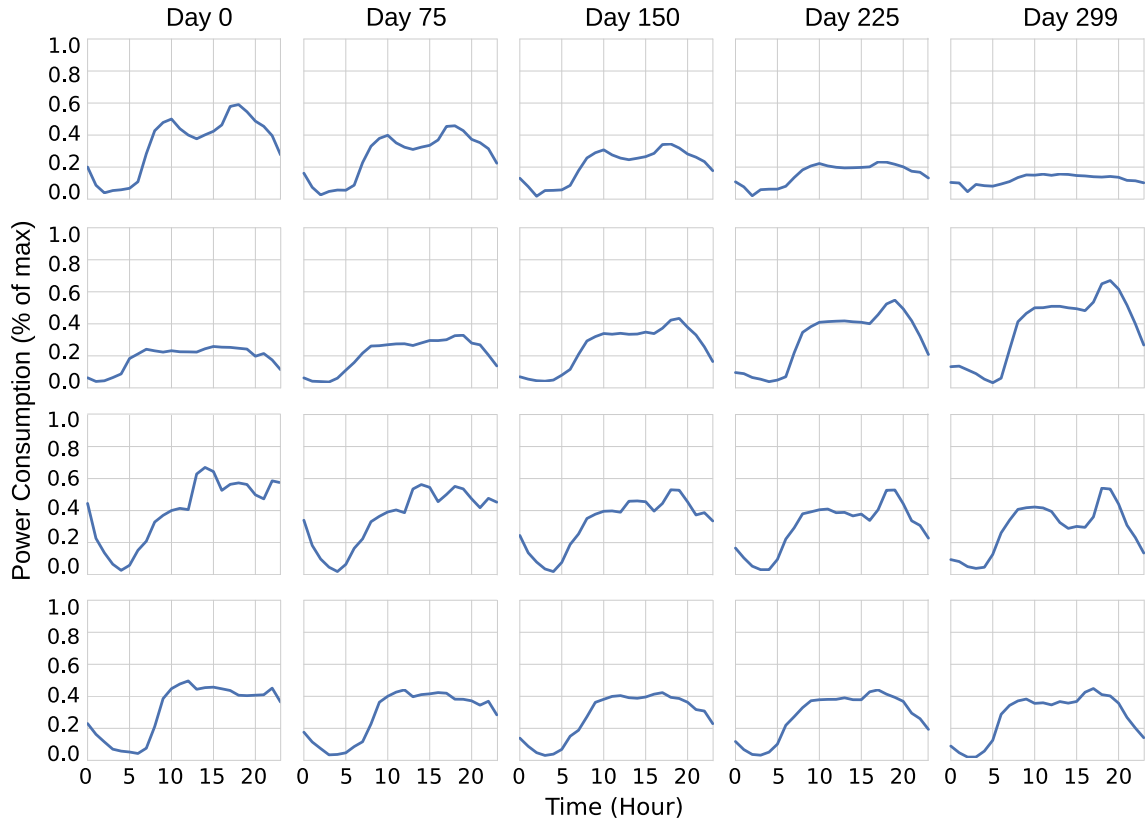
4.1. Data generation

The following simulation aims at demonstrating that the proposed algorithm can both handle slow changing typical load profiles as well as disruptive behaviors (i.e. unobserved in the previous steps). To do so four slow changing and one disruptive typical load profiles are created based on nine typical load profiles from ENTSO-E data base displaying different load behaviors at hourly resolution [23]. The four slowly changing typical load profiles are generated using four pairs of typical load profiles from ENTSO-E. The daily typical load profiles $\mathbf{Y}_k$ are synthetically created as weighted averages

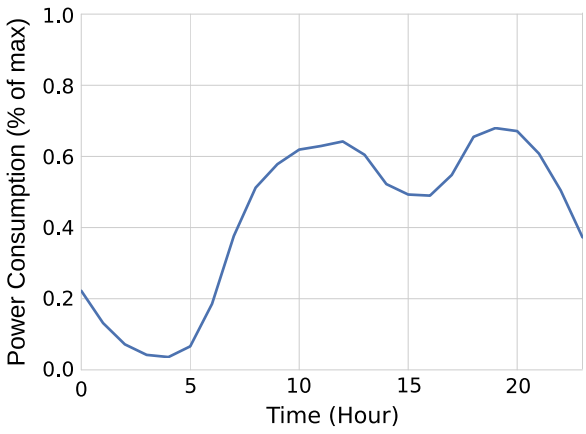
$$\mathbf{Y}_k^t = \left(1 - \frac{t}{300}\right) \mathbf{Y}_k^0 + \frac{t}{300} \mathbf{Y}_k^{299} \tag{13}$$

Table 2
Influence of the parameters on the clustering process.

Parameter	Definition	Value	Influence
K^0	Initial number of cluster centers	$n \in \{2, \ldots, N\}$	On the first iterations
λ	Exponential forgetting	$[0, 1]$	Smooth the clustering
C^f	Facility cost	Relative to C^a	Size (radius) of the clusters
d_{min}	Minimum distance between cluster centers	Relative to the expected number of clusters	Size (radius) of the clusters
γ_{min}	Number of disruptive load needed to create a new cluster	$n \in \{2, \ldots, 10\}$	Limits fluctuations of K



(a)



(b)

Fig. 2. Synthetic data generation based of ENTSO-E load profiles. Morphing of the four typical load profiles (a) and the new load behavior appearing at day 200 (b).

of $\mathbf{Y}_k^0$ and $\mathbf{Y}_k^{299}$ respectively the starting (Day 0) and ending (Day 299) profiles exhibited by typical load profile k. Fig. 2(a) presents the evolution of these typical load profiles. From the four slow changing typical load profiles, 1000 individual daily loads are sampled. Each of the 1000 simulated customers is randomly assigned to one of the four slow changing typical load profiles $\boldsymbol{\Upsilon} =$

$\{\mathbf{Y}_1, \ldots, \mathbf{Y}_4\}$. The simulated daily loads

$$\mathbf{X}_i^t = \mathbf{Y}_k^t + \mathcal{N}(0, \boldsymbol{\Sigma}) \tag{14}$$

are sampled from the typical load profiles $\mathbf{Y}_k^t$ by adding a multivariate Gaussian noise $\mathcal{N}(0, \boldsymbol{\Sigma})$. The covariance matrix

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2\rho_{11} & \ldots & \sigma_1\sigma_l\rho_{1m} & \ldots & \sigma_1\sigma_M\rho_{1M} \\ \vdots & \ddots & & & \vdots \\ \sigma_l\sigma_1\rho_{l1} & & \sigma_l^2\rho_{lm} & & \sigma_l\sigma_M\rho_{lM} \\ \vdots & & & \ddots & \vdots \\ \sigma_L\sigma_1\rho_{L1} & \ldots & \sigma_L\sigma_m\rho_{Lm} & \ldots & \sigma_L^2\rho_{LM} \end{bmatrix} \tag{15}$$

is stochastically generated. $\boldsymbol{\sigma} = \{\sigma_1, \ldots, \sigma_{24}\}$ is a normalized random vector of standard deviation and ρ is a matrix of coefficient decreasing exponentially from the diagonal

$$\rho_{lm} = \exp\left(\frac{-|l-m|}{\tau}\right), \tag{16}$$

which adds some small time shifting to the patterns. The daily load data were then normalized to [0, 1] by dividing them by their reference power.

At time step 100, 250 of the 1000 customers generated disrupt from their slow transitioning typical load profile to the fifth typical load profile (Fig. 2(b)). The number of clusters is in the same time expected to change automatically, generating an extra cluster center which groups the 250 disruptive loads. At time step 200, the 250 customers are catching back their original slow transitioning typical load profiles. The number of clusters is then expected to change from five to four clusters and reassigning the loads to the cluster corresponding to their original typical load profile.

4.2. Performance evaluation

The performance of the algorithm is evaluated using the Normalized Mutual Information (NMI) which is a criterion based on entropy and cross-entropy as defined in information theory. It actually evaluates how much information is shared between two vectors of labels. The entropy

$$\mathrm{H}(\mathbf{U}) = \sum_{i=1}^{I} p(u_i)\log(p(u_i)), \tag{17}$$

is the amount of disorder in vector $\mathbf{U}$ where each element can take a value in $\{u_1, \ldots, u_I\}$ and the probabilities $p(u_i)$ represent the probabilities that an object picked at random from $\mathbf{U}$ has value u_i. The Mutual Information (MI)

$$\mathrm{MI}(\mathbf{U}, \mathbf{V}) = \sum_{i=1}^{I}\sum_{j=1}^{J} p(u_i, v_j)\log\left(\frac{p(u_i, v_j)}{p(u_i)p(v_j)}\right) \tag{18}$$

is basically the entropy of the joint probability between vector of labels $\mathbf{U} = \{u_1, \ldots, u_I\}$ and $\mathbf{V} = \{v_1, \ldots, v_J\}$. In the context of this work, the joint probability $p(\mathbf{Y}_j, \boldsymbol{\Pi}_k)$ is the probability that a load is both from typical load profile $\mathbf{Y}_j$ and assigned to cluster $\boldsymbol{\Pi}_k$. MI is then normalized,

$$\mathrm{NMI}(\mathbf{U}, \mathbf{V}) = \frac{\mathrm{MI}(\mathbf{U}, \mathbf{V})}{\sqrt{\mathrm{H}(\mathbf{U})\mathrm{H}(\mathbf{V})}} \tag{19}$$

and takes value in [0, 1] where 1 is a perfect match between $\mathbf{U}$ and $\mathbf{V}$.

4.3. Online clustering setup and results

The evaluation focuses on the online adaptive clustering algorithm. Hence the consensus clustering is not performed on the synthetic data and the first partition $\boldsymbol{\Pi}^0$ is actually the real assignment of the customers to typical load profiles in $\boldsymbol{\Upsilon}$. The facility cost is fixed to $C^f = 100$, the number of points γ_{min} required to generate a new cluster is set to one, the exponential smoothing is $\lambda = 0.85$ and the minimum distance between cluster centers is $d_{min} = 0.07$.

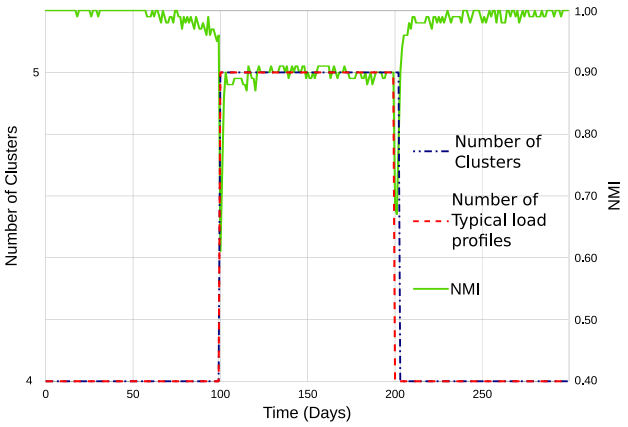


Fig. 3. Mode of the NMI and the number of clusters over the test period.

The results of the analysis are presented in Fig. 3 with the evolution of the number of clusters (blue dotted line) and the NMI (green solid line) over the test period. The number of typical load profiles used to sample the daily load curves is displayed with a red dashed line. It shows that the number of clusters generated by the online adaptive clustering algorithm follows precisely the number of typical load profiles until day 200. Even when at time step 100, the number of typical load profiles is changing from four to five. When the number of typical load profiles used to sample the loads is decreasing from five to four, we observe a delay of 2 days before the algorithm adjusts the number of clusters. Besides detecting correctly the number of clusters until day 100, the algorithm groups correctly the loads that are sampled from the same typical load profile in the same cluster as the NMI stays close to 1.0 until day 50. At day 50, a decrease of the NMI is observed. This can be explained by the real partition information provided at $t = 0$ fading away in the exponential smoothing as well as a convergence of the typical load profiles (see Fig. 2) which can engender some misclassifications. At day 100, we observe a drop to approximately 0.6 and the NMI comes quickly back to around 0.9 in the next day and oscillate around 0.9 until day 200. At day 200, the NMI drops again down to 0.75 and comes quickly back to 0.95 the next day and continues to increase until it oscillates between 0.99 and 1.0. At the end of the test period, the algorithm classifies correctly the loads despite the change from four to five and back to four typical load profiles.

The online adaptive clustering performed accurately on the synthetic dataset as it generates the correct number of clusters and groups correctly customers generated from the same typical load profile in the same cluster.

5. Applications to real-world data

The online adaptive clustering algorithm has been tested on two real-world datasets, (i) central district heating loads from 97 buildings in Copenhagen at hourly resolution for a month, (ii) 13 241 electrical loads from industries, businesses and households with PV (i.e. for billing purposes) at hourly resolution for an entire year. They exhibit different characteristics which can be observed when profiling loads and demonstrate a wider range of applications of the online adaptive load profiling clustering to energy systems. The code of the online adaptive clustering algorithm has been made available on *GitHub*[1] to the interested reader.

[1] https://github.com/gleray/Online-Kmeans.

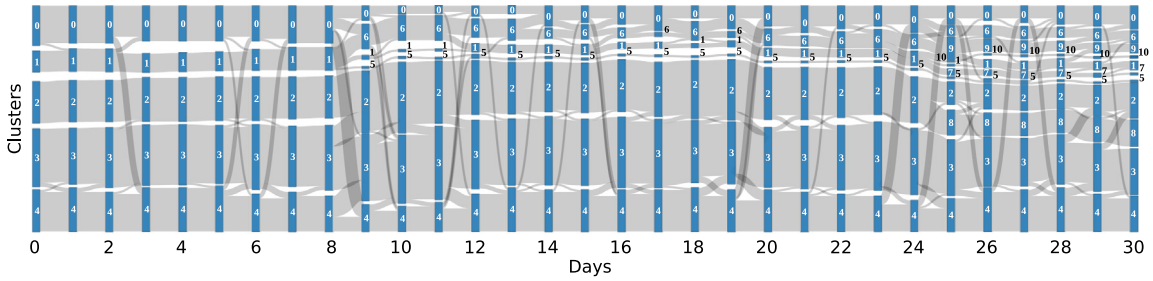


Fig. 4. Sankey graph displaying the flow of buildings in the different clusters from one day to another during the test period.

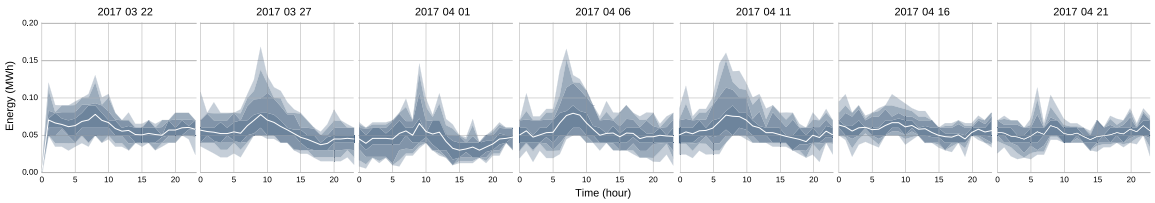


Fig. 5. Evolution of cluster 4 along the test period.

5.1. Central district heating data

HOFOR, the operator of the central district heating in Copenhagen area, provided data from 97 building over a period of a month (31 days in March–April) with hourly resolution. A block $\mathbf{\Omega}$ groups a day of data ($M = 24$) and no preprocessing has been operated on the data as the consumptions are all of the same magnitude [0, 1].

The consensus clustering is using a modified version of the K-means algorithm that uses d_{DTW} as a distance metric and is applied with $K_q = \{2, \ldots, 10\}$ on data generated at day 0. From the dendrogram obtained with the hierarchical ascending clustering, a partition $\mathbf{\Pi}^0$ into five clusters is generated. The online adaptive clustering uses the centroids $\mathbf{\Upsilon}^0$ and $\mathbf{D}^0$ obtained from $\mathbf{\Pi}^0$ as initial parameters and runs over 31 days.

The online adaptive clustering algorithm starts with $K = 5$ clusters, a facility cost $C^f = 7.5$, it needs only one building to create a new cluster center, the exponential smoothing is $\lambda = 0.85$ and the minimum distance between cluster centers is $d_{min} = 0.01$. The evolution of the clusters' composition is summarized in a Sankey diagram (Fig. 4) that displays the flow of customers between clusters from one day to the next. The clusters are stable over time, besides some adjustments with few buildings changing cluster at each iteration. At iteration 9, cluster 1 splits into cluster 1, 6 and 5. Cluster 5 is actually a single building with a pattern different from the rest of the pool. We observe again a splitting at day 25, cluster 1 splits into cluster 1, 10, 7 and cluster 9 is taking elements from cluster 1, 0 and 6. The evolution of cluster 4 is presented in Fig. 5 and shows how the typical load profile's shape changes slowly over the period.

No predefined classification is available to assess the quality of the partition, hence the number of clusters and the RMSE between individual loads and their assigned typical load profile over the test period (Fig. 6) have been used to evaluate the partition. As expected, the RMSE does not depend on the number of clusters and globally decreases over the period. It confirms first that the online adaptive clustering algorithm summarizes accurately the loads using a limited number of typical load profiles in the case of stable typical load profiles; second, it is fast as it uses only one K-means iteration at each time step and last but not least it demonstrates successful application of the online adaptive clustering algorithm for profiling non electric loads.

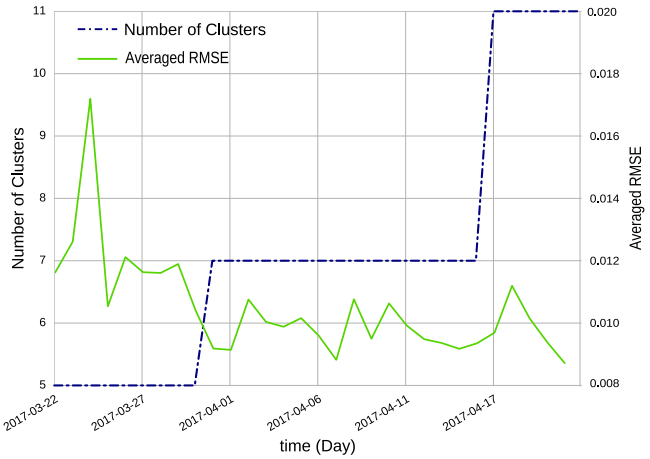


Fig. 6. Number of clusters and Averaged Root Mean Square Error (RMSE) between daily loads and their assigned typical load profile over the test period for the central district heating data.

5.2. Electrical load data

Radius, the DSO in Copenhagen area, provided the second data set. It consists of a year (240 days after removing missing data) of hourly power consumption data from $N = 13\,241$ customers. The customers metered at hourly resolution by Radius are businesses, industries and households with PVs. The blocks $\mathbf{\Omega}$ consist of a day ($M = 24$) and the loads were preprocessed by dividing them by the peak over the period as explained in Section 2.

The consensus clustering is using the same modified version of the K-means algorithm presented in Section 5.1 with $K_q\{10, \ldots, 100\}$ on day 0: 2015-01-12. The partition $\mathbf{\Pi}^0$ is obtained by cutting the dendrogram into seven clusters. It is intentionally underestimating the number of clusters present in the dataset (approx. 20).

The online adaptive clustering algorithm starts with $K = 10$ clusters centers, a facility cost set to $C^f = 950$, the minimum number of customers to create a new cluster center is $\gamma_{min} = 5$, an exponential smoothing coefficient of $\lambda = 0.85$, and a minimum distance between two cluster centers set to $d_{min} = 0.13$. A standard K-means algorithm using Euclidean distance and a Self Organizing Map (SOM) applied on the entire set of daily load profiles with K

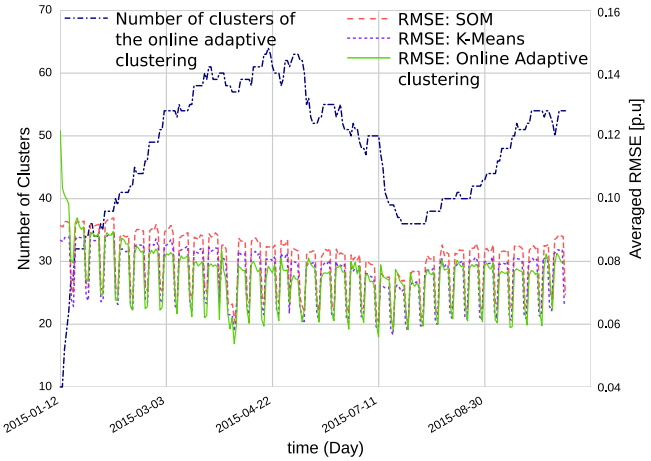


Fig. 7. Number of clusters and Averaged Root Mean Square Error (RMSE) per unit of peak load, between daily loads and their assigned typical load profile over the test period for Benchmarks: SOM (7.9% RMSE overall score), K-means (7.5% RMSE overall score) and the online adaptive clustering algorithm (7.2% RMSE overall score).

$= 50$ have been used to benchmark the online adaptive clustering algorithm [5].

The results are presented in Fig. 7 with the number of clusters and the averaged RMSE between daily loads and their assigned typical load profile for the benchmarks and the online adaptive clustering algorithm. The number of clusters generated by the online adaptive clustering algorithm first increases up to 64 around 2015-04-22 before going down to 36 during the summer and up again to 54. The averaged RMSEs show weekly periodicity which results from the customers' activity in the study, mostly industries and businesses, that are usually closed on weekends and thus displays more homogeneous patterns. The number of clusters seems to influence marginally the average RMSE of the online adaptive clustering algorithm, besides the first 14 days, which means that the algorithm is finding accurately the number of latent typical load profiles at each iteration. When the number of clusters in the online clustering algorithm gets close to 50, as set in the benchmark, the average RMSE of the benchmarks is getting closer (or equal) to the RMSE of the online adaptive clustering algorithm. The K-means actually beats the online adaptive clustering algorithm when the K^t is under 40. It stabilizes the RMSE between 0.08 (weekdays) and 0.06 (weekends) when benchmarks have a tendency to fluctuate. After removing the 30 first days of convergence to a stable solution, the overall RMSE score of the online adaptive clustering algorithm is 7.2% against 7.5% and 7.9% respectively for the K-means and SOM. In terms of computational time, on a desktop computer equipped with Intel Xeon CPU 3.50 GHz $\times$ 8 cores, the consensus clustering (computed in parallel on 7 cores) takes 15 min to be completed, and each iteration of the online adaptive clustering (on a single core) takes approximately 1 min 20 s as the overall process on 240 days took 6 h. As comparison, the SOM takes 16 h to be completed and the K-means 20 h. These results confirm the necessity of having an adaptive clustering process that estimates correctly the number of clusters in the data at time t. The difference of performance between the online adaptive clustering algorithm and the benchmark would be larger with a dataset displaying more non-stationary behavior like loads involved in Demand Response (DR) programs or with large amount of PVs.

Fig. 8 gives an example of the evolution of a cluster over the period. The shape of the typical load profile as well as the lowest value are changing over the period. The cluster groups mostly restaurants, which are active daily from lunch to dinner time with a peak at dinner time and along the year at high activity periods (Christmas holidays, Saturdays, and from April to August).

From the application on real data the algorithm has fulfilled expectations, it handles correctly slow changing and fast changing profiles, splitting and merging to keep the same overall accuracy with a low computational cost at every iteration.

6. Conclusions and future works

In this paper, we have presented an online adaptive clustering methodology for load profiling. We have demonstrated its efficiency on both synthetic and real datasets. It is more agile than traditional clustering based load profiling as it is recursive and uses the most recent data to update daily typical load profiles and computationally efficient as it processes only short time series at each time step of the online clustering. In this framework, the typical load profiles are generated every day, hence a customer's profile can be summarized by concatenating the typical load profiles assigned for each day to generate its load profile over a period of interest.

The methodology establishes a first step toward dynamic profiling of electricity consumption patterns. The deployment of smart meters associated with the increasing share of renewables will make dynamic load profiling compulsory for future grid management as fast decision making under uncertainty is becoming the common situation. A dynamic clustering methodology is then more suited for handling balancing between generation and consumption which is a dynamic problem both on production and demand sides.

