A data driven domestic simulator based on smart meter data.

Karen Yadira Gonzalez-Trapero

A Thesis presented for the degree of
Master of Science by Research in Engineering

Durham Energy Institute
Department of Engineering
University of Durham
England

January 2020

Dedicated to

My parents

(Florencio González Armenta and Elizabeth Trapero Ruíz)

Developing a data driven domestic simulator based on Smart Meter Data.

Karen Yadira Gonzalez-Trapero

Submitted for the degree of MSc by Research in Engineering
January 2020

Abstract

There have been numerous international efforts to reduce carbon emissions in recent years. The success of this endeavour, however, would be better achieved if we understood more greatly how people use energy, and how to be better equipped to experiment with it. The CLNR, a Durham University Project, generated Smart Meters Data from 14 thousand domestic customers which is being used to create a simulator that will help to experiment with different energy use scenarios, one instance of which is Demand-Side Management. The purpose of this research is to develop a tool which can help to better understand and manage how electricity is used within a the domestic environment of a home. The simulator was developed using the CLNR data as a base to simulate the behaviour of 280 houses that are connected to the energy network. From the CLNR database we collected and processed the data of individual energy consumption for every appliance in a household. It is based on this that we can simulate when an appliance is likely to be ON or OFF with an estimation of how much energy will be used. The development of a simulator like this means that we can have a virtual lab for testing the demand-side management and new houses' appliances.

Declaration

The work in this thesis is based on research carried out at the Department of Engineering in Durham University, England. No part of this thesis has been submitted elsewhere for any other degree or qualification and it is all my own work unless referenced to the contrary in the text.

Acknowledgements

I would like to thank the following people, without whom I would not have been able to complete this research, and without whom I would not have made it through my masters degree!

First of all I would like to thank my supervisor Dr. Peter Matthews, for his enthusiasm for the project, for his support, encouragement and patience.

My studies at Durham University would not have been possible without the aid of the CONACYT-SENER fund, which sponsored my masters studies. Thanks to these two institutions, the Mexico's Ministry of Energy (SENER) and the National Council for Science and Technology (CONACYT), I was awarded with a scholarship to study a postgraduate degree in England. This invaluable opportunity allowed me to discover a new country, learn about its energy matters, challenge myself and open my mind about strategies that can be applied in Mexico.

I also desire to express my sincere and total gratitude to Professor Teresa Alonso and Marcial Arrambi for their dedicated support and guidance. To the company Hunabsys R&D, where I was previously employed and who encouraged me to do my masters abroad and were amazingly supportive, especially during the application process at Durham University.

To Durham University and the Engineering Department for being so welcoming and always willing to help during this journey. To Fiona O'Carroll, Evelyn Tehrani, the International Office and all the team who made the Pre-sessional at Mexico possible in 2017.

To Maura, who I met in Durham and has been a great friend, a friend with whom I spent many nights and many cups of coffee working till late. To my lovely boyfriend and proofreader who knows how to make me happy with silly jokes, food and brownies... thank you James!

Last but not least, I want to thank Dr. Clemente García, Dr. Lucia Barrón, Dr. Leopoldo Zepeda and Dr. Ricardo Quintero at Instituto Tecnologico de Culiacan for their recommendation and high quality teaching, which was crucial for the development of this project. To my parents and brothers, who have been incredibly supportive and encouraging for this.

Contents

List of Figures

List of Tables

List of Abbreviations

ABM Agent-Based Model

ABS Agent-Based System

CER Commission for Energy Regulation

CLNR Customer-Led Network Revolution

DDDS Data Driven Domestic Simulator

DLC Direct Load Control

DSM Demand Side Management

EPRI Electric Power Research Institute

EVs Electric Vehicles

HVAC Heating, Ventilation and Air Conditioning

ILC Indirect Load Control

JSON JavaScript Object Notation

kW kilowatts

kWh kilowatt-hour

LCN Low Carbon Network

NESEMP North East Scotland Energy Monitoring Project

Wh watt-hour

PV Solar Photovoltaics

REFIT Renewable Energy Feed-in Tariff

RDBMS Relational Database Management System

SMEs Small and Medium-sized Enterprises

SQL Structured Query Language

USA United States of America

Wmin watts-minute

Chapter 1

Introduction

One of the United Kingdom's Government goals by 2050 is to reduce the carbon emissions by at least 100 percent when compared to the levels in 1990 [2]. The goal is set to be achieved by finding suitable strategies to deal with these new demands. The procedure of implementing strategies to deal with these new demands are twofold: firstly, this can be solved by following the traditional method of reinforcing the energy network, or secondly, by using the Customer-Led Network Revolution (CLNR) method (the "smart" method). The latter involves the question "how to deliver the most of the current network capacity at a low cost?".

According to Goran Strbac the UK's infrastructure is more than 50 years old. UK generation, transmission and distribution systems were considerably expanded in the late 1950s and early 1960s. It should be possible to keep using that infrastructure by implementing smart solutions like Demand Side Management (DSM) [3]. DSM can be defined as the modification of electricity demand through consumer interactions, e.g. by giving financial incentives or behavioural education. The traditional approach has been letting the demand set the target that has to be satisfied and schedule energy generation based on it, i.e. scheduling energy generation to meet demand. It is necessary to reverse that approach by scheduling the energy generation, letting it run in the lowest carbon mix possible before modifying the demand to match the current generation mix. Parameters like storage, people behaviour and the weather affect directly the demand for energy.

Powells et al [4] argue that social practices shape electricity demand curves and that in order to reduce the energy demand it is important to have a clear understanding of how and why people use energy and the possibilities and limits regarding "behaviour change". As a way of understanding social practices, energy behaviour and consumption in a household, energy companies started to introduce Smart Meters. Smart Meters are a device that provides the users with almost near real time information about their energy use, so that they can make decisions about how they manage the use of their electronic devices [5]. These efforts are being made with the hope of reducing energy demand, to save money and to reduce carbon emissions.

Since the introduction of Smart Meters to households a few experiments have taken place in many parts of the world, including the United Kingdom. In 2014 Northern Power Grid along with other partners, including Durham University, concluded a Smart Grid Project demonstration that involved Smart Meters, entitled 'The Customer-Led Network Revolution project' (CLNR). The CLNR Project generated smart meters data from thousands of real domestic customers. This data was generated to analyse the customers load and generation profile which is essential for decision making when changes to the energy network have to be planned. Currently, the CLNR data is still available in the project website for further research and analysis, and it is the same data that this work has used as a base of research.

Having an agent-based simulator driven by Smart Meter Data can help to better understand and manage how electricity is used, stored and delivered. The work presented in this dissertation represents the results of the efforts for creating a **Data Driven Domestic Simulator (DDDS)** and it is divided in six additional chapters: **Literature Review** (where smart meter trials are addressed), **Methodology** (where the approach to create the simulator is discussed, along with the methods for data selection), **CLNR Data Pre-processing** (because it was necessary to deal with the data before it was ready to be used in the simulator), **Simulator** (the explanation of how the created agent-based simulator works), **Results** (the main simulation results and a decarbonisation experiment) and finally **Conclusions and future work**.

Chapter 2

Literature Review

In this Chapter some smart meter trials are described briefly, emphasis is made in the trial undertaken by the CLNR. In addition, Simulation and Agent Based Simulation are addressed.

2.1 Smart Metering Trials

Smart Meters were created as a way to take control of energy use. Currently there are many types of smart meters in the market which all share the collective purpose to generate real time data in order to help people reduce their energy use, help energy suppliers deliver better customer service and accurate as well as bills and provide households with the knowledge they need to be more active and engaged consumers. Smart meters are central to move to a smarter and more flexible energy system of the future [6]. Between 2009 and the present year (2019) approximately a dozen smart meter trials have been in operation around the world on a large scale. In particular, some of the smaller trials focused on residential consumers, small to medium-sized enterprises Small and Medium-sized Enterprises (SMEs) or public buildings, meanwhile other larger trials covered all of these categories within large geographic areas of the countries in question, as is the case of the CLNR project in the UK. These trials helped to generate data at a specific interval rate regarding how electricity is being used for lighting, running appliances, electronics and often for heating and cooling.

2.1.1 Previous Smart Meter Trials

CER in Ireland.

The CER covered 4,225 residential consumers and 2,210 small to medium-sized enterprises between July 2009 and December 2010. The data collected in this trial corresponds to electricity consumed during a thirty minutes interval measured in kilowatt-hour (kWh). Heating and cooling systems were not considered when the source of energy came from gas or oil. [7] [8]

Public Buildings in Japan.

This trial has been running from 2012 to the present (2020) in a suburban city near Tokio, Japan. The sample involves thirty-four public buildings: one office building, ten cultural/social facilities and twenty-three elementary or junior high schools. In this trial, demographic data is being considered such as type of building and type of HVAC (heating, ventilation and air conditioning) system and floor areas. Data concerning local outdoor temperature is also being collected besides electricity demand data [9].

EPRI in Austria.

Between December 2009 and November 2010 the Electric Power Research Institute (EPRI) made a trial with more than 1,500 households in Linz, Austria. The data collected in this trial corresponds to energy consumption for the following appliances: boiler, dishwasher, dryer, freezer, refrigerator and television. No households using electricity for heating were included in the sample due to the fact that electricity is hardly used for heating or cooling purposes in Austria. The EPRI kept a record of some other parameters like household characteristics (income, level of education and number of household members by age), income (where every household belongs to one of the groups: 1.- Below 1,500, 2.- Between 1,500 and 2,500, 3.- Above 2,500) and level of education (represented by a boolean value, which is true when the survey respondent received at least ten years of education). The collected data about electricity consumption was available for 1,525 households, where 775 were pilot group households and 750 control group households. The data collected was stored in the smart meter and then transferred to a data concentrator once a day [10].

Smart Meters in the United States of America (USA).
By the end of 2016 the electric companies had installed 65 million smart meters, covering more than 50 percent of the USA households and more than thirty electric companies in the country had fully deployed smart meters [11].

The REFIT electrical load measurements dataset in Glasgow, UK.
This trial consisted of a two year study made between 2012 and 2015. The experiment involved twenty households where the whole house and “appliance by appliance” energy consumption was measured. This trial had as an output a dataset that comprises 1,194,958,790 readings, that represent over 250,000 monitored appliance uses. This comprised the active power measurements of the household aggregate as well as nine appliances, all recorded at eight-second intervals. It was only possible to track nine appliances because of the limitations of the hardware used in the trial. The collected data has been used for different purposes like designing algorithms to disaggregate the load, managing temporal dynamics of demand response, appliance modelling, finding usage patterns and linking domestic routines to their energy implications [1].

There was a total of ten energy sensors per household. Nine of these sensors were used to measure each of the nine appliances per house, and the last one was used to measure the energy consumption as a whole. Each reading taken from the sensors was sent to an“Energy aggregator" which subsequently sends the collected data to a gateway. Once the data was received by the gateway this then sends it through the internet, so that it can be stored in a MySQL database and accessed via a web portal.

2.1.2 CER Smart Metering Project.

This trial was initiated by the Commission for Energy Regulation (CER) in Ireland. It took place during 2009 and 2010 and generated a dataset which contains energy consumption measurements of over 5,000 Irish households and businesses. The measurements recorded energy consumption in kWh every thirty minutes. Besides energy consumption this dataset also contains information about survey results, which includes consumers' demographics (occupation, family type, etc.), house information (ownership, age, floor area, etc.), and appliance usages (dishwasher, television, water pump, etc.) [8].

2.1.3 CLNR

This trial took place between 2011 and 2015 in the United Kingdom and consisted in a four year smart grid demonstration: the CLNR. It was conducted by the Northern Power Grid and partners like Low Carbon Network (LCN) Fund, British Gas, Durham University, Newcastle University and EA Technology. The trial was performed across four strategically selected areas of the North East of England: Northumberland, Tyne and Wear, Durham and Cleveland. More than 13,000 domestic, SMEs took part in the CLNR project, where customers were equipped with smart meters. Valuable results were generated from the trial and it also highlighted areas for further study. [12].

CLNR test cells

The data from customer trials led by the CLNR was gathered in nine different test cells. Eight of the test cells belong to a domestic customer type and one belonging to SMEs. The eight test cells with a domestic customer type are the following: 1) *TC1a* includes data of basic profiling of 8,798 domestic smart meter customers, 2) *TC2a* includes enhanced profiling data of 199 domestic smart meter customers, 3) *TC3* includes enhanced information of 89 domestic customers with air source heat pumps, 4) *TC5* contains enhanced information about profiling of 155 domestic customers with Solar Photovoltaics (PV), 5) *TC6* includes enhanced profiling information of 144 domestic customers with Electric Vehicles (EVs), 6) *TC9a* contains information about 665 domestic smart meter customers on time of use tariffs, 7) *TC20Auto* includes information of 98 domestic PV customers with automatic in-premise balancing for hot water charging and 8) *TC20IHD* contains information about 147 domestic PV customers using in-home displays for manual in-premise balancing. The test cell related to SMEs is the *TC1b* and it is a dataset that contains basic profiling information of 1,783 customers [13]. All of those datasets which contain information about the nine explored test cells represent the work and efforts of the LCN Fund and its partners during a period of four years which has allowed for appropriate learning in order to have insights of how future network and generations costs could be reduced.

Test cells: *TC1a* and *TC2a*

The datasets for the test cells *TC1a* and *TC2a* contain basic information of households. Neither *TC1a* nor *TC2a* contain information of households with air source heat pumps, solar photovoltaics or electric vehicles. The dataset related with TC1a stores information of over 300 million recordings where every recording represents energy consumption measurements made by an electricity supply meter and it is measured in kWh. There are two datasets related with the test cell *TC2a*: *tc2amicro* and *tc2apassiv*, the first one is a circuit monitoring by the company Microwatt and the second one is a smart plug monitoring by Passiv Systems. Datasets *tc2amicro* and *tc2apassiv* store more than 180 and 382 million recordings respectively. Records in *tc2apassiv* represent the average power [measured in kilowatts (kW)] while records in tc2amicro represent energy consumption in periods [measured in watt-hour (Wh)], both in time intervals of one minute. Every record in these datasets corresponds to particular appliances as well as the whole home. Since *tc2amicro* and *tc2apassiv* datasets were generated by different companies they consequently measured the appliances in different ways, also creating different channels names for the appliances. Dataset *tc2amicro* contains forty-four different channel names (see Appendix B.1) whilst dataset *TC2apassiv* contains thirty-six channels (see Appendix B.2).

Demographics

In order to have more detailed information the CLNR project decided to implement a households segmentation system. This segmentation consisted in using and applying a geodemographic classification made in 2009 by the company Experian. Experian created a system for classification of UK households, called Mosaic UK, according to its creator this system provides an accurate understanding of the demographics, lifestyles and behaviour of all individuals and households in the UK. This geodemographic segmentation system considered factors such as demographics, property value, socioeconomic and consumption, property characteristics, location and financial measures. Mosaic UK classified the UK population into fifteen main groups [14] and within this, sixty-seven different types. Every household in the CLNR trial belongs to one of the fifteen main groups which are described as follows:

A. Alpha Territory

   It contains many of the most wealthy and influential people in Britain, people who have risen to positions of power in the private and public sectors.

B. Professional Rewards

   They are the UK's executive and managerial classes. Often in their 40s, 50s or 60s, some may be owners of small or medium sized businesses whilst others will have risen to senior positions in large multinational organisations.

C. Rural Solitude

   It contains people who live in small villages, isolated farmhouses or cottages where farming and tourism are the pillar of the economy.

D. Small Town Diversity

   This group has people who live in medium sized and smaller towns in neighbourhoods of older housing where there is relatively little change in the population from one year to the next.

E. Active Retirement

   These neighbourhoods contain people aged over 65 whose children have grown up and, on retirement, have decided to live in a community among people of similar ages and incomes.

F. Suburban Mindsets

   They are mostly married people of middle age, living together with their children in family houses.

G. Careers and Kids

   This people are young couples, married or living with their partner whose lives are focused on the needs of their growing children and the creation of a comfortable family home.

H. New Homemakers

They live in homes which are likely to have been built only in the last five years.

I. Ex-Council Community

These neighbourhoods are populated by people who are practical and enterprising, rather than well-educated, who have created a comfortable lifestyle for themselves through their own hard work.

J. Claimant Cultures

They are some of the most disadvantage d people in the UK including significant numbers who have been brought up in families that have a history of dependency on the state for their welfare.

K. Upper Floor Living

This people are on limited incomes and rent small flats from local councils or housing associations.

L. Elderly needs

They are pensioners whose faculties are now fading and who can no longer easily manage the responsibility of looking after a house and garden.

M. Industrial Heritage

This people are traditional and conservative, living in communities that historically have been dependent on mines, mills and assembly plants for their livelihood.

N. Terraced Melting Pot

This people work in relatively menial, routine occupations and are poorly educated. The majority are young, some still single, others living with a partner with children of nursery and primary school age. These people live close to the centres of small towns or, in London, in areas developed prior to 1914.

O. Liberal Opinions

They are young, professional, well educated people, cosmopolitan in their tastes, liberal in their views, who enjoy the vibrancy and diversity of inner city living. These neighbourhoods also contain a high proportion of the country's students living in term-time accommodation, whether in halls of residence or shared accommodation.

Therefore, when conducting learning based on data, it is possible to make hypotheses based on demographics, lifestyles and behaviour.

Peak electricity demand

The UK's network energy system is designed to meet peaks in demand for electricity and guarantee energy for any scenario. Currently, the demand peaks in the late afternoon (between 4.00pm and 8.00pm), and the whole electrical distribution system is designed around this single peak [15]. However, there are two challenges that can change this situation: decarbonisation, and energy security. There is uncertainty in serving peak demand under this logic (network design around one single peak) and at the same time having no constraints in consumers energy consumption because the extent that networks grant unlimited energy resources is unknown. Managing power and its peak demand is important to guarantee affordable energy provision and discovering the potential of renewable sources [4].

One of today's energy challenges is in fact flattening the demand load curve. It has been argued that this could be achieved starting firstly by enabling an understanding of how load is constituted and the ways in which electricity use may be flexible. The study of the CLNR Dataset through an examination of the ways in which electricity is being used in domestic and industrial environments could help to understand the extent that electricity use might be flexible. In order to achieve such a challenge like that it is necessary to develop new forms of network control and flexibility.

According to the CLNR data there is a peak in demand between 4.00pm and 8.00pm. That is the time in the United Kingdom when a socio-materially structured reproduction of practices occurs. It is this performing of a practice that matters and affects directly the peak demand. The time when an activity is performed also matters in contributing to the peak demand. However, disruptions have revealed that there is a flexible side in habits and routines that were imagined as stable, continuous and persistent. That is the reason why there is interest in investigating peak flexibility with trials such as the CLNR that gave detailed insight as an output through four data types: Power systems monitoring allowed to have Network Performance Data and 12,000 electricity customers provided Consumption Data (Smart Meters) whilst social technical surveys and face to face visits to domestic customers and small organisations collected qualitative data about current and emerging practices.

2.1.4 Conclusion

Smart Meters are devices of great importance due to the fact that allow to acknowledge and explore energy consumption in a household. The CLNR is one of the biggest Smart Meter Trials worldwide and consisted in a four years project where the output was detailed insight about energy consumption at domestic and business environments. Having energy consumption data can support and benefit the managing of power and its peak demand that can lead to guarantee affordable energy provision.

2.2 Simulators

A simulation is an approximate imitation of the operation of a process or system [16]. There are many types of simulation, the focus of this work will be computer simulation. The purpose of computer simulation is being a solution for the representation and analysis of complex systems. When using computer simulation the system under study is often a complex nonlinear system for which simple, intuitive analytic solutions are not readily available. Rather than deriving a mathematical analytic solution to the problem, experimentation with the model is accomplished by adjusting the parameters of the system in the computer, and studying the differences in the outcome of the experiments. In consequence of this operation theories of the model can be derived, or deduced from these computational experiments.

2.2.1 Agent-Based Simulation

Agent Based Simulation Modelling is a type of computational model and social simulation which can be achieved with the implementation of an Agent-Based Model (ABM). An ABM is a computer simulation that involves multiple entities (the agents) which interact between themselves based on their programmed behaviour [17]. Agents can be used to represent organisations, animals, people, houses, bank accounts, etc. ABM are useful and can help to better understand the behaviour of a system when the behaviour of the individual components is known (and vice-versa), and to discover the behaviour of the individual components and how they explain the overall system behaviour.

Today's organisations accumulate large amounts of data about their processes in databases. Agent based modelling is an effective way to put data to work and an agent based simulation model can be fed by real, personalised, properties and behaviours taken from databases. By simulating the actions and interactions of autonomous agents a model can be generated as an output which can help to capture the details of the dynamics of a system.

2.2.2 An Agent-Based and Data Driven Simulator Based on Smart Meter Data

Having an agent-based simulator driven by Smart Meter Data can provide a computational laboratory environment where new DSM strategies can be tested out. Something key to understand is the concept of the Smart Meter, that is, the designated usage of these devices to provide to the user, with almost near real time information, data about energy use, so that they can make decisions regarding how they manage the use of their electronic devices with the hope of reducing their energy demand, to save money and also to reduce emissions [5].

2.2.3 Conclusion

A simulation can provide us with a model of a process or system. A type of simulation is Computer Simulation. Rather than deriving simulation from a mathematical analytic solution, computer simulation bases its behaviour on experimentation and the adjusting of parameters of the system. Agent Based Simulation is a type of Computer Simulation and it allow us to represent entities of interest to help us understand the behaviour of a system or its individual components. Moreover, an Agent Based Simulator can be fed by Smart Meter Data and work as a virtual lab where DSM strategies can be tested.

2.3 Related work

In the energy and environmental research fields, many agent-based models and data-driven approaches have been proposed. As an example of a data-driven approach Dent suggests that the overall network can be benefited by implanting DSM programmes that can transform the existing consumer behaviour. According to Dent identifying and exploring repeated patterns of behaviour within households is important to accomplish that objective, his work is aimed to demonstrate this by using an approach of clustering techniques and a database with electricity meter data from the North East Scotland Energy Monitoring Project (NESEMP) that collected data during a year at a 5 minutes resolution in parts of Scotland between 2011 and 2012 [18].

There are data driven approaches where the data used for simulation does not come from a database, it comes from a survey. Examples of those approaches are the ones developed by Johnson et al [19], Wang and Paranjape [20] and Muratori et al [21]. Johnson et al propose a simulation tool developed in MATLAB to estimate demand response. In this simulation tool the individual residential loads are modelled with the objective of understanding the impact of demand response strategies such as Direct Load Control (DLC) and Indirect Load Control (ILC). Johnson et al used the U.S. Census Bureau in the American Time Use Survey (ATUS) to model the occupant behaviour, and based on that build a model for the load distribution and finally the residential load [19]. Wang and Paranjape created an agent-based model to evaluate household energy management system and test demand response strategies and techniques by utilizing the output of the UK 2000 Time Use Survey, a comprehensive survey of how people spend their time in the UK [20]. This survey has a large-scale household scope that featured self-completing diaries where users had to input the amount of time spent in domestic activities, face to face interviews and over the phone. It was measured between June 2000 and September 2001 and allowed to give inputs to three mathematical models in this work: a Richardson Model, a Dynamic Price Model and a Home Energy Management System Model. The agent-based model consisted in a three layer system architecture (GUI, Communication and model). Wang and Paranjape concluded that their development can be a testbed to evaluate demand response strategies and in the simulation results it can be appreciated that after implementing the Energy Management System the peak demand, energy bills and generation cost are reduced. Muratori et al created a simulation of the power demand in a household with multiple individuals based on the 2003-2009 American Time Use Survey [21] where the data was collected from working and non-working males and females (18-85 years old) and children (15-17 years old). The purpose of this work is to simulate an average household in the US by using a regression model with 400 households and considering average power consumption data given by the 2012 American Appliance stock by the US Department of Energy. A Markov Chain was used to create and activity pattern of individuals and a physically-based model was used to simulate the appliances in the household. In this model the energy consumed in the household can be due to cold appliances (fridge and freezer), HVAC System, activities of the household members (cooking, use the dishwasher, watch tv), lighting and finally a concept to include others such as lights that are always ON and appliances' standby power, it is also considered that an occupant can be doing only one activity at the same time. This can be a tool to simulate the status quo of the residential sector, it is useful as well to evaluate new energy policies and technology. At the end the proposed model generates electricity demand profiles with the same statistical features as residential metered data.

Tian and Shang developed an agent-based household energy consumption model to capture the group habits of different stakeholders in order to assess the potential of clean energy promotion to contribute to better policy decisions and measures in China. This model was developed in Java with the Repast Symphony Platform [22] with the purpose of getting more insights of behaviour and group habits as means to increase clean energy utilisation [23]. The authors consider that it is worthy of simulation whether the increase of residents' income will increase the proportion of clean energy consumption. For the development of this agent-based model the hourly energy consumption of the devices was given by the "China Household Energy Consumption Report 2016". They introduce the concept of region, which corresponds geographically to an area where many households belong and share similar energy consumption habits and income. There are 5 types of agents: region (climate condition so heating requirements), household, operator, device, and fuel. They consider 35 kinds of devices in the categories of Cooking, Heating, Water Heating, Cooling, Appliance, Freezing and Clothes cleaning, and 10 kinds of fuels like electricity, gas and coal. A device ownership is associated with the income level of a household. Tian and Shian findings show indications of as living standards improve, residents will spontaneously move toward clean energy consumption.

Some other approaches are data-driven an implement an agent-based model, however the data used for the simulation is collected from a small number of households. Dhar et al model and predict the energy consumption of a house in India [24]. This simulation used as a dataset a collection of appliance level data at a 1Hz frequency from an Indian household in Gandhinagar, where it stored the usage time and frequency of each appliance. It is applied a NILM Algorithm to calculate electricity consumption profile of appliances. In this approach the agent is a house, this house has as attributes different appliances such as refrigerator, television, washing machine, water purifier, dishwasher, air conditioner and rice cooker. Weather variations are also considered.

Although there have been numerous achievements in multiple fields, only a few studies have analyzed household behavior and household energy consumption using a data-driven approach combined with an agent-based model.

2.4 Conclusion

There have been about a dozen significant smart meter trials worldwide, some in a bigger scale than others but all them with the same purpose, generate data at a specific interval rate about how electricity is used within a building, household, small to medium-sized enterprises (SMEs) or all of these categories together. The CLNR Project was one of those trials in the UK that covered all the categories mentioned, one of the greatest trials in households due to the involvement of fourteen thousand domestic customers which generated Smart Meters Data. It has been said that social practices shape electricity demand curves and that in order to reduce the energy demand it is important to have a clear understanding of how and why people use energy and the possibilities and limits regarding “behaviour change”. With the rich data available from the CLNR and the agent based technology it can be developed a high resolution simulator which can help to better understand how electricity is used within a home. This means that would be possible to have a virtual lab for testing demand-side management and new houses’ appliances.

Chapter 3

Methodology

In this Chapter the methodology approach to create the simulator is discussed and methods for data selection are described. Methods for data selection describe how the energy profile was obtained for each appliance in a household. Moreover it is explained how the likelihood of usage was calculated. Finally, the technology and tools used for simulation are listed.

3.1 Methodology approach

The data used to help this research was taken from the CLNR project, and only the data related to Test Cell 2 was considered. This collected data included enhanced profiling data of 280 domestic smart meter customers, which can be accessible to the public and can be downloaded from `http://www.networkrevolution.co.uk/resources/project-data`. With the collected data there were calculated likelihood values, ***likelihood*** is defined as the result of a function where the number of successes and the number of tries of previous experiments are given. The value for a likelihood can be from 0 to 1. The aim of calculating the likelihood is looking for the probability of a success given previous data from observed results. A clear example of the calculation of the likelihood is when tossing a coin ten times and trying get 'tails' for each result, but being successful with only seven. Given that seven successes were observed in ten tries then the likelihood would be 0.7.

Despite there being nine test cells in the CLNR Project, only one was selected for this research due to the specific interest in exploring energy consumption in domestic houses and because test cell 2 contained detailed energy consumption data in an appliance level for every household in the experiment. The CLNR Data contains years of measurements of appliances' energy consumption at a 1 minute resolution. Using this data it was possible to generate multiple likelihood values for every appliance in every household, so that it was possible to say how likely it is for an appliance in a household to be either ON, OFF or on a *stand by* status at different times during the day.

Based on the CLNR Data it was also possible to calculate an average energy consumption for every appliance in a given household. The data was extracted and managed with queries and stored procedures in SQL and processed with Python. Once the likelihood values and averages of energy consumption were calculated for each and every appliance it subsequently was possible to use that generated data as an input for simulation. An Agent-Based System (ABS) was created using Python and a framework named MESA. The ABS was created in a way that every appliance was represented by an Agent and with every step in time this agent could either go ON, OFF, or remain in a *stand by* status. The agent would consume energy from the network if the current state was ON or in *stand by* (also if the *stand by* energy value attributed to this appliance is greater than zero, which could be the case of only certain appliances, such as a microwave, due to consuming energy not only when it is being used but also by showing the time on a screen, for instance). These changes made by every agent were dictated by the data stored in both the likelihood files and energy consumption profiles.

The energy consumption for the energy network is analysed and compared against the real data from the CLNR after the simulation of an average day of the year and after running experiments. This comparison is done through the plotting of the energy consumed during the day in a line graph and visually analysing the pattern and similarity of the plotted lines. Using these line graphs reveal trends and progress over time.

3.2 Methods of data selection.

There is data about energy profiles for an average appliance available online such as the ones listed in the Household Electricity Use Survey in the UK [25], however the CLNR project has provided real data for energy consumption of different appliances in real households which is going to be used to generate a more realistic energy profile per appliance per household. This data that the CLNR project has provided has to be selected and processed to create the new real energy profiles and according to the times of use calculate values that represent how likely is for an appliance to be turned ON, OFF or remain in a STAND BY during the day.

3.2.1 Energy consumption calculation.

Energy profiles were created for every physical appliance in a household based on the energy consumption data in TC2. It can be appreciated in Eq. 3.2.1a and Eq. 3.2.1b that average of energy consumption was calculated for every appliance considering only the values where consumption is greater than the *stand by* value calculated for that appliance. It was important to remove the zero values from being considered in the average so that the result represents only the average consumption for when the appliance is being used. For *tc2amicro* it was necessary to make an extra operation due to energy value captured in the unit Wh, considering that, and the resolution of 1 minute, it was possible to convert this to the unit kW. The output for this operation resulted in one of the main input files used during simulation.

Getting energy profile from dataset *tc2amicro*:

$$Input\_file_1^j = (loc\_id_j, appliance_j, max\_energy\_kw_j, avg\_energy\_kw_j)_j \tag{3.2.1a}$$

$$avg\_energy\_kw_1^j = \left(\frac{\sum_{i=1}^{m} \frac{energy\_wh_i}{1000 \cdot \left(\frac{1}{60}\right)}}{m} \right)_j \tag{3.2.1b}$$

$$max\_energy\_kw_1^j = \left(\frac{\max\{energy\_wh_1, ..., energy\_wh_i\}}{1000 \cdot \left(\frac{1}{60}\right)} \right)_j \tag{3.2.1c}$$

Where:

j is the number of unique physical appliances distributed among different households.

m is the number of energy measurements greater than the *stand by* value registered by the unique appliance.

n is the number of measurements equal to zero for a unique appliance.

i is the total of measurements for a unique appliance in a household.

SQL Query:

```
SELECT tm.loc_id,
       tm.meas_desc AS appliance_name,
       AVG(NULLIF(tm.energy/(1000*(1/60)), 0)) as energy_kw,
       MAX(tm.energy/(1000*(1/60)))
  FROM tc2amicro tm
 INNER JOIN customers_res cr
    ON tm.loc_id = cr.loc_id
 INNER JOIN channels ch
    ON tm.meas_desc = ch.meas_desc
 WHERE cr.tc_id = '2a' AND
       tm.date_time >= cr.date_start AND
       tm.date_time <= cr.date_end AND
       tm.energy/(1000*(1/60)) >= ch.stand_by_kw
 GROUP BY tm.meas_desc,
```

```
        tm.loc_id;
```

Getting energy profile from dataset *tc2apassiv*:

$$Input\_file_1^j = (loc\_id_j, appliance_j, max\_energy\_kw_j, avg\_energy\_kw_j)_j \tag{3.2.2a}$$

$$avg\_energy\_kw_1^j = \left(\frac{\sum_{i=1}^{m} energy\_kw_i}{m}\right)_j \tag{3.2.2b}$$

$$max\_energy\_kw_1^j = (\max\{energy\_kw_1, ..., energy\_kw_i\})_j \tag{3.2.2c}$$

Where:

j is the number of unique physical appliances distributed among different households.

m is the number of energy measurements greater than the *stand by* value registered by the unique appliance.

n is the number of measurements equal to zero for a unique appliance.

i is the total of measurements for a unique appliance in a household.

SQL Query:

```
SELECT tm.loc_id,
       tm.meas_desc AS appliance_name,
       AVG(NULLIF(tm.energy, 0)) AS energy_kw,
       MAX(tm.energy)
  FROM tc2apassiv tm
 INNER JOIN customers_res cr
    ON tm.loc_id = cr.loc_id
 INNER JOIN channels ch
    ON tm.meas_desc = ch.meas_desc
 WHERE cr.tc_id = '2a' AND
       tm.date_time >= cr.date_start AND
       tm.date_time <= cr.date_end AND
       tm.energy >= ch.stand_by_kw
 GROUP BY tm.meas_desc,
          tm.loc_id;
```

3.2.2 Likelihood Functions.

Likelihood values for every physical appliance were calculated, stored in a file and used during simulation. The purpose of calculating these likelihood values is being able to decide if an appliance should go ON or OFF, to consume electricity or remain in a *stand by* state during the simulation. These values represent how likely it is for an appliance to be turned ON or OFF during the day, and furthermore a likelihood value is being calculated for every minute of the day starting at minute 00:00 and ending at minute 23:59. It is in consequence of this that there are 1440 likelihood values attributed to every appliance in the simulation.

The following Structured Query Language (SQL) Query extracts the data from the SQL tables and calculates the value of how likely is for an appliance to be ON or OFF in a specific minute of the day. There is an explanation as well for every field that the SQL Query is retrieving.

Fields retrieved by query:

- *loc_id*: It is the unique identifier for the household.
- *appliance_name*: The name of the appliance which can be any of the channels listed in Appendix. B.
- *time*: The minute of the day in format HH:mm.
- *stand_by_kw*: It is the energy consumed by an appliance when it is not being used but still connected to electricity.
- *date_start*: Points out when the energy readings started and from where the data is valid.
- *date_end*: Points out the date of the ending of energy readings.
- *readings*: It is the number of valid number of readings of an appliance filtered by an specific minute of the day.
- *greater_than_stand_by_value*: Number of readings where the energy consumed in kw was greater than the calculated *stand by* value.
- *below_stand_by_value*: Number of reading where the energy consumed in kW was below the *stand by* value limit.
- *equal_to_zero*: Number of energy reading where the energy consumed by an

appliance was equal to zero.