From a widespread power system perspective, the methodology can be a systematic tool providing insights at a reasonable time scale for demand side management programs. It can also be used to develop new dynamic tariffs that would reflect the marginal cost each customer generates by shifting or synchronizing their peak load to the overall peak load. In electricity demand analysis a dynamic load profiling method will provide information on the stability of customers load patterns over time and whether it displays some periodicity (e.g. day open and close for a supermarket). Classic load profiling applications like estimating load for planning of the grid will also benefit from the output of the online adaptive clustering algorithm. Sampling of the most probable weekly or yearly load profiles can be generated by concatenating the daily load profiles and provides information on the uncertainty of the load behavior *via* confidence intervals or probability density distribution around the curve.

Several possible extensions of the work can be considered; the methodology could be extended to multi-energy profiling which is lacking at the moment and will be needed as multi-energy solutions exploiting volatility of different markets are being deployed (e.g. fuel shifting solution). It could also be tested on a dataset with large amounts of non-stationary loads, typically households subject to DR or equipped with PVs, EVs and batteries. As the code has been made available on *GitHub*,[2] we invite readers with such dataset to download the code and test it on their data. On a more technical aspect, the methodology could be transposed in a Bayesian framework, which will provide direct evaluation of the uncertainty.

Acknowledgments

The authors thank EnergyLab Nordhavn partners, Radius and HOFOR for supporting this work by providing data, as well as the EUDP, Denmark for funding through the EnergyLab Nordhavn project (EUDP 64015-0055).

[2] https://github.com/gleray/Online-Kmeans.

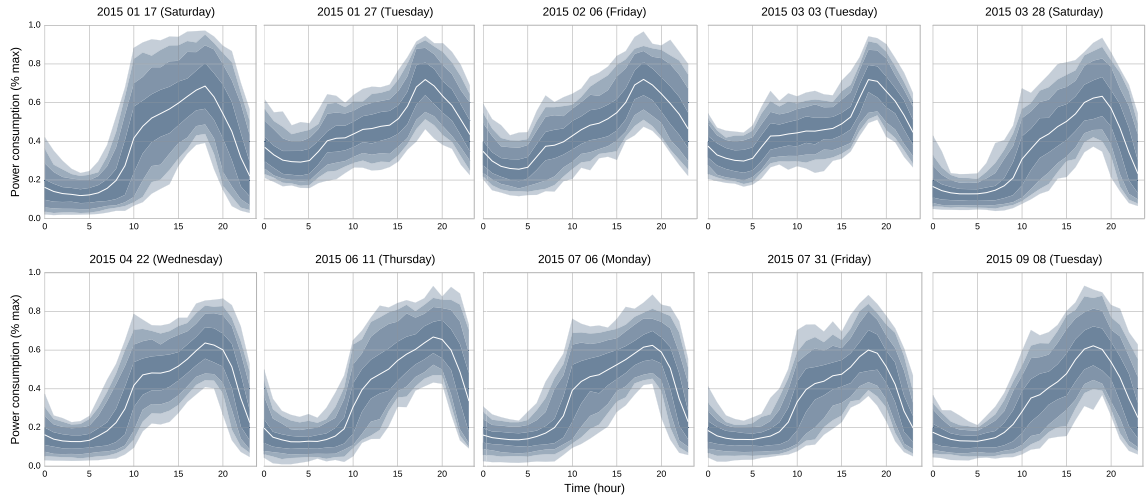


Fig. 8. Evolution of cluster 15 along the test period.

References

[1] EU. 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, 2009.

[2] American Recovery and Reinvestment Act of 2009, Public Law 1115, 2009.

[3] Yi Wang, Qixin Chen, Chongqing Kang, Mingming Zhang, Ke Wang, Yun Zhao, Load profiling and its application to demand response: A review, Tsinghua Sci. Technol. 20 (2) (2015) 117–129.

[4] Kai-le Zhou, Shan-lin Yang, Chao Shen, A review of electric load classification in smart grid environment, Renew. Sustain. Energy Rev. 24 (2013) 103–110.

[5] Gianfranco Chicco, Roberto Napoli, Federico Piglione, Comparisons among clustering techniques for electricity customer classification, IEEE Trans. Power Syst. 21 (2) (2006) 933–940.

[6] Fintan McLoughlin, Aidan Duffy, Michael Conlon, A clustering approach to domestic electricity load profile characterisation using smart metering data, Appl. Energy 141 (2015) 190–199.

[7] I P Panapakidis, M C Alexiadis, G K Papagiannis, Enhancing the clustering process in the category model load profiling, IET Gener. Transm. Distrib. 9 (7) (2015) 655–665.

[8] Kevin Mets, Frederick Depuydt, Chris Develder, Two-stage load pattern clustering using fast wavelet transformation, IEEE Trans. Smart Grid 7 (5) (2016) 2250–2259.

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

[10] Stephen Haben, Colin Singleton, Peter Grindrod, Analysis and clustering of residential customers energy behavioral demand using smart meter data, IEEE Trans. Smart Grid 7 (1) (2016) 136–144.

[11] Ramon Granell, Colin J. Axon, David C.H. Wallom, Clustering disaggregated load profiles using a Dirichlet process mixture model, Energy Convers. Manage. 92 (2015) 507–516.

[12] Ignacio Benítez, Alfredo Quijano, José Luis Díez, Ignacio Delgado, Dynamic clustering segmentation applied to load profiles of energy consumption from Spanish customers, Int. J. Electr. Power Energy Syst. 55 (2014) 437–448.

[13] Ignacio Benítez, José-Luis Díez, Alfredo Quijano, Ignacio Delgado, Dynamic clustering of residential electricity consumption time series data based on Hausdorff distance, Electr. Power Syst. Res. 140 (2016) 517–526.

[14] Charu C Aggarwal, Chandan K Reddy, Data Clustering: Algorithms and Applications, CRC press, 2013.

[15] Thomas Grotkjær, Ole Winther, Birgitte Regenberg, Jens Nielsen, Lars Kai Hansen, Robust multi-scale clustering of large DNA microarray datasets with the consensus algorithm, Bioinformatics 22 (1) (2006) 58–67.

[16] T. Warren Liao, Clustering of time series data - A survey, Pattern Recognit. 38 (11) (2005) 1857–1874.

[17] John Paparrizos, Luis Gravano, k-shape: Efficient and accurate clustering of time series, in: Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data, ACM, 2015, pp. 1855–1870.

[18] Ahlame Douzal Chouakria, Panduranga Naidu Nagabhushan, Adaptive dissimilarity index for measuring time series proximity, Adv. Data Anal. Classif. 1 (1) (2007) 5–21.

[19] Alon Efrat, Quanfu Fan, Suresh Venkatasubramanian, Curve matching, time warping, and light fields: New algorithms for computing similarity between curves, J. Math. Imaging Vision 27 (3) (2007) 203–216.

[20] Kuldip K Paliwal, Anant Agarwal, Sarvajit S Sinha, A modification over Sakoe and Chiba's dynamic time warping algorithm for isolated word recognition, Signal Process. 4 (4) (1982) 329–333.

[21] Adam Meyerson, Online facility location, in: Foundations of Computer Science, 2001. Proceedings. 42nd IEEE Symposium on, IEEE, 2001, pp. 426–431.

[22] Edo Liberty, Ram Sriharsha, Maxim Sviridenko, An Algorithm for Online K-Means Clustering, CoRR, abs/1412.5, 1–13, 2014.

[23] ENTSO-E. Consumption Data Hourly load value of a specific country for a specific day March 27th 2013, 2013. URL https://www.entsoe.eu/data/data-portal/consumption/Pages/default.aspx.

[Paper C] Unsupervised energy disaggregation: From sparse signal approximation to community detection.

Unsupervised Energy Disaggregation: From Sparse Signal Approximation to Community Detection

Pierre Winkler, Guillaume Le Ray, *Student Member, IEEE,* and Pierre Pinson, *Senior Member, IEEE*

***Abstract*—Data collected from smart meters is to be potentially used as a basis for many different purposes e.g. demand response management and pricing. A popular problem to tackle using such smart meter data is Non-Intrusive Load Monitoring (NILM), which is a deconvolution process to separate appliances and their consumption based on aggregated measurements only. State-of-the-art NILM approaches are supervised algorithms based on e.g. Factorial Hidden Markov Models (FHMM) or Artificial Neural Networks (ANN). They require large training datasets and may have limited generalization ability when used for new data and situations not seen during the training process. In contrast here, we propose an unsupervised NILM approach that combines (i) sparse signal approximation into a sum of boxcar functions by Orthogonal Matching Pursuit (OMP), (ii) a Gaussian Mixture Model (GMM) reducing redundancies in the boxcar functions, (iii) community detection to obtain appliance signatures from association of boxcar functions. The algorithm shows performance in the same range as the FHMM and NN at high resolution (6 seconds) but can also perform well at lower resolution (1 minute). As the approach is generic and unsupervised, it fits the requirements for a real-world implementation with standard metering data.**

***Index Terms*—Clustering, Community Detection, Non-Intrusive Load Monitoring (NILM), Orthogonal Matching Pursuit (OMP), Unsupervised Learning.**

I. Introduction

LARGE scale deployment of smart meters and, more generally, information and communication technologies (ICT) open up towards intelligent networks that will comprise the informational backbone of future smart grids. A prospect is to benefit from Demand Side Management (DSM) concepts to integrate Renewable Energy Sources (RES) efficiently and economically by harnessing consumer flexibility. With that objective in mind, smart meters connecting consumers and utilities (plus possibly some other agents in the system e.g. flexibility aggregators) with a two-way communication protocol allow utilities to get near real-time feedback on electricity consumption, while consumers can receive incentives e.g. dynamic tariffs, and modify their energy consumption behavior accordingly.

The exponentially growing amount of metering data brings new perspectives in terms of solving challenging R&D problems while allowing for new business models. These generally relate to developing new insights about electric demand characteristics relevant to DSOs and aggregators (e.g. consumer behavior, detection of electric vehicles and heat pumps, estimation of PV capacity), to the consumers (more control on electricity consumption) and to electricity providers (energy efficiency programs) [1]. Non-Intrusive Load Monitoring (NILM), which consists in separating single appliances and their electricity consumption signal from the overall consumption signal of a household thanks to their specific signatures, is probably the most known approach to extract information on what happens behind the meter. NILM was first described by Hart in the 90's [2] but applications became only possible with the booming of high frequency metering and the increase of computing resources as the process potentially requires heavy computational resources.

Hidden Markov Models (HMM) [3] and Artificial Neural Networks (ANN) [4] are the two dominant approaches used in the NILM literature and thus form the state-of-the-art [5]. However they are both supervised approaches which present certain drawbacks: (i) they are computationally expensive to train, (ii) a representative training (labeled) dataset should be available, and (iii) they cannot cope with new appliances (unseen during the training). Hence they perform well on overall electricity consumption signals that consist of the sum of appliance consumption signals observed during the training phase and that have a data granularity high enough to detect transient signatures. These are fairly strong restrictions for real-world implementation [5].

A few examples of unsupervised NILM approaches using event detection can be found in the literature [6], [7]. However they can only disaggregate relatively simple appliances with on/off states or multi-states with strong dependency on the labeling process. Another unsupervised approach consists in generating an over-complete dictionary of boxcar functions using a sparse decomposition algorithm to approximate the overall consumption [8]. Thereafter the boxcar functions are then labeled using a classifier which is trained on single appliances consumption patterns.

In contrast with the existing methods, we propose here an unsupervised NILM approach that takes advantage of (i) a sparse decomposition algorithm based on Orthogonal Matching Pursuit (OMP) principles, as presented in [8], (ii) a Gaussian Mixture Model (GMM) to reduce statistical redundancy of boxcar functions in the dictionary, and (iii) a community detection approach to obtain appliance signatures from association of boxcar functions. The performance is first benchmarked against existing methods (FHMM, ANN) on the UK-DALE dataset [9]. The aim of the benchmarking is to compare the performance of our algorithm against state-of-

P. Winker, G. Le Ray and P. Pinson are with the Centre for Electric Power and Energy, Technical University of Denmark, Kgs. Lyngby, Denmark (email: {gleray,ppin}@elektro.dtu.dk).

This work was partly supported by EUDP through the EnergyLab Nordhavn project (EUDP 64015-0055).

the-art methodologies. However, as it is not fair to compare supervised and unsupervised approaches, on such grounds only, we do not expect to outperform then, but at least to reach similar performance in an unsupervised environment. We subsequently evaluate how the performance of our approach is affecting by the granularity i the data collected. This is done by changing the resolution from six seconds to one minute. Having access to data at the minute resolution, or coarser, is more in line with real-world implementation that can be envisaged today.

The remainder of this paper is organized as following. After this introduction, Section II gives the overall sketch of the methodology as well as its individual components. Then, Section III concentrates on its application on the UK-DALE dataset, with both a benchmarking exercise (against state-of-the-art supervised approaches) at 6 seconds, and an analysis of the performance obtained when degrading the temporal resolution in the data at hand. Eventually, we gather a set of conclusions and perspectives for future work in Section V.

II. Methodology

The energy disaggregation problem can be mathematically formulated as

$$x^t = \sum_{n=1}^{N} y_n^t + \varepsilon^t, \quad \forall t \tag{1}$$

where y_n^t is the individual consumption of appliance n considered in the model and ε^t are the residuals which correspond to the error and the consumption of appliances not considered in the model [10]. Our approach to unsupervised NILM is illustrated in Figure 1, where the various blocks involve coarse decomposition approximation, dimension reduction, community detection and labelling. These are all covered individually and successively in the following.

A. Power Signal Sparse Approximation

Power signal sparse approximation uses a large set functions φ stored into a dictionary $\mathbf{\Phi} = \{\varphi_1, \ldots, \varphi_k, \ldots, \varphi_K\}$ to approximate the aggregated power consumption (top rectangle in Figure 1). The observation of load behavior and the over-representation of type I (i.e. ON/OFF) and type II (i.e. multistate) appliances dictate the choice of translation-invariant boxcar functions to fill the dictionary [8]. For each index k, the boxcar function is characterized by two parameters l and w, such that

$$\varphi_k^t = \frac{1}{\sqrt{w}} \, \Pi_{l-w/2, l+w/2}^t, \tag{2}$$

where l is the translation of the boxcar function, w the width and $\Pi_{a,b}^t = H(t-a) - H(t-b)$, where H is a Heaviside step function. The general shape of a boxcar function and the different parameters involved in the design are presented in Figure 2.

In a power signal sparse approximation framework, (1) is reformulated as

$$x^t = \sum_{k=1}^{K} \alpha_k \varphi_k^t + \varepsilon^t, \quad \forall t \tag{3}$$

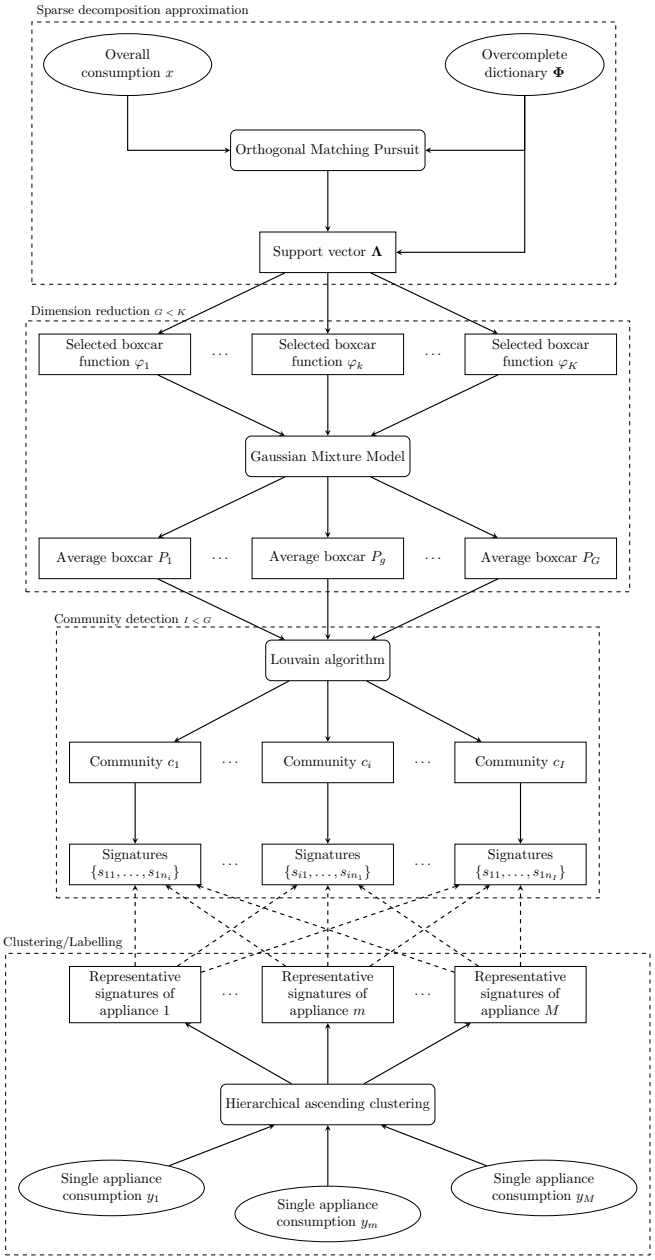


Fig. 1. Algorithm overview.

where α_k denotes the activation coefficient of the boxcar function φ_k. A compact vectorial form of the equation is, $\mathbf{x} = \boldsymbol{\alpha}\, \mathbf{\Phi} + \boldsymbol{\epsilon}$ where $\mathbf{x}$ is the load vector of all time indices, $\boldsymbol{\alpha}$ is the vector of coefficients α_k and $\boldsymbol{\epsilon}$ a vector of residuals.

The overcomplete dictionary $\mathbf{\Phi}$, with all the possible boxcar functions, is generated ahead of the approximation [11]–[13]. Hence only a subset of J boxcar functions with $J \ll K$ is used during the approximation. Subsequently, $\boldsymbol{\alpha}$ is a sparse matrix as most of the coefficients are equal to zero. A support vector Λ regroups the non-zero entries in $\boldsymbol{\alpha}$ and restricts the dictionary $\mathbf{\Phi}$ to the subset $\mathbf{\Phi}_\Lambda$. The aim of our implementation is to adjust sparsity by setting a fixed or an adaptive threshold corresponding to the minimum variation of power to consider as a change of state, in order to eliminate transient states or

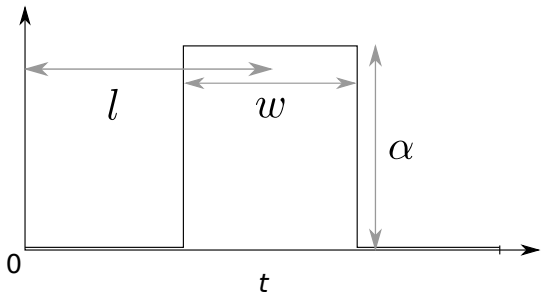


Fig. 2. Representation of the general shape of a boxcar function

internal states fluctuation of appliances.

Direct approaches to sparse approximation, like combinatorial optimization, are complex and require important computing resources to perform the approximation. A greedy algorithm using an iterative process is implemented instead [14]. Matching pursuit, the most used greedy algorithm, finds locally the optimal solution that is also close enough to a global optimal solution at each iteration. Orthogonal Matching Pursuit (OMP) outperforms matching pursuit by updating all activated coefficients simultaneously in generating the orthogonal projection of the selected boxcar functions at every iteration [15]. However OMP requires more computational resources due to extra calculations.

At iteration $j = 0$, the residuals are set to $\mathbf{r}_0 = \mathbf{x}$, the coefficients are set to $\boldsymbol{\alpha}_0 = 0$ and the support vector Λ_0 is empty. At each iteration j, a single element k_j that maximizes

$$k_j = \underset{k}{\operatorname{argmax}} ||\mathbf{r}_{j-1}\varphi_k||_2, \tag{4}$$

and corresponds to the element of the dictionary with the largest width w that fits under $\mathbf{r}_{j-1}$. The last selected function k_j is then added to the support vector

$$\Lambda_j = \Lambda_{j-1} \cup k_j. \tag{5}$$

The coefficients $\boldsymbol{\alpha}_j$ are computed as least square estimates, i.e,

$$\boldsymbol{\alpha}_j = \underset{\alpha}{\operatorname{argmin}} ||\mathbf{x} - \boldsymbol{\alpha}_{\Lambda_j}\boldsymbol{\Phi}_{\Lambda_j}||_2 = \boldsymbol{\Phi}^{\dagger}_{\Lambda_j}\mathbf{x}, \tag{6}$$

where $\boldsymbol{\Phi}^{\dagger}_{\Lambda_j}$ is the Moore-Penrose pseudo-inverse of $\boldsymbol{\Phi}_{\Lambda_j}$ [16]. The approximated signal $\widehat{\mathbf{x}}_j$ is then updated, $\widehat{\mathbf{x}}_j = \widehat{\mathbf{x}}_{j-1} + \alpha_{k_j}\varphi_{k_j}$ where α_{k_j} is the activation coefficient of φ_{k_j}. Finally, the residuals are updated as $\mathbf{r}_j = \mathbf{x} - \widehat{\mathbf{x}}_j$. The output is the support vector Λ_J restricting the overcomplete dictionary to $\boldsymbol{\Phi}_{\Lambda_J}$.

B. Dimension Reduction

After sparse signal approximation, the load signature of an appliance is formed by a combination of boxcar functions which are alike. Hence these boxcar functions appear close to each others at every activation event. They have a strong cross-time dependence. But the OMP selects almost all boxcar functions only once and the community detection relies on the assumption of cross-time dependency of boxcar functions to form the multi-state signatures. To create redundancy and cross-time dependencies between boxcar functions generated by the same appliance, a clustering is implemented on the J selected boxcar functions. The input of the clustering is the parameters $(\boldsymbol{\alpha}, \mathbf{w})$ that describe the shape (power amplitude, operation time) of boxcar functions. The shape of clusters in the 2D space $(\boldsymbol{\alpha}, \mathbf{w})$ is not known *a priori*. A Gaussian mixture model fits 2D Gaussian distributions to form the clusters which means that they can have round or ellipsoidal shape in the 2D space depending on whether their covariance matrix structure [17]. Selected boxcar functions in Λ_J have coordinates $\zeta_j = (\alpha_j, w_j)$ and all the points form $\boldsymbol{\zeta} = \{\zeta_1, \ldots, \zeta_J\}$. A mixture of G clusters is then defined as

$$p(\boldsymbol{\zeta}) = \sum_{g=1}^{G} \phi_g \, \mathcal{N}(\boldsymbol{\zeta} \mid \boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g), \tag{7}$$

where $\boldsymbol{\mu_g}$ is the mean, $\boldsymbol{\Sigma_g}$ is the covariance matrix and ϕ_g is the mixture coefficient of the g^{th} cluster. The probability density function $p(\boldsymbol{\zeta})$ integrates to one with $\phi_g \geq 0$ and $\mathcal{N}(\boldsymbol{\zeta}|\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g) \geq 0$. The mixture coefficients ϕ_g are constrained by

$$\sum_{g=1}^{G} \phi_g = 1 \quad \text{and} \quad 0 \leq \phi_g \leq 1. \tag{8}$$

Boxcar functions in the same cluster are considered as the same boxcar function in the community detection. G, the number of clusters, is determined empirically with backwards and forward fine-tuning as the performance of the community detection depends on G.

C. Community Detection

The GMM generates cross-temporal dependencies between boxcar functions that are parts of the same signature. Graph theory is a field of mathematics that analyzes relationships (connections) between objects in a network (graph). A graph $\Gamma(V, E)$ is a mathematical representation of the pairwise relations between objects, where interactions are expressed with a set of vertices (i.e. boxcar functions) V and a set of edges (cross-temporal dependency) $E = \{e(u, v) \ : \ u, v \in V\}$ [18]. The relations $e(u, v)$ and $e(v, u)$ can be considered different (edges are oriented in a direction) or equivalent (edges are undirected or bidirectional). In this work we consider that $e(u, v) = e(v, u)$.

Community detection consists in forming strongly interconnected subsets from the graph [19]. In the context of this work, a community is a set of boxcar functions with strong cross-temporal dependencies as they appear repeatedly adjacent.