- *likelihood_on*: This value is calculated by the division of the number of readings that are greater than the *stand by* value between the total number of readings.

$$likelihood\_on = \frac{greater\_than\_stand\_by\_value}{readings} \quad (3.2.3a)$$

- *likelihood_off*: This value is calculated by the division of the number of readings that are less than the *stand by* value value between the total number of readings.

$$likelihood\_off = \frac{below\_stand\_by\_value}{readings} \quad (3.2.4a)$$

SQL Query:

```
SELECT m.loc_id,
       m.meas_desc AS appliance_name,
       DATE_FORMAT(date_time, '%H:%i') AS time,
       COALESCE(ch.stand_by_kw, 0) AS stand_by_kw,
       date_start,
       date_end,
       COUNT(TIME(date_time)) AS readings,
       SUM(CASE WHEN m.energy > COALESCE(ch.stand_by_kw, 0)
                THEN 1 ELSE 0 END) AS greater_than_stand_by_value,
       SUM(CASE WHEN m.energy <= COALESCE(ch.stand_by_kw, 0)
                THEN 1 ELSE 0 END) AS below_stand_by_value,
       SUM(CASE WHEN m.energy = 0
                THEN 1 ELSE 0 END) AS equal_to_zero,
       SUM(CASE WHEN m.energy > COALESCE(ch.stand_by_kw, 0)
                THEN 1 ELSE 0 END)/COUNT(TIME(m.date_time))
                AS likelihood_on,
       SUM(CASE WHEN m.energy <= COALESCE(ch.stand_by_kw, 0)
                THEN 1 ELSE 0 END)/COUNT(TIME(m.date_time))
                AS likelihood_off
  FROM tc2apassiv m
 INNER JOIN customers_res cr
    ON cr.loc_id = m.loc_id
```

```
INNER JOIN channels ch
   ON m.meas_desc = ch.meas_desc
WHERE cr.tc_id = '2a' AND
      m.date_time >= cr.date_start AND
      m.date_time <= cr.date_end
GROUP BY m.loc_id,
         m.meas_desc,
         just_time;
```

The following charts shown in Fig. 3.1 and Fig. 3.2 are examples of the likelihood distribution of some appliances which were plotted using the 1440 likelihood values calculated for every minute of the day. In these charts it is displayed the household 10017 from dataset *tc2amicro*, the reason to show this household is that it illustrates the likelihood really well due to having more appliances than most of the households. The household 10017 reported energy consumption data from eight appliances: two electric heaters, lighting, two immersion heaters, shower and cooker.

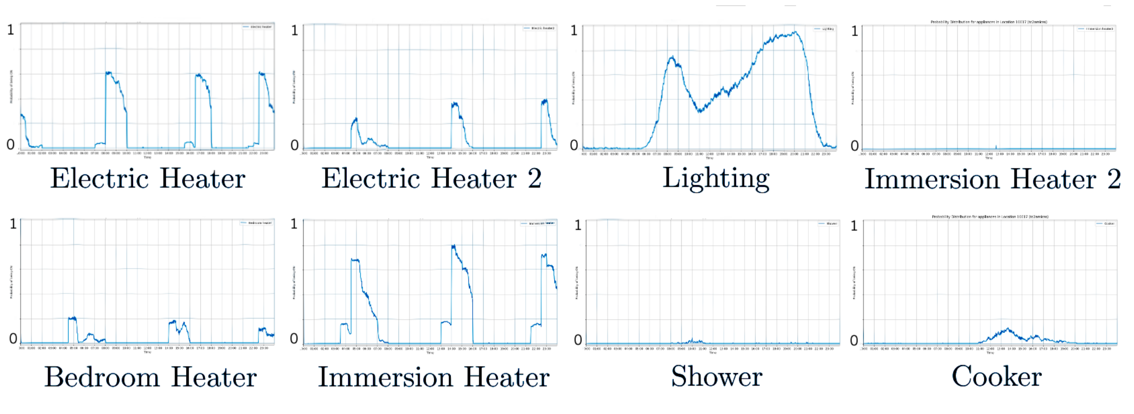

Figure 3.1: Likelihood Distribution Functions for appliances in household 10017.

In Fig. 3.1 it can be appreciated one chart for each of these appliances, the axis x represents the time of the day and the axis y the likelihood which can go from zero to one. It can be seen that appliances related to heating (electric heater, electric heater 2, immersion heater and bedroom heater) share a similar behaviour, they have a constant high likelihood of being turned ON during a period of 2-4 hours of the day, then it stops and repeats after 5-6 hours. The lighting is likely to be turned ON when people are awake (from 6.00am to 11.00pm) but with a high likelihood

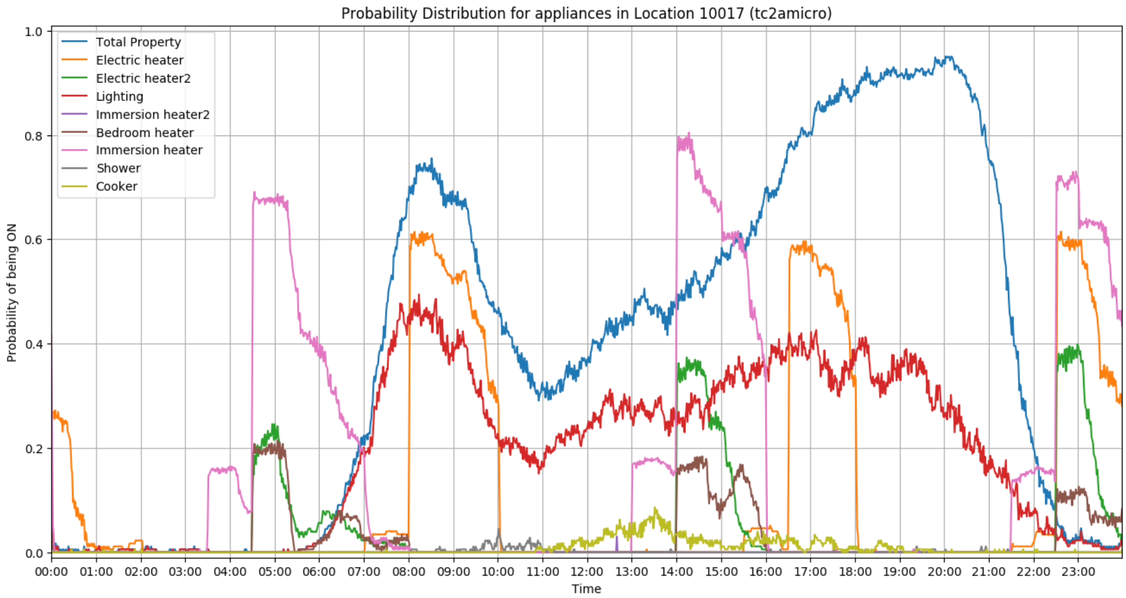

Figure 3.2: Likelihood Distribution Functions for appliances in household 10017.

between 7.00am and 9.00am and between 5.00pm and 9.00pm. The shower is more likely to be turned ON between 8.00am and 11.00am. Cooking is an activity that is very likely to be done between 11.00am and 8.00pm, showing a high likelihood between 1.00pm and 2.00pm. All these likelihood values were calculated from real energy consumption data in household 10017 and are considered in the simulation of every appliance of the same house.

3.3 Enabling Technologies

3.3.1 Python

This project has been developed with Python in its version 3.5.2. Python is a clear and powerful object-oriented programming language which has become an increasingly popular language for exploratory, interactive, and computation-driven scientific research [26]. Some notable features of Python are its independence of the platform (runs anywhere, including Mac OS X, Windows, Linux, and Unix), its scalability (is easily extended by adding new modules implemented in a compiled language such as C or C++) its extensive standard library which provides thousands of components that can be used within python scripts and lastly Python has a wide community support.

3.3.2 MESA

MESA [27] is a useful framework for agent-based modelling in Python. It allows the user to create, analyse and visualise agent-based models (ABM).

MESA's architecture was designed with modularity. This was achieved by separating the functionality of MESA into independent, interchangeable modules that can also work together. The modules are grouped into three categories.

1. Modelling: Modules used to build the models themselves.
   - Model: The *Model* class is the core of the simulation. It holds the model-level attributes, manages the agents, and generally handles the global level of the model. It also contains a scheduler to handle time (what order the agents act in), and a space for the agents to inhabit and move through.
   - Agent: The *Agent* class provides a mechanism to define the agent behaviour in a simulated model. It can also contain logic related to where the agent is placed in space.
   - Scheduler: The *scheduler* is a special model component which controls the order in which agents are activated in.
   - Space: It refers to where the agents are situated and where they perform their actions, which is defined by means of a grid with coordinates (x,y).
2. Analysis: Tools to collect data generated from the model, or to run it multiple times with different parameter values.
3. Visualisation: Classes to create and launch an interactive model visualisation, using a server with a JavaScript interface.

MESA is the main python package that was used to create the simulation of the appliances in the DDDS which is explained in Chapter 5.

3.3.3 MariaDB

MariaDB [28] is an open-source Relational Database Management System (RDBMS) made by the original developers of MySQL [29]. It was created from a version of MySQL and as a replacement for it. MariaDB provides an SQL interface for accessing data and it is also used because it is fast, scalable and robust, with a rich ecosystem of storage engines, plugins and many other tools make it very versatile for a wide variety of use cases.

MariaDB is the RDBMS that was used to store the CLNR Data. The extraction of data and some operations with it were performed with SQL stored procedures and SQL queries to the MariaDB database.

3.4 Conclusion

In this Chapter the methodology approach to create the simulator was discussed and methods for data selection were described. The DDDS is now ready thanks to the implementation of an ABS in Python, supported by the framework MESA. Energy profiles were created, likelihood values per appliance usage were calculated and they are the core of the DDDS.

Chapter 4

CLNR Data Pre-processing

In this Chapter it is explained how the CLNR data was processed in order to be used as an input for simulation. Firstly, datasets *tc2amicro* and *tc2apassiv* are explored, it will be noted that there are significant differences between these two datasets and it is not possible to match and analyse the data as it is. Secondly, efforts to create a merging strategy for these two datasets are described in order to deal with data and appliances labels.

4.1 CLNR Datasets

The CLNR data has been processed so far by exploring Test Cell 2 which corresponds to the enhanced profiling of domestic smart meter customers. There is a total of 280 locations in Test Cell 2. Every location in Test Cell 2 represents a household. The data in Test Cell 2 is accessible from two tables in SQL, *tc2amicro* and *tc2apassiv.* Table *tc2amicro* contains information about 199 locations. Figure 4.1 shows how some locations appear in both tables *tc2amicro* and *tc2apassiv.*

4.1.1 Table: *tc2amicro*

The structure of table *tc2amicro* can be seen in Fig. 4.2. This table contains data of 168 locations and its set of appliances. Every record in this table represent a measurement of energy consumed by one of the forty-four possible appliances in a household.

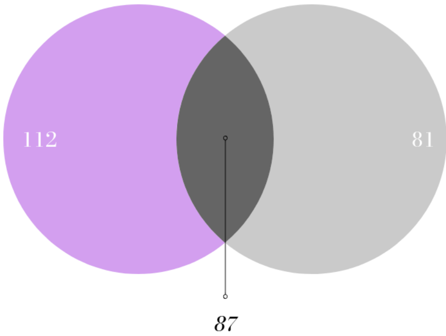

Figure 4.1: Intersection of locations between *tc2amicro* and *tc2apassiv*. There is a total of 280 unique locations.

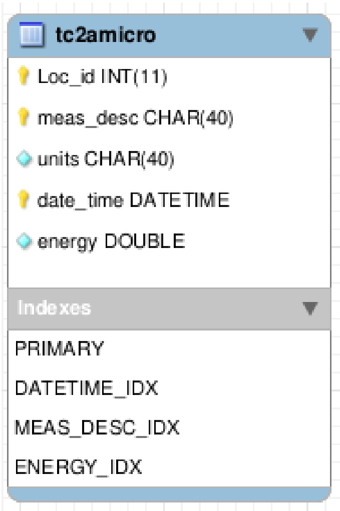

Figure 4.2: Structure of table *tc2amicro*.

4.1.2 Table: *tc2apassiv*

This table contains thirty-six channels, which refers to thirty-six appliances measured among almost two hundred locations. The structure of table *tc2apassiv* can be seen in Fig. 4.3.

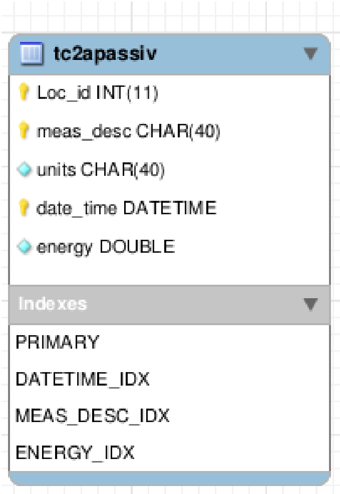

Figure 4.3: Structure of table *tc2apassiv*.

4.2 Consumers behaviour during night using table *tc2amicro*.

As a way to validate the data from the CLNR and have a better understanding of it here is a comparison against the data in the Electrical Load Measurements dataset collected by the Renewable Energy Feed-in Tariff (REFIT) project team. In [1] it is presented as a chart that shows the power demand for a specific house during the night of April 22nd, 2015 (Figure 4.4). In consequence to this it has been plotted with the same scenario using the same date of that year with the CLNR data.

There are 168 households in dataset *tc2amicro*, only 41 of those households have records for the evening of April 22nd, 2013. As shown in Table 4.1, these 41 households have a number of channels (appliances) between one and seven. It can be seen in Fig. 4.6 and Fig. 4.5, the two households with more channels among the 41 of interest, the switching on and off events of appliances during a night in spring. Figure 4.5 shows a sharp peak at 6.53pm, this is due to an error in the CLNR data.

Figure 4.6 shows seven channels (appliances). Three channels are related to lighting (Cellar lights in blue, Downstairs lights in orange and Upstairs lights in pink), two of them are about showering (Shower in red, Shower2 in purple), one channel is related to the freezer (Freezer in green) and lastly there is the Total Property chan-

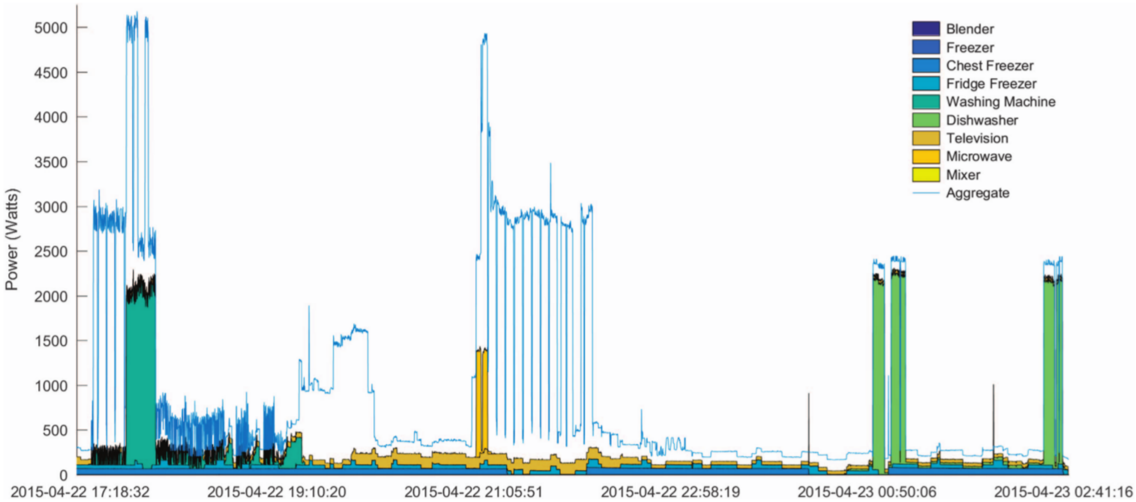

Figure 4.4: Power demand for House 10 during the evening of April 22nd 2015 (from [1]).

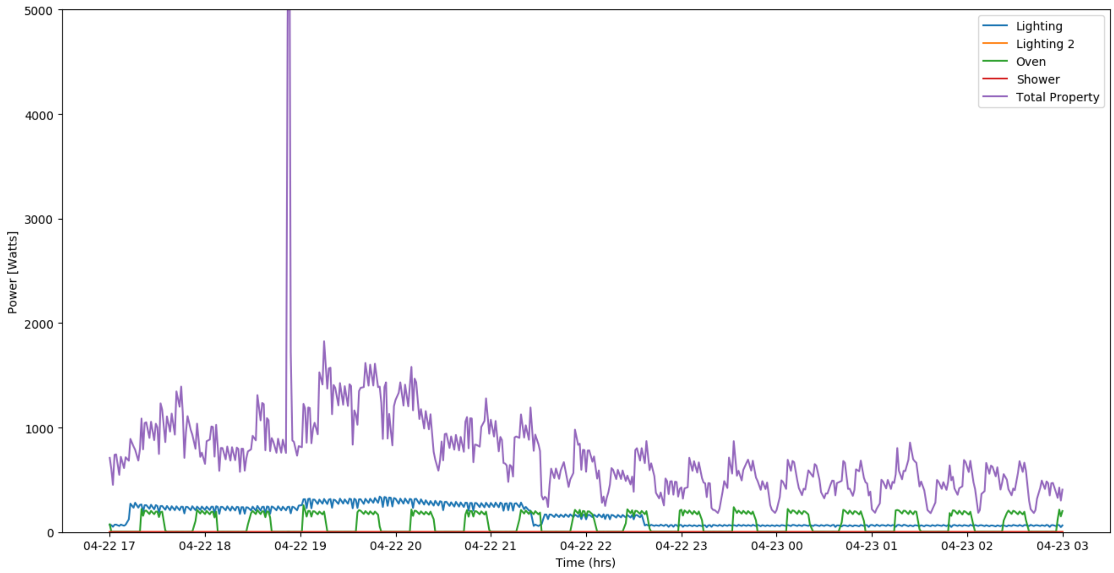

Figure 4.5: Power demand for household 10028 during the evening of April 22nd 2013.

nel which is supposed to measure all the energy consumed in a household (brown). According to the chart, cellar lights were turned on for some minutes around 8.00pm and after 10.00pm, downstairs lights were turned off from 5.30pm and went on before 8.00pm. The downstairs lights were used actively between 8.00pm and 11.00pm and kept at a low energy consumption profile afterwards (below 2kW). The upstairs lights remained turned off most of the time during the observation period, and were only turned on five times for a few minutes, and no energy consumption was registered because of a shower (Shower and Shower2 channels). The Freezer shows a normal and constant behaviour in energy consumption, it goes on and off around every 40 minutes. There is only a different behaviour between 8.30pm and 9.30pm, the freezer went on and off regularly during that period which might mean it was in use by people in the household. It is despite of having a Total Property Channel, it can be seen in the chart that the amount of energy consumed by the household does not represent the aggregation of the energy consumed by all the tracked channels. Thus, there is a difference of energy being used which is not being considered as part of any channel.

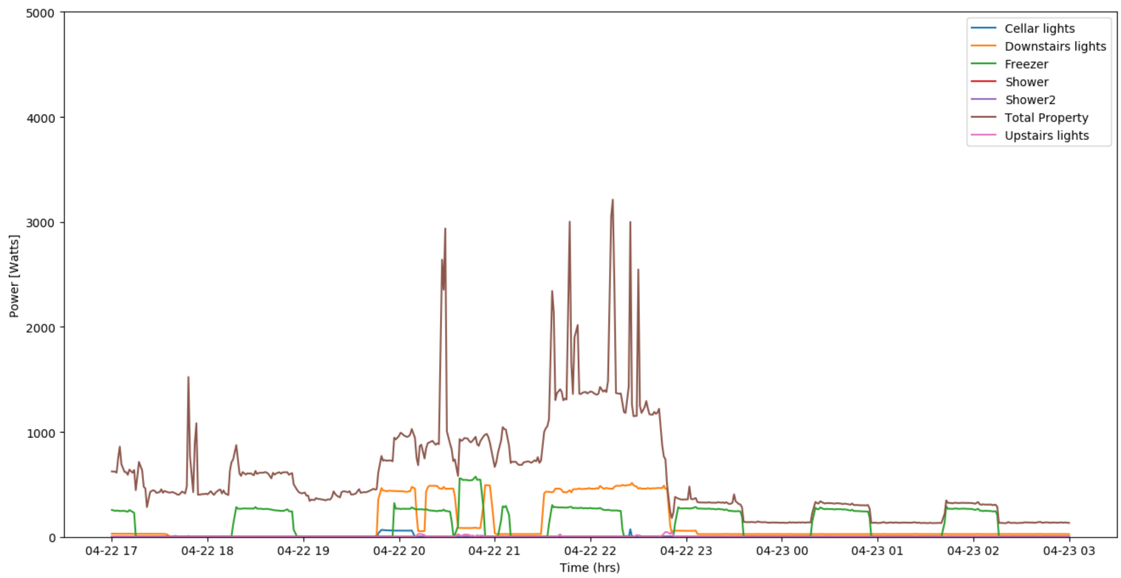

Figure 4.6: Power demand for household 10059 during the evening of April 22nd 2013.

Figure 4.7 shows the behaviour of only three channels among the seven which had been tracked. Between 5.00pm and 3.00am there was no energy consumed by the specific channels of the Cooker, Immersion heater, Shower and Shower2. Lights were turned on between 7.30pm and 11.00pm.

The comparison between REFIT and CLNR data helps to understand that the energy consumption peaks in the evening and although the activities in these house-

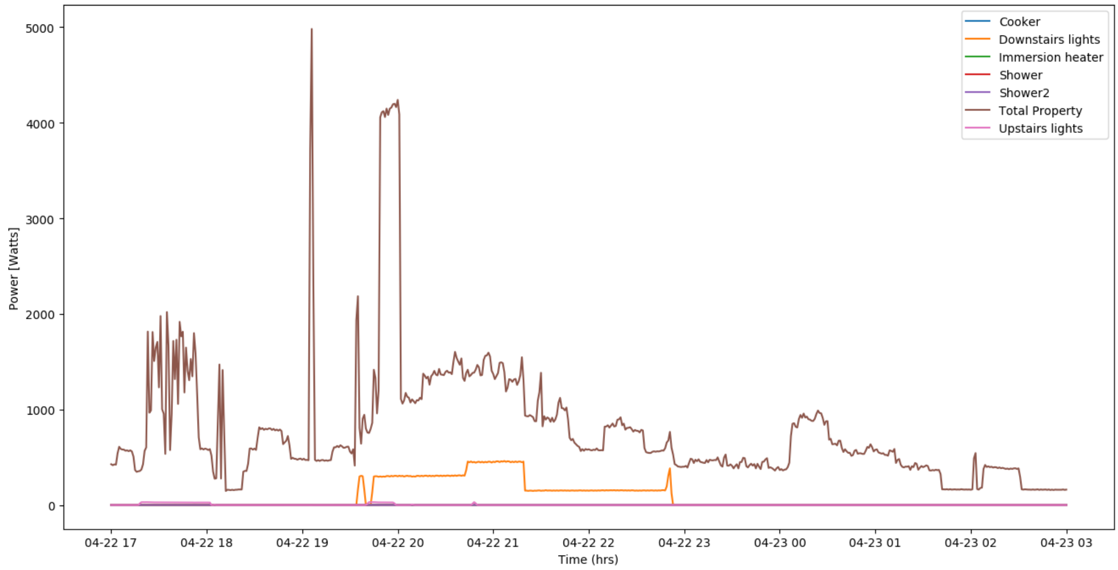

Figure 4.7: Power demand for household 10242 during the evening of April 22nd 2013.

holds are different they consume a similar amount of energy, reaching at some points of the night till 5000 Watts of power demand. The fact that during night the houses in REFIT and CLNR reach similar demand for power gives reliability to the measurements given by the smart meters in the households involved in the CLNR Project.

4.3 Consumers behaviour during night using table *tc2apassiv*.

There are 199 households in the dataset *tc2apassiv*, only 106 of those households have records for the evening of April 22nd, 2013. As shown in Table 4.1, these 106 households have a number of channels (appliances) between one and seven. It can be seen in Fig. 4.8 and Fig. 4.9 the switching on and off events of appliances during a night in spring in two households with seven channels each (among the 106 households) of interest.

Figure 4.8 shows cooking activities that involve the use kettle, microwave, dishwasher

Locations	Number of channels
10059, 10242	7
10157, 10091, 10105, 10234	6
10252, 10257, 10156, 10028, 10181, 10049, 10216, 10121,	5
10251, 10153, 45653, 10039, 10223	4
10011, 10023, 10164, 10165, 10173, 10225	3
7310, 10132, 10137, 10026, 45596, 45617, 10021, 10047	2
10158, 10030, 10032, 10268, 10192, 10064, 10074, 10228	1

Table 4.1: Channels per location.

power and even the main freezer power. There are also three more channels in the chart that show the energy consumption of the main television, the tumble dryer and the washing machine. The microwave was turned on for some minutes around 6.00pm and the kettle was used twice around 8.00pm and 10.30pm, dishes were washed in the dishwasher in two periods from 10.30pm to 11.20pm and from 12.10am to 12.20am. The main freezer power remained constant during the observation period, showing an energy consumption of less than 0.2kW. The main television power shows energy consumption during all the observation period but there is an increase between 8.00pm and 10.00pm. The washing machine was working from 8.00pm to 10.00pm. The tumble dryer was not used during the observation period.

Figure 4.9 includes the behaviour of seven channels (appliances) during a night between 5.00pm and 3.00am. During that period the tumble dryer, dishwasher and washing machine were not used. The microwave and kettle were used around 6.00pm, the microwave was used twice and the kettle once, although the kettle was used only once during a few minutes it consumed almost two times the microwave energy. It can be seen that the fridge and freezer power peaks for 1.5 hours and then goes off for 2 hours; it is only around 10.30pm when an unexpected peak appears might due to frequent door-opening by people in the household, the amount of energy consumed by the fridge and freezer is two times the amount of energy consumed when the fridge compressor is actually running. The Main TV was turned on from

5.40pm to 10.20pm showing a very stable energy consumption during those near-to five hours.

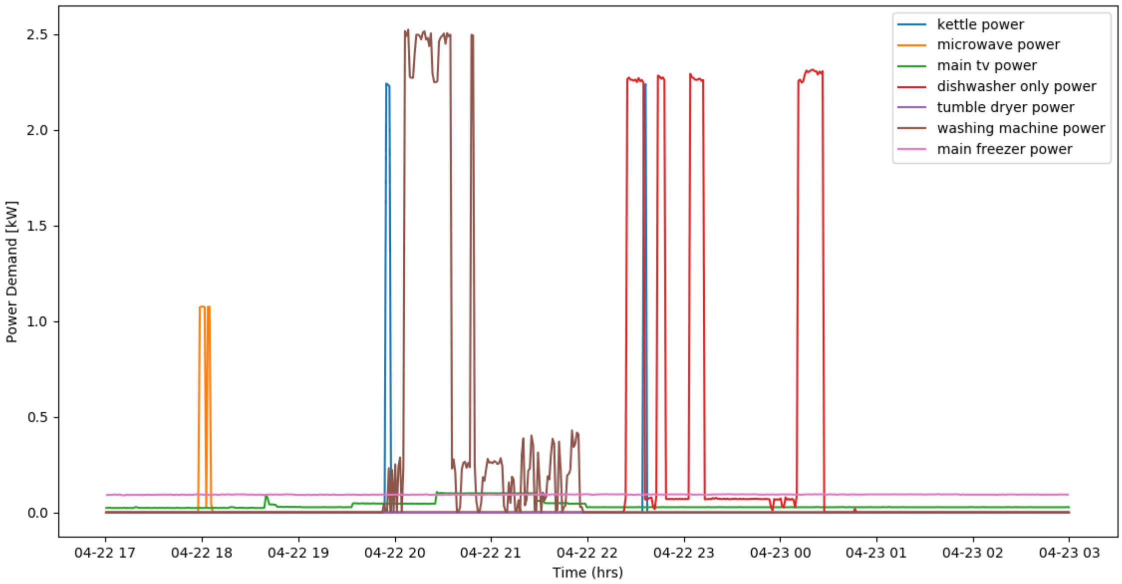

Figure 4.8: Power demand for household 46653 during the evening of April 22nd 2013.

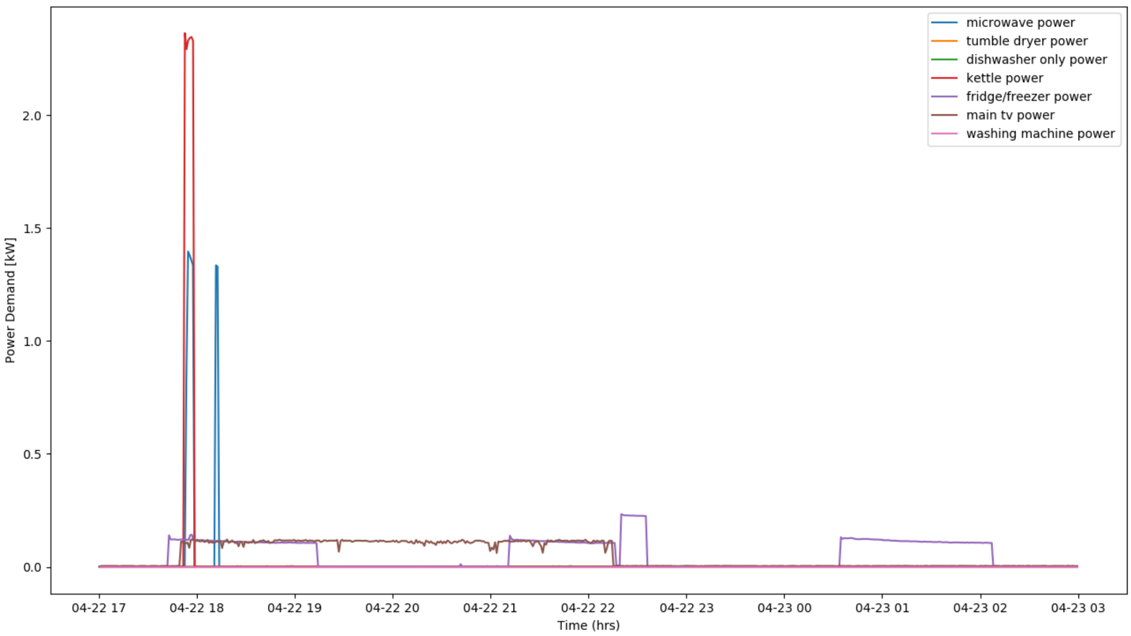

Figure 4.9: Power demand for household 10189 during the evening of April 22nd 2013.

4.4 Channels Merge Process.

The data necessary to analyse domestic energy use belongs to dataset *tc2amicro* and *tc2apassiv*. Every record in each dataset is associated with a channel, where a channel can be understood as a measurement type for the energy that is being used, i.e. the bathroom heater channel in dataset *tc2amicro* or the water heater channel in dataset *tc2apassiv*. Despite some coincidences in the labelling of channels between the two mentioned datasets, they have completely different channels because they were designed by different companies. Dataset *tc2amicro* has 44 channels whilst *tc2apassviv* has 36. There are channels in dataset *tc2amicro* that are not considered in *tc2apassiv* and viceversa. For instance, the appliances kettle, microwave and dishwasher are not part of *tc2amicro* whilst hall, kitchen and cellar lights are not part of *tc2apassiv*. There are some equivalences between channels in the datasets but still, it is not possible to match and analyse the data like that, it is necessary to create a mechanism to merge the data at least logically.

4.4.1 Dataset: *tc2amicro*

It contains data from 168 different households. Not all the households in dataset *tc2amicro* have the same channels, even the amount of channels being measured differ from one house to another one. An example of that can be seen between locations 10196 and 10017, both locations have the same number of channels which is nine, but only location 10196 has information regarding the fridge energy consumption (channel:“fridge") and only location 10017 has information about the heating in the bedroom (channel: “Bedroom heater"). Thus, it is not possible to match and compare the same parameters individually between households.

Table 4.2 shows the representation in number between households and channels. It can be seen that the greatest number of channels presented in a house is ten whilst the total number of channels in the entire dataset is forty-four. The statistical mode has three channels with almost the 20 percent of households having only one channel (which in most cases corresponds to the “Total Property” channel, a channel that represents the entire energy consumption in a household).

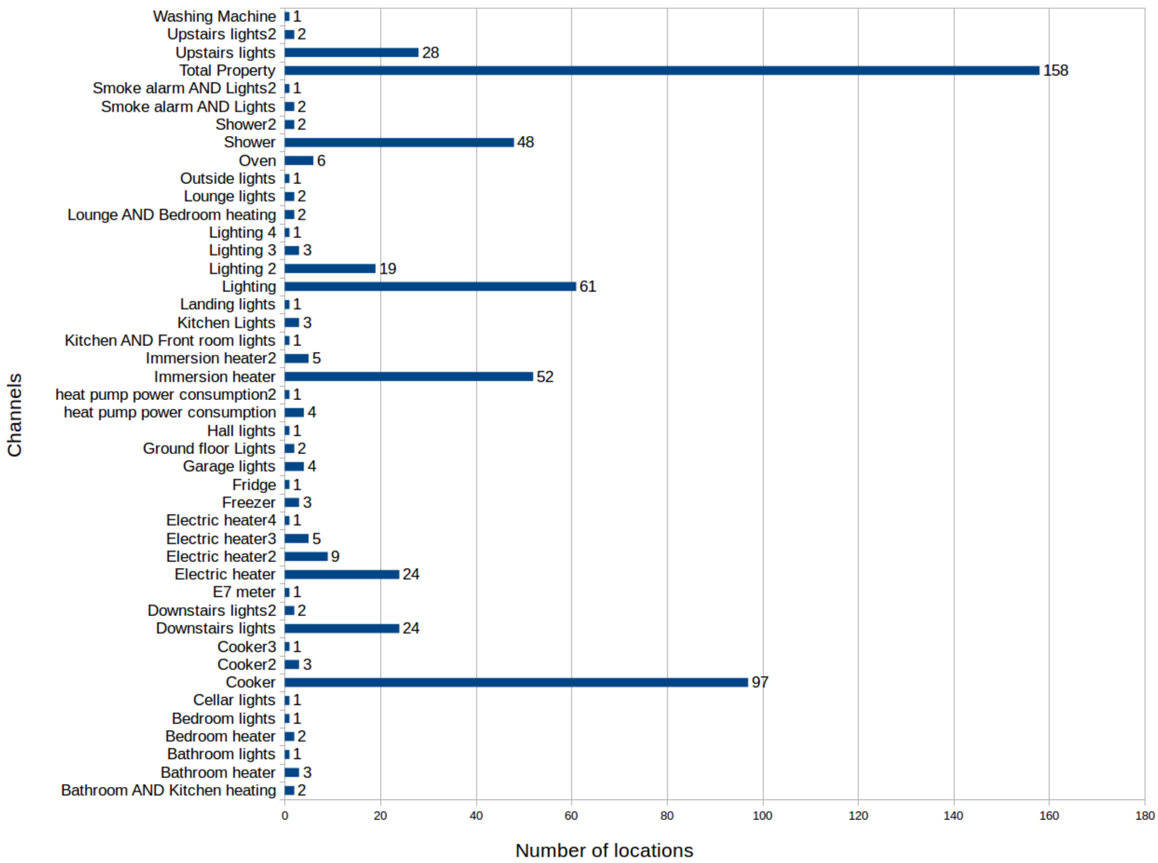

Figure 4.10: Channels representation in trialled locations (Dataset: *tc2amicro*).

Figure 4.2 shows a list of all the appliances in dataset *tc2amicro* and the amount of locations in which every appliance is being measured. The most popular appliance between households is the cooker, with 97 locations measuring the energy consumption for this appliance which corresponds to the 48.74 percent of houses in dataset *tc2amicro*.

4.4.2 Dataset: *tc2apassiv*

It contains data from 199 different households. Not all the households in dataset *tc2apassiv* have the same channels, even the amount of channels being measured differ from one house to another. An example of this can be seen between locations 10216 and 10107, both locations have the same number of channels which is two, but whilst location 10216 has information regarding kettle power and microwave power

Number of channels	Number of houses	Percentage
10	1	0.59%
9	2	1.19%
8	1	0.59%
7	7	4.16%
6	16	9.52%
5	26	15.47%
4	24	14.28%
3	35	20.83%
2	24	14.28%
1	32	19.04%
	168 houses	100%

Table 4.2: Appliances representation in households in dataset *tc2amicro*.

consumption (channels: “kettle power” and “microwave power") location 10107 has information about unknown appliances power consumption (channels: “Appliance 1 power consumption" and “Appliance 2 power consumption"). Thus, it is not possible to match and compare the same parameters (channels) individually between households.

Number of channels	Number of houses	Percentage
7	13	6.53%
6	23	11.55%
5	42	21.10%
4	63	31.65%
3	34	17.08%
2	10	5.02%
1	14	7.03%
	199 houses	100%

Table 4.3: Appliances representation in households in dataset *tc2apassiv*.

Table 4.3 shows the representation in number between households and channels.

It can be seen that the greatest number of channels presented in a house is seven whilst the total number of channels in the entire dataset is thirty-six. The statistical mode is having four channels per household and 7 percent of households only have one channel (which in most cases corresponds to the "solar power" channel).