A weight $\omega_{u,v}$ sampled from a Gaussian distribution is assigned to each edge $e(u, v)$

$$\omega_{u,v} = \exp\left(\frac{-(t_u - t_v)^2}{2\sigma^2}\right), \tag{9}$$

σ denotes the scaling parameter and t_u and t_v are respectively the position of the center of boxcar functions u and v. It measures the strength of the cross-temporal dependency between u and v.

Concretely, the community detection algorithm forms disjoints groups $\{c_1, \ldots, c_i, \ldots, c_I\}$ that regroup boxcar functions with strong cross-temporal dependencies. The core of

the community detection process relies on the definition of the objective function that evaluates the aggregation into communities. The objective function defines the notion of community as groups of vertices with better internal connections than external [20]. The most used objective function for community detection is the modularity Q of the partition

$$Q = \sum_{i=1}^{I} \left[\frac{\Sigma_{in}^{c_i}}{2m} - \frac{{\Sigma_{tot}^{c_i}}^2}{4m^2} \right], \tag{10}$$

where m is the sum of all the weights in the network, $\Sigma_{in}^{c_i}$ is the sum of the weights from the internal edges of community c_i and $\Sigma_{tot}^{c_i}$ the sum of the weights from the edges incident to a vertex in community c_i [21]. Q takes values in $[-1, 1]$ where a high value reveals strong interconnections between the vertices in the same community and less dense connections between the vertices of the neighboring communities.

Computing the modularity of the communities is an NP-complete optimization problem and thus computationally expensive. The Louvain method is a heuristic approach: in practice it is iterative and a greedy-type algorithm (inspired by hierarchical ascending clustering) which can solve the problem in $O(n \log n)$ [19], [20]. As in the hierarchical ascending clustering, the first phase starts with each vertex u as its own community and then for each vertex the algorithm calculates the gain of modularity of moving u from its community to the community c_i of a neighbor vertex v

$$\Delta Q_{u,c_i} = \frac{\sum_{v \in c_i} \omega_{u,v}}{2m} - \frac{\Sigma_{tot}^{c_i} \omega_u}{2m^2}, \tag{11}$$

where, ω_u is the sum of the weights of the edges incident to the vertex u, $\sum_{v \in c_i} \omega_{u,v}$ is the sum of the weights of the edges from the vertex u to vertices in community c_i (only one v at this stage). u gets assigned to the community that maximized $\Delta Q_{u,c_i}$ only if it is positive. This stage is operated until the communities get stable, when no individual move can improve the modularity [19]. The second step builds a network of the communities obtained after convergence, using the sum of the weights of the edges between vertices of the communities. Then the first step is reapplied on the network and so on until no improvement on the modularity is observed.

D. Labelling Using Clustering

The disaggregation is unsupervised which means that the appliance consumptions have been individualized but not identified. Hence a post-processing labelling of the signatures in each community is necessary to identify them. Appliances have generally several programs (i.e. cycles of a fridge or programs of a washing machine) which means that an appliance exhibits many signatures. To identify the different signatures after disaggregation, we propose to generate sets of signatures from the single appliance consumption signals that summarize the different load behaviors corresponding to different programs. To do so, a clustering process is implemented on the activation events extracted from a training set of single appliance consumption signals completely disjoints in time from the test set. Activation event are aligned to all start at $t = 2$ and approximated using OMP. As the activation events are time series of different length, the standard Euclidean distance cannot be used to cluster them. it is used to measure distance between time series that may be shifted in time but it is also convenient to measure distance between time series of different length [22]. Clustering using a Dynamic time Warping (DTW) based distance was used successfully to cluster load profiles to align patterns that may be shifted in time [23]. DTW is a one-to-many points distance metric where each point of one time series is compared to many of the second time series [24]. It then form a matrix and the shortest path from the bottom left corner to the top right corner of the matrix is called warping path [25]. The sum of the Euclidean distances on the path is the DTW distance. The main benefits of DTW compared to Euclidean is to compensate for temporal translations in the patterns and it can also calculate distance between activation events of different durations (length of the time series). The pairwise DTW distance matrix of all the activation events is computed for each appliance a hierarchical ascending clustering using the Ward criterion is computed on the DTW distance matrix. The dendrogram of the hierarchical ascending clustering is used to generate the best partition into $3 \leq K \leq 6$ clusters which centroids form the representative signatures.

The identification of activation events from the communities is then done by calculating the DTW distances with the set of representative signatures of each appliances. The label of the closest representative signature is then assigned to the activation event. The labelling is a mapping process between labels and the shape of the signature thus it needs *prior* information about how the signature of the appliances of interest looks like. In other words, it cannot be done in a totally unsupervised manner.

III. Case Study

The algorithm performance is evaluated on the UK-Dale dataset [9] and benchmarked against the results obtained by [26] and [4] using respectively Factorial Hidden Markov Model (FHMM) and a Recurrent Neural Network (RNN) with a Long Short Term Memory (LSTM).

A first implementation with the same dataset at 6 seconds resolution is done to compare our results to the ones in [4], [26]. The consumption from one household over three months with five single consumption appliances: dishwasher, washing machine, kettle, microwave and fridge. This implementation is detailed in Section III-A.

A second implementation is performed on the same data at one minute resolution to evaluate how the performances of the algorithm evolves with the change of resolution.

A. Implementation

The overcomplete dictionary stores all the possible boxcar functions based on the combination of translation l and width w. Hence the size of the dictionary increases exponentially with the length of the signal to approximate. In practice, it is then not realistic to implement the OMP algorithm on large portions of data (i.e. days), thus our consumption signal has to be split into smaller batches of data. We fixed the maximum

length of a batch to 1000 time steps, our dictionary consists of all the boxcar functions with length between one to 1000 and represents 7.4 GB. The way batches are formed influences the output of OMP as it approximates the signals independently over each batch. In our implementation, batches have variable durations as the cut of the consumption time series is made in periods of lower consumption (less than 70 Watts) which are materialized with vertical gray lines in Figure 5. For the entire evaluation periods, the operation time of batches ranges from 30 seconds to 1 hour 20 minutes, with an average of 40 minutes which is in accordance with the duration of the activation events of the selected appliances. Before performing

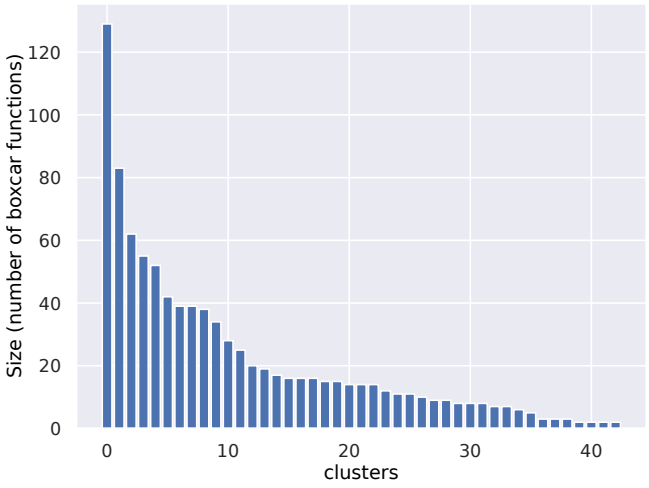


Fig. 3. Histogram of the clusters generated by the GMM.

any operation, a low pass filter sets power consumptions lower than 5 Watts to zero for all the datasets used for test, training or performance evaluation. The OMP has been applied using an error tolerance of 0.055 defined empirically. It eliminates high frequency variations and keeps only the large variations materializing the change of states of appliances as shown in red dashed line in Figure 5. For example the averaged RMSE generated by the approximation in Figure 5 is of 77 Watts, which is larger than the averaged RMSE over the entire period (53 Watts). The output of OMP, presented for the sample in the bottom of Figure 5, is the support vector Λ which consists of 6660 unique boxcar functions selected from the over-complete dictionary. The elements in Λ can have approximately the same amplitude α and width w but a different translation l (Figure 2). The boxcar functions are alike and we can suppose that they are produced by the same process. Hence we assume in this work that two boxcar functions which have similar shapes are likely to be generated by the same appliance.

A first filtering of the type-I appliances (e.g. fridge, microwave, kettle) is operated. The single boxcar functions which distance to a pattern from the labelling library (Figure 4) is under a defined threshold are labeled with the corresponding type-I appliances. The reason to execute the filtering at this stage is that type I appliances in this test have a high frequency of activation, especially the fridge, which means that they are often activated at the same time that larger appliances (e.g. washing machine or dishwasher). They remain unrelated

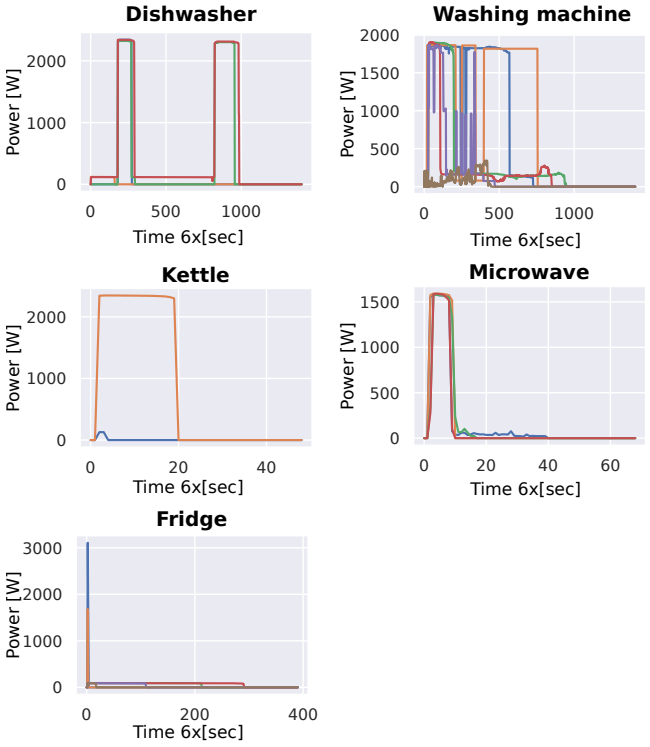


Fig. 4. Representative signature of each appliances

events but the community detection would gather them as a single activation event from the same appliance, forming a corrupted signature.

The GMM generates clusters of boxcar functions which have approximately the same shape $(\boldsymbol{\alpha}, \mathbf{w})$ and assumed to be generated by the same appliance (see Figure 1). The GMM uses only the number of clusters to be formed as parameter, which is set to 43 in the present work. Figure 3 shows the histogram of the cluster sizes.

The community detection requires two parameters, the modularity threshold set to 1 and the scaling parameter for the weights set to 0.95. The training set is sampled from the same household and corresponds to a year of data. The clustering process is run on all the activation events collected from the individual appliances consumption signals after sparse signal approximation. The hierarchical ascending clustering generates the following number of representative signatures; dishwasher: 5, washing machine: 6, microwave: 4, kettle: 2, fridge: 6. Figure 4 gives an overview of the representative signatures per appliance.

B. Performance Evaluation

The performance of our disaggregation algorithm is evaluated using both classification performance metrics and the estimation accuracy commonly used in the NILM literature [4], [27]. With classification algorithms, the performance of the prediction is evaluated in comparison to the ground truth. The recall or True Positive Rate (TPR), the precision, also known as Positive Predictive Value (PPV), and the accuracy (ACC) are defined as

$$\text{TPR} = \frac{\text{TP}}{\text{TP + FN}}, \quad \text{PPV} = \frac{\text{TP}}{\text{TP + FP}}, \quad \text{ACC} = \frac{\text{TP+TN}}{\text{P + N}}$$

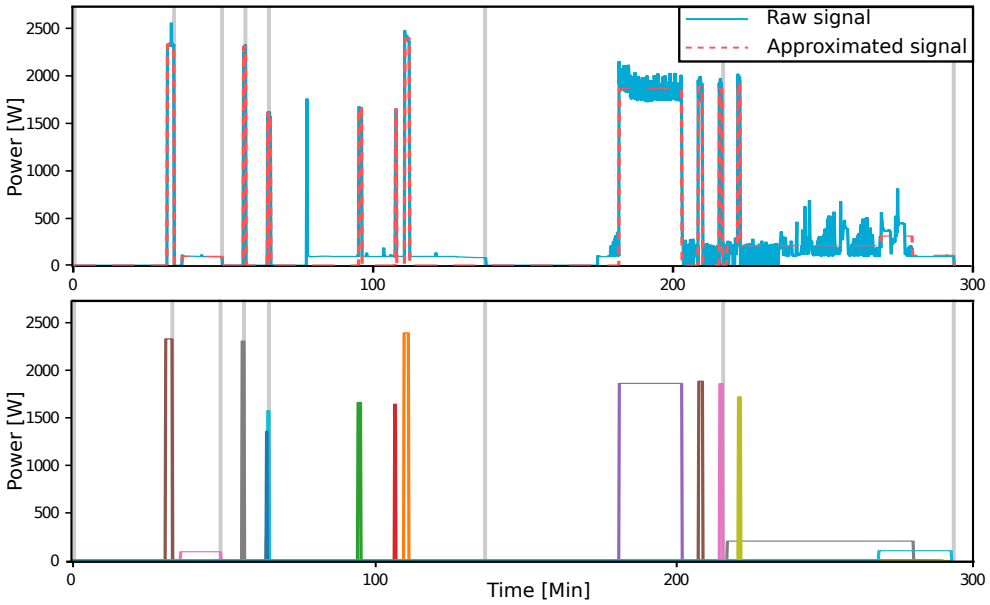


Fig. 5. Top: Raw consumption signal (blue) and the OMP approximated signal (red dashed). Bottom: detailed boxcar functions used to approximate the signal. the vertical gray lines correspond to the limits of the batches.

and are calculated from the output of the confusion matrix,

		Actual: Positive	Actual: Negative
Predicted	Positive	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)
		P	N

where True Positive/Negative (TP/TN) represents the number of times a disaggregated signal from a single appliance being ON/OFF is correctly assigned, False Positive (FP) represents the number of times a disaggregated signal from a single appliance is considered ON (consumption significantly larger than zero) but was actually OFF and False Negative (FN) is a disaggregated signal from a single appliance considered as OFF (consumption close to zero) but was ON. From the Precision and the Recall, the F_1-score

$$F_1 = 2\,\frac{\text{PPV TPR}}{\text{PPV + TPR}} \tag{12}$$

can be calculated. In the NILM literature, the estimation accuracy (EC)

$$\text{EC}^n = 1 - \frac{\sum_{t=1}^{T} |y_n^t - \hat{y}_n^t|}{2\sum_{t=1}^{T} y_n^t}, \tag{13}$$

where t is the time index and n is the appliance index, is used to estimate how accurately the consumption from each appliance n has been estimated [4], [6], [27]–[29]. The overall estimation accuracy (EC) is given by

$$\text{EC} = 1 - \frac{\sum_{t=1}^{T}\sum_{n=1}^{N} |y_n^t - \hat{y}_n^t|}{2\sum_{t=1}^{T}\sum_{n=1}^{N} y_n^t}. \tag{14}$$

and tells how much of the overall signal is kept after summing up the disaggregated signals. All the introduced performance evaluation measures are taking values in $[0, 1]$ and are read the same way: the higher the value, the higher the performance.

IV. Results and Discussions

The results of the implementation at six seconds are first evaluated against the benchmarks. Afterward, the evolution of the performance with the degradation of the data resolution is reported.

A. Benchmarking Against State-of-the-art Methodologies at High Resolution

The results of our methodology based on OMP are benchmarked against the results of FHMM as implemented in [26] and the RNN with LSTM, simply noted LSTM afterward, as implemented in [4] (Table I). In the first column 'Across Appliances', the only performance metric that is calculated across appliances is the estimation accuracy, the other performance metrics are simply the average across all appliances.

Looking first at the performance evaluation across appliances, the OMP approach has a slightly better EC that the FHMM and the LSTM. Regarding the classification performance, our approach outperforms the benchmarks for all of them besides the TPR. For the TPR our algorithm (0.55) performs slightly better than the FHMM (0.53) but is far from the LSTM (0.85). LSTM detects most of the activation events (high TPR) but have also a high false positive rate (lower PPV) compared to OMP.

The reading of the detailed performances of each appliances unveils large variability. It appears that the dishwasher is a difficult appliance to detect accurately, as all methods perform poorly. The OMP algorithm detects it less often than the FHMM and LSTM (lower TPR) but makes less false positives,

TABLE I
DISAGGREGATION PERFORMANCE OF THE OMP-GMM-COMMUNITY DETECTION APPROACH COMPARED TO FHMM [26] AND RNN-LONG SHORT TERM MEMORY (LSTM) [4].

Metrics	Accross Appliances			Dishwasher			Washing machine			Kettle			Microwave			Fridge		
Methods	OMP	FHMM	LSTM	OMP	FHMM	LSTM	OMP	FHMM	LSTM	OMP	FHMM	LSTM	OMP	FHMM	LSTM	OMP	FHMM	LSTM
EC	**0.93**	0.91	0.92	0.26	**0.91**	0.86	0.71	0.84	**0.88**	0.56	0.92	**0.99**	0.67	0.84	**0.98**	0.81	0.94	**0.97**
F_1-score	**0.54**	0.18	0.38	**0.26**	0.05	0.08	**0.61**	0.08	0.03	0.47	0.19	**0.93**	**0.54**	0.01	0.13	**0.83**	0.55	0.74
PPV	**0.62**	0.12	0.36	**0.24**	0.03	0.04	**0.73**	0.04	0.01	0.33	0.14	**0.96**	**0.79**	0.01	0.07	**0.99**	0.40	0.72
TPR	0.55	0.53	**0.85**	0.29	0.49	**0.87**	0.53	0.64	**0.73**	0.82	0.29	**0.91**	0.41	0.34	**0.99**	0.71	**0.86**	0.77
ACC	**0.96**	0.70	0.66	**0.96**	0.33	0.30	**0.96**	0.79	0.23	0.99	0.99	**1.00**	**0.99**	0.91	0.98	**0.88**	0.50	0.81

considering it is ON when it is actually OFF (higher PPV), which leads to a higher F_1-score for the OMP. For the washing machine, the EC of the OMP is higher than for the dishwasher. Against the benchmarks, its TPR is lower but its PPV is higher which generates again a better F_1-score for the OMP (0.61). For the kettle, the LSTM is almost perfect as it displays high performance for all metrics. Nevertheless the OMP performs better than the FHMM for the activation events detection (higher F_1-score, PPV, TPR). For the microwave, the detection of activation events (TPR) of the OMP is higher than the FHMM but lower than the LSTM which detects almost all activation events (TPR=0.99). However in terms of F_1-score the performance of the FHMM is again better. For the fridge the performance of all the methodology is high, the FHMM has the highest TPR (0.86). The OMP and the LSTM are performing similarly (respectively 0.71 and 0.77). The OMP has again a lower rate of false positive which generates a higher F_1-score for OMP compare to the benchmarks.

The performances for the individual appliances differ largely depending on the complexity of the load behavior. As the OMP performs an approximation of the signal into a square signal (Figure 5), appliances which show complex transient load behaviors, will yield poorer results. For the same reason appliances displaying similar average power consumptions and operation times are hard to separate. Comparing supervised against unsupervised learning algorithms is not a fair comparison and it is not expected that our algorithm outperform the benchmarks but just reaches similar performance.

B. Evolution of The Performance With the Degradation of the Resolution

The state-of-the-art approaches require high resolution to identify signatures based on transient states which means that their performance drops really quickly with the degradation of the resolution. As the OMP does not rely on the transient states, it is expected to perform correctly at lower resolution. An implementation of the OMP algorithm was done on the same dataset at one minute resolution to compare the performance and evaluate how it evolves with a degradation of the resolution. one minute resolution, divide already the number of points by ten and brings us closer to what is implemented at large scale.

The performance metrics across appliances displays that the performance at six seconds is better for all performance criteria besides the PPV (Table II). Looking closely at the values, it appears that they are close for the EC and which means that the performance do not reduce much with the change of resolution. The PPV at one minute is higher than at 6 seconds and the TPR is higher at 6 seconds that at 1 minute which results into similar F_1-score performances.

The individual appliances performance metrics, are showing that the change of in performances between six seconds and one minute are again depending on the appliance and its load behavior. For the dishwasher, The performance at one minute resolution is actually better than at six seconds resolution. The EC is doubled which means that the recovery of the signal is better. For the detection the F_1-scores are equivalent, but there are no false positive at one minute (PPV=1), but the TPR is lower than at six seconds (0.15 against 0.29). For the washing machine, the implementation at six seconds outperforms the implementation at one minute for all the metrics. For the kettle, the implementation at one minute recover better the signal that the one at six seconds (larger EC). Regarding the classification metrics, the F_1-score is also better at one minute as it shows a high PPV. However, the TPR is better at six seconds. For the microwave, the performance of the implementation at six seconds is better for all metrics but the TPR where the implementation at one minute is slightly better. For the fridge, the performances are both high.

Between six seconds and one minute, the number of data points is divided by ten, yet the performance of the OMP is not much deteriorated by this change of resolution besides the performance for the washing machine and the microwave. nevertheless, state-of-the-art methodologies relying on signatures in transient states would have their performances seriously altered by this change of resolution.

V. CONCLUSIONS AND FUTURE WORKS

The method presented in this paper is an application oriented and unsupervised approach to NILM. Indeed the state-of-the-art NILM algorithms suffer from a lack of generalization to any household and ability to perform at lower resolutions. As the OMP approach is unsupervised, no training using individual appliances consumption signals is required before running the algorithm. Hence it generates no training bias. It comes at the cost of a lower performance in terms of estimated accuracy compare to state-of-the-art approaches. It is a generic tool which can be used for different purposes at different resolution: high resolution (appr. 1 minute) to provide feedback to customers on their consumption (as presented in this work), medium (5-15 minutes) in a DSM framework to separate the consumption of appliances that could provide some services by moving or decreasing its consumption and provide feedback to aggregators or DSOs. However the benefit

TABLE II
COMPARISON OF THE PERFORMANCES OF THE OMP APPROACH WITH DATA AT SIX SECONDS AND ONE MINUTE RESOLUTION

Metrics	Accross Appliances		Dishwasher		Washing machine		Kettle		Microwave		Fridge	
Resolution	6s	1min	6s	1min	6s	1min	6s	1min	6s	1min	6s	1min
EC	**0.93**	0.91	0.26	**0.57**	**0.71**	0.58	0.56	**0.73**	**0.67**	0.35	**0.81**	**0.81**
F_1-score	**0.54**	0.53	**0.26**	0.25	**0.61**	0.54	0.47	**0.60**	**0.54**	0.44	**0.83**	0.82
PPV	0.62	**0.78**	0.24	**1.00**	**0.73**	0.59	0.33	**0.89**	**0.79**	0.47	**0.99**	0.98
TPR	**0.55**	0.45	**0.29**	0.15	**0.53**	0.51	**0.82**	0.45	0.41	**0.42**	**0.71**	**0.71**
ACC	**0.96**	**0.96**	0.96	**0.98**	**0.96**	0.95	**0.99**	**0.99**	**1.00**	0.99	**0.88**	0.87

of disaggregating down to the smallest appliances has a limited impact compare to the resources needed to execute it. Hence it is pragmatic to focus only on the detection and/or disaggregation of specific appliances with large consumptions (e.g. PV, EV, heat pump if controllable) that provides flexibility to the grid.

Such an approach is then better suited for real-world implementation as it can be used at various resolution. Subsequently it is also less computationally expensive as it handles less data. It will also reduce ethical concerns raised by NILM algorithm that could spy on high resolution energy consumption unveiling human behaviors.

As an extension of this work could be to apply the methodology to the use case of heat pumps, EVs or electric heater for example. These appliances have potential to be provide service to the grid and represent an increasing part of the load.

ACKNOWLEDGMENT

The authors would like to thank the EUDP for funding through the EnergyLab Nordhavn project (EUDP 64015-0055).

REFERENCES

[1] K. C. Armel, A. Gupta, G. Shrimali, and A. Albert, "Is disaggregation the holy grail of energy efficiency? the case of electricity," *Energy Policy*, vol. 52, pp. 213–234, 2013.