Figure 4.11 shows a list of all the appliances in dataset *tc2apassiv* and the amount of locations in which every appliance is being measured. It can be noted that the most popular appliances being measured at houses are the kettle and microwave.

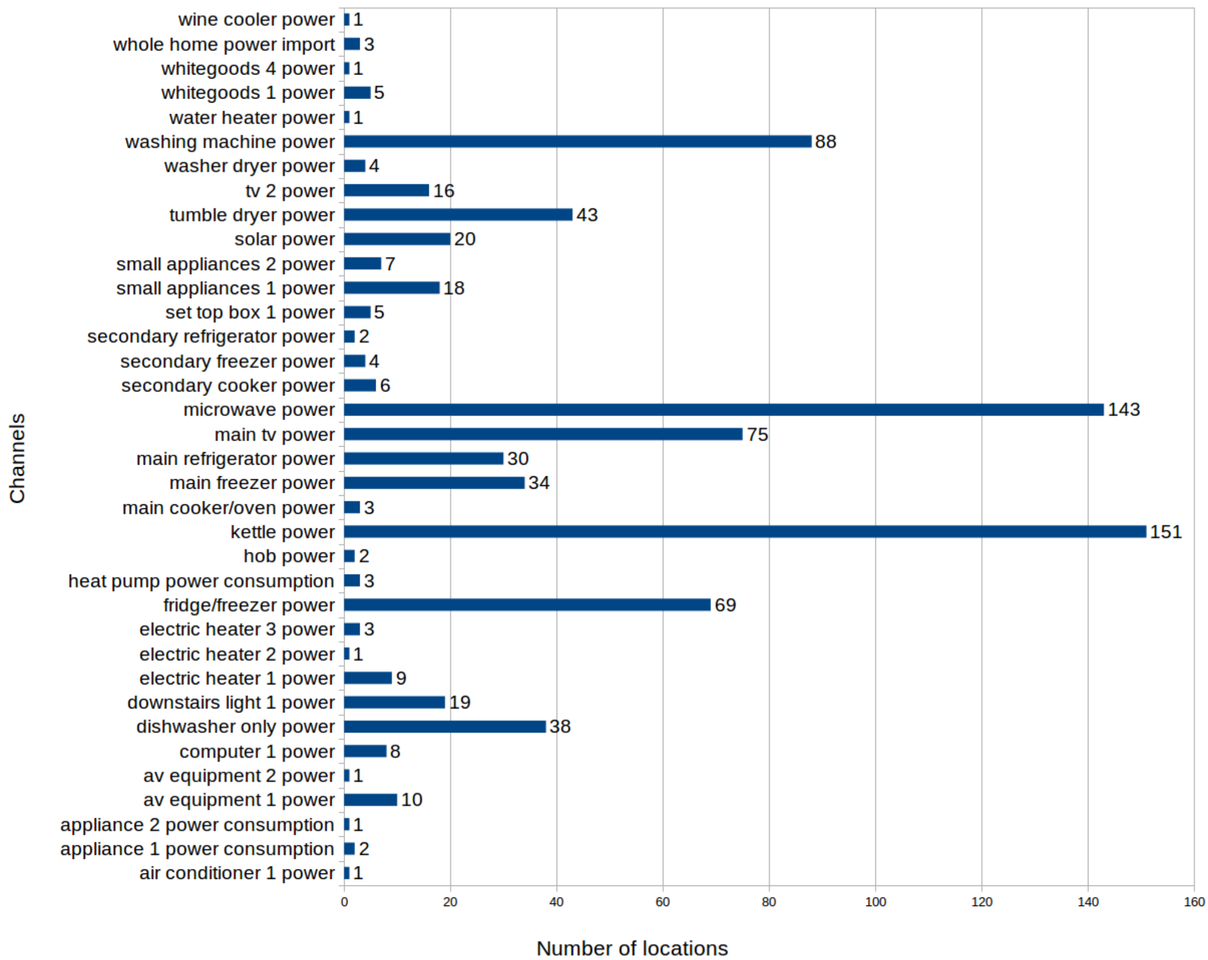

Figure 4.11: Channels representation in trialled locations (Dataset: *tc2apassiv*).

Dataset: *tc2amicro*	Dataset: *tc2apassiv*
Freezer	Main freezer power
Cooker	Main cooker/over power
Downstairs lights	Downstairs light 1 power
Electric heater	Electric heater 1 power

Table 4.4: Examples of equivalences between channels.

4.5 Channels Categorisation

The goal of this section is creating an abstraction tree diagram for the appliance labels in dataset *tc2amicro* and *tc2apassiv* so that the processing of both datasets can be done as one. Due to these datasets using a different set of channels to refer to the appliances in a household and also due to the disparity between households and which appliances were being measured in each of these, it was necessary to design a strategy to be able to evaluate the existing data about energy consumption as it came from only one source. This strategy would help to see the data in different abstraction levels. Analysing information in different abstraction levels allows to go from a generic perspective about energy consumption in a household to particular matters. A generic perspective could be the total energy consumption in a household, and a less generic perspective case could be the energy consumption only by cooking, and a particular perspective could be the energy consumption by a kettle. It is important to have a categorisation system that allows you to manage the data in datasets *tc2amicro* and *tc2apassiv* as one data source besides exploring the energy consumption by channels, its behaviour among the trialled locations, and the different levels of abstraction.

The next algorithm states the steps followed to create the abstraction tree:

1. Get a list of all the channels in datasets *tc2amicro* and *tc2apassiv*.
   Both lists can be seen in the Appendix section B.

2. Find equivalences between channels in datasets *tc2amicro* and *tc2apassiv*.
   Some equivalences between channels can be seen in Table 4.4.

3. Begin the creation of the tree diagram starting by adding its root: the Domestic House node. Here, the parent node is a Domestic House and the leaf nodes represent a channel or appliance which belongs to either *tc2amicro* or *tc2apassiv*. Intermediate nodes are abstract categories created to contain subtrees or groups of channels that have something in common i.e. the Node “Bathroom” which comes directly from the parent node, has as direct nodes the abstract category “Immersion Heater” and “Shower”, along with the channels “Bathroom lights” and “Bathroom Heater”. The aim is to place all the channels below categories.

The complete result of the categorisation process can be appreciated in Fig. 4.12. Due to the number of categories in the complete diagram it may be difficult to appreciate it. It is easier to see this diagram through the individual eleven diagrams that represent the main categories that are direct nodes from the root node “Domestic House” (see Fig. C.2, Fig. C.3, Fig. C.4, Fig. C.5, Fig. C.6, Fig. C.6, Fig. C.7, Fig. C.8, Fig. C.9, Fig. C.10, Fig. C.11 and Fig. C.12 in the Appendix C.1).

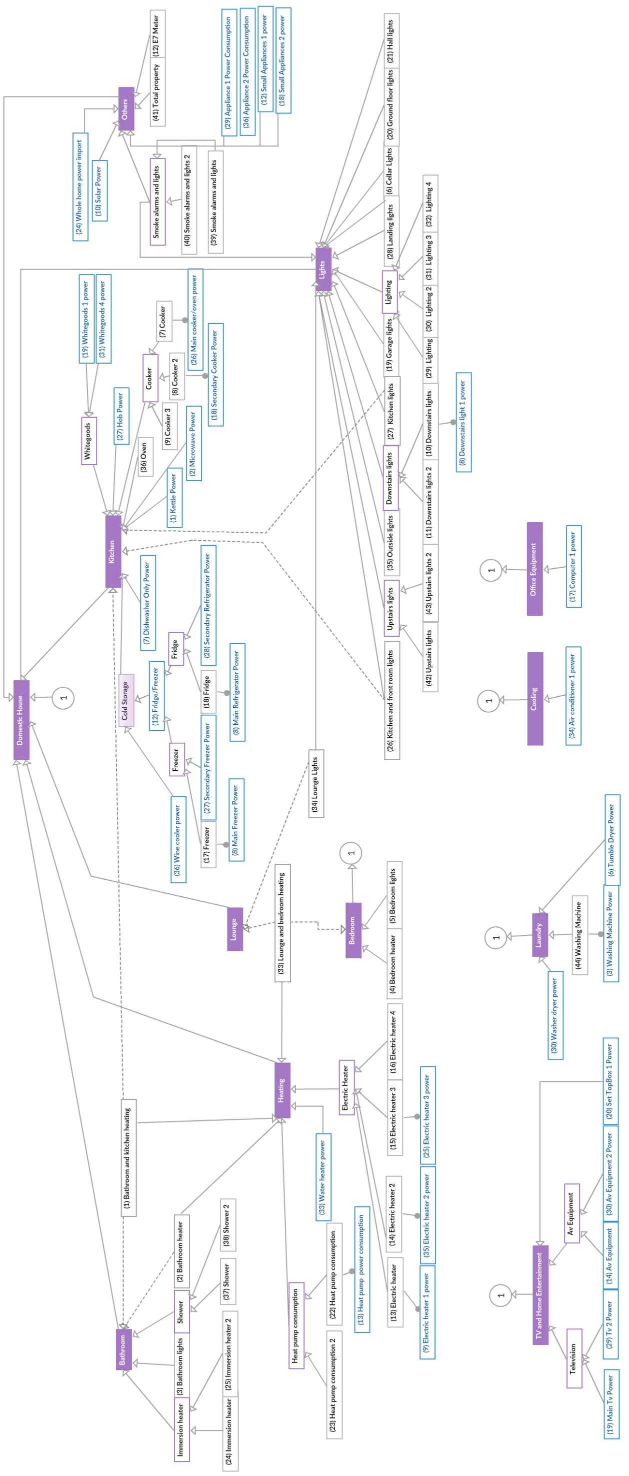

Figure 4.12: Full view of abstraction tree diagram.

Explanation of symbols in the tree diagram:

Nodes

- Filled purple node: These nodes are the main categories of the tree, together they represent the level zero and one in the tree diagram. There are eleven main categories that were created as a way of abstraction of the principal activities, actions or rooms in a house (see Appendix C.1).
- Light purple node: It is a middle category between a main category and another category. There is only one node of this kind in the tree diagram, it is the “Cold Storage” node whose direct parent is the “Kitchen” node.
- Straight purple line node: This node works as a combiner, it was created specially to group those appliances that are equivalents between them i.e. “Fridge” and “Main Refrigerator Power”.
- Straight blue line node: These nodes are the last ones in the tree, they have no children. There are thirty-six nodes of this kind and they represent the appliances names in dataset *tc2apassiv* which can be seen in Appendix B.2.
- Straight grey line node: These nodes are exactly the names of the appliance channels in dataset *tc2amicro* (listed in Appendix B.1). They can also be found as leaves in the tree and there are forty-four nodes of this kind.

Relations

- Solid line: Shows a strong relationship between a node and a direct parent node. A node could fit into more than one category but only one is considered as the best fit, and that category is considered as the strongest direct parent node. An example of this relationship can be seen in Fig. 4.13.

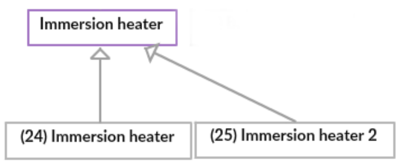

Figure 4.13: Showing parent node and two children.

- Dotted line: Shows a weak relationship between a node and a potential but not strongest parent node (see Fig. 4.14).

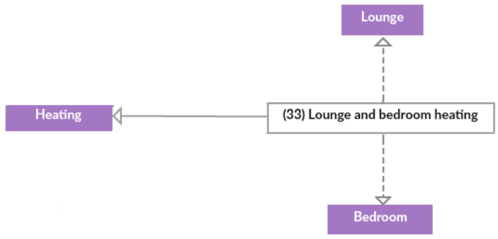

Figure 4.14: Example of relation with dotted line in abstraction tree diagram.

- Solid dot line: Shows an equivalence between channels from different datasets. It can be seen in Fig. 4.15 that channels *Cooker2* and *Secondary cooker power* belong to different datasets (*tc2amicro* channels are in grey and *tc2apassiv* are in blue).

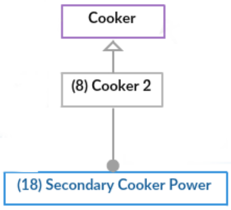

Figure 4.15: Equivalences between channels in datasets *tc2amicro* and *tc2apassiv*.

4.6 Conclusions

Through the creation of an abstraction tree diagram (which was later converted to a logical structure in a relational database) it is going to be possible to use dataset *tc2amicro* and *tc2apassiv* as if it were only one source of data and be able to filter information by appliance, without any problem due to the labelling of these.

Chapter 5

Simulator

In this Chapter it is described how the simulator was implemented. Firstly, an overview is presented along with a components diagram. Secondly, the model, agents and space where the agents interact are explained. Thirdly, configuration files needed for the simulation are described, and finally, the main functionality is presented.

5.1 Overview

The modelling of the domestic house environment was made based on a Diagram of Components as shown in Fig. 5.1. The main modules are: (1) the **agents**, which represent the appliances in a household; (2) the physical **space**, which corresponds to the households that contains a number of appliances; (3) a **model**, which initialises and controls simulation by following the defined parameters in the configuration files; (4) a **configuration** component, which helps setting up the energy profile and use likelihoods of the agents and finally (5) a **visualisation** component that allow us to appreciate the simulation in real time and the results generated from running it.

5.2 Model

The Model is the simulation core, most of the interactions are made through it. The fundamental task that the model is responsible for is the initialisation of the

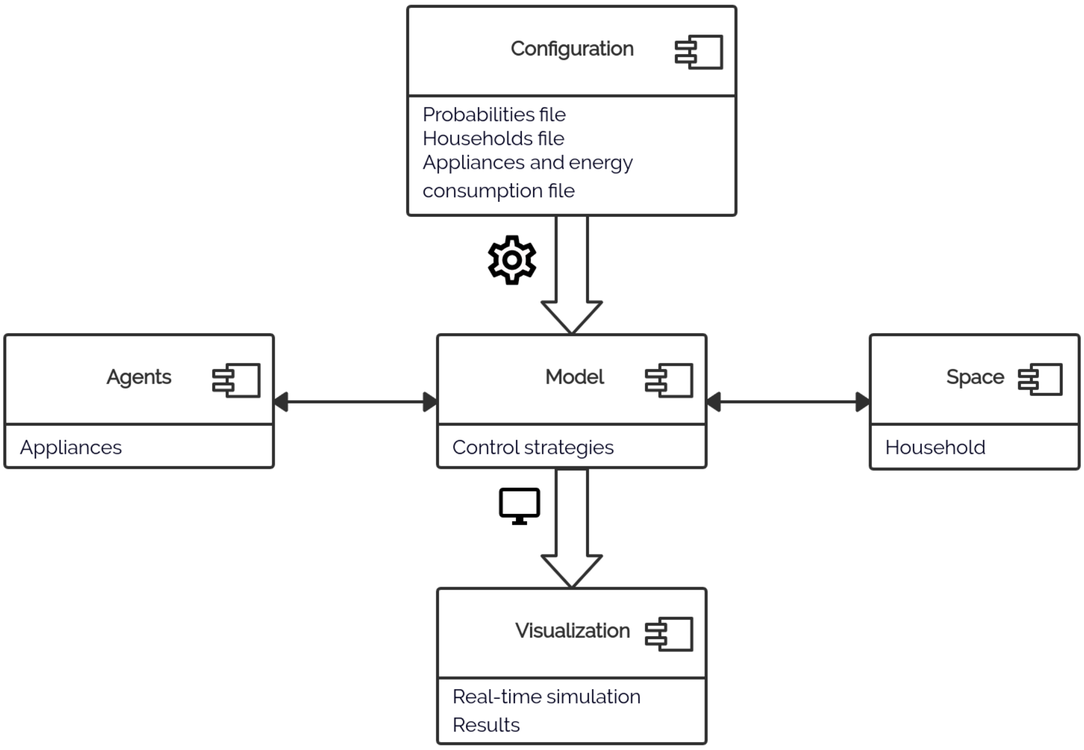

Figure 5.1: Components of Simulation Model.

simulation, this includes the space (households) and the 1145 agents which represent the appliances in different households such as lighting, cooking, watching television, showering, etc. Moreover, the model can compute the amount of energy that has been consumed during the day, calculate the energy demand for every time of the day, collection of data and even stop the simulation.

A timing module is used inside the model due to the fact that it is important to define the appliances' behaviour in accordance to the time. This module controls seconds, minutes and hours through the simulation.

5.3 Agents

The control of the simulation is made by the model but the simulation performance is responsibility of the agents. The agents represent the **appliances** which have been modelled using energy demand profiles and a state estimation based on likehihood

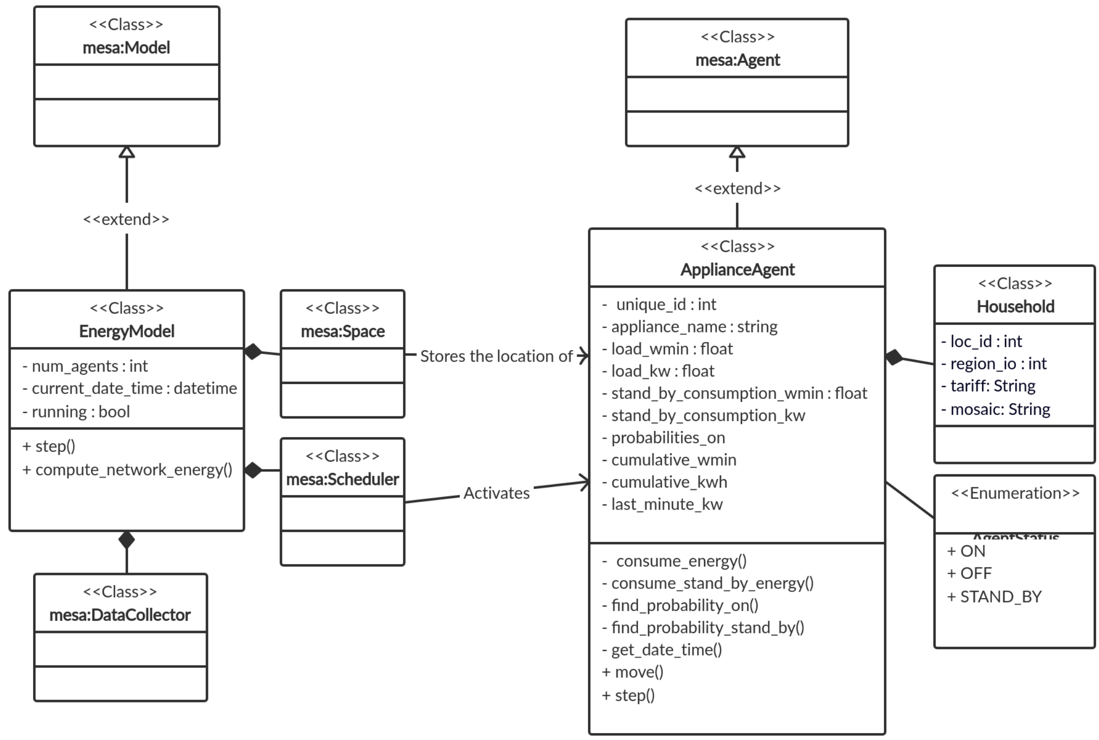

Figure 5.2: Simulator class diagram.

values. Each state represents an appliance's mode. An appliance can have three states: *ON*, *OFF* or *STAND BY*. The state ON refers to when an appliance is being used properly to a high potential whilst the state *STAND BY* refers to when an appliance is consuming energy but just a small amount due to the effort of having the appliance available to switch on very quickly or power a display. Some examples of *STAND BY* states in appliances are a small display in a microwave or oven that shows the time and the infrared light in a television that allows the remote control to turn it ON at any time. Other appliances that normally have a *stand by* state include dishwashers, washer machines, microwaves, anything with a AC/DC charger like laptops, printers and desktop computers.

Every appliance is given a unique ID to identify them during the simulation, it has also a name that can be any in the list of channels listed in Appendix B and an appliance belongs to one of the 280 simulated households. Regarding the energy profile, an appliance agent has four attributes that contain energy consumption data. The first two attributes are *load wmin* and *load kw*, they contain the average energy used to power the appliance when it has the state *ON*. The next two attributes are *stand by consumption wmin* and *stand by consumption kw*, they contain the average energy used by the appliance when it has the state *STAND BY*. The energy profile demand data has been stored considering the measurement units of kilowatts (kW) and also watts-minute (Wmin). The unit Wmin is being considered due to two facts: 1) The CLNR energy data was taken at a 1 minute resolution and 2) Every step that the simulator makes represents also one minute that has passed in time. A clear case where these values can be appreciated is with the appliance “tv 2 power” which has a *STAND BY* energy power of 0.056 kW or 56 Wmin of energy consumption. Having the data in the unit of Wmin allow us to think about energy that is being used constantly over a period of time that equals one minute, that is a more specific value than thinking about kilowatts hour (kWh).

An agent has also as an attribute the likelihoods of being ON during the day. These values are accessed during the simulation to get specific likelihood given a time. Each appliance agent has two accumulative attributes that keep track of how much energy has been consumed by it since the beginning of the simulation, these values are stored in the unit of kilowatt-hour and watt-minute.

Conversion from kW to Wmin

A watt minute is a unit of energy, just like the kWh or Wh, the difference is the period of time over which energy is being used in a constant rate. As pointed out in Eq. 5.3.1a, energy can be calculated if power and time is known. The unit for energy is kWh and having this value is possible to calculate Wmin. Some examples of appliances in the simulation and how load and energy are different for every state can be seen in Table 5.1.

$$energy = power \cdot time \tag{5.3.1a}$$

$$kWh = kW \cdot h \tag{5.3.1b}$$

$$Wmin = kWh \cdot tm \tag{5.3.1c}$$

Where:

tm is 60000, the number of minutes in an hour multiplied by a thousand.

0.056 kW over a 1 min period

$$kWh = 0.056kW \cdot \frac{1}{60}h \tag{5.3.2a}$$

$$Wmin = kWh \cdot tm \tag{5.3.2b}$$

$$Energy = 0.000933333kWh \cdot 60000 = 55.99998Wmin \tag{5.3.2c}$$

$$\tag{5.3.2d}$$

Where:

tm is 60000, the number of minutes in an hour multiplied by a thousand.

Agent	State	Power usage	Energy (Wmin)
Tv 2 power	ON	0.062 - 2.109 kW	62 - 2109 Wmin
	OFF	0 kW	0 Wmin
	STAND BY	0.056 kW	56 Wmin
Microwave power	ON	0.107 - 2.508 kW	107 Wmin - 2508 Wmin
	OFF	0 kW	0 Wmin
	STAND BY	0.104kW	104Wmin
Oven	ON	0.035 - 2.747 kW	35 - 2747 Wmin
	OFF	0 kW	0 Wmin
	STAND BY	0.00282 kW	2.82 Wmin

Table 5.1: Appliances states, power usage and energy consumption values per minute.

5.3.1 Step method

As can be seen in the class diagram in Fig. 5.2, the class *ApplianceAgent* has one public method named **step**. This method is run in each simulation step to make a change of state, it starts with the increment of the time by one minute. In the real world, with the time passing by during the day our activities also change and so the likeliness of performing them. In the simulator every step or minute that passes by changes the likelihood of an appliance being used, and with that change the appliance can be turned on or off, or even adopt a stand by state (see Fig. 5.3).

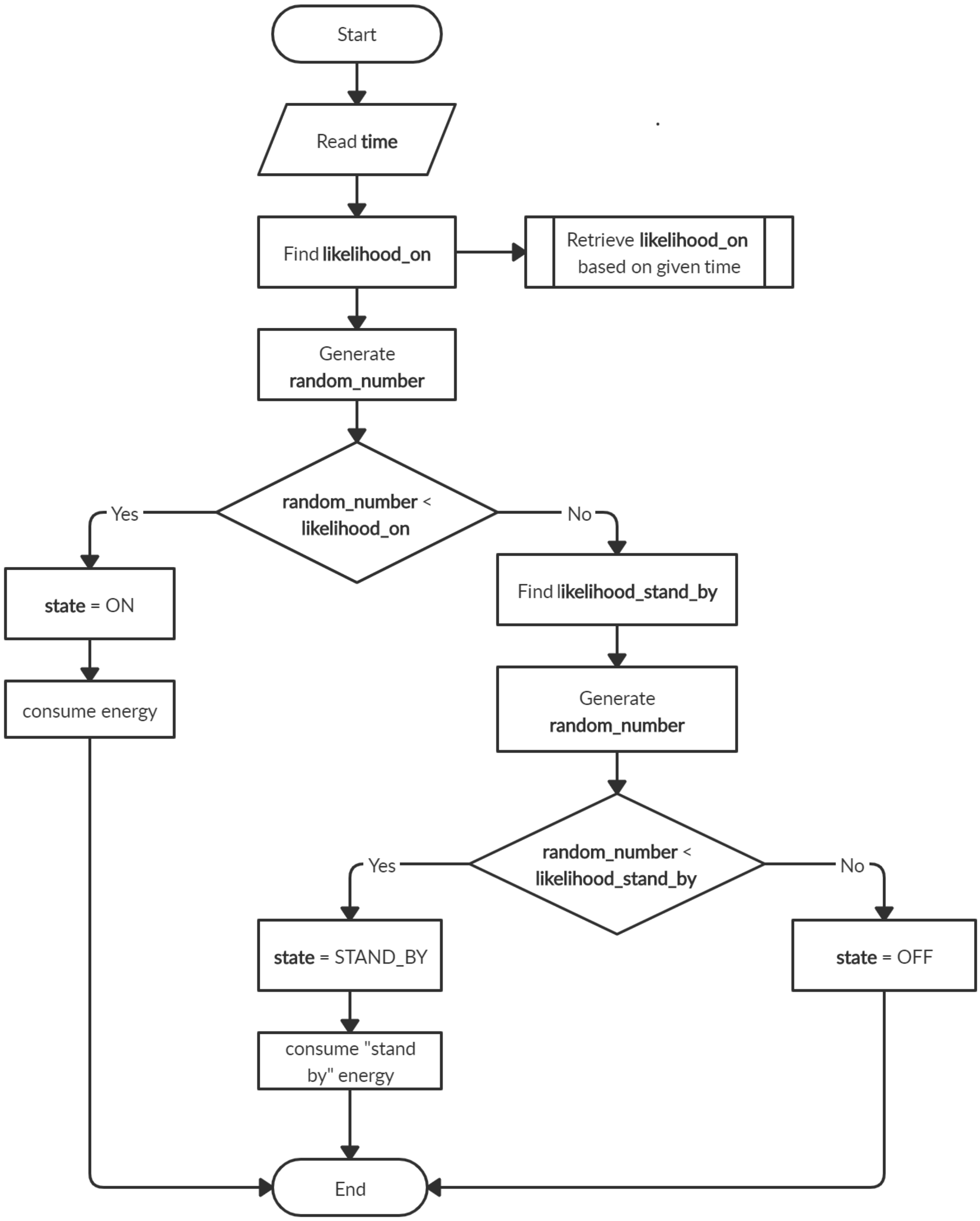

Figure 5.3: Flowchart of agent step method.

5.4 Configuration files

All the values mentioned previously are defined in **configuration files**. These configuration files enable the simulator to recognise the households and the appliances with their energy profile demand data along with their likelihood ratios for different times of the day.

The simulation is configured by four setting files. First, the *households configuration* file enables to load the information of identifier, region, tariff and mosaic class for every simulated house. Second, the *unique appliances* file is used to initialise each of the agents. Finally, the *likelihood appliances* files for *tc2amicro* and *tc2apassiv* enables to specify the likelihood ratios for every physical appliance during the day. The configuration files are given in a JavaScript Object Notation (JSON) which is an open-standard file format that allows to store data using a structure of a collection of name/value pairs and an ordered list of values [30]. The following is a more detailed explanation of the configuration files:

1. **Households**

   Using a JSON object for every household in the simulation it is given:
   (a) The unique identifier for the household.
   (b) The region of the household which indicates whether it is in or out of Northern Powergrid's region.
   (c) A description of the tariff type of the customer.
   (d) The assigned mosaic class, which is a geodemographic segmentation system for UK households designed by Experian.

```
[
    {
        "Loc_id": 1,
        "region_io": 1,
        "tariff": "Flat␣rate␣tariff",
        "mosaic": "I␣Ex-Council␣Community"
    }
    ...
]
```

2. **Unique appliances**

   These files define every unique appliance in the simulation that will be represented as an agent. For each unique appliance it is given a JSON with its unique identifier, its name, the household where it belongs, its energy consumption values in units of watt per-minute and the kilowatt for when the appliance is turned *ON* or on the *STAND BY* mode.

```
[
    {
        "unique_id": 1,
        "loc_id": 832,
        "meas_desc": "Immersion␣heater",
        "min_power_wh": 0.002,
        "avg_power_wh": 51.99445082,
        "max_power_wh": 58.17,
        "source": "tc2amicro",
        "Wmin": 3119.667174,
        "stand_by_consumption_kw": 0,
        "avg_power_kw": 3.119667174,
        "new_avg_power_wh": 51.9944529,
        "stand_by_consumption_wmin": 0
  }
  ...
]
```


3. **Likelihood appliances**

   A JSON is used for every minute of the day related with a unique appliance, every entry represented as a JSON contains the following:

   (a) The unique identifier of the household where the appliance belongs.
   (b) The name of the appliance.
   (c) The time of the day for which the likelihoods were calculated.
   (d) The date from where the readings used to calculate the likelihoods started.
   (e) The date from where the readings used to calculate the likelihoods ended.
   (f) The number of existing readings for the appliance that were captured during an specific minute of the day.
   (g) The number of readings at an specific minute of the day which value was below the defined stand by value.
   (h) The number of readings at an specific minute of the day which value was equal to zero.
   (i) The number of readings at an specific minute of the day which value was greater than the defined stand by value.
   (j) The likelihood that the appliance is not ON during an specific minute of the day.
   (k) The likelihood that the appliance is ON during an specific minute of the day.

(l) The likelihood of the appliance (once it is not ON) that it is completely off.

(m) The likelihood of the appliance (once it is not ON) that it is in a stand by mode.

(n) The kilowatts used by the appliance when it is in a stand by mode.


```
[
   {
      "loc_id": 10017,
      "meas_desc": "Bedroom␣heater",
      "just_time": "00:00",
      "date_start": "2013-05-02",
      "date_end": "2014-03-31",
      "readings": 178,
      "below_stand_by_value": 164,
      "equal_to_zero": 164,
      "greater_than_stand_by_value": 14,
      "likelihood_off": 0.9213,
      "likelihood_on": 0.0787,
      "if_not_on_then_prob_off": 1,
      "if_not_on_then_prob_stand_by": 0,
      "stand_by_kw": 0
    }
  ...
]
```


5.5 Visualisation

For evaluation, comprehension and debugging of the simulator it is possible to see a real-time visualisation of its running. There are two modes of running the simulator in the background, the first one only obtains the result data from the simulation in a csv format, and the second one generates charts for energy consumption in the network per household. The uniqueness of the third mode is due to the visualisation component that has been created, as can be seen in Fig. 5.4 and Fig. 5.5, it is suitable for analysis and improving the understanding of energy consumption in a network level and per appliance in a household.

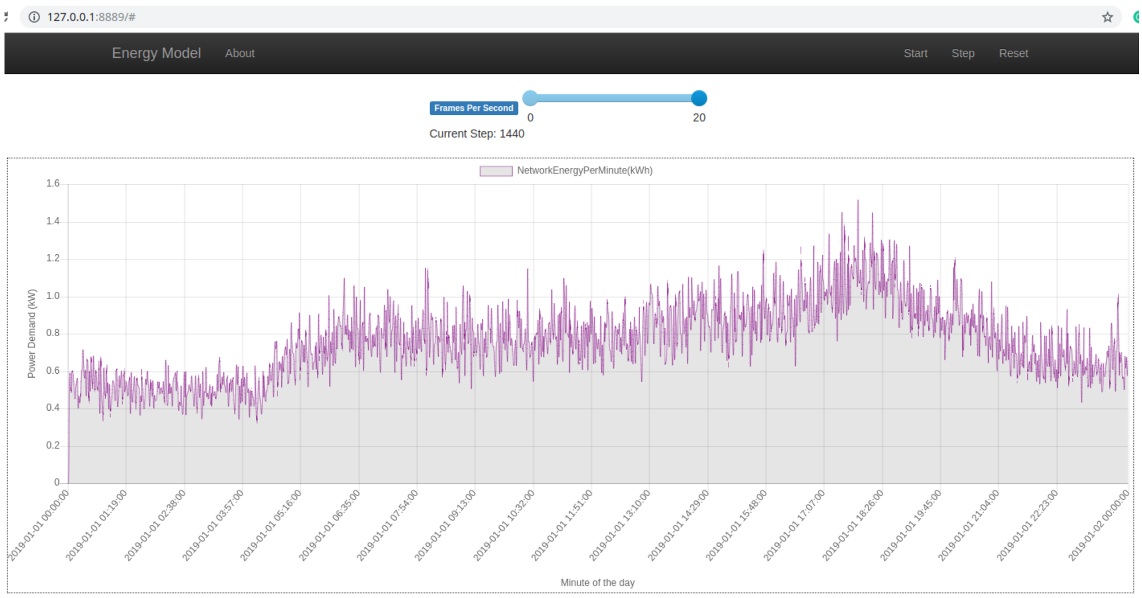

Figure 5.4: Real time simulation showing energy consumption in network.

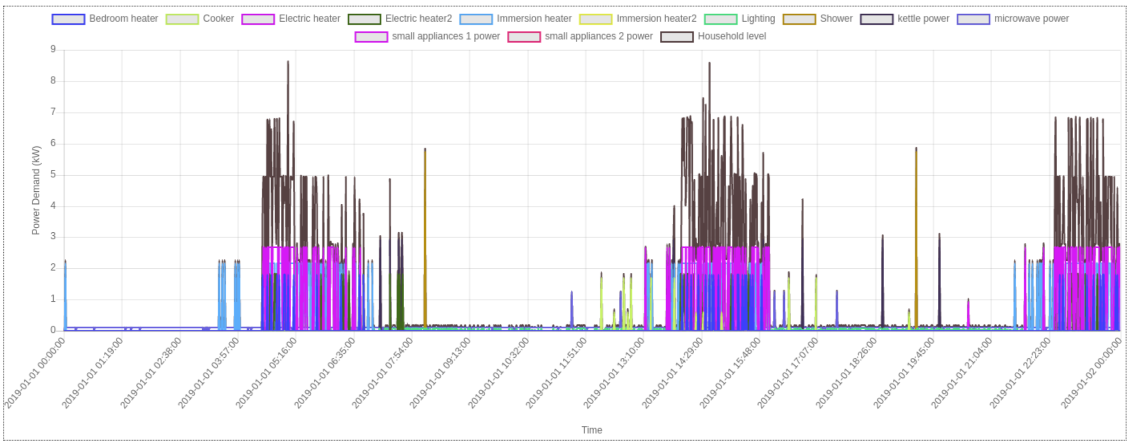

Figure 5.5: Real time simulation showing energy consumption in a household.

Chapter 6

Results

6.1 Simulation results

In Fig. 6.1 it can be appreciated the main results of the simulation: The simulator against the real data. The blue line shows the daily mean demand of a household, which was generated as an output of the simulation. The orange line shows the daily mean demand of a household which was plotted with real data taken from the CLNR project. We can see that the simulation plot is similar in shape to the real data from the CLNR. However, the power demand of simulation output and observation is different in quantitative terms. This is due to the fact that the number of times that an appliance is switched ON is not yet being controlled. The Likelihood for an appliance going ON at certain times during the day was calculated but that same likelihood is not changing once the appliance goes on, nor is the number of minutes that the appliance should remain on yet considered once it has been turned ON. This issue combined with a fixed *stand by* value for a type of appliance results in a considerable increase of power demand in a household.