[2] G. W. Hart, "Nonintrusive appliance load monitoring," *Proceedings of the IEEE*, vol. 80, no. 12, pp. 1870–1891, 1992.

[3] J. Z. Kolter and T. Jaakkola, "Approximate inference in additive factorial hmms with application to energy disaggregation," in *Artificial Intelligence and Statistics*, 2012, pp. 1472–1482.

[4] J. Kelly and W. Knottenbelt, "Neural nilm: Deep neural networks applied to energy disaggregation," in *Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments*. ACM, 2015, pp. 55–64.

[5] A. Zoha, A. Gluhak, M. A. Imran, and S. Rajasegarar, "Non-intrusive load monitoring approaches for disaggregated energy sensing: A survey," *Sensors*, vol. 12, no. 12, pp. 16 838–16 866, 2012.

[6] N. Henao, K. Agbossou, S. Kelouwani, Y. Dubé, and M. Fournier, "Approach in nonintrusive type i load monitoring using subtractive clustering," *IEEE Transactions on Smart Grid*, vol. 8, no. 2, pp. 812–821, 2017.

[7] B. Zhao, L. Stankovic, and V. Stankovic, "Blind non-intrusive appliance load monitoring using graph-based signal processing," in *Signal and Information Processing (GlobalSIP), 2015 IEEE Global Conference on*. IEEE, 2015, pp. 68–72.

[8] S. Arberet and A. Hutter, "Non-intrusive load curve disaggregation using sparse decomposition with a translation-invariant boxcar dictionary," in *Innovative Smart Grid Technologies Conference Europe (ISGT-Europe), 2014 IEEE PES*. IEEE, 2014, pp. 1–6.

[9] J. Kelly and W. Knottenbelt, "UK-DALE: A dataset recording UK Domestic Appliance-Level Electricity demand and whole-house demand," *ArXiv*, vol. 59, 2014.

[10] A. Faustine, N. H. Mvungi, S. Kaijage, and K. Michael, "A Survey on Non-Intrusive Load Monitoring Methodies and Techniques for Energy Disaggregation Problem," *arXiv*, no. March, 2017.

[11] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4311–4322, 2006.

[12] L. Rebollo-Neira and Z. Xu, "Sparse signal representation by adaptive non-uniform B-spline dictionaries on a compact interval," *Signal Processing*, vol. 90, no. 7, pp. 2308–2313, 2010.

[13] J. Xu, H. He, and H. Man, "Active Dictionary Learning in Sparse Representation Based Classification," *arXiv*, no. 1, pp. 1–7, 2014.

[14] L. Daudet, "Audio sparse decompositions in parallel," *IEEE Signal Processing Magazine*, vol. 27, no. 2, pp. 90–96, 2010.

[15] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Proceedings of 27th Asilomar conference on signals, systems and computers*. IEEE, 1993, pp. 40–44.

[16] R. Gribonval, H. Rauhut, K. Schnass, and P. Vandergheynst, "Atoms of all channels, unite! average case analysis of multi-channel sparse recovery using greedy algorithms," *Journal of Fourier analysis and Applications*, vol. 14, no. 5-6, pp. 655–687, 2008.

[17] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.

[18] S. Fortunato, "Quality functions in community detection," in *Noise and Stochastics in Complex Systems and Finance*, vol. 6601. International Society for Optics and Photonics, 2007, p. 660108.

[19] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2008, no. 10, p. P10008, oct 2008.

[20] X. Que, F. Checconi, F. Petrini, and J. A. Gunnels, "Scalable community detection with the louvain algorithm," in *Parallel and Distributed Processing Symposium (IPDPS), 2015 IEEE International*. IEEE, 2015, pp. 28–37.

[21] M. E. J. Newman, "Analysis of weighted networks," *Phys. Rev. E*, vol. 70, no. 5, p. 056131, Nov. 2004.

[22] S. Aghabozorgi, A. S. Shirkhorshidi, and T. Y. Wah, "Time-series clustering–a decade review," *Information Systems*, vol. 53, pp. 16–38, 2015.

[23] G. Le Ray and P. Pinson, "Online adaptive clustering algorithm for load profiling," *Sustainable Energy, Grids and Networks*, vol. 17, p. 100181, 2019.

[24] T. Warren Liao, "Clustering of time series data - A survey," *Pattern Recognition*, vol. 38, no. 11, pp. 1857–1874, 2005.

[25] J. Paparrizos and L. Gravano, "k-shape: Efficient and accurate clustering of time series," in *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data*. ACM, 2015, pp. 1855–1870.

[26] N. Batra, J. Kelly, O. Parson, H. Dutta, W. Knottenbelt, A. Rogers, A. Singh, and M. Srivastava, "Nilmtk: an open source toolkit for non-intrusive load monitoring," in *Proceedings of the 5th international conference on Future energy systems*. ACM, 2014, pp. 265–276.

[27] S. Makonin and F. Popowich, "Nonintrusive load monitoring (NILM) performance evaluation: A unified approach for accuracy reporting," *Energy Efficiency*, vol. 8, no. 4, pp. 809–814, jul 2015.

[28] J. Z. Kolter and M. J. Johnson, "REDD: A public data set for energy disaggregation research," in *Workshop on Data Mining Applications in Sustainability (SIGKDD), San Diego, CA*, vol. 25. Citeseer, 2011, pp. 59–62.

[29] B. Zhao, L. Stankovic, and V. Stankovic, "On a Training-Less Solution for Non-Intrusive Appliance Load Monitoring Using Graph Signal Processing," *IEEE Access*, vol. 4, pp. 1784–1799, 2016.

[Paper D] Detection and Characterization of Domestic Heat Pumps.

Detection and Characterization of Domestic Heat Pumps

Guillaume Le Ray, Morten Herget Christensen, Pierre Pinson
Technical University of Denmark
Centre for Electric Power and Energy
Kgs. Lyngby, Denmark
email:{gleray, mhchris, ppin}@elektro.dtu.dk

***Abstract*—Smart meters allow utilities to gain access to tremendous amount of metering data, from which they can improve their knowledge on customers. In this paper, we propose a method to detect and characterize domestic heat pumps from household power consumption data. Appliance detection is not a trivial task, owing to the variability in heat pump load dynamics and to the distortion of their power consumption signature by other appliances. Compared to State-of-the-art methodologies that relies on energy disaggregation with supervised learning and high-resolution data (e.g., 10 seconds), our novel approach uses lower resolution data. Also, it is semi-supervised and relies on a Bayesian framework allowing to continuously learn as new data becomes available. To do so, the overall power consumption signals are decomposed and approximated into a dictionary of boxcar functions using sparse signal approximation to isolate heat pumps activity. The learning phase consists then to generate distributions summarizing power consumption, operation time and frequency of heat pumps activation events summarized as boxcar functions after approximation. During the test phase, the distributions are used as prior to calculate the likelihood that a boxcar function is generated by a heat pump. Using standard classification performance measure and an application to data from the EcoGrid EU project, the methodology reaches high performance in heat pump detection.**

***Index Terms*—Bayesian framework, energy analytics, heat pump detection, semi-supervised learning.**

I. Introduction

The deployment of two-way communication smart meters and more generally information and communication technologies form the basis of a smart grid. Its aim is to manage and operate the grid in a more effective manner through an optimized integration of Renewable Energy Sources (RES) as well as decentralized generation. The intermittency of RES generation can be accommodated through the implementation of Demand Side Management (DSM), allowing for a paradigm shift with consumption partially following generation [1].

At the same time, large electrical appliances, like heat pumps, electric vehicles or air coolers, are an increasing part of the domestic customers electricity consumption [2]. In a DSM context, they can also be an asset for aggregators or Distribution System Operators (DSOs) as they can be used as virtual generator by moving their consumption in time and provide flexibility [3] or services [4], [5] to the grid. Furthermore for marketing purposes it would drastically reduce costs to target customers with these appliances to enroll them in a more energy efficient program or propose a flexibility contract for example [6]. Hence it is strategically important to know at a distribution level which households are equipped with those appliances.

In recent years, an approach called Non Intrusive Load Monitoring (NILM), first described in 1992 by Hart [7], became reality thanks to increasing power consumption data collection and computational power. It consists of disaggregating overall metered power consumption into individual appliances' power consumption. State-of-the-art NILM algorithms require high data resolution (down to the second) and training (supervised) which makes it complicated to implement [8]. From a utility perspective, the only data available are metering data, mostly power consumption, at a resolution from 1 minute to 1 hour. Hence NILM as it is, is not possible to implement at large scale. In a previous work, an unsupervised NILM methodology has been presented which have generated performances in the same magnitude as the state-of-the-art NILM algorithm but it proved to perform as well with 1 minute and 6 seconds resolution [9]. The methodology relies on a sparse approximation of the signal into a set of boxcar functions selected from an overcomplete dictionary. After clustering the boxcar functions according to their shape, a community detection algorithm is run to group clusters with strong cross-temporal dependencies that will form the load signature of multilevel appliances.

In this work we propose an application of the framework to a specific use case: the detection of heat pumps. Indeed beyond providing feedback to customers, little work has been done to provide specific application of NILM [10]. From the original methodology in [9], the community detection has been removed and a Bayesian framework has been added to generate distribution describing the behavior of heat pumps.

The remaining of the paper is organized as follows: Section II presents the empirical framework; Section III introduces the different elements of the methodology; Section IV reports the results of the methodology on different resolutions and Section V provides conclusions and ideas for future works.

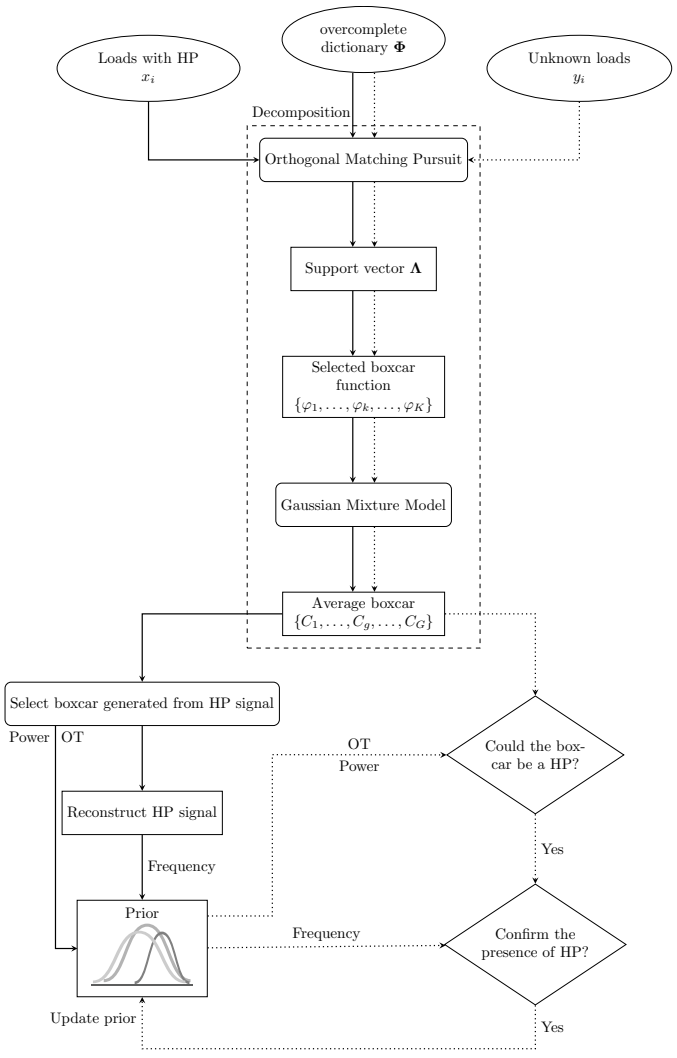


Fig. 1. Algorithm overview.

II. Data Presentation

The data used have been collected during the EcoGrid EU project, which was completed in 2015 [11]. Only a small section of the data collection, corresponding to the coldest period (i.e. where we know heat pumps are active), ranging from the 26th of December 2014 to the 31st of December 2014 is used in this demonstration. The resolution of the data is 5 minutes, which means that heat pumps when active are easily identified from a plot of the power consumption.

A part of the households in the dataset is known to be equipped with a heat pump and labelled as such. The other part is unlabelled, thus we do not know which households are equipped with a heat pump. Among the households equipped with a heat pump, 50 are selected for the training set. The test set consists of 10 households randomly picked among the households with a heat pump as well as 65 unlabelled households for a total of 75 households.

III. Methodology

The signal decomposition (dashed box in Figure 1) is a simplified version of the work implemented in [9]. Indeed, heat pumps, electric heaters, air coolers or electric vehicles (without smart charging station) power consumption are type I appliances (ON/OFF). Hence using power signal sparse approximation and clustering is sufficient to isolate heat pumps power consumption signal.

The methodology is a semi-supervised approach, with a training step, described in Section III-C, where a prior is learned from the training dataset (solid line in Figure 1). The second step, developed in Section III-D, is the heat pump detection of unlabelled households using the prior which is inferred after every iteration (dotted line in Figure 1).

A. Power Signal Sparse Approximation

Signal approximation takes a signal and approximated it as the sum of boxcar function as presented in Figure 2. In practice it takes boxcar functions φ_k subjects to

$$\begin{bmatrix} x^1 \\ x^2 \\ \vdots \\ x^T \end{bmatrix} \approx \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_K \end{bmatrix} \begin{bmatrix} \varphi_1^1 & \varphi_2^1 & \cdots & \varphi_K^1 \\ \varphi_1^2 & \ddots & & \\ \vdots & & \ddots & \vdots \\ \varphi_1^T & & \cdots & \varphi_K^T \end{bmatrix}, \tag{1}$$

with α_k the activation coefficient of boxcar function φ_k, from a dictionary $\mathbf{\Phi} = \{\varphi_k\}_{k=1}^K$ (Figure 1) and generates a sparse support vector $\mathbf{\Lambda}$ with the non-zero elements in $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_k, \ldots, \alpha_K]$ of boxcar function selected to approximate the signal $\mathbf{x}$.

The dictionary, generated beforehand, is said to be over-complete as it consists of all the possible translation-invariant boxcar functions in the signal and is of size K. From the dictionary only J selected boxcar functions are used to approximate the signal with $J \ll K$.

The dictionary summarizes what we know about the appliances load behavior and approximates their signatures both accurately and in a simple manner. Appliances are mostly of two types, I and II, which have respectively two states (ON/OFF) and multi-states (OFF and several activation states)appliances. The simplest way to approximate such signals is using translation-invariant boxcar functions as building blocks

$$\varphi_{l,w}^t = \frac{1}{\sqrt{w}} \, \Pi_{l-w/2,l+w/2}^t, \tag{2}$$

where l is the boxcar function translation and w the width [12]. The boxcar function is made of two Heaviside step function H, $\Pi_{a,b}^t = H(t-a) - H(t-b)$.

Practically, the approach flags changes in power due to an ON/OFF switch or change in state with a fixed or adaptive threshold. The aim is to approximate the power consumption signal so that the internal states fluctuation are removed resulting in simplified the power signal averaged over the activation state (Figure 2).

Iterative processes using a greedy algorithm can be implemented to avoid complexity of a direct sparse approximation,

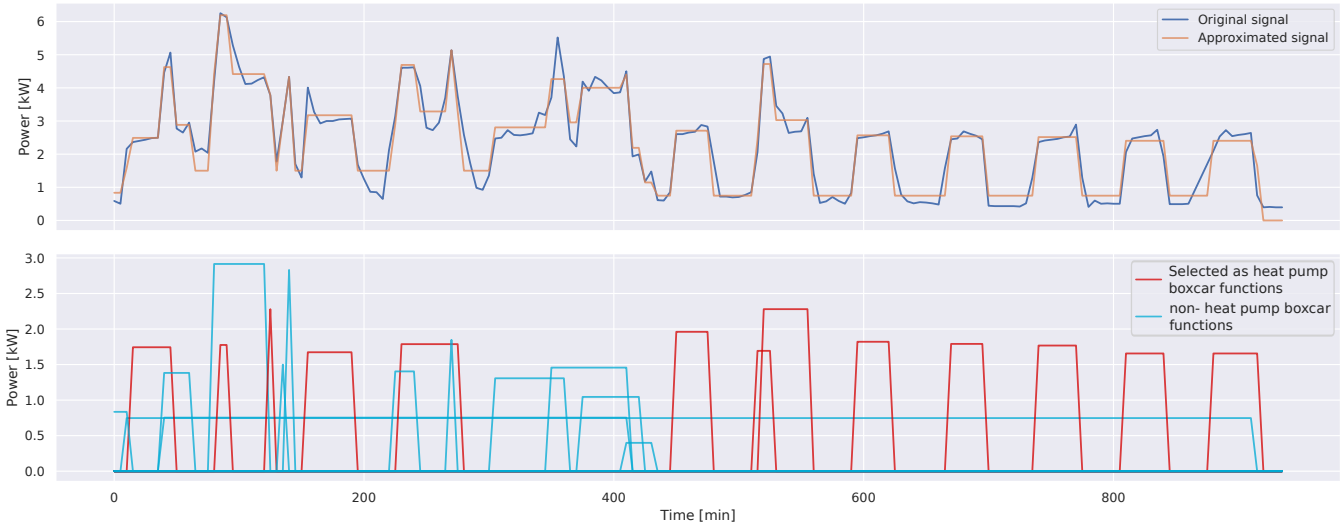


Fig. 2. Raw and approximated power consumption (top) and corresponding boxcar functions of an household equipped with a heat pump.

i.e. using combinatorial optimization [13]. At each iteration, a local optimal solution, also close enough to a global one (i.e. overall approximation), is found. The most used greedy algorithm is matching pursuit; it selects a function in the dictionary according to the function contribution in the sparse signal recovery. Orthogonal matching pursuit is an improved version of matching pursuit that updates all the activated coefficients at every iteration in computing the orthogonal projection of the selected functions. Hence it iteratively estimates the coefficients, and discards them when needed. Orthogonal matching pursuit is computationally more expensive but also returns more accurate results.

The orthogonal matching pursuit algorithm starts with the residuals $R^0 = \mathbf{x}$, the coefficient $\alpha^0 = \mathbf{0}$ and $\Lambda^0 = \emptyset$. At each iteration j the residual $R^j = \mathbf{x} - \widehat{\mathbf{x}}^j$ is updated by subtracting $\widehat{\mathbf{x}}^j$ the approximated signal. The next function selected, $\varphi_{k_{max}^{j+1}}$, in the dictionary $\mathbf{\Phi}$ is the one that maximizes

$$k_{max}^{j+1} = \underset{k}{\operatorname{argmax}} \left| R^j \varphi_k^\top \right|_2 \quad \forall k \notin \Lambda^j, \tag{3}$$

the ℓ_2-norm of the inner product $R^j \varphi_k^\top$. At each iteration, a single element $\varphi_{k_{max}^{j+1}}$ is added to the signal recovery $\widehat{\mathbf{x}}^j$ simultaneously updating the sparse support vector

$$\Lambda^{j+1} = \Lambda^j \cup k_{max}^{j+1} \quad \forall k \notin \Lambda^j, \tag{4}$$

The coefficients $\boldsymbol{\alpha}^j$ are then computed as least square

$$\boldsymbol{\alpha}^j = \underset{\boldsymbol{\alpha}}{\operatorname{argmin}} \left\| \mathbf{x} - \boldsymbol{\alpha}_{\Lambda^j} \mathbf{\Phi}_{\Lambda^j} \right\|_2 = \mathbf{\Phi}_{\Lambda^j}^\dagger \mathbf{x}, \tag{5}$$

where $\mathbf{\Phi}_{\Lambda^j}^\dagger$ is the Moore-Penrose pseudo-inverse of $\mathbf{\Phi}_{\Lambda_j}$ the subset of selected boxcar functions in the dictionary. The last iteration generates the sparse support vector $\Lambda^J = \{k^j\}_{j=1}^J$ restricting the dictionary to $\mathbf{\Phi}_{\Lambda^J}$. The sparse approximated signal can be reconstructed using $\mathbf{x} \approx \boldsymbol{\alpha}_{\Lambda^J} \mathbf{\Phi}_{\Lambda^J}$ (Figure 2).

Concretely the output of the power signal sparse approximation is a set of boxcar functions with different width $\mathbf{w}$ and height $\boldsymbol{\alpha}$ and their location in time l approximating the power consumption signal when summed up as showed in Figure 2.

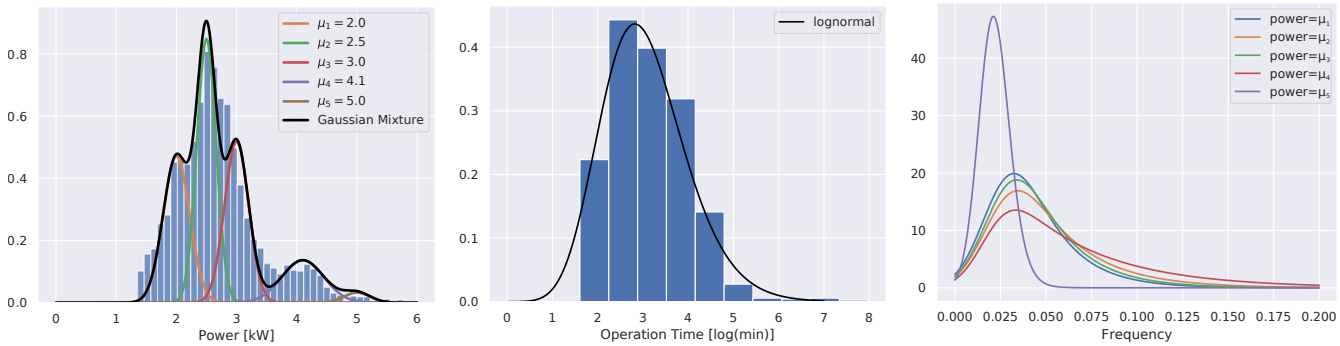
B. Clustering

A direct consequence of the dictionary being overcomplete is that most of the boxcar functions are used only once. However many are displaying similar height α (i.e. power) and width w (i.e. operation time). Indeed they are often generated by the same appliance (e.g. heat pump) and the sparse approximation estimates them differently due to the presence of other appliances or a variation in the load signature. In order to reconstruct the power signal of a heat pump, all the boxcar functions from the heat pump signal have to be flagged. It is then a long and tedious task to analyze the boxcar functions one-by-one. To speed up the process, a clustering is implemented on the selected set of boxcar function $\mathbf{\Phi}_\Lambda$ with power amplitude $\boldsymbol{\alpha}$ and operation time $\mathbf{w}$ (see Figure 1).
The clusters may vary in amplitude and operation time which means that they have various shapes in the 2D space $(\boldsymbol{\alpha}, \mathbf{w})$. Hence Gaussian mixture model that can form round or ellipsoidal clusters is a suitable algorithm [14]. Each boxcar function φ_j of $\mathbf{\Phi}_\Lambda$ has then coordinates $y_j = (\alpha_j, w_j)$ and $\mathbf{y} = \{y_1, \ldots, y_J\}$. The Gaussian mixture model of G clusters is then written

$$p(\mathbf{y}) = \sum_{g=1}^{G} \phi_g \, \mathcal{N}(\mathbf{y} \mid \boldsymbol{\mu}_g, \mathbf{\Sigma}_g) \quad \forall g : \phi_g \geq 0, \tag{6}$$

where $\boldsymbol{\mu_g}$ is the mean, $\mathbf{\Sigma}$ is the covariance matrix and ϕ_g is the mixture coefficient of the g^{th} component. The total probability density function sums up to one with $p(\mathbf{y}) \geq 0$ and $\mathcal{N}(\mathbf{y} \mid \boldsymbol{\mu}_g, \mathbf{\Sigma}_g) \geq 0$. It implies that the mixture coefficients ϕ_g must satisfy

$$\sum_{g=1}^{G} \phi_g = 1 \quad \text{and} \quad 0 \leq \phi_g \leq 1. \tag{7}$$



(a) Power consumption priors

(b) Duration or operation time prior

(c) Frequency of activation prior

Fig. 3. Prior generated during the training phase.

The number of components G is determined empirically, according to how much the dictionary should be reduced. In this specific implementation, $G = 40$ for every household.