Figure 6.2 shows the contrast regarding the daily mean demand for power in a household between the simulation, the real data from the CLNR Project (taken from test cell *2a*) and test cell *1a*. The test cell *1a* (green line) only gathered data of the total power import per house, thus it is more defined and represents a more realistic behaviour of the energy consumption. The simulation (blue line) and data gathered

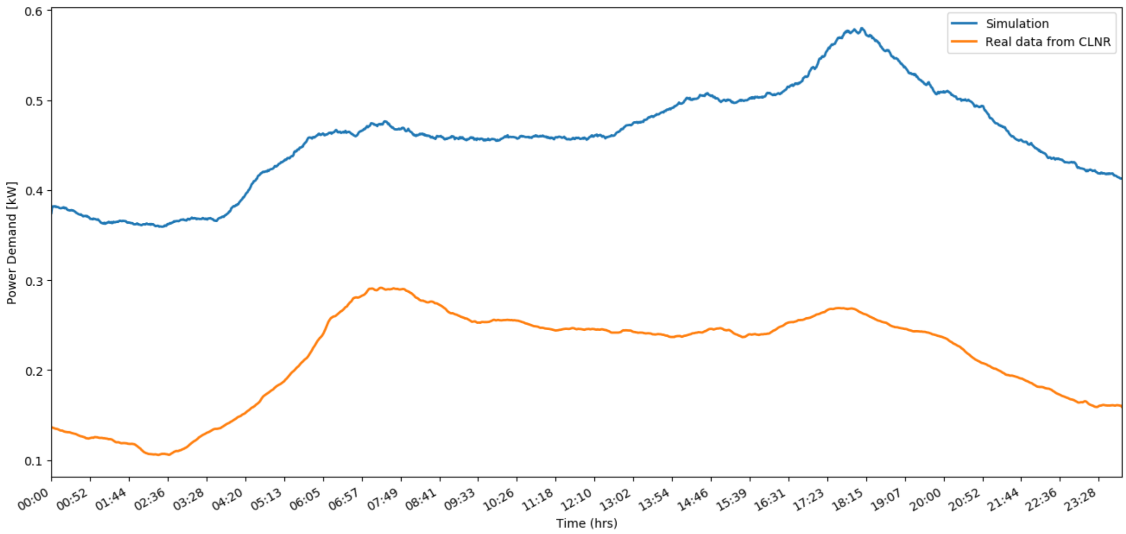

Figure 6.1: Daily mean demand comparison between Simulation and real data from the CLNR Project.

in test cell *2a* (orange line) was plotted considering the mean energy consumption per household, considering as total the sum of the individual power demand of only the set of appliances being measured, this means that the total was not absolute due to certain appliances not being measured in every household during the smart meter trial carried out by the CLNR.

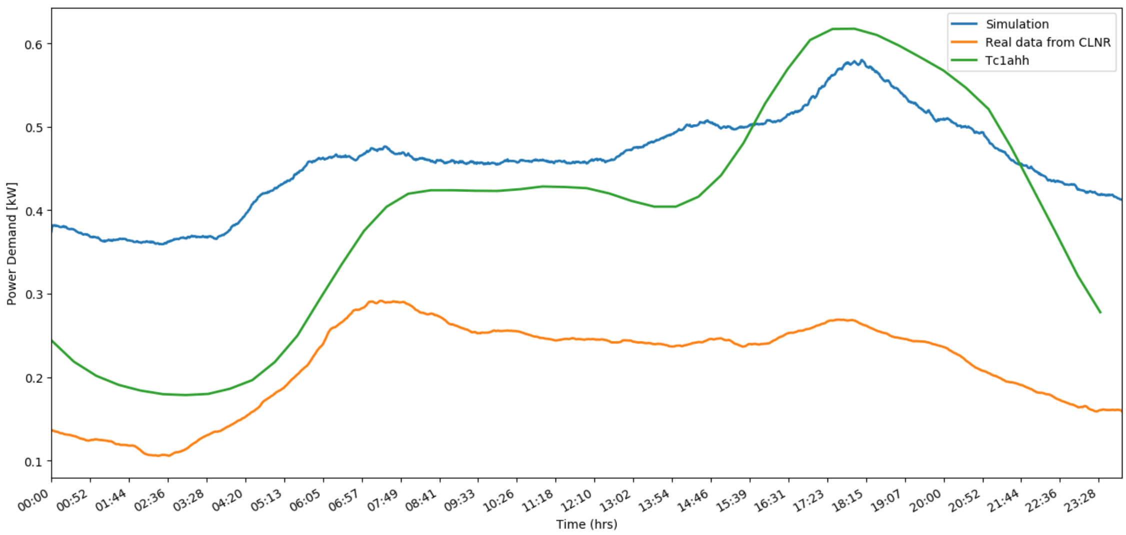

Figure 6.2: Daily mean demand comparison against test cell 1.

6.1.1 Energy results

In section 4.5 of Chapter 4 it was described an abstract tree diagram that classifies the appliances in a household and allocate them into eleven categories. This classification allowed to match different appliances among datasets and bring homogeneity to the data. Figure 6.3 shows the result of the simulation of one day of energy consumption for the entire network, it can be appreciated by category. The entire network in this case is composed by the 280 households which appliances are consuming energy through a simulated day. The **kitchen** category consumes more energy due to it is composed by the most popular appliances shared among households. This category involves cooking activities (cooker, oven and microwave), keeping food fresh (fridge and freezer), washing the dishes (dishwasher) and boiling water (kettle). A more significant result could be seen if the simulated appliances were the same for every simulated household.

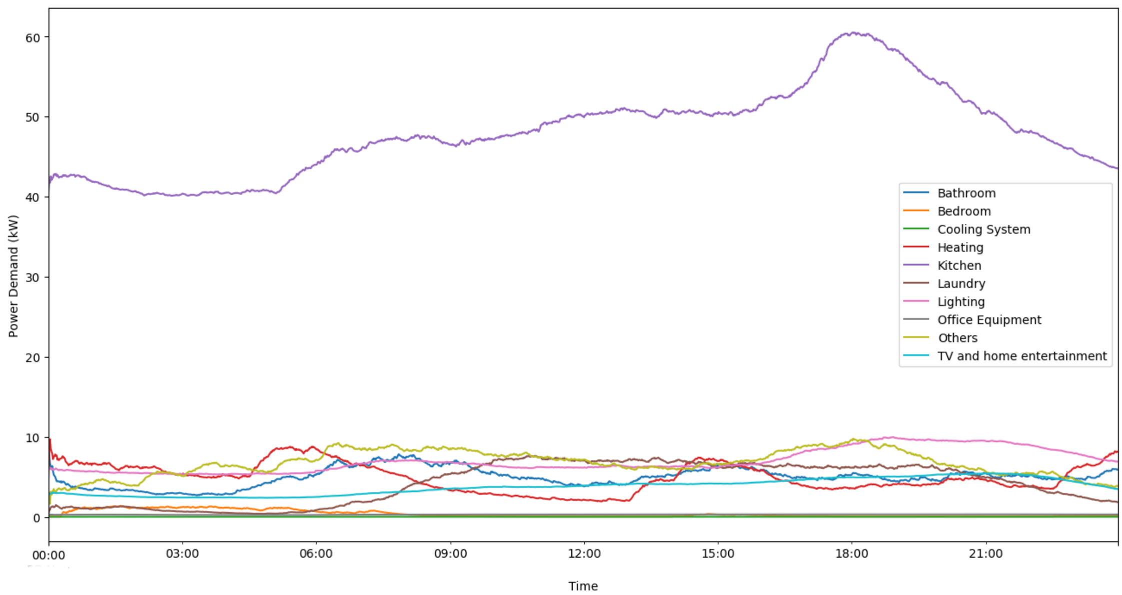

Figure 6.3: Simulated energy demand in network categorised by main nodes.

Figure. 6.4 shows the mean energy demand by category. In order to plot this chart it was calculated the average energy consumption per category among households and the results were stacked in twenty-four bars that correspond to periods of one hour. The values considered for plotting these results was taken from the data collected during simulation. It can be appreciated that in an average household the energy consumed in heating and lighting increases after 5:00p.m. According to the chart, the peak for energy in a household appears during 5:00p.m. and 6:00p.m. It is noted again that the kitchen category consumes more energy due to it is composed by the most popular appliances shared among households.

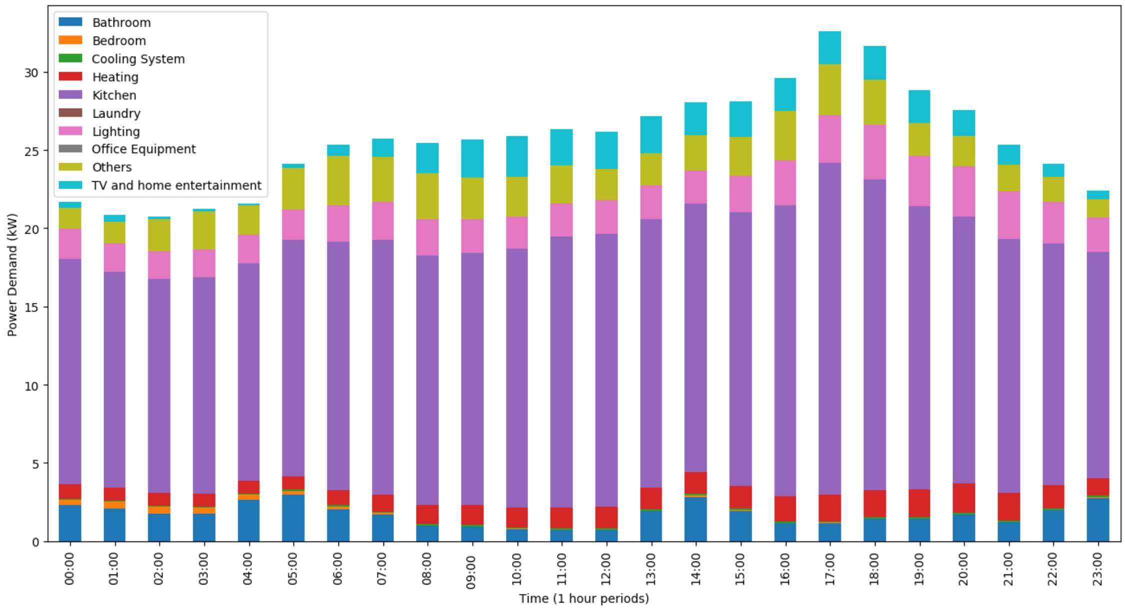

Figure 6.4: Mean energy demand in households, filtered by main appliance category.

6.2 Experiments

6.2.1 Decarbonisation of heat.

Decarbonisation involves the reducing of carbon dioxide emissions, with the decarbonisation of the grid it will be necessary the electrification of heat and finding the most efficient ways to use electricity for heating [31]. With the decarbonisation of the grid happening so rapidly in the UK it is necessary to estimate the power load that the grid will have to satisfy in an scenario where each household in the country or an specific geographic area has replaced the energy source of heating for electricity.

According to the real data collected by the CLNR in test cell *2a*, 131 households out of 280 do not have appliances related to heating, which suggests that the heating in the 46.78 percent of the households is not powered by electricity.

Using the simulator, a set of appliances for heating, and along with a generic energy profile and likelihood values of energy consumption given a time, it was possible to simulate that each of the households (which were powered by gas) were demanding for energy to satisfy heating needs. The selected profile for the simulated decarbonised households consisted in one bathroom heater, three additional electric heaters and two immersion heaters. The energy consumption values and likelihoods values (for when an appliance is turned *ON/OFF* or remains in *STAND BY*) used for these appliances are a duplicate of the values generated for household 10198 from the CLNR data. Household 10198 was selected due to the quality of the smart meter readings during the CLNR project and for the reason that is a household that consistently registered energy use of electric heaters during night and early morning. With the simulation of decarbonisation in households in this particular experimental scenario higher levels of power demand were reached, from having a peak of 111.64 kW at 5:50p.m to 162.11 kW at 11:38p.m.

Figure 6.5 shows the comparison chart of running the simulation considering the real appliances in the households (according to test cell *2a*), against the experimental scenario where all the households have heating powered by electricity. In this experiment it can be seen that between 12:00am and 7:00am the "after decarbonisation" line is under the "before decarbonisation" line. It has to be pointed out that the chart presented in Figure 6.5 corresponds to the whole network and during the running of the simulation some households spend less energy in heating than others due to a heating appliance not necessarily being switched on at the same time than in previous runs of the simulation. The action of switching on or off an appliance is decided during the step method of the agent (see flowchart in Fig. 5.3) and its execution in each agent during simulation can result in a slightly different line when plotting the power demand every time the simulation is run. However, when running the simulation two peaks appeared every time between the periods from 7:00am to 11:00am and from 4:30pm to 9:00pm when the simulation was run.

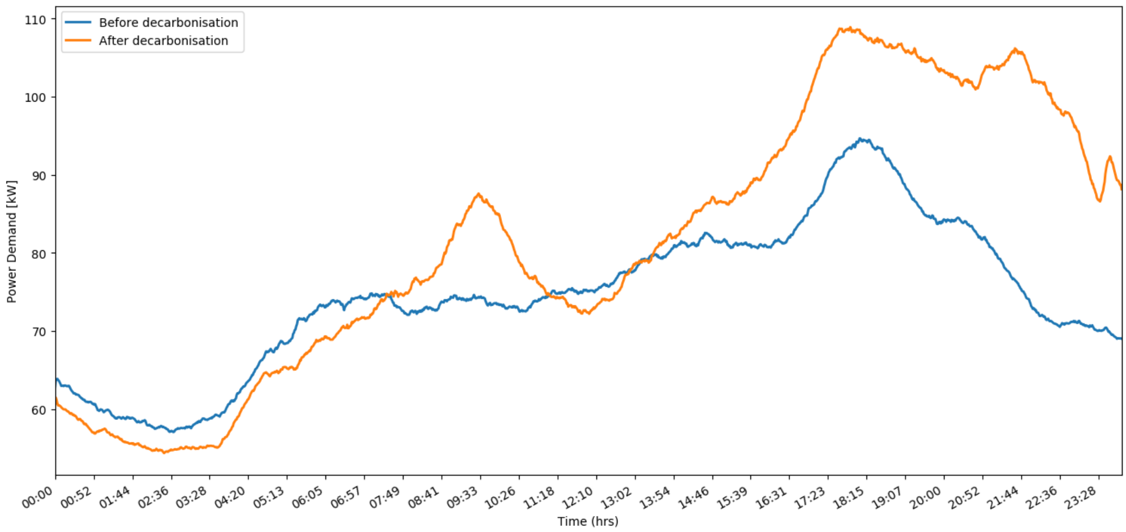

Figure 6.5: Simulation of decarbonisation by replacing gas heating with electricity in households.

6.2.2 Constraints for appliances during the peak period.

The next experiment was motivated on simulating moving the load of certain appliances out of certain hours, with the objective of decreasing the energy peak between 5.00pm and 9.00pm. Although the load in this experiment is not moved because the simulator does not allow yet to control the number of times that an appliance goes on during the day. With the current approach in the development of the simulator it was only possible to restrict the selected appliances from switching ON during a defined period. A solution for a complete representation of moving the load is to increment the likelihood for the selected appliances, in the simulator perspective the likelihood values for an appliance going on during the constrained period would be moved to the hours of the day that are not affected by constraints. That solution would allow the simulator to move loads by internally moving the likelihoods of the selected appliances. During this experiment the defined constraint consisted in laundry activities banned from the energy peak period as well as turning ON the dishwasher

Appliances that do not go ON when it is peak period:

- Washer dryer power
- Washing Machine
- Washing machine power
- Tumble dryer power
- Dishwasher only power

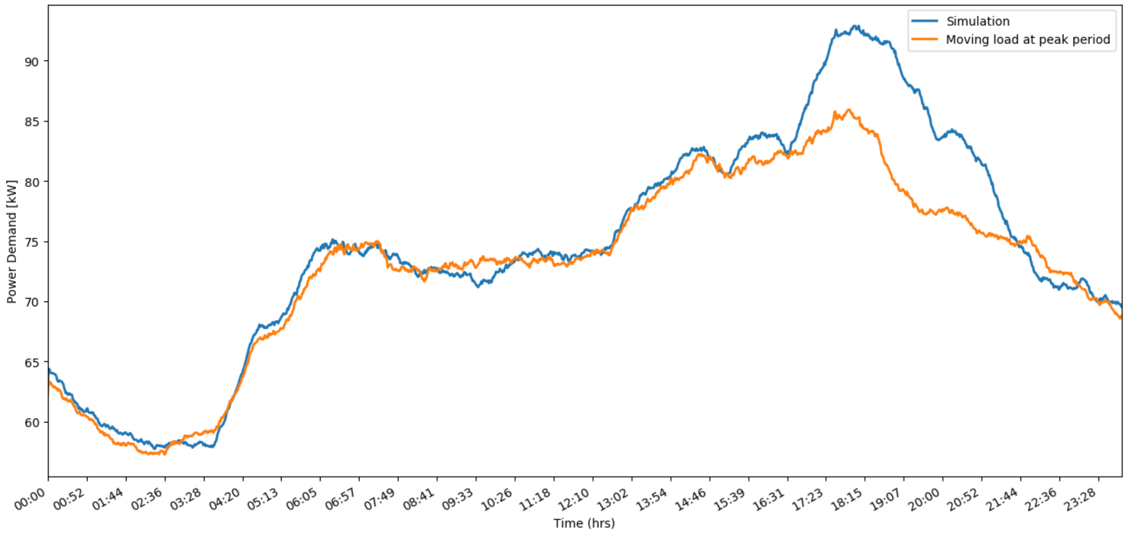

Figure 6.6: Comparison of simulation results against simulation of moving load out of the energy peak period .

The results of running this experiment can be appreciated in Fig. 6.6. This chart compares the result from running the simulation without any constraint, against running the simulation when between 5.00pm and 9.00pm is not possible to perform laundry and dishwasher activities. With the simulation of moving the load of only those activities out of the peak period the peak decreased from 92.87kW at 18.07pm to 85.92kW at 17.54pm, that is a difference of 6.95kW in a set of 280 households. The total energy used through the day for the entire network of 280 household went from 966.85 kWh to 931.044 kWh after adding the constraint for washing activities during the peak period.

6.3 Conclusions

The main results of the simulation against the real data of energy consumption are similar in shape which suggests that the behaviour of the appliances being turned ON and OFF are correct. However, further tuning for the energy profile of the appliances has to be done to increase the accuracy as compared to the real energy demand. During the simulation and in the real data as well, the *kitchen* category consumes most of the energy in the network due to the majority of households had at least a cooker and stove in its set of appliances during the CLNR project. A more significant result could have been obtained if the simulated appliances were the same for every simulated household. Nevertheless, there is a good indication that the simulator could still be of assistance as it is when some typical patterns of behaviour regarding energy consumption can be appreciated, especially with heating and lighting.