C. Training Phase: Prior Definition

The solid line on the left hand side of Figure 1 represent the training phase of the methodology. During the training phase, the subset of households that are equipped with a heat pump is processed through signal sparse approximation and clustering. From the approximate signal and the centroids of the clusters, heat pumps' power consumption is visually assessed, hence clusters with the corresponding power consumption are flagged (Figure 1 & 2). Thereafter the power α (height) and the operation time w (width) of each boxcar function in the flagged clusters are extracted. The heat pump power consumption signal is then reconstructed with the set of flagged boxcar functions. The frequency of activation of each heat pump can be computed by differentiating the heat pump consumption signal and counting how many times the differentiated signal is positive over a rolling window of half a days.

Using the empirical distribution of the power consumption, operation time and the frequency, Probability Density Functions (PDF) (Gaussian, Lognormal and Weibull) are then fitted. The best fit for each feature is presented in Figure 3. The power consumption $\boldsymbol{\alpha}$ is actually a Gaussian mixture where the different Gaussian distributions $p(\boldsymbol{\alpha}) = \sum_{p=1}^{P} \phi_p \ \mathcal{N}(\boldsymbol{\alpha} \mid \boldsymbol{\mu}_p, \boldsymbol{\sigma}_p^2)$ with $\boldsymbol{\mu}_p = \{2.0, 2.6, 3.0, 4.0, 5.0\}$ corresponds to the average consumptions of each size of heat pump (Figure 3(a)). The operation time is actually transformed using the natural logarithm of the distribution to fit PDFs. Both the Lognormal and the Weibull fit well the empirical distribution as it exhibits a long tail on the high end side. The Lognormal PDF is then chosen as it is simpler to handle (Figure 3(b)). The fit of the frequency is a conditional PDF to each power consumption $\boldsymbol{\mu}_p$. The empirical distributions of the frequency fits a Lognormal PDF as they exhibit also a long tail on the high end side (Figure 3(c)).

The set of three PDFs forms the prior definition of heat pumps' load behavior which can be used to identify the presence of heat pump from the overall electricity consumption of unlabelled households.

D. Test Phase: Heat Pump Detection and Characterization

The heat pump detection phase is presented with a dotted line on the right hand side of Figure 1. As for the learning phase, the overall consumption signal of each household is processed through sparse signal approximation and Gaussian mixture model. From the PDFs of the power consumption, the likelihood

$$\mathcal{L}(\mu_p \mid \alpha) = f(\alpha \,;\, \mu_p, \sigma_p^2) = \frac{1}{\sigma_p \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\alpha - \mu_p}{\sigma_p}\right)^2} \tag{8}$$

of each centroid power consumption α being sampled from a heat pump with power consumption μ_p is calculated as well as the likelihood of the power consumption not being sampled from a heat pump that follows a lognormal distribution

$$\mathcal{L}(\mu_0 \mid \alpha) = f(\alpha \,;\, \mu_0, \sigma_0^2) = \frac{1}{\alpha} \frac{1}{\sigma_0 \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\ln \alpha - \mu_0}{\sigma_0}\right)^2}. \tag{9}$$

The maximum likelihood decoding is obtained by calculating

$$P[\alpha = \mu_p] = \frac{\mathcal{L}(\mu_p \mid \alpha)}{\sum_{p=0}^{P} \mathcal{L}(\mu_p \mid \alpha)} \tag{10}$$

for each value of μ_p and returns the probability that each centroid is sampled for a specific distribution.

For each household, a counting with weighting by the sizes of each cluster is done to determine which μ_p has the most often the highest probability. The clusters with the selected μ_p as potential heat pump power consumption are then kept and the likelihood, the maximum likelihood decoding as well as as the counting is then implemented on individual boxcar functions. The households which have μ_0 as the most probable power consumption are then considered not being a domestic heat pump and thus left aside.

The lognormal PDF fitted on the natural logarithm of the operation time is then used to calculate the likelihood of each boxcar function being generated by a heat pump from an operation time perspective. As with the power, the PDF of the natural logarithm of operation times of boxcar function not

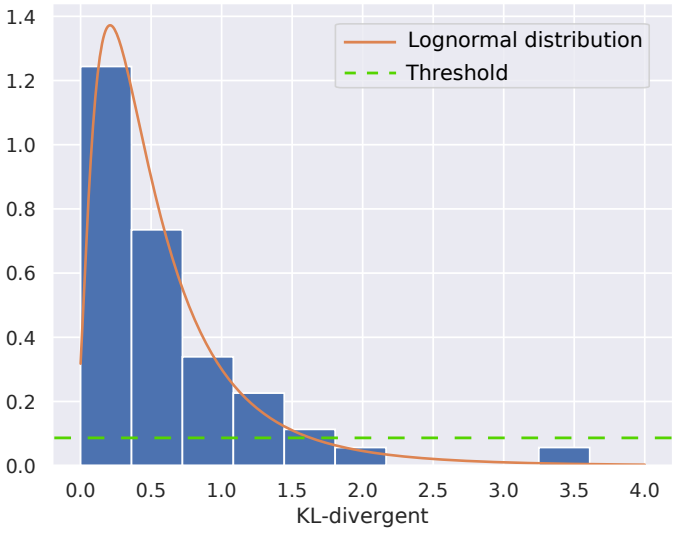


Fig. 4. Distribution of the Kullback-Leibler divergence calculated between the frequency PDF fitted on individual households in the trainning set and the corresponding frequency PDF in Figure 3(c). The orange line is the fitted lognormal PDF and the green dotted line is the threshold under which it is unlikely that a distribution of frequencies is sampled from a heat pump.

generated by heat pumps is used to calculate the complementary likelihood of each boxcar function not being generated by a heat pump. After decoding of the maximum likelihood, the discrimination of the household possibly equipped with heat pump and the one without is operated by counting which of the options has the most often the hightest probability.

The last step consists of recreating the consumption signal of a probable heat pump using the set of boxcar functions that are likely to be produced by a heat pump. As in Section III-C, the consumption signal is differentiated to calculate, using a rolling window of half a day, the frequencies of activation. The frequency PDF is conditional to the power consumption of the heat pump, μ_p, hence the PDF used to calculate the likelihood of each consumption signal being generated by a heat pump is conditional to its evaluated power consumption μ_p. In this case, the decoding of the maximum likelihood cannot be done as it will require high computing resources to decompose all the signal to calculate the likelihood of frequencies generated by other appliances. Hence, the Kullback-Leibler divergence

$$D_{KL}(P \,||\, Q) = -\sum_i P(i) \ln\left(\frac{P(i)}{Q(i)}\right) \qquad (11)$$

between the conditional PDFs of the frequency generated during the training phase, and each individual heat pump frequency PDF of the test set is calculated. The output, Figure 4, is another lognormal PDF, which allows us to define a threshold of 0.1 under which we assume it is unlikely that the frequency is sampled from a heat pump power consumption signal.

E. Performance Metrics

The performance of our detection algorithm is evaluated using standard classification performance measures, the performance of the prediction is evaluated in comparison to (what we suppose to be) the ground truth. From the output of the

TABLE I
CONFUSION MATRIX

		Actual	
		Positive	Negative
Predicted	Positive	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)
		P	N

confusion matrix (Table I), the recall or True Positive Rate (TPR), the precision, also known as Positive Predictive Value (PPV), and the accuracy (ACC)

$$\text{TPR} = \frac{\text{TP}}{\text{TP + FN}} \;,\; \text{PPV} = \frac{\text{TP}}{\text{TP + FP}} \;,\; \text{ACC} = \frac{\text{TP+TN}}{\text{P + N}}$$

are calculated.

IV. Results

The performance measures are calculated at each step of the detection as well as for the overall performance of the algorithm (Table II). The results are also presented with the outlook of what we know a priori is the ground truth (left hand side of the table) and after corrections checking for unlabelled heat pumps detected and thus considered false positive (right hand side of the table). Indeed, as explained in Section II uncertainties exist in the subset of unlabelled households on the presence of heat pumps. Therefore, household loads that seem to be misclassified as equipped with a heat pump have to be visually checked. After visual checking of all the false positives and false negatives, three households which were not labelled as equipped with a heat pump were displaying patterns that were identified as heat pump and were actually correctly detected. No unlabelled heat pumps were placed with the false negatives.

TABLE II
PERFORMANCE OF THE DETECTION AT EACH STEP, WITH OR WITHOUT CORRECTING FOR FALSE POSITIVES BEING ACTUAL HEAT PUMPS.

	A priori			Corrected		
	TPR	PPV	ACC	TPR	PPV	ACC
Power	0.80	0.11	0.15	0.84	0.16	0.19
Operation Time	1.0	0.23	0.56	1.0	0.30	0.60
Frequency	0.88	0.32	0.21	0.91	0.45	0.35
Overall	0.70	0.31	0.75	0.77	0.46	0.80

The first step of the detection assigns a potential heat pump power consumption to each loads, hence the classification performance is not high and only the TPR is actually relevant as we do not want to discard households labelled with a heat pump. Nevertheless two households with heat pumps are discarded at this stage of the detection. They both present consumptions which are on the extremities of the range

$[1.2, 6.0]$ that we consider possible for domestic heat pumps based on the PDFs in Figure 3(a). All the other households discarded at this step have power consumption under 1.5kW which has a low probability for the presence of a heat pump.

From an operation time perspective, the detection assigns well all households with heat pumps as having a heat pump. This process is discriminating with a correct accuracy (0.56), without misclassifying households labelled with a heat pump (TPR=1.0).

The frequency step is simply refining the classification by reducing the false positive rate as it has the highest PPV of the different steps. It has discarded only one labelled heat pump which presents a relatively high frequency for a heat pump.

The Overall performance of the algorithm is correct in term of detection of heat pumps labelled as such. However the false positives are still relatively high which leads to low PPV performance event after correction. A thorough examination of the set of false positives at the end of the process reveals that it consists of households equipped with large electric heaters which in these case have a load behavior similar to heat pumps, in terms of amplitude and frequency. Their operation times are in contrary relatively short (5 minutes or less) but they were not discarded as the operation time prior is not strict on the low side of the distribution (see Figure 3(b)). These information are valuable as they help us to update the prior, to improve the performance of the algorithm through inference.

V. Conclusions and Future Works

This paper presents a Bayesian framework to detect domestic heat pumps and characterize their load behaviour. After generating a first definition of heat pumps load behaviour using PDFs summarizing the range of possible power consumption, operation time and frequency, a first inference is implemented and its performance evaluated. It proved to be performing well despite a small training set resulting in a relatively slack prior definition. The results of this work could be generalized to electric vehicles charging (unless equipped with smart charging facilities), electric heaters and air coolers which are also large ON/OFF appliances. Auto-consumption from PVs or batteries behind the meter could in principle generates problems to the methodology. Nevertheless as long as sections of the signal can be recovered with good accuracy, the general statistics of the HP can be extracted. In other words if the PV or battery capacity is small compare to the overall consumption and/or their activation states are short (some hours) it should not be a major obstacle to the algorithm. Furthermore it is possible to observe combination of these appliances in households (e.g. HP and electric vehicle), as long as the characteristics of their activation states are different, the algorithm should manage to separate them.

The methodology could bring valuable insights for targeted marketing strategies regarding efficiency programs or flexibility contract. Indeed, heat pumps, electric vehicles charging, air coolers, electric heaters are potentially flexible as they are not continuously related to human activities like stoves or light can be. The possibility of knowing where these appliances are geographically placed at a grid distribution level is actually fundamental for efficient DSM and distributed energy generation management. Indeed, the smartness of the grid is not in the meter or in the technology but in the information obtained from processing data and used to take decision. The prior generated could also be used to simulate heat pumps power consumption behavior at a population level and understand how the synchronicity of the peak load can affect the grid.

Future works would consist of updating the prior definition, observe its evolution and study the evolution of the performance of the detection. As smart charging facilities are being promoted by electricity suppliers, it would also be interesting to test the detection of electric vehicles charging to enroll them in the corresponding program.

Acknowledgment

The authors thank EcoGrid EU partners for supporting this work by providing the dataset, as well as the EUDP for funding through the EnergyLab Nordhavn project (EUDP 64015-0055).

References

[1] G. Strbac, "Demand side management: Benefits and challenges," *Energy policy*, vol. 36, no. 12, pp. 4419–4426, 2008.

[2] P. Mancarella, "MES (multi-energy systems): An overview of concepts and evaluation models," *Energy*, vol. 65, pp. 1–17, 2014.

[3] D. Papadaskalopoulos, G. Strbac, P. Mancarella, M. Aunedi, and V. Stanojevic, "Decentralized participation of flexible demand in electricity marketspart ii: Application with electric vehicles and heat pump systems," *IEEE Transactions on Power Systems*, vol. 28, no. 4, pp. 3667–3674, 2013.

[4] T. Masuta and A. Yokoyama, "Supplementary load frequency control by use of a number of both electric vehicles and heat pump water heaters," *IEEE Transactions on smart grid*, vol. 3, no. 3, pp. 1253–1262, 2012.

[5] W. Kempton and J. Tomić, "Vehicle-to-grid power implementation: From stabilizing the grid to supporting large-scale renewable energy," *Journal of power sources*, vol. 144, no. 1, pp. 280–294, 2005.

[6] H. Fei, Y. Kim, S. Sahu, M. Naphade, S. K. Mamidipalli, and J. Hutchinson, "Heat pump detection from coarse grained smart meter data with positive and unlabeled learning," in *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2013, pp. 1330–1338.

[7] G. W. Hart, "Nonintrusive appliance load monitoring," *Proceedings of the IEEE*, vol. 80, no. 12, pp. 1870–1891, 1992.

[8] A. Zoha, A. Gluhak, M. A. Imran, and S. Rajasegarar, "Non-intrusive load monitoring approaches for disaggregated energy sensing: A survey," *Sensors*, vol. 12, no. 12, pp. 16 838–16 866, 2012.

[9] P. Winkler, G. Le Ray, and P. Pinson, "Unsupervised energy disaggregation: From sparse signal approximation to community detection," *IEEE Transaction on Smart Grid*, Submitted.

[10] S. Barker, S. Kalra, D. Irwin, and P. Shenoy, "Nilm redux: The case for emphasizing applications over accuracy," in *NILM-2014 workshop*. Citeseer, 2014.

[11] Y. Ding, S. Pineda, P. Nyeng, J. Østergaard, E. M. Larsen, and Q. Wu, "Real-Time Market Concept Architecture for EcoGrid EU - A Prototype for European Smart Grids," *IEEE Transactions on Smart Grid*, vol. 4, no. 4, pp. 2006–2016, 2013.

[12] S. Arberet and A. Hutter, "Non-intrusive load curve disaggregation using sparse decomposition with a translation-invariant boxcar dictionary," in *Innovative Smart Grid Technologies Conference Europe (ISGT-Europe), 2014 IEEE PES*. IEEE, 2014, pp. 1–6.

[13] L. Daudet, "Audio sparse decompositions in parallel," *IEEE Signal Processing Magazine*, vol. 27, no. 2, pp. 90–96, 2010.

[14] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.

[Paper E] Evaluating price-based demand response in practice with application to the ecogrid eu experiment.

Evaluating Price-Based Demand Response in Practice—With Application to the EcoGrid EU Experiment

Guillaume Le Ray, Emil Mahler Larsen, and Pierre Pinson, *Senior Member, IEEE*

***Abstract*—Increased emphasis is placed today on various types of demand response, motivated by the integration of renewable energy generation and efficiency improvements in electricity markets. Some advocated for the development of price-based approaches, where the conditional dynamic elasticity of final users is exploited in the power system, e.g., for system balancing. However, very few real-world experiments have been carried out and price-based demand response has consistently been found difficult to assess and quantify. It is our aim here to describe an approach to do so, as motivated by the large-scale EcoGrid EU experiment. In this project, 1900 houses were equipped with smart meters and other automation devices in order to adapt consumption to real-time electricity prices every 5 min, while monitoring it with the same resolution. Our approach first relies on the clustering of residential load observations that behave similarly within a given experiment. Then, a clinical testing approach, based on a test and a control group, is adapted to assess whether price-responsive loads were actually responsive or not. Interestingly, in the deployment phase of the project, the results show that houses could be deemed price-responsive on some test days, while results were inconclusive on some others.**

***Index Terms*—Clustering, demand response, electric load modelling, smart grid, time-series analysis.**

I. Introduction

TARGETS to increase the proportion of renewable energy production to 27% by 2030 across all 28 EU member states [1] present significant technical challenges, since existing markets, services and technologies are unlikely to be robust enough to cope with the expansion of variable power generation, also with limited predictability. Among the various options to support large-scale renewables penetration like wind and solar power, Demand Response (DR) has emerged as a popular approach, with its natural advantages and caveats [2]. Applications of DR as a system service are multiple, frequency control [3], congestion management [4], distribution grid services [5] and overall system balancing [6]. Some also discuss long term impacts on the grid planning [6]. The Ecogrid EU project is actually only using DR for overall system balancing. Recent developments in that direction follow the concepts of *(i)* direct control, where a higher-level operator would somehow operate these electric loads, and *(ii)* control by price, where advantage is taken of the elasticity and cross-elasticity of electric power consumers. There obviously are obstacles in rolling out DR, including the non-flexibility of demand [7] and the low participation due to information asymmetries [8]. Control by price has additional difficulties over direct control due to the complexity in predicting response to price variations [9], although forecasting models and control schemes that make effective use of them have been researched [10].

From articles presenting early DR program [11] to those reviewing DR state-state-of-the-art deployment [12]–[15], some highlight problems of communication between the different units during the implementation of DR programs resulting in non-responsiveness of the controllers. However techniques that can readily identify whether a set of electric loads is price-responsive remain lacking, while this may be crucial in practice. This issue is of particular relevance during the deployment phase of demand response equipment and programs. Indeed a logical subsequent step after deploying necessary hardware and software is to control that the different elements communicate as expected, react to the right information, or simply to verify that the overall concept functions.

The present paper introduces a proposal test-control method to assess whether or not electric loads are price-responsive or not. The principle of comparing control and test groups has been extensively used in the medical industry to evaluate the efficiency of a treatment for over 200 years [16], and more recently in the electricity field, industrials working on load research practices have been using this approach to develop Customer Base Line (CBL) and evaluate candidate customers under DR [17]. This method has the advantage of having both the candidate customers and CBL to be exposed to the same weather conditions. Such an approach aims at assessing through hypothesis testing whether loads are responsive or not, which is a basic question to answer before aiming for a quantification and characterization of that response.

Prior to undergoing this test-control analysis, electrical loads are clustered based on similar behavior within a given experiment (i.e., a test day with a given price profile). This allows to identify electric loads that do not respond as expected,

Manuscript received May 11 2016; revised August 4 2016; accepted September 12 2016. Date of publication September 19, 2016; date of current version April 19, 2018. This work was supported by the European Commission through the Project EcoGrid EU under Grant ENER/FP7/268199. Paper no. TSG-00644-2016.

The authors are with the Centre for Electric Power and Energy, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark (e-mail: gleray@elektro.dtu.dk; emlar@elektro.dtu.dk; ppin@elektro.dtu.dk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSG.2016.2610518

while sorting subgroups of responsive households. Note that here the terms 'household' are used with the same meaning as 'load' and do not consider the consumer behavior as the response is automated by a controller; when we write 'responsive loads' or 'responsive households' the reader should read 'responsive controllers which have modified significantly the electricity consumption'. The value of the clustering step of our methodology also lies in the dimension reduction of the problem since, instead of trying to assess whether each and every household in a large-scale demand response experiment (with 1000 households or more) is responsive or not, a fully data-driven clustering step narrows down the analysis by focusing on a low number of subgroups of households with similar dynamic characteristics. This may also be seen as having the side benefit of pinpointing electric loads that could be useful in providing specific grid services such as balancing and congestion management, in view of the characteristics of their response.

Existing literature related to clustering applications (also referred to as segmentation) focuses on profiling, to group the consumers with similar energy consumption patterns [18], [19], or on modelling, to obtain more homogeneous data to improve forecasting accuracy [20], [21]. However, similar approaches using clustering to exclude electric loads that are not responding to the price have not been found in the literature, despite interest from industry in knowing whether a smart controller is responsive or not [17].

The development of this methodology was originally motivated by, and then applied to, the EcoGrid EU demand response experiment, in which 1900 houses and 100 industrial loads receive new electricity prices every five minutes [22]. On the Danish island of Bornholm where the experiment takes place, the majority of the participants have resistive electric heating and heat pumps installed. Their controllability, combined with the heat capacity of the buildings, yields virtual electric power storage. Houses are equipped with smart meters reporting consumption in real-time, as well as a range of automated controllers that make provision of DR convenient by enabling controllability of a wide range of small-scale Distributed Energy Resources (DERs) in a cost-efficient manner. The automated controllers are proprietary and were developed by different companies. In this study, they are therefore considered as black boxes. However, it is known that these rely on state-of-the-art control techniques used for DR, like hysteresis control and economic model predictive control, allowing to schedule consumption optimally considering weather and price forecasts, as well as customer preferences in terms of comfort.

The prices seen by these electric loads originate from the EcoGrid EU market. It was primarily designed to support balancing when larger shares of renewables are present in the power system, yielding additional and more variable balancing needs. In EcoGrid EU, knowledge of the power system state is updated every five minutes. This higher temporal resolution, compared to the hourly time units broadly used in deregulated power systems today, naturally allow to better adapt to dynamic balancing needs. Another key aspect of the market is that it is bidless for demand, hence reducing risk and increasing convenience for small customers who would not otherwise participate. A full introduction to the market behind price generation in the EcoGrid EU experiment is given in [23]. The first phase of the EcoGrid EU project was completed in early 2014, where price-responsive controllers from two different manufacturers were installed in 1200 houses. The price-responsiveness of participants was analysed and eventually validated using the clustering and test-control methods presented here.

The paper is structured as following. Section II presents the empirical framework of the experiment, with particular emphasis on the data and various test-cases to be analyzed. Our methodology is described in Section III, by first introducing the clustering approach for identifying fully non-responsive households and subgroups of responsive electric loads, followed by the test-control method to assess whether these reponsiveness can be seen as genuine price-responsiveness. The results for the roll-out phase of the EcoGrid EU experiment are used as an illustration in Section IV. The paper ends with conclusions and perspectives for future work in Section V.

II. Empirical Framework

The analysis is data-driven and the processes of generation of the data used in this paper are described in the following section. Details on how the controllers operate were kept confidential by the manufacturers.

A. Data Presentation

The datasets consist of electricity consumption for each candidate household with a resolution of 5 minutes. Real-time price series have the same temporal resolution, allowing for the joint analysis of the dynamics of both price and consumption series. Only consumption related to space heating varies as a function of prices based on the controllers deployed for heat pumps and resistive electric heating.

B. Customer Base Line Generation

Throughout the initial phase of the demonstration, households were recruited and then made price-responsive gradually. Some households had their automation disabled deliberately by the central operator, while others had their automation disabled due to reported technical problems. These households were gathered and averaged to form the CBL. Due to the random nature of technical problems, the composition of the test and CBL groups varied from one test-case to the next. Test and CBL groups also varied according to the number of households using one of two control-equipment types and according to different heating types (heat pump or resistive electric heating). As the size of the CBL and participant groups differ throughout the overall experiment, this influences the resulting data analysis and especially the estimated confidence intervals and hypothesis tests performed.

C. Test Cases Presentation

In order to test the controllers, test cases were designed to stress and assess their price-responsiveness with extreme

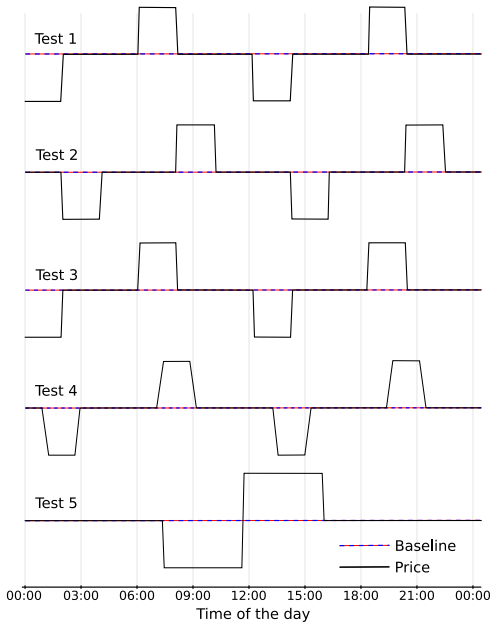


Fig. 1. Price signals broadcast during the different test cases.

TABLE I
PRICE VARIATIONS DURING THE TEST-CASES [24]

Test period	Min / Max (€/MWh)	Baseline (€/MWh)	Test signal
25/10/2013	-53.7 / 148.0	47.2	Test 1
07/11/2013	-53.7 / 148.0	47.2	Test 2
21/11/2013	-61.3 / 140.3	39.4	Test 3
27/11/2013	37.4 / 41.4	39.4	Test 3
06/12/2013	-134.5 / 134.5	0	Test 5
10/12/2013	-134.5 / 134.5	0	Test 5
11/12/2013	-134.5 / 134.5	0	Test 5
12/12/2013	-134.5 / 134.5	0	Test 5
20/01/2014	-134.5 / 134.5	0	Test 5
21/01/2014	-134.5 / 134.5	0	Test 5
22/01/2014	-134.5 / 134.5	0	Test 5
23/01/2014	-134.5 / 134.5	0	Test 5
08/03/2014	-61.3 / 140.3	39.4	Test 3
11/03/2014	37.4 / 41.4	39.4	Test 3
09/04/2014	-61.3 / 140.3	39.4	Test 4
13/04/2014	-61.3 / 140.3	39.4	Test 4

price variations. As the energy consumption should be a function of the price, a significant change in the electricity consumption is expected when such extreme price variations occur [24]. More precisely, a variation in price is to be seen as an incentive for modification of electricity consumption: upwards when the price goes down, and downwards when the price goes up. Table I and Fig. 1 gives a summary of the price variations applied during each test case. All the test cases have the same duration of 24 hours. The negative prices are strong incentives to consume electricity send to controllers as they are making money by doing so. In contrary high positive prices are design to decrease immediately the consumption.

III. METHODOLOGY FOR ASSESSING PRICE-RESPONSIVE BEHAVIOR

A. A Non-Supervised Classification for Dimension Reduction and to Identify Sub-Groups

A natural way to reduce dimension and to extract information from a large and noisy dataset is to group it into more homogeneous clusters. Each of these clusters exhibit more homogeneous characteristics of its individuals than the overall dataset does [25]. Consequently, clustering can be used to exclude groups which could be considered as outliers [17]. In addition, it emphasizes characteristic patterns in consumption, which may implicitly include the consumption variations due to changes in price.

As it is most likely the case for any real-world experiment, it was observed within the EcoGrid EU demonstration that uncertainty existed in the actual price-responsiveness of heat appliance controllers during DR experiments. This may be due to customers being able to interact with controllers - turning them off or changing comfort settings. Other issues, e.g., bad choice of location for temperature sensors used by controllers, can also result in households not being responsive (or just a little) at certain times. A number of other punctual technical problems can affect the responsiveness of these heat appliance controllers. Therefore, employing clustering for identifying and isolating these outliers can focus our analysis on the DR of well-functioning installations. On a more practical level it generates a list of targets to troubleshoot for the technicians. The time period (24 hours), the replications (16 tests) in the experiment, the amplitude of the incentives and consequently the amplitude of the responses leave little doubt that the largest part of the consumption variation is from variations in the controllers and not from changes in consumers preferences.

Clustering approaches have been extensively described in the literature. The interested reader is for instance referred to [26] for an overview of clustering algorithms and [25] for applications in electric load analysis. Out of this wealth of algorithms, the most suitable one to be used depends upon the data setup and our a priori knowledge of the expected output (e.g., the number of clusters to be obtained) [27]. Hierarchical clustering permits to effectively choose the number of clusters, a posteriori, according to the so-called dendrogram, which is a clustering tree where the level of details (and the number of clusters) is increasing as its branches are further divided. An example dendrogram used to cluster 35 households in one of the EcoGrid EU experiment is shown in Fig. 2. Hierarchical clustering is a non-supervised classification method where individuals are grouped according to their relative distances in a similarity space determined by a set of variables [28]. Hierarchical clustering can be performed in an agglomerative or divisive manner. The former approach starts with each household as a cluster and ends up with one cluster (bottom-up approach), while the latter one sees the whole set of households as one cluster to start with and eventually ending with each household as a cluster (top-down approach). Their outputs are similar, but Hierarchical Agglomerative Clustering (HAC) is known to be faster to compute.

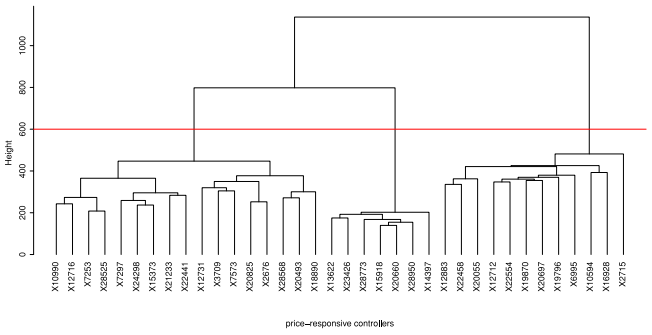


Fig. 2. Dendrogram for the clustering of 35 households in one of the EcoGrid EU experiment. The red line indicates the cut to be made to obtain 3 clusters.

Here our households may naturally have different average consumption levels depending on the house types, number of inhabitants and human behavior. Consequently, some form of alignment is needed to make them all comparable in order to measure some kind of distances between them. However the variance σ_i^2 of the time-series from each household i should not be affected, as the variability in amplitude of the adjustment in consumption during DR events is of high importance. For each and every test case in the experiment, electric power consumption series were centered on their average consumption, by subtracting the mean consumption on a per-household basis, over the entire test case. Considering the original power consumption series $\mathbf{x}'_i = \{x'_{i,1}, \ldots, x'_{i,t}, \ldots, x'_{i,T}\}$ for household i ($i = 1, \ldots, I$), with t the time index, this reads

$$x_{i,t} = x'_{i,t} - \frac{1}{T}\sum_{t=1}^{T} x'_{i,t}, \quad i = 1, \ldots, I, \ t = 1, \ldots, T \tag{1}$$

$\mathbf{x}_i = \{x_{i,1}, \ldots, x_{i,t}, \ldots, x_{i,T}\}$ is the resulting centered power consumption series for household i, with I the number of households at time t. The series $\mathbf{x}_i$ has the same dynamics and amplitude as $\mathbf{x}'_i$, though centered on 0, thus allowing to better compare the higher-order dynamics of the various households [29], [30].

In our experimental framework, the hypothesis is that if a household is active and receives a price variation during a DR event, the consumption should be affected. The variation in consumption is not expected the same for all houses because of their prior status (e.g., temperature, controller setup), nevertheless it should be possible to cluster similar patterns of consumptions' variation as they are expected to react. In that context, the chosen distance for the clustering approach ought to account for covariances between the consumption series. In our experimental framework, the space we have to explore has the dimension of the number of measurements performed in time. With a temporal resolution of 5 minutes and a test case duration typically of 24 hours, this translates to fairly large dimensions. However, it is expected that power consumption observations are serially correlated, i.e., not independent from one time instant to others. In other words, the effective dimension of the space within which the consumption patterns are observed is clearly less than the number of time steps T. The chosen distance for the clustering approach ought to reflect that aspect. The Mahalanobis distance [31], which fulfills this requirement, is then adopted. For two series $\mathbf{x}_i$ and $\mathbf{x}_j$, it is defined as

$$d(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i - \mathbf{x}_j)^\top S_{ij}^{-1} (\mathbf{x}_i - \mathbf{x}_j) \tag{2}$$

where S_{ij} is the covariance matrix between the two time-series. However, the covariance matrix S_{ij} may happen to be singular when the number of households (I) is smaller or about the same as the number of data points (T) in the time-series [32]. This problem arises often while working with time-series as the number of data points can be extensive compared to the number of households. To prevent such issues with singularity, S_{ij} is replaced in (2) by a shrunk covariance matrix S^*_{ij}. Shrinkage is an efficient way to obtain a non-singular closest estimate of the original covariance matrix S_{ij}. It is calculated as

$$S^*_{ij} = \lambda T_{ij} + (1 - \lambda) S_{ij} \tag{3}$$

where T_{ij}, commonly referred to as the target, is a diagonal matrix formed with the element on the main diagonal of the original covariance matrix S_{ij} [32]. λ is the shrinkage coefficient. S^*_{ij} is a trade-off between a highly-structured matrix (T_{ij}) and a non-organized one (S_{ij}), while λ allows controlling the balance between the two [33]. We set

$$\lambda = \begin{cases} \lambda^*, & \text{if } \lambda^* \leqslant 1 \\ 1, & \text{otherwise} \end{cases} \tag{4}$$

with

$$\lambda^* = \frac{\sum_{m \neq n} \widehat{\sigma}(c_{mn})}{\sum_{m \neq n} c_{mn}^2} \tag{5}$$

where c_{mn} are the components of the (sample) covariance matrix S_{ij} and $\widehat{\sigma}(c_{mn})$ their estimated variance [34].

HAC is a fairly general framework, given a metric suitable for the data at hand (e.g., the Mahalanobis one used here). Similarly, one may flexibly choose the way to regroup individuals within clusters. The most common one is the Ward's method, also known as minimum-variance method. It aims to minimize the increase of the within-cluster sum of squared distances, E, at each iteration of the agglomerative process [26],

$$E = \sum_{k=1}^{K} \sum_{\mathbf{x}_i \in C_k} d(\mathbf{x}_i, \mathbf{g}_k)^2 \tag{6}$$

where K is the number of clusters, $\mathbf{x}_i \in C_k$ the households in cluster C_k and $\mathbf{g}_k$ the center of gravity of cluster C_k, defined as

$$\mathbf{g}_k = \frac{1}{n_k} \sum_{\mathbf{x}_i \in C_k} \mathbf{x}_i \tag{7}$$

where n_k is the number of households in C_k.

Following [35], the total variance of a set of households, after clustering, can be expressed as the sum of the within-cluster variance plus the between-cluster's center variance. Consequently, since the Ward's method aims at minimizing the increase of within-cluster variance at each iteration, it also maximizes the variance between cluster centers. The resulting clusters can then be seen as the most homogeneous possible subgroups from the set of households. The HAC algorithm

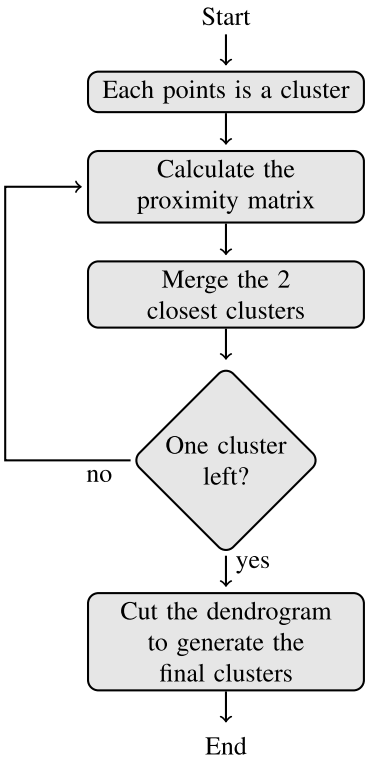


Fig. 3. The Hierarchical Agglomerative Clustering (HAC) algorithm.

is illustrated in Fig. 3, starting with each household being its own cluster. It then iterates until all households are merged into a single cluster. The result of the HAC is conveniently represented in a dendrogram such as that in Fig. 2. The dendrogram is a basis to decide on how many clusters should be chosen. The decision of where to cut the tree depends on the structure of the tree and the goal of clustering. If the goal is to have a clear and precise information on each cluster, a higher number of cluster will be favored. Conversely, if the goal is to isolate outliers, a lower number of clusters will be favored. It is then difficult to implement an automated routine to select the number of clusters. The decision is based on our expertise in interpreting the structure of the tree and thus subjective [36].

After computing the HAC, the information contained in the different clusters should be summarized. When it comes to time-series, the clusters' averaged time-series is a suitable way to represent the specificities of each cluster. One of our test cases, with 5 averaged cluster time-series identified from the dendrogram, is shown in Fig. 4 together with the averaged time-series of the CBL, as well as the corresponding price signal. Such representation allows clusters with reactive adjustment to the price variations (if compared to CBL) to be sorted apart from those that do not adjust during the DR event or show erratic patterns (e.g., due to technical problems). These are consequently not considered in the subsequent analysis. In the example of Fig. 4, the households from the clusters 2 and 5 are to be excluded from the test group, since cluster 2 follows the CBL while cluster 5 has no daily variations which most likely means that the households are empty. As these

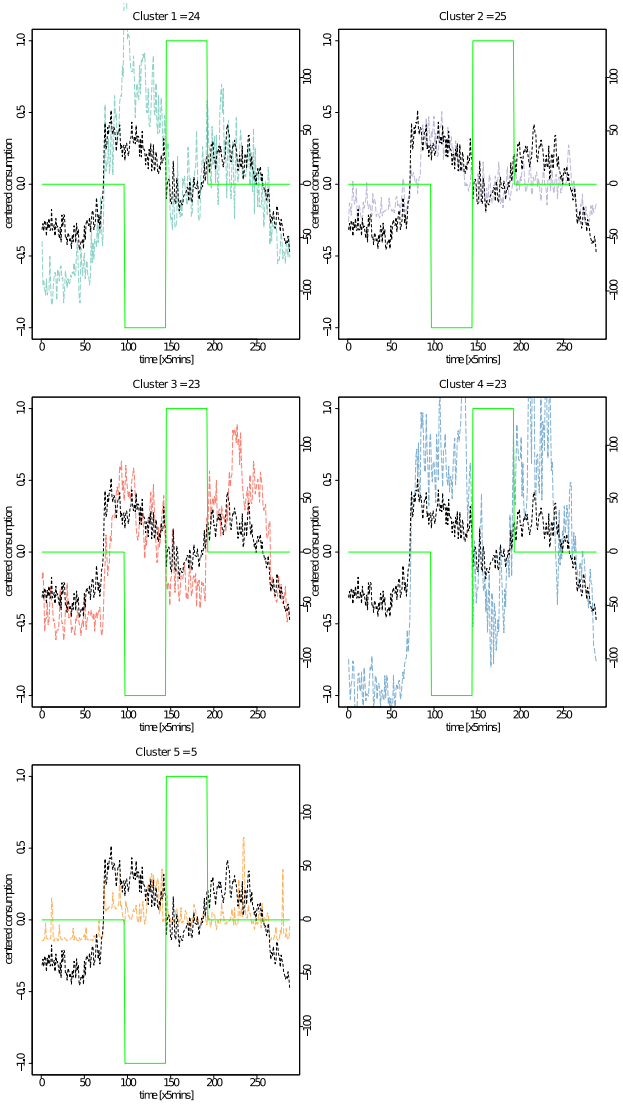


Fig. 4. The averaged time-series is calculated for each cluster in the test group and displayed as a colored dashed line. The black dashed line is the averaged time-series of the CBL and the green line is the price.

outliers are removed, the data quality of the treatment group is improved and eases the subsequent qualitative and quantitative analysis. When mentioning test groups in the remainder of the paper, we refer to those subgroups selected after the clustering was performed.

B. A Clinical Trial Test Approach

Clinical trials were historically developed in the pharmaceutical industry. Owing to the variety of potential responses of biological organisms as individuals, it became common to perform tests on populations instead, thereby smoothing the potential negative effect of individual features on an overall assessment. In the present case, we can employ a similar clinical trial test approach since our data comes from a reference (CBL) and a test group, while our interest lies in the difference in consumption between these two

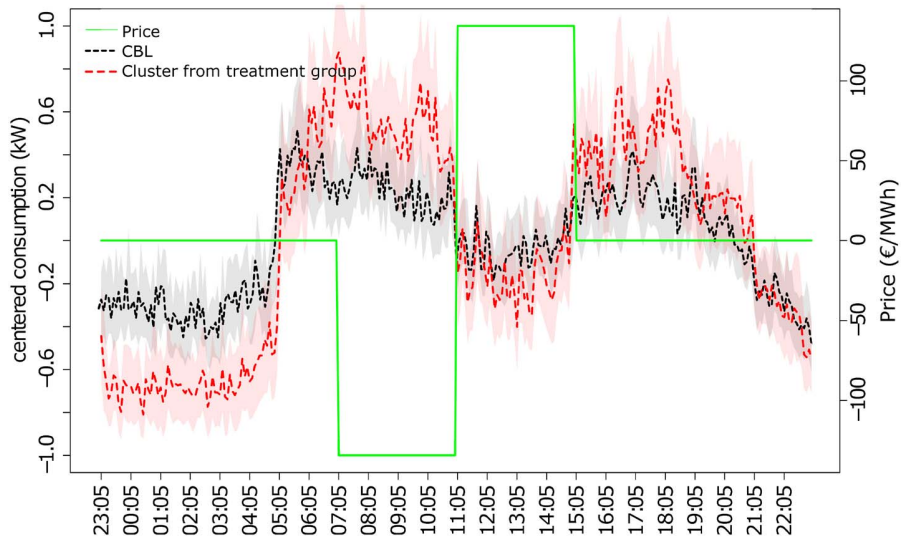


Fig. 5. The average time-series from the CBL and a cluster from the test group with their respective 95% confidence intervals generated from the bootstrap. The green line is the price.

TABLE II
NUMBER OF HOUSEHOLDS IN VARIOUS TEST CASES: NUMBER IN THE TEST GROUP (NUMBER DEEMED PRICE-RESPONSIVE AFTER CLUSTERING) / NUMBER IN THE CBL

Date	Manufacturer 1 Electric Heating	Manufacturer 1 Heat Pump	Manufacturer 2 Electric Heating
25/10/2013	68 (48) / 288	36 (24) / 197	88 (55) / 82
07/11/2013	65 (58) / 289	36 (33) / 197	88 (61) / 92
21/11/2013	67 (61) / 292	36 (36) / 200	87 (75) / 94
27/11/2013	66 (55) / 292	36 (34) / 201	89 (78) / 91
06/12/2013	—	—	115 (74) / 99
10/12/2013	—	—	103 (84) / 91
11/12/2013	—	—	100 (70) / 86
12/12/2013	—	—	106 (69) / 89
20/01/2014	—	—	230 (194) / 105
21/01/2014	—	—	223 (121) / 100
22/01/2014	—	—	230 (110) / 104
23/01/2014	—	—	229 (107) / 105
08/03/2014	30 (30) / 324	20 (17) / 236	237 (125) / 75
11/03/2014	38 (38) / 317	24 (18) / 229	232 (171) / 76
09/04/2014	101 (99) / 249	58 (43) / 188	249 (197) / 109
13/04/2014	38 (38) / 311	24 (22) / 222	269 (188) / 114

groups [37]. Moreover, the inherent uncertainty on the responsiveness (response to a treatment) resulting from the absence of homogeneity in the test group as well as in the reference group (e.g., behavior of the user, thermal comfort setup, energetic profile of the buildings), supports the idea that a clinical trial approach is relevant here. The question we aim to answer can be formulated as *Do price variations induce significant changes in power consumption patterns?*

The results of the clustering on households recruited in the DR program from the different groups exposed in Table II, can be analyzed in two different ways. On the one hand, one can visually assess whether the average response of the selected clusters from each group is responsive by direct comparison with the CBL during the DR event. The purpose of visual inspection is to show that DR works for some clusters, and not for others, to a non-scientific audience who is not familiar with more objective statistical methods. However such an analysis cannot conclude on the significance of the response observed, while relying on expert knowledge at evaluating variation in patterns. On the other hand, this can be tested more rigorously in an hypothesis testing framework (see Section III-C). The purpose of hypothesis testing is to satisfy a scientific audience who requires a degree of objectivity when the results (e.g., lower unit electricity costs for the consumer) are presented.

Fig. 5 is an example of a case used for visual evaluation of a given test case. By observing the dynamics of the mean consumption series of the test group compared to the CBL, one may conclude on the responsiveness of that test group based on confidence intervals. Experience with such consumption data shows that it does not follow a Gaussian distribution. Hence, a nonparametric approach (Non-Studentized pivotal method) is used to obtain confidence intervals. More specifically, we employ a common resampling technique known as bootstrap [38] to generate them. From all the 5000 resampled average time-series, 95% confidence intervals defined by 2.5% and 97.5% quantiles of the distributions of bootstrap samples are obtained.

From visual inspection of Fig. 5, one may infer that the behavior of the test group is different from that of the CBL when the confidence intervals are not overlapping (for example, from 7:05 to 8:05). In other situations, when the confidence intervals overlap or when the average time-series lies within the confidence intervals of another one, one cannot conclude. A more detailed analysis of Fig. 5 shows that the test group exhibits higher consumption during the low price period and lower consumption in the high price period with respect to the CBL. The lower consumption in the period 23:05 to 5:05 is induced by the smart controllers in the experiment shifting load to the lower price period that starts at 07:05.

Smart controllers receive a day-ahead price forecast (as well as an hour-ahead price forecast every half hour) allowing them to schedule consumption in an optimal manner. The value of the relative real-time price with respect to recent and limited forecasted prices therefore contributes to visual estimation of whether a test group is price response or not. For example, in Fig. 5, the relative price is high in the period 23:05 to 5:05, so it is expected that a price-responsive cluster would have lower consumption than the CBL during this period.

C. Hypothesis Testing to Assess Price-Responsiveness

A standard way to assess results in a clinical trial test is to employ hypothesis testing. The hypothesis obviously depends on the question, e.g., is the test group's consumption different than that of the CBL during a DR event? In this question it can even be specified lower or higher instead of 'different'. Based on this hypothesis, a test is formulated and applied to the data. The method used to analyze the hypothesis test should be chosen according to the assumptions on the sample values' distribution. The aim of the EcoGrid EU DR program is to displace electric power consumption from periods with higher prices to periods with lower prices. Whether this goal is achieved or not can then be determined based on the economic value to the households, i.e., in relation to cost per unit of electricity consumed. Consequently here, hypothesis testing may allow us to objectively state whether a test group is price responsive or not. We use a framework similar to that of conventional clinical trial tests, with a type I error threshold α of 0.05.

A hypothesis test can be formulated, since the average cost of a kWh of electricity consumed during a test-case by the test group should be lower than the cost for the CBL during the same period. The average unit cost $\bar{C}_i$, for a test case with T time steps, is calculated as

$$\bar{C}_i = \frac{\sum_{t=1}^{T} C_{ti} P_t}{\sum_{t=1}^{T} C_{ti}} \tag{8}$$

where C_{ti} is the consumption of electricity from household i at time t and P_t the price at time t. A simple observation of the average unit cost distributions tells us that the variances of the 2 samples are different and that they may have heavy tails. Therefore, standard parametric tests are excluded. The Mann-Whitney test (also known as the Wilcoxon rank sum test) is a convenient solution, since the number of households in each of the subgroups is large. A one-sided Mann-Whitney test is performed on the ranks. The hypotheses are the following,

$$\begin{aligned} H_0 &: \mu_{test} \geq \mu_{CBL} \\ H_1 &: \mu_{test} < \mu_{CBL} \end{aligned} \tag{9}$$

where μ corresponds to the sum of ranks, H_1 is the one-sided tailed alternative hypothesis and H_0 is the null hypothesis. The null hypothesis means that the activity of the price-responsive controllers is not significantly modifying the average unit cost, so that it could be considered lower than the control group average unit cost. If the H_0 is rejected, the alternative hypothesis is confirmed statistically.

The one-sided tailed alternative hypothesis is more restrictive than the two-sided tailed standard hypothesis test, as it specifies that the samples should not only be different, but that the test sample's mean should be lower than the control sample's mean. The Mann-Whitney test defines the statistic U with the following formula

$$U = \min\left(n_1 n_2 + \frac{n_1(n_1-1)}{2} - R_1,\ n_1 n_2 + \frac{n_2(n_2-1)}{2} - R_2\right) \tag{10}$$

where n_1, n_2 are the size of the 2 samples and R_1, R_2 are the sum of the ranks for these two samples respectively. U follows a normal distribution and we can calculate the p-value as

$$P(U \geq U_{1-\alpha} | \mu_{test} \geq \mu_{CBL}) \tag{11}$$

The P-value can be seen as the probability of obtaining a test statistic result at least as extreme or as close to the one that was actually observed, assuming that the null hypothesis is true. The test is considered significant when the p-value is lower than the type I error threshold α, which is the chance that we mistakenly reject the null hypothesis (that the samples' means are different). All the details related to statistical aspects can be found in [39].

IV. Results

A. Clustering Results

The cluster analysis aims to identify the price-responsive participants in the test group (possibly in the form of various subgroups) and to separate them from obviously non-responsive households. Emphasis is placed here on how many households are kept in the analysis from the original subgroups after computation of the HAC, as it influences the subsequent analysis.

The clusters are visually selected by comparing the averaged consumption time-series of each cluster to the averaged CBL consumption time-series during event with price variations (Fig. 4). If the averaged consumption time-series of a cluster seems to be flat (no activity, e.g., as for cluster 5 in Fig. 4), following the same pattern as the averaged CBL time-series (e.g., as for cluster 2 in Fig. 4) or showing unexpected pattern, it will be excluded from the dataset used in the evaluation of the price responsiveness. When all the clusters are non price-responsive, only the aberrant ones will be removed.

Table II gives a summary of the clustering selection; the range of the selection from the original data goes from 52% to 100%. In other words, a maximum of a half (48%) of the smart controllers were in the test group, but did not visually appear to be price-responsive. The graphical representation of clusters is also useful for identifying different types of price-responsive behavior. For example, in Fig. 4, cluster 1 gathers the controllers which have been stimulated by the first price variation, while cluster 3 gathers the ones which have been stimulated by the second price variation and cluster 4 gathers the ones which have reacted to both stimuli. It also illustrate the differences of behavior between the manufacturers as the

TABLE III
THE COLOR OF THE CELL RETURN THE RESULTS OF THE VISUAL EVALUATION; GRAY IS RESPONSIVE, LIGHT GRAY IS NON-RESPONSIVE. THE FIGURE IS THE P-VALUE FROM THE MANN-WHITNEY TEST. SIGNIFICANT TEST AT $\alpha = 5\%$ ARE SHOWN IN BOLD AND ITALIC

Date	Manufacturer 1 Electric Heating	Manufacturer 1 Heat Pump	Manufacturer 2 Electric Heating
25/10/2013	0.98	0.44	0.20
07/11/2013	0.39	***0.0013***	0.88
21/11/2013	0.95	0.20	0.09
27/11/2013	0.34	0.34	0.81
06/12/2013	—	—	0.13
10/12/2013	—	—	***0.00022***
11/12/2013	—	—	***0.0015***
12/12/2013	—	—	0.12
20/01/2014	—	—	0.99
21/01/2014	—	—	0.22
22/01/2014	—	—	0.21
23/01/2014	—	—	0.63
08/03/2014	0.54	0.59	***0.0068***
11/03/2014	0.86	0.80	0.28
09/04/2014	***0.0034***	***0.0096***	0.21
13/04/2014	0.96	0.12	***0.0014***

price-response strategies and constraints are implemented differently. Such information was not known beforehand, and brought more insight on how a set of controllers behave at the occasion of large price variations. However, this paper does not focus on this aspect, but it worth mentioning it as it is a good way to illustrate it.

B. Results of the Clinical Trial Test Approach

The chosen clusters are used to generate graphical overviews of each group during the different test-cases (Fig. 5). Table III summarizes the visual evaluation of the graphs displaying the averaged time-series associated with the 95% confidence intervals of the treatment and CBL groups for each manufacturer, equipment type and for different test-days. Results here should be interpreted as, for each experiment, whether it was possible to find one or more clusters that could be seen as price responsive, or not.

In the roll-out phase of the EcoGrid EU demonstration, controllers and other infrastructures were continually developed and improved, which explains the improvement of the DR as the heating period went on.

C. Results of the Hypothesis Testing

The main goal of the EcoGrid EU project is to push electricity consumption during periods of high prices to periods of low electricity prices. This means an economic evaluation can be done, by comparing the average unit cost of selected test groups to the CBL. In this case, hypothesis testing could be applied to each and every identified clusters, or only to those where visual assessment indicated that price response may be present. As for the visual assessment before, the test is applied to all clusters that were not discarded through the clustering analysis, for instance since deemed as outliers or clearly non-responsive.

Table III shows the Mann-Whitney test's results for the different test periods. A standard type I error threshold is chosen ($\alpha = 5\%$). The significant tests are shown in bold and italic. The comparison between the results from visual evaluation and the hypothesis testing in Table III exposes the difference between price-responsiveness which can be visually noticed but not statistically validated using the measure of unit cost, and the price-responsiveness that does have a significant economical impact on the average unit cost. The results show that towards the end of the roll-out of the EcoGrid EU project, it was possible to visually and rigorously find differences between CBL and test groups (manufacturer 1 electric heating, manufacturer 1 heat pump and manufacturer 2 electric heating), indicating a price-responsive behavior overall. The improvement of the responsiveness over time is a direct consequence of the tunning operated on the controllers during the experiment. Further steps in such an evaluation work would consist in quantifying and characterizing this price-responsiveness, while also assessing if this corresponds to the maximum response that could be provided by these groups of households.

V. CONCLUSION

The method presented in this paper shows how a systematic evaluation of DR can be done even with datasets that contain outliers, noise, and other undesirable effects. The clustering can easily be generalized to other time series classification, although scalability to data with more observations remains an area for inquiry. We have successfully applied it to 2 weeks data with a resolution of 5 minutes, but further work should investigate clustering of time-series with more observations. Clustering based on the coefficients of an auto-regressive model of each subject may be viable.

The methodology established provides a springboard to further understand the different types of DR present in residential loads. User interaction with DER controllers is expected to have a large impact on the DR available, and the HAC used to separate useful households from those which do not appear extremely effective in this circumstance.

From a widespread power system perspective, being able to identify which customer segments exhibit a price response is important for grid operators looking to identify and invest in customers to participate in new DR schemes. Such clustering may also be a useful technique to decide additional financial reward for customers who perform best, in the form of a capacity payment, perhaps funded by the same public service obligations (PSOs) that support renewable generation.

Comparing treatment subgroups to the CBL graphically is also useful for presenting the differences in consumption to a broad audience in an intuitive manner. However, visual interpretation is not a statistically valid way of confirming a response. Therefore, the 2-sample Mann-Whitney test comparing the averaged unit cost of price-responsive and non price-responsive subgroups supplements the graphical approach well, as it allows us to validate or reject hypothesis

for each test-case. This analysis answers one of the key points of the demonstration: cost can be reduced for some consumers. Obviously, a necessary further step is to characterize and quantify the responsiveness of electric loads. This has been the focus of our further research over a 8 month live experiment in the EcoGrid EU project, which kicked off after the first assessment results presented here allowed to verify the demand response potential in our set of electric loads.

ACKNOWLEDGMENT

The authors would like to thank all their partners in the EcoGrid EU project for their contribution to the methodological and experimental work. Special thanks go to Maja Felicia Bendtsen (Østkraft A/S) and her team for their hard work on the field, Dieter Gantenbein and his colleagues at IBM and Martin Bo Sjøberg and Andreas Arendt at Siemens for providing equipment and expertise, and finally Niels Ejnar Helstrup Jensen, Stig Holm Sørensen and Niels Per Lund from Energinet.dk for help with the data and constant feedback on this work.

REFERENCES

[1] European Council, "Conclusions on 2030 climate and energy policy framework," Eur. Council, Brussels, Belgium, Tech. Rep. SN 79/14, Oct. 2014.

[2] N. O'Connell, P. Pinson, H. Madsen, and M. O'Malley, "Benefits and challenges of electric demand response: A critical review," *Renew. Sustain. Energy Rev.*, vol. 39, pp. 686–699, 2014.

[3] Z. Xu, J. Ostergaard, and M. Togeby, "Demand as frequency controlled reserve," *IEEE Trans. Power Syst.*, vol. 26, no. 3, pp. 1062–1071, Aug. 2011.

[4] A. Yousefi, T. T. Nguyen, H. Zareipour, and O. P. Malik, "Congestion management using demand response and FACTS devices," *Int. J. Elect. Power Energy Syst.*, vol. 37, no. 1, pp. 78–85, 2012.

[5] J. Medina, N. Muller, and I. Roytelman, "Demand response and distribution grid operations: Opportunities and challenges," *IEEE Trans. Smart Grid*, vol. 1, no. 2, pp. 193–198, Sep. 2010.

[6] R. Hledik and A. Faruqui. (2015). *Valuing Demand Response: International Best Practices, Case Studies, and Applications, Brattle Group Website*. [Online]. Available: http://www.brattle.com/system/publications/pdfs/000/005/343/original/Valuing_Demand_Response_-_International_Best_Practices__Case_Studies__and_Applications.pdf?1468964700

[7] G. Strbac, "Demand side management: Benefits and challenges," *Energy Policy*, vol. 36, no. 12, pp. 4419–4426, Dec. 2008.

[8] J. Torriti, M. G. Hassan, and M. Leach, "Demand response experience in Europe: Policies, programmes and implementation," *Energy*, vol. 35, no. 4, pp. 1575–1583, Apr. 2010.

[9] J. L. Mathieu, D. S. Callaway, and S. Kiliccote, "Examining uncertainty in demand response baseline models and variability in automated responses to dynamic pricing," in *Proc. IEEE Conf. Decis. Control Eur. Control Conf.*, Orlando, FL, USA, Dec. 2011, pp. 4332–4339.

[10] O. Corradi, H. Ochsenfeld, H. Madsen, and P. Pinson, "Controlling electricity consumption by forecasting its response to varying prices," *IEEE Trans. Power Syst.*, vol. 28, no. 1, pp. 421–429, Feb. 2013.

[11] D. M. Keane and A. Goett, "Voluntary residential time-of-use rates: Lessons learned from pacific gas and electric company's experiment," *IEEE Trans. Power Syst.*, vol. 3, no. 4, pp. 1764–1768, Nov. 1988.

[12] P. Jazayeri *et al.*, "A survey of load control programs for price and system stability," *IEEE Trans. Power Syst.*, vol. 20, no. 3, pp. 1504–1509, Aug. 2005.

[13] V. Giordano *et al.*, "Smart grid projects in Europe: lessons learned and current developments 2012 update," Publications office EU, 2013.

[14] P. Siano, "Demand response and smart grids—A survey," *Renew. Sustain. Energy Rev.*, vol. 30, pp. 461–478, Feb. 2014.

[15] V. S. K. M. Balijepalli, V. Pradhan, S. Khaparde, and R. M. Shereef, "Review of demand response under smart grid paradigm," in *Proc. IEEE PES Innov. Smart Grid Technol. India (ISGT India)*, 2011, pp. 236–243.

[16] S. B. Harvey, "The Harvard medical school guide to men's health," *Publishers Weekly*, vol. 249, no. 31, pp. 31–34, 2002.

[17] P. Bartholomew *et al.* (2009). *Demand Response Measurement & Verification, AEIC Website*. [Online]. Available: http://aeic.org/wp-content/uploads/2013/07/AEIC-MV-Whitepaper-Rev-051613.pdf

[18] J. Kwac, J. Flora, and R. Rajagopal, "Household energy consumption segmentation using hourly data," *IEEE Trans. Smart Grid*, vol. 5, no. 1, pp. 420–430, Jan. 2014.

[19] S. Haben, C. Singleton, and P. Grindrod, "Analysis and clustering of residential customers energy behavioral demand using smart meter data," *IEEE Trans. Smart Grid*, vol. 7, no. 1, pp. 136–144, Jan. 2016.

[20] M. Chaouch, "Clustering-based improvement of nonparametric functional time series forecasting: Application to intra-day household-level load curves," *IEEE Trans. Smart Grid*, vol. 5, no. 1, pp. 411–419, Jan. 2014.

[21] F. L. Quilumba, W.-J. Lee, H. Huang, D. Y. Wang, and R. L. Szabados, "Using smart meter data to improve the accuracy of intraday load forecasting considering customer behavior similarities," *IEEE Trans. Smart Grid*, vol. 6, no. 2, pp. 911–918, Mar. 2015.

[22] *EcoGrid EU Website*. Accessed on Apr. 10, 2016. [Online]. Available: www.eu-ecogrid.net

[23] Y. Ding *et al.*, "Real-time market concept architecture for EcoGrid EU—A prototype for European smart grids," *IEEE Trans. Smart Grid*, vol. 4, no. 4, pp. 2006–2016, Dec. 2013.

[24] N. E. Helstrup, P. Lund, M. F. Bendtsen, D. Gantenbein, and A. Arendt, "Deliverable 6.3—System operation and monitoring: Large-scale smart grids demonstration of real time market-based integration of DER and DR," EcoGrid EU, Tech. Rep. D6.3, Fredericia, Denmark, 2013.

[25] G. Chicco, R. Napoli, and F. Piglione, "Comparisons among clustering techniques for electricity customer classification," *IEEE Trans. Power Syst.*, vol. 21, no. 2, pp. 933–940, May 2006.

[26] R. Xu and D. C. Wunsch, II, *Clustering*. New York, NY, USA: Wiley, 2008.

[27] L. Kaufman and P. J. Rousseeuw, "Finding groups in data: An introduction to cluster analysis," *J. Amer. Stat. Assoc.*, vol. 86, no. 415, pp. 830–832, 1991.

[28] M. L. Zepeda-Mendoza and O. Resendis-Antonio, "Hierarchical agglomerative clustering," in *Encyclopedia of Systems Biology*, W. Dubitzky, O. Wolkenhauer, K.-H. Cho, and H. Yokota, Eds. New York, NY, USA: Springer, 2013, pp. 886–887.

[29] B. Efron and R. Tibshirani, "Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy," *Stat. Sci.*, vol. 1, no. 1, pp. 54–75, 1986.

[30] R. Yaffee and M. McGee, "Introduction to time series analysis and forecasting with applications of SAS and SPSS," *Int. J. Forecast.*, vol. 17, no. 2, pp. 301–302, 2001.

[31] P. Mahalanobis, "On the generalized distance in statistics," *Proc. Nat. Inst. Sci.*, vol. 2, no. 1, pp. 49–55, 1936.

[32] Z. Prekopcsák and D. Lemire, "Time series classification by class-specific Mahalanobis distance measures," *Adv. Data Anal. Classif.*, vol. 6, no. 3, pp. 185–200, 2012.

[33] O. Ledoit and M. Wolf, "Honey, I shrunk the sample covariance matrix," *J. Portfolio Manage.*, vol. 30, no. 4, pp. 110–119, 2004.

[34] J. Schäfer and K. Strimmer, "A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics," *Stat. Appl. Genet. Mol. Biol.*, vol. 4, no. 1, 2005, Art. no. 32.

[35] J. P. Benzécri, *Analyse des données, Tome 1, La Taxinomie*, 1st ed. Paris, France: Dunod, 1976.

[36] H. C. Romesburg, *Cluster Analysis for Researchers*. Raleigh, NC, USA: Lulu, 2004.

[37] C. L. Meinert and C. Buck, *Clincal-Trials Design, Conduct, and Analysis*, vol. 78. New York, NY, USA: Oxford Univ. Press, 1987.

[38] B. Efron and R. Tibshirani, *An Introduction to the Bootstrap*, 1st ed. Boca Raton, FL, USA: CRC Press, 1993.

[39] R. R. Wilcox, *Applying Contemporary Statistical Techniques*. Houston, TX, USA: Gulf Prof., 2003.

Guillaume Le Ray received the M.Sc. degree in applied statistics from Agrocampus Ouest, Rennes, France. He is currently pursuing the Ph.D. degree in energy analytics with the Technical University of Denmark. He has been working in different fields using data analysis since then, ranging from sensory analysis in the audio industry to electricity consumption analysis through nudging and human behavior analysis.

Emil Mahler Larsen received the M.Sc. degree in wind energy and the Ph.D. degree from the Technical University of Denmark, in 2011 and 2016, respectively. He is a Consultant with the Danish Energy Association, modeling future energy systems with a high proportion of renewable energy sources. He also participates in smart grid Research and Development Projects and enjoys applying machine learning techniques to different energy market problems.

Pierre Pinson (M'11–SM'13) received the M.Sc. degree in applied mathematics from the National Institute for Applied Sciences, INSA Toulouse, France, and the Ph.D. degree in energetics from Ecole des Mines de Paris, France. He is a Professor with the Technical University of Denmark, Centre for Electric Power and Energy, Department of Electrical Engineering, also heading a group focusing on Energy Analytics and Markets. His research interests include among others forecasting, uncertainty estimation, optimization under uncertainty, decision sciences, and renewable energies. He acts as an Editor for the *International Journal of Forecasting* and *Wind Energy*.

[Paper F] Data-driven demand response characterization and quantification.

Data-driven Demand Response Characterization and Quantification

Guillaume Le Ray, Pierre Pinson
Technical University of Denmark
Centre for Electric Power and Energy
Kgs. Lyngby, Denmark
email: {gleray, ppin}@elektro.dtu.dk

Emil Mahler Larsen
Danish Energy Association
Frederiksberg, Denmark
email: eml@danskenergi.dk

***Abstract*—Analysis of load behavior in demand response (DR) schemes is important to evaluate the performance of participants. Very few real-world experiments have been carried out and quantification and characterization of the response is a difficult task. Nevertheless it will be a necessary tool for portfolio management of consumers in a DR framework. In this paper we develop methods to quantify and characterize the amount of DR in a load. The contribution to the aggregated load from each household is quantified on a daily basis, showing the potential variability of the response in time. Clustering on the average values and standard deviation of the contribution regroups households with the same average response. Independent Component Analysis (ICA) is used to characterize different DR delivery profiles.**

***Index Terms*—Demand Response (DR), DR characterization, DR quantification, smart grid, energy analytics.**

I. Introduction

Research and the application of controlling electric loads in useful ways continues to grow as system operators look for new ways to balance production when a large share of renewable energy sources (RES) is present. Demand Response (DR) answers this problem by moving the load in time or shedding it. DR's rise has coincided with large-scale smart meter roll-outs across the world, which provide more precise information about the demand status. Utilities often consider two applications of DR: emergency use, which is a safety net in case of unexpected outages, and economic, which reduces the use of energy-only ancillary services [1].

Different DR schemes have been designed and implemented, with wide-ranging benefits and drawbacks, although how to quantify DR during real operation remains underexplored [2], [3]. Indirect control through dynamic tariffs has emerged as a popular option due to its ability to respond quickly to variation in production (i.e. wind, solar) and the simplicity of financial settlement, i.e. reward for consumers providing flexibility [4]. Indirect control does not require each load to respond to price signals in the same way, but the population's response as a whole remains nevertheless statistically predictable.

The evaluation of how much DR is in a load is often made by comparison to a baseline. The baseline can be generated using a model fitted on consumption data from historical non-DR days [5]. It can also be simultaneously collected from a reference sample of households which has the same characteristics (i.e. number, size, temperature) as the DR sample of households but which do not receive DR incentives [6]. However, sufficient data to generate reliable forecasts of loads or representative reference set are not always available. In the literature, little work has been done on quantifying and characterizing flexible loads from residential consumers in a DR framework when a baseline is not available.

In this paper we propose: 1) a methodology to retrospectively evaluate how much DR was delivered, on average, in a group of similar loads without estimating a baseline; 2) a method which gives information about the characteristics of the response in terms of amplitude and rate of response, i.e. how large the response is and how often a load is able to deliver the response. When managing a portfolio of households in a DR framework, it is the combination of both quantification and characterization of the response that is important when generating the future prices. Furthermore, combining the average response from groups of consumers, each receiving a different price signal, can then be a possible way to obtain a desirable aggregated load.

The paper is organized as follows: Section II presents the data and the EcoGrid EU DR implementation; Section III introduces the methodology to categorize, characterize and quantify the response to price; Section IV presents the results of the method applied to data from EcoGrid EU and Section V provides conclusions and ideas for future work.

II. Empirical Framework

The dataset used comes from the EcoGrid EU project, which was completed in 2015. The goal of EcoGrid EU was to exploit the flexibility of residential consumers in a large scale real-time market with a large share of RES in the energy mix [7].

A. Hypothesis on quantifying real time pricing DR

Houses are equipped with automated devices from different manufacturers, which in turn can be controlled by different algorithms. Each house has its own threshold for which it reacts to price, according to the comfort settings determined by the family living there. We suppose that the controller's algorithm does not affect the amplitude of the response but the pace and the frequency at which the response is delivered.

The behavior of a load is dynamic and depends on the household activity. It is therefore stochastic and should be analyzed as such. Each house may respond differently from one day to another, depending on the activity of the inhabitants. The categorization of the house's DR delivery should therefore be based on both the average response and its variability from day-to-day, since the latter describes the reliability of a portfolio.

B. Real Time Pricing Market Presentation

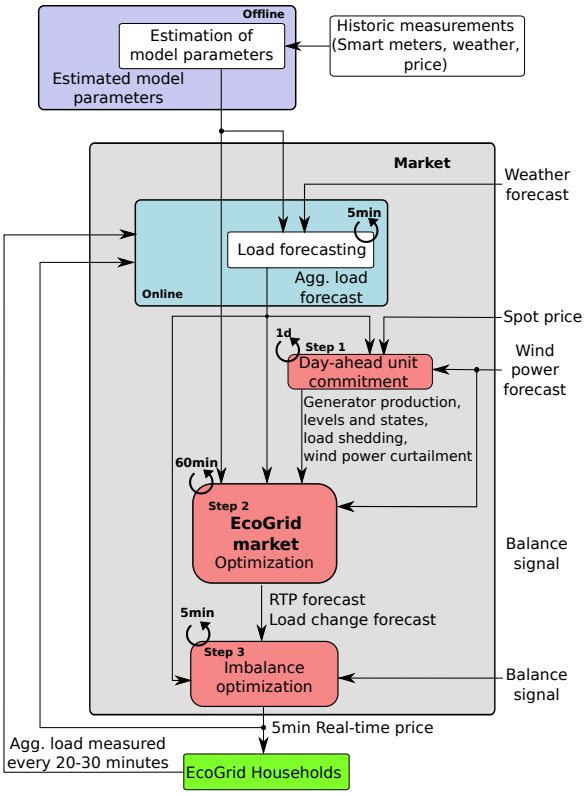


Fig. 1: Overall representation of EcoGrid EU market and price generation.

The DR framework in EcoGrid EU is a hardware-in-the-loop market platform (Figure 1). The market design can be divided into four blocks:

1) **The forecasting module** computes the load forecast based on historical consumption, weather forecasts prices.
2) **The Day-ahead unit commitment** schedules generation according to known spot prices and the wind power forecast.
3) **The EcoGrid market** is the core market clearing that finds 5-minute prices for the next hour. Its main objective is to maintain the system balance and maximize the social welfare by dispatching conventional balancing power and DR (via 5-minute prices).
4) **The 5-min imbalance optimization** makes minor adjustment to prices before committing the real-time price.

The Real-Time Price (RTP) generally follows the daily load pattern, since the spot price (a component of the RTP) is positively correlated with the consumption profile - periods of high demand cause a higher spot price. When a load is reactive, the consumption moves in an opposite direction to the price variation (i.e. consumption increases as the price decreases and vice versa). A description of the market clearing mechanism that determines the real-time price is described in [8].

C. Electric Load Data Collection

EcoGrid EU relied mostly on heating devices that could deliver DR by exploiting the implicit thermal storage of households. Out of 1900 houses in total, the houses with good quality data are listed in Table I. There were four groups that were fitted with automated equipment (groups 3–6). group 1 was designed as a reference group, but was shown to be non representative when attempting to generate a reliable baseline [9]. group 2 was a manual group with no automation of devices and participants were asked to modify their consumption through SMS. Two types of heating were installed in the other households: Electric Heating (EH) or Heat Pump (HP); the equipment (i.e. controller) installed in the households of groups 3–5 comes from the same manufacturer, with different algorithms for group 5. Group 6 is equipped with a different hardware and software. The data was collected from September 2014 to February 2015 with temperature spanning from -10.6 °C to 18.0 °C. The resolution of the electric load is 5 minutes, with data transferred to a central database every 20–30 minutes. Houses and time periods with more than 10% values missing were filtered out.

TABLE I: TEST GROUP DESCRIPTIONS.

Group	type of heating	manufacturer	type of DR	number
Group 1	mixed	none	none	253
Group 2	mixed	none	manual	455
Group 3	HP[a]	manuf. 1	auto.	195
Group 4	EH[b]	manuf. 1	auto.	322
Group 5	HP	manuf. 2	auto.	84
Group 6	EH	manuf. 3	auto.	398

[a]Heat Pump
[b]Electric Heating

III. METHODOLOGY FOR QUANTIFYING AND CHARACTERIZING RESPONSE TO RTP

The behavior of a load can change from extremely reactive to not reactive from one day to another as the inhabitants can change the thermostat temperature settings (affecting the comfort zone by extension). We evaluate the DR delivered by each house on a daily basis and summarized it using the average and standard deviation per household rather than evaluating it over the entire heating season.

A. Response Categorization

As previously described, the RTP follows the average daily pattern of the load. The less a load provides DR, the closer it will correlate with the RTP. This is captured by a positive correlation between a load x_n and RTP p as presented in (1). When a load is reactive to a price variation and show some DR, the variation of the load is opposite to the price and the correlation between the load and RTP is then negative.

$$Cor(p,x_n)=\frac{\sum_{t=1}^{T}(p_t-\overline{p})(x_{n,t}-\overline{x_n})}{\sqrt{\sum_{t=1}^{T}(p_t-\overline{p})^2}\sqrt{\sum_{t=1}^{T}(x_{n,t}-\overline{x})^2}} \quad (1)$$

$$d.Cor(p,x_n)=\sqrt{\left(\frac{1-Cor(p,x_n)}{1+Cor(p,x_n)}\right)} \quad (2)$$

Using the correlation-based distance (d.Cor) in (2) between the individual loads x_n and the RTP p of a single day, the response during that day can be quantified [10]. Note that an average correlation of 0 over a day corresponds to a d.Cor of one and the more negative the correlation is, the larger the d.Cor.

The d.Cor is calculated for each household and every day of the heating season. The mean and standard deviation is calculated per household in order to cluster them according to their average distance from the RTP and how it varies during the season. The cluster algorithm used to split the data is K-means (k=5) run with 25 different sets of starting points to validate the stability of the partition.

B. Response Quantification

The response in each cluster, is estimated using Finite Impulse Responses (FIR) for price [9]. FIRs are obtained from a general linear model of x_t the observed load,

$$x_t=\widetilde{\lambda_t}^{\top}\theta_\lambda+\widetilde{z_t}^{\top}\theta_z+\widetilde{\chi_t}^{\top}\theta_\chi+\epsilon_t \quad (3)$$

where $\widetilde{\lambda_t}^{\top}$ is a vector including forecast, real-time and historic electricity prices (e.g. day-ahead, hour-ahead and RTP), $\widetilde{z_t}^{\top}$ is a vector of exogenous variables (e.g. solar radiance, wind speed, exterior temperature, Fourier series of the daily independent base load), $\widetilde{\chi_t}^{\top}$ is a vector of interactions of some of the precedent variables and ϵ_t is the normally distributed error.

The FIRs correspond to the vector of relative real-time price coefficients θ_λ in (3). It is calculated based on the relative price which is the difference, at each time step, between the RTP and the hour-ahead price. The fitting of the model is done using a Lasso penalization [11].

C. Response Characterization

The RTP exhibits many local extremii during a single day,to which households do not always react. Clustering is a good way to group the households according to their responses on average, but when more details are required to characterize the response, the limits of the method are reached. Each cluster is the average reaction of a pool of households and, although confidence intervals can be calculated, they obscure an otherwise describable, diverse set of behaviors. To characterize the loads, let us first consider that each consumption time-series is a linear combination of I different consumption pattern s_i and that the vector of time-series consumption $\boldsymbol{x}=\{x_1,...,x_n,...,x_N\}$ can be written,

$$\boldsymbol{x}=\sum_{1}^{I}a_i s_i \quad (4)$$

where a_i is a vector of coefficients. Each component a_i represents a response to the price or heating preferences setting and, by combining and weighting them, the consumption of the time-series can be reconstructed. The method is called Independent Component Analysis (ICA) [12]. It was first used in signal processing to solve the problem of blind source separation, like the *cocktail party problem*. (4) can also be written,

$$\boldsymbol{x}=\boldsymbol{a}\boldsymbol{s} \quad (5)$$

where $\boldsymbol{a}=\{a_1,...,a_i,...,a_I\}$ is a matrix of coefficients and $\boldsymbol{s}=\{s_1,...,s_i,...,s_I\}$ is a vector representing a base of synthetic signals with no units, called independent components (IC). After estimating $\boldsymbol{a}$, its inverse $\boldsymbol{w}$ is computed and the IC can be obtained by,

$$\boldsymbol{s}=\boldsymbol{w}\boldsymbol{x} \quad (6)$$

and corresponds to weights of consumption time-series $\boldsymbol{x}$ on ICs $\boldsymbol{s}$. The only assumptions made to perform an ICA are that the components should be statistically *independent* and so they must have *non-gaussian* distributions. The ICA estimation is based on measure of non-gaussianity (e.g. kurtosis, negentropy) of $w_i^T\boldsymbol{x}$ where w_i^T is a transpose row w_i.

Before running an ICA, the signals are centered and whitened using Principle Component Analysis (PCA). Thus all signals have a zero mean and a unit variance.

The *fastICA* is used to perform the ICA and is presented in Algorithm 1 [12]. First $\boldsymbol{w}$ is initiated randomly by sampling from Gaussian distribution with $\mu=0$ and $\sigma^2=1$. A fixed point iteration process then finds the maximum non-gaussianity by approximating the negentropy using its approximative Newton iteration. In the presented implementation, the function g and g' are respectively the derivative and second derivative of the non-quadratic function G,

$$G(u)=-\exp(-u^2/2) \quad (7)$$

Algorithm 1 FastICA

Initiate:
$\boldsymbol{w}\leftarrow\{w_1,...,w_i,...,w_I\}\sim\mathcal{N}(\mu,\sigma^2)$
for i in $\{1,...,I\}$ with $I\leq N$ **do**
  while not converged **do**
    $w_i^+=E\{\boldsymbol{x}g(w_i^T\boldsymbol{x})\}-E\{g'(w_i^T\boldsymbol{x})\}w_i$
    $w_i=w_i^+/||w_i^+||$
  end while
end for

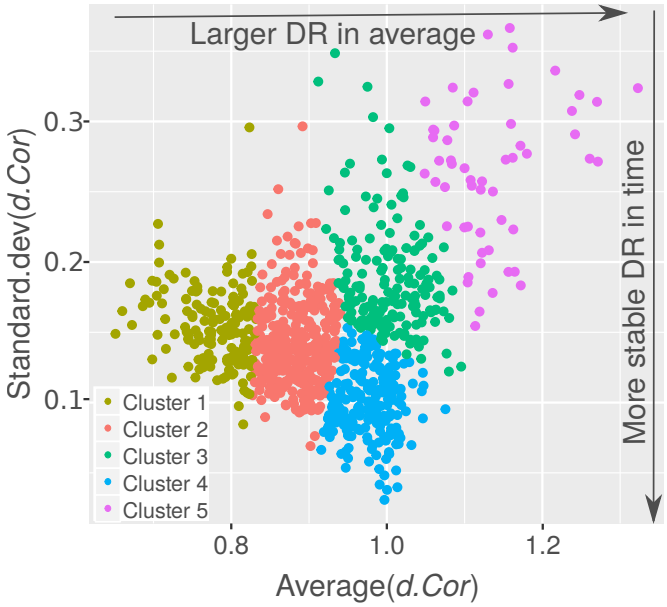


Fig. 2: Clusters generated with K-means (k=5) on the mean and standard deviation of the $d.Cor$ distance

IV. Results

A. Response Categorization Results

The results of the clustering are shown in Figure 2. Each point represents an household in the two dimensional space defined by the average d.Cor and the standard deviation of the d.Cor over the heating season. The color codes the partition into clusters. An increase of standard deviation with average d.Cor results from the non-linearity of d.Cor [10]. The standard deviation is mostly used to discriminate between houses with the same average participation but different variability over the heating season.

The analysis of repartition of the groups in each cluster (Figure 3 on the left) reveals that besides cluster 5, which is mostly composed of group 6 households, the manufacturer or the type of heating does not influence the average response. Indeed cluster 5 which is the farthest from the RTP, is mostly composed of households from the group 6 and 4 which are both equipped with EH but it could be the result of another factor as group 4 and 6 are also present in other clusters (e.g. cluster 3). With the exception of cluster 5, these results appear to support the hypothesis that the group and subsequently the technology is not the main driver of the average response of the load as the groups are distributed in the different clusters.

The repartition of the dwelling types in each cluster is presented in Figure 3 on the right. The three clusters with the highest response to RTP on average present also the largest shares of holiday houses. On the one hand the fact that inhabitants are not present in the dwelling is a problem to d.Cor as the base load is then not following the base pattern of the price and it increases the value of the distance correlation. On the other hand the absence of interaction with inhabitants may also be an asset in the planning of the heating which

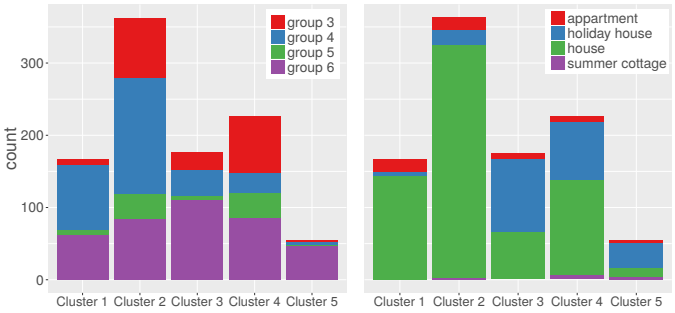


Fig. 3: Split of the groups (left) and the type of dwelling (right) into the K-means clusters

gives the controller the full control on planning the heating over a whole day.

B. Response Quantification Results

The quantification of the response has been done based on the clusters generated in Section IV-A. Figure 4 presents the FIR of each cluster, which is a traditional time-series analysis tool that represents the average response over the entire test period (not just a daily metric). The results are given in percentage of the maximum load.

With this traditional time-series analysis approach, cluster 4 exhibits the smallest response, peaking at 10% of the load 35 minutes after the price change, followed by a stiff and large rebound (21%) where the energy to be recovered is actually larger than the initially delivery of energy. Cluster 4 shows a small increase in consumption when the change in price first arrives, which may be a time-delay caused by minimum run-times, since this coincides with the slow DR delivery. Cluster 1 and 2 exhibit the next smallest DR volume, at 13%, with a

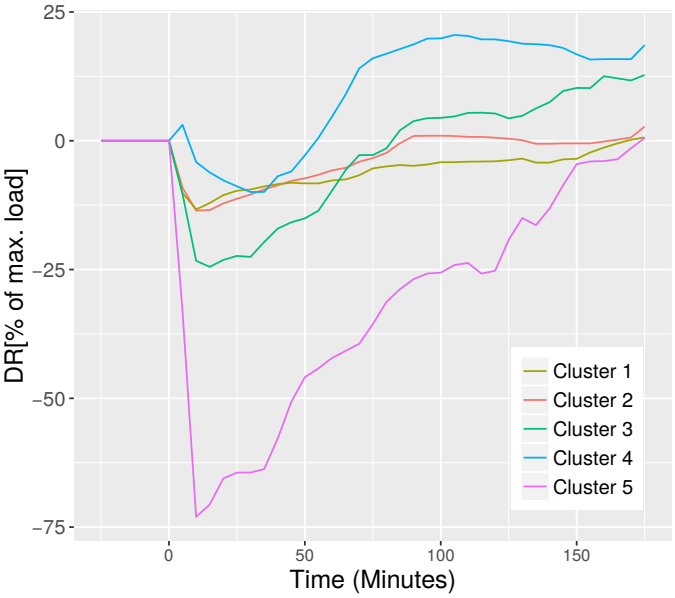


Fig. 4: FIR for the clusters obtained based on their average response and variability in the response.

TABLE II: DISTRIBUTION OF THE GROUPS ABOVE THE 95% QUANTILE AND UNDER THE 5% QUANTILE OF THE DENSITY OF WEIGHTS FOR EACH IC

(a) under the 5% quantile

IC #	1	2	3	4	5	6	7	8	9	10
Group 3	33	5	**10**	5	14	4	0	5	4	5
Group 4	12	21	**29**	7	16	21	0	10	**27**	23
Group 5	2	4	4	1	7	3	14	7	2	4
Group 6	3	20	7	**37**	13	22	36	28	**17**	18

(b) above the 95% quantile

IC #	1	2	3	4	5	6	7	8	9	10
Group 3	12	9	10	7	19	6	15	11	8	31
Group 4	20	27	10	**26**	15	17	17	20	**27**	9
Group 5	4	1	0	0	7	6	7	6	1	3
Group 6	14	13	**30**	17	9	21	11	13	**14**	7

long-lasting response (85 min. for cluster 2 and 175 min. for cluster 1) followed by no significant rebound.

Cluster 3 delivers 24% DR, with maximum delivery after 15 minutes. It exhibits a slow but larger rebound after 85 minutes.

The most homogeneous cluster, composed mostly of group 6 houses, is cluster 5, which delivers by far the most DR. Its DR amplitude reaches 72% of its maximum load, with a delay of just 10 minutes, signifying that this small cluster is composed of high performing houses. It does not present any rebound effect after the DR event. This result tentatively confirms the order of response from not much responsive (cluster 1) to large response (cluster 5) observed in Section IV-A.

C. Response Characterization Results

The ICA has been implemented on a single day; Tuesday the 6^{th} of January 2015. After whitening the consumption time-series using PCA, the I=10 first principal components are kept to perform the ICA. The ten first components of the PCA collect 72% of the total variance of the data. The decomposition of the signals into ICs separates the independent observed behaviors of the loads. In other words, any load generated by a household (i.e. real or fictive) with the characteristics of the pool of households used to generate the ICs can be represented as a linear combination of the ICs as presented in (4). ICA projects each load pattern into a space with the ICs as a base of dimension I.

It is important to note that the ICA does not prioritize an IC over another when generation the base, which means that the ICs cannot be ranked according to a specific metric (e.g. % of explain variance in the case of PCA). The IC, when represented, corresponds to one possible direction (positive weights) of the base vector, which makes the interpretation non-direct. The extreme 5% quantiles (above 95% and under 5%) of the density distribution of weights, which correspond to the most contributive points, are analyzed for each IC (Table II & III) to evaluate which groups and which clusters contribute highly (positively or negatively) to each IC.

From Table II, Figure 5 and the knowledge gathered about the possible settings of the different equipment, an interpretation of the IC numbers 3, 4 and 9 is given.

TABLE III: DISTRIBUTION OF THE CLUSTERS ABOVE THE 95% QUANTILE AND UNDER THE 5% QUANTILE OF THE DENSITY OF WEIGHTS FOR EACH IC

(a) under the 5% quantile

IC #	1	2	3	4	5	6	7	8	9	10
Cluster 1	4	9	**16**	1	8	7	23	1	**27**	8
Cluster 2	23	26	**26**	8	30	19	20	11	**17**	20
Cluster 3	10	5	3	**24**	3	15	6	18	2	12
Cluster 4	11	7	5	3	6	4	0	13	4	9
Cluster 5	2	3	0	**14**	3	5	1	7	0	1

(b) above the 95% quantile

IC #	1	2	3	4	5	6	7	8	9	10
Cluster 1	16	13	0	**23**	9	10	1	26	**13**	6
Cluster 2	17	24	10	**16**	21	19	24	22	**28**	19
Cluster 3	4	3	**14**	7	7	11	13	2	3	6
Cluster 4	13	8	**8**	3	12	3	7	0	5	19
Cluster 5	0	2	**18**	1	1	7	5	0	1	0

A close look at IC 3 reveals that the load described by the component responds for every RTP variation, inversely if the weights is positive (above 95% quantile) or following if the weight is negative (under 5% quantile). From Table IIa, groups 3 and 4 are the one following the variation of the price. From Table IIb group 6 is the one reacting inversely to every variation of the price.

IC 4 has a pattern which is totally independent from the price. When the weight is positive, it displays an increase of the consumption from 08:00 to 12:00. When the weight is negative, it is the opposite, the consumption drops from 08:00 to 12:00. This pattern seems to depend on the presence of somebody during the day (positive) or an 'economy setting' (negative).

IC 9 displays a rebound, it anticipates the peak in price (at 07:00) by heating up and using the thermal inertia of the dwelling and cuts heating for the next two hours and half (positive weight) or inversely (negative weight). This is a sign that the controller makes its decision based on information obtained from the day-ahead price or hour-ahead price. The results from Table II clearly shows that groups 6 and 4 are both divided almost into the same proportion between large positive and large negative weights. It is also interesting to note that both groups are equipped with electric heating.

The results shown in Table III are not as clear as the one in Table II but some tendencies can be observed. On IC 3, the most negative weights are mostly from clusters 1 and 2 and the highest positive are from clusters 3, 4 and 5 which confirm their active response (on this date). On IC 4, the largest positive weights are from clusters 1 and 2 which have the largest share of standard houses. On the negative side, the clusters 3 and 5 which have the largest shares of summer houses. Clusters 1 and 2 combined are approximately in the same numbers under 5% and above 95%; it is a consequence of the composition of these clusters which have a majority of households from groups 4 and 6 which presents the same cleavage into half rebound before and half rebound after.

The analysis was also run on the type of housing and only

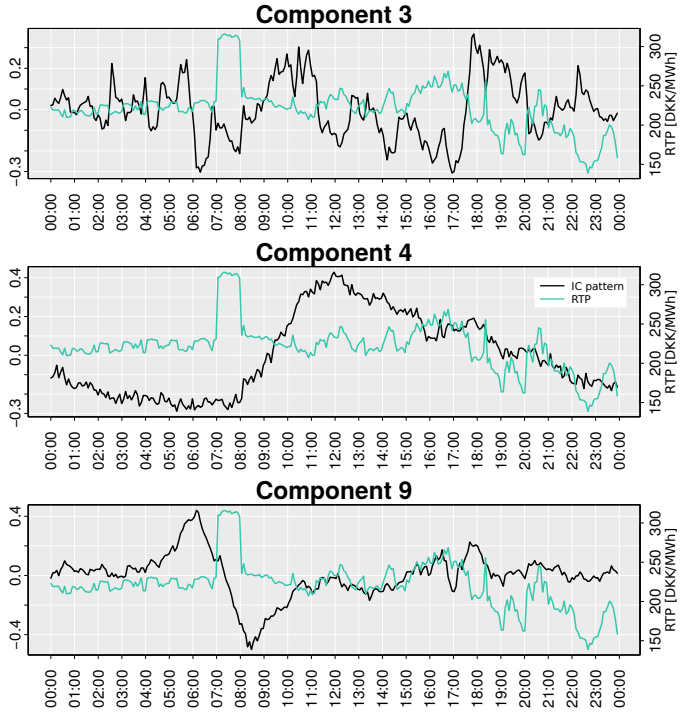


Fig. 5: Representation of the IC number 3, 4 and 9 with the RTP in the same period (06-01-2015)

IC 4 splits clearly between summer houses (negative) and standard houses (positive), which confirms that IC 4 pattern is an 'economy setting'. It also rejects the idea that the type of housing is a factor facilitating the response to a small variation of price, which corresponds to IC 3.

It appears that ICA provides meaningful results when studied in details.The IC 3, 4 and 9 illustrate three different types of observed behaviors, response to RTP (IC 3), setting of the controller by the user (IC 4),and anticipation of a price variation due to controller scheduling (IC 9). The other ICs expose other various behaviors.

V. Conclusion and Future Work

The results from the clustering (Section IV-A) combined with the results from the FIR analysis (Section IV-B) and the ones from the ICA (Section IV-C) investigate the problem from 2 different angles. The first one intends to give an overall estimation of the average response of the households along the heating season as well as the variability by categorizing them and quantifying the response of each cluster. The second aims at characterizing the response as a linear combination of independent signals (i.e. IC). This way the details of the response of each household can be comprehended as the sum of different load behaviors resulting from interactions with the inhabitants, heating scheduling or response to price incentives. Combining this information provides valuable insights for a system operator who might want to activate different clusters at different times to deliver a more finely-tuned response than activating a single cluster would achieve.

The ICA has shown promising results for this specific application and it could theoretically be used to create specific aggregated responses by sampling households with specific weights. The FIR analysis can actually be used as a validation tool after sampling process, as it uses historical data and models the response of a pool of households. Subsequently, different prices could be sent to households to obtain a predefined aggregated load. This approach assumes a certain stationarity of the time-series which is known to not be true when looking at single consumption time-series, but which can be an acceptable assumption when looking at aggregated consumption time-series. In the continuity of this idea, a unit commitment model with different prices per subgroup of households could be implemented to evaluate the benefit of broadcasting different prices.

The drawback of ICA is that is does not produce good results on long time-series. An extensions of this work could be to implement an online ICA or a similar method which would provide results on the entire period instead of only one day.

Acknowledgment

The authors thank our EcoGrid EU partners for supporting this work, as well as the EUDP for funding through the EnergyLab Nordhavn project (EUDP 64015-0055).

References

[1] P. Palensky and D. Dietrich, "Demand side management: Demand response, intelligent energy systems, and smart loads," *IEEE transactions on industrial informatics*, vol. 7, no. 3, pp. 381–388, 2011.

[2] P. Siano, "Demand response and smart grids - A survey," *Renewable and Sustainable Energy Reviews*, vol. 30, pp. 461–478, 2014.

[3] V. Giordano, F. Gangale, and Others, "Smart Grid projects in Europe: lessons learned and current developments," *JRC Reference Reports, Publications Office of the European Union*, 2011.

[4] N. O'Connell, P. Pinson, H. Madsen, and M. O'Malley, "Benefits and Challenges of Demand Response: A Critical Review," *Renewable and Sustainable Energy Reviews*, vol. 39, pp. 686–699, 2013.

[5] J. L. Mathieu, P. N. Price, S. Kiliccote, and M. A. Piette, "Quantifying changes in building electricity use, with application to demand response," *IEEE Transactions on Smart Grid*, vol. 2, no. 3, pp. 507–518, 2011.

[6] G. Le Ray, E. M. Larsen, and P. Pinson, "Evaluating price-based demand response in practice - with application to the EcoGrid EU Experiment," *IEEE Transactions on Smart Grid*, pp. 1–1, 2016.

[7] P. Lund, P. Nyeng, R. N. Grandal, S. H. Sørensen, and Others, "Deliverable 6.7 Overall evaluation and conclusion integration of DER and DR," EcoGrid EU Final Report, Tech. Rep., 2016.

[8] Y. Ding, S. Pineda, P. Nyeng, J. Østergaard, E. M. Larsen, and Q. Wu, "Real-time market concept architecture for EcoGrid EU - A prototype for European smart grids," *IEEE Transactions on Smart Grid*, vol. 4, no. 4, pp. 2006–2016, 2013.

[9] E. M. Larsen, P. Pinson, F. Leimgruber, and F. Judex, "Demand response evaluation and forecasting - methods and results from the EcoGrid EU experiment," *Sustainable Energy, Grids and Networks*, in press.

[10] P. Montero and J. Vilar, "TSclust: An R Package for Time Series Clustering," *JSS Journal of Statistical Software*, vol. 62, no. 1, pp. 1–43, 2014.

[11] G. James, D. Witten, T. Hastie, and R. Tibishirani, *An Introduction to Statistical Learning*, 2013.

[12] A. Hyvärinen and E. Oja, "Independent component analysis: Algorithms and applications," pp. 411–430, 2000.

Department of Electrical Engineering
Center for Electric Power and Energy (CEE)
Technical University of Denmark
Elektrovej, Building 325
DK-2800 Kgs. Lyngby
Denmark

www.elektro.dtu.dk/cee
Tel: (+45) 45 25 35 00
Fax: (+45) 45 88 61 11
E-mail: cee@elektro.dtu.dk