Two experiments were performed with the simulator as a base. The first one addressed the decarbonisation problem through the electrification of heat and simulated an scenario when each household in the simulation had electric heaters. The result of the electrification of the heat showed that the peak in energy demand between 5.00pm and 9.00pm will definitely increase. The second experiment addressed moving power load out of the network during the peak period. That consisted in adding a constraint to the simulator where washing activities could not be performed during 5.00pm and 9.00pm. As a result of the second experiment, the electricity consumption peak was considerably flatten which suggest that restricting and rescheduling some activities that consume energy to another time can indeed help with flattening peak electricity demand in households and the entire network. These experiments are just two energy use scenarios out of the many that could be tested with the DDDS.

Chapter 7

Conclusions and future work

In this project a simulator has been developed that uses the CLNR data as a base to simulate the behaviour of 280 houses that are connected to the energy network. Based on the data of individual energy consumption of appliances in a household, which was generated every minute, it is possible to simulate when an appliance is likely to be either turned *ON* or *OFF*, or on *STAND BY*, and an estimation of how much energy will be used.

We can see the main results charts have the same properties and share the same pattern. It can therefore be said that with a some additional tuning, validation and verification this simulator could be a valid tool to start experimenting with energy use scenarios to better understand energy use so that we can consequently have a better transition into a decarbonised grid.

This approach has its limitations, they main limitations are listed below. With future work the simulator can overcame this limitations and become a better tool to help other researchers develop new methods or algorithms to improve the efficiency and reliability of future power systems.
Limitations that need future work:

- There is not a homogeneous electric profile for households.
  The households in the simulation are based on real data, the problem is that in the real data households do not share a set of appliances in common. The results shown do not contribute with many insights and can be difficult for

interpretation, i.e. it could be said that cooking activities is where most of the energy is used because it is one of the activities which was popular with measurements in different households during the CLNR Project. Work has to be done to define a representative electric profile for the households in the simulation. A more detailed energy profile can be defined for appliances that work with different power levels (low/medium/high), such as a washer machine that uses different power modes per cycle. It is needed an electric profile that also allows to add an arbitrary number of new virtual households to the simulation in order to emulate a more realistic electric grid, where the power demand is created by thousands and not hundred of houses.

- *Stand by* power.
  The stand by values for energy consumption is fixed for each type of appliance. In real life it is not the same *stand by* power for a big television at home in comparison to a small one or the lighting in a six bedroom house in comparison to a students' flat. Work has to be done to define and manage *stand by* power values for appliances.

- The times an appliance is switched on during the day.
  The number of times an appliance is turned ON during the day is not controlled. The likelihood for an appliance to switch ON is considered at all times during the simulation although that likelihood does not change once the appliance has been turned ON during the simulated day. The likelihood value should change every time an appliance is turned ON or OFF. Work has to be done to calculate new likelihood values when the state of an appliance changes and also to decide how much time an appliance should remain ON once it reached that state.

- Power demand and likelihood values for appliances.
  The current energy values for appliances were calculated as the average per year, that is a limitation because it means an average day of the year can only be simulated. Currently, likelihood values for every appliance were calculated according to the time during the day, and that does not change during the

simulation of a day for another season, the simulation results are the same for a day during the summer than for a day during winter. Thus, it is important to calculate likelihood values associated with the day of the year, day of the month, day of the week, season and consider special days such as bank holidays in order to have better and more realistic results.

- Energy consumption not considered.

  In a real household there are sockets that are not exclusive from an appliance but when used they consume energy that is not allocated to any appliance listed in this simulator. The energy that is not considered in a total of energy consumption per household affects the value of energy consumed in the electric network. This problem could be solved by considering the *Total Property* channels in the CLNR datasets. Given that in the CLNR datasets households have *Total Property* channels but the sum of the energy demand of its appliances does not match this total, it is important to consider the difference in energy between them and create an energy concept during simulation called “energy used but not considered". Work has to be done as well to include in the simulator entities like batteries, electric vehicles and even solar power sources to experiment with them to find and optimum mix of these solutions.

Bibliography

[1] Lina Stankovic Vladimir Murray, David Stankovic. An electrical load measurements dataset of united kingdom households from a two-year longitudinal study. *Scientific Data*, 4(160122), 2017.

[2] Climate change act 2008, 2008. Available at: `http://www.legislation.gov.uk/ukpga/2008/27/contents` [Accessed: 11.09.19].

[3] Goran Strbac. Demand side management: Benefits and challenges. *Energy Policy*, 36(12):4419 – 4426, 2008. Foresight Sustainable Energy Management and the Built Environment Project.

[4] Gareth Powells, Harriet Bulkeley, Sandra Bell, and Ellis Judson. Peak electricity demand and the flexibility of everyday life. *Geoforum*, 55:43 – 52, 2014.

[5] Transition to smart meters, October 2018.

[6] Smart metering implementation programme: progress report 2018. *Department for Business, Energy and Industrial Strategy*, 2018.

[7] Rob J. Hyndman Marc G. Genton Souhaib Ben Taieb, Raphaël Huser. Forecasting uncertainty in electricity smart meter data by boosting additive quantile regression. *IEEE Transactions on Smart Grid*, 7(5), September 2016.

[8] Dipanjan Chakraborty Karl Aberer Deva P. Seetharam Tri Kurniawan Wijaya, Tanuja Ganu. Consumer segmentation and knowledge extraction from smart meter and survey data. *SIAM International Conference on Data Mining*, April 2014.

[9] Osamu Kimura, Hidenori Komatsu, Ken-ichiro Nishio, and Toshihiro Mukai. A prototype tool for automatically generating energy-saving advice based on smart meter data. *Energy Efficiency*, 11(5):1247–1264, Jun 2018.

[10] Joachim Schleich, Marian Klobasa, Sebastian Gölz, and Marc Brunner. Effects of feedback on residential electricity demand—findings from a field trial in austria. *Energy Policy*, 61, 10 2013.

[11] Adam Cooper. Electric company smart meter deployments: Foundation for a smart grid. Technical Report October 2016, The Edison Foundation: Institute for Electric Innovation, October 2016.

[12] Customer-led network revolution - customer trials. Available at: `http://www.networkrevolution.co.uk/customer-trials/`, 2018. [Accessed: 2018-04-01].

[13] CLNR-Project. Specification of domestic and sme customer data. Technical report, Customer Led Network Revolution Project, September 2015.

[14] Experian Mosaic. Optimise the value of your customers and locations, now and in the future. mosaic uk - the consumer classification of the united kingdom. It was accesible at: `http://www.experian.co.uk/assets/business-strategies/brochures/mosaic-uk-2009-brochure-jun10.pd`, June 2010.

[15] Energy Futures Lab. Energy infrastructure, research overview, 2017.

[16] J. Banks; J. Carson; B. Nelson; D. Nicol. *Discrete-Event System Simulation*. Prentice Hall, 2001. ISBN 978-0-13-088702-3.

[17] Stefania Bandini, Sara Manzoni, and Giuseppe Vizzari. Agent based modeling and simulation: An informatics perspective. *Journal of Artificial Societies and Social Simulation*, 12(4):4, 2009.

[18] Ian Dent. *Deriving knowledge of household behaviour from domestic electricity usage metering*. PhD thesis, 07 2015.

[19] B. J. Johnson, M. R. Starke, O. A. Abdelaziz, R. K. Jackson, and L. M. Tolbert. A dynamic simulation tool for estimating demand response potential from residential loads. In *2015 IEEE Power Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, pages 1–5, 2015.

[20] Z. Wang and R. Paranjape. Agent-based simulation of home energy management system in residential demand response. In *2014 IEEE 27th Canadian Conference on Electrical and Computer Engineering (CCECE)*, pages 1–6, 2014.

[21] Matteo Muratori, Matthew C. Roberts, Ramteen Sioshansi, Vincenzo Marano, and Giorgio Rizzoni. A highly resolved modeling technique to simulate residential power demand. *Applied Energy*, 107:465 – 473, 2013.

[22] Michael J . North, Nicholson T. Collier, Jonathan Ozik, Eric R. Tatara, Charles M. Macal, Mark Bragen, and Pam Sydelko. Complex adaptive systems modeling with repast simphony. *Complex Adaptive Systems Modeling*, 1, March 2013.

[23] Shanjun Tian and Shiyan Chang. An agent-based model of household energy consumption. *Journal of Cleaner Production*, 242:118378, 09 2019.

[24] Sohini Dhar, Babji Srinivasan, and Rajagopalan Srinivasan. Simulation and analysis of indian residential electricity consumption using agent-based models. In Anton Friedl, Jiří J. Klemeš, Stefan Radl, Petar S. Varbanov, and Thomas Wallek, editors, *28th European Symposium on Computer Aided Process Engineering*, volume 43 of *Computer Aided Chemical Engineering*, pages 205 – 210. Elsevier, 2018.

[25] Jean-Paul Zimmermann, Matt Evans, Jonathan Griggs, Nicola King, Les Harding, Penelope Roberts, and Chris Evans. Household electricity survey. *Department of Energy and Climate Change*, 2012.

[26] Michael Aivazis K. Jarrod Millman. Python for scientists and engineers. *Computing in Science & Engineering*, 13(2):9–12, March 2011.

[27] Project Mesa Team. *Mesa Documentation*, release .1 edition, October 2019.

[28] About mariadb. Available at: `https://mariadb.org/about/` [Accessed: 21.10.19].

[29] Paul DuBois and Michael Widenius. *Mysql*. New Riders Publishing, USA, 1999.

[30] T. Bray. The javascript object notation (json) data interchange format. rfc 7159, rfc editor, march 2014. *Internet Engineering Task Force (IETF)*, 2014.

[31] Jesus Lizana, Daniel Friedrich, Renaldi Renaldi, and Ricardo Chacartegui. Energy flexible building through smart demand-side management and latent heat storage. *Applied Energy*, 230:471 – 485, 2018.

Appendix A

Replicating consumers behaviour during night.

A.1 Dataset: *tc2amicro*

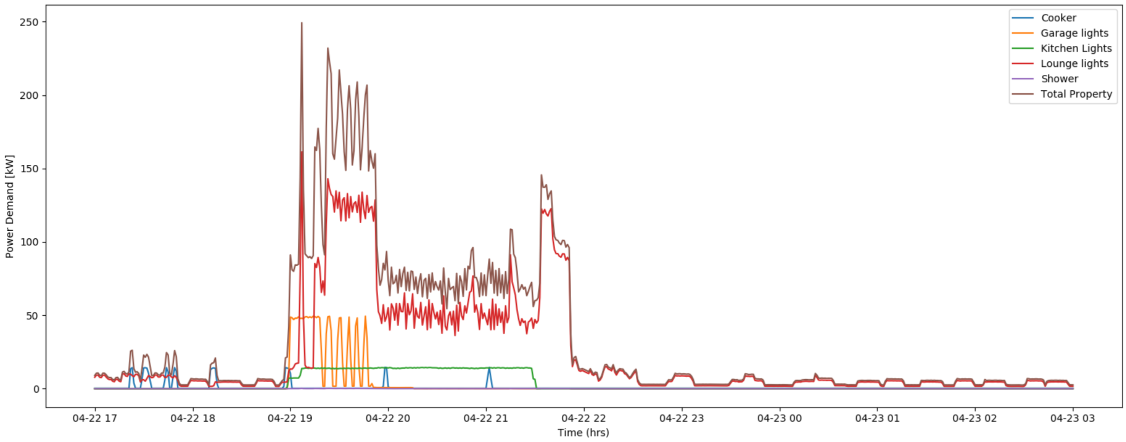

Figure A.1: Power demand for household 10234 during the evening of April 22nd 2013.

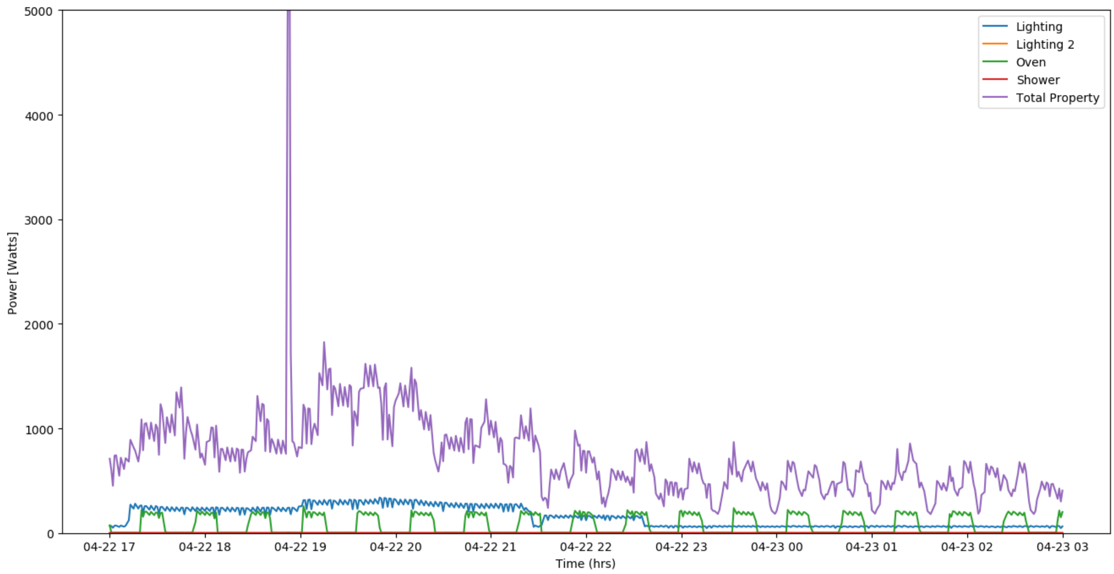

Figure A.2: Power demand for household 10028 during the evening of April 22nd 2013.

A.2 Dataset: *tc2apassiv*

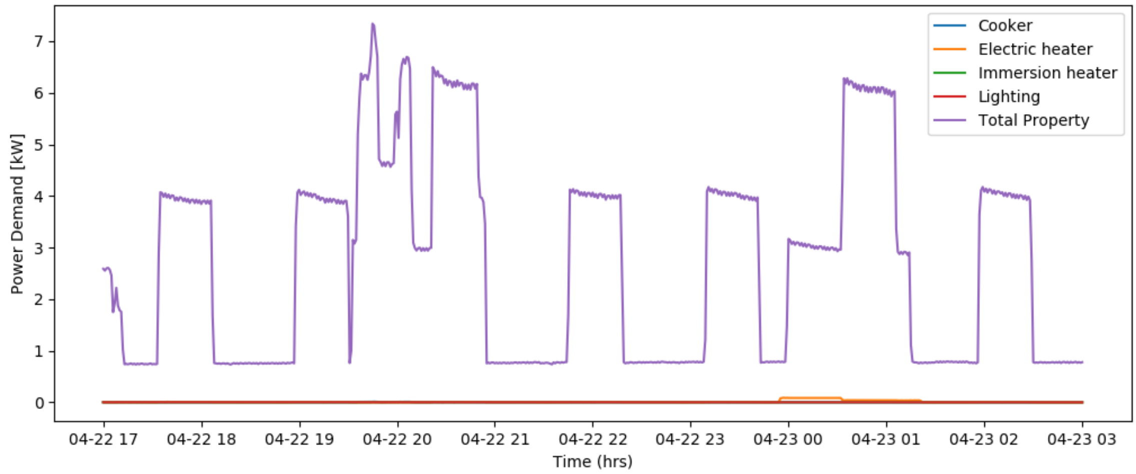

Figure A.3: Power demand for household 10049 during the evening of April 22nd 2013.

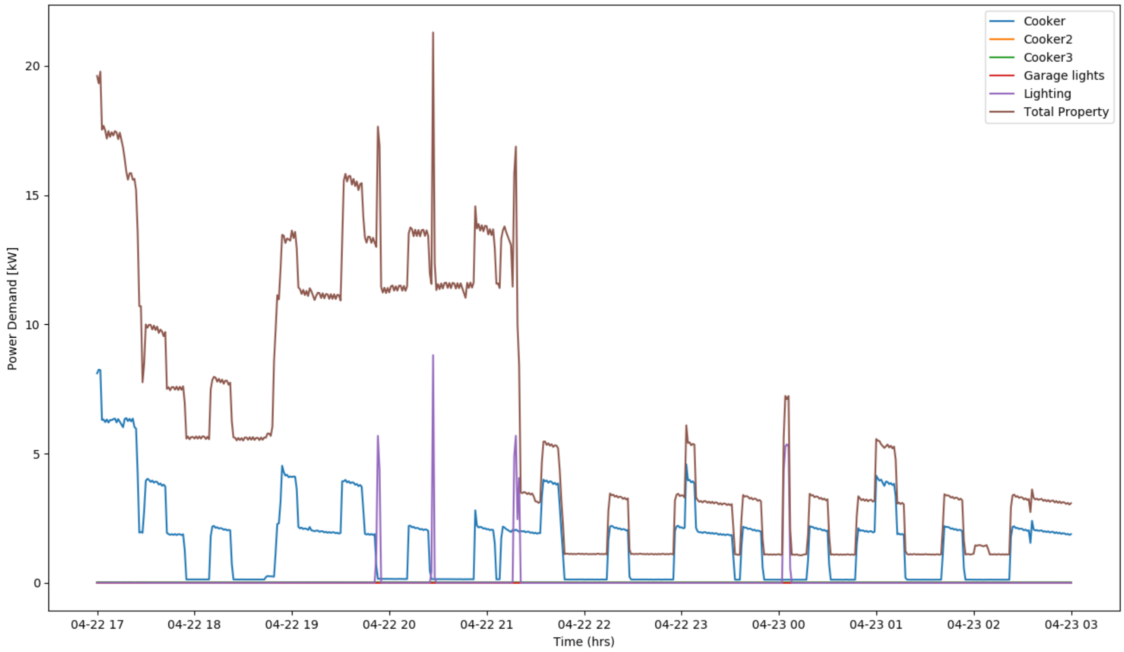

Figure A.4: Power demand for household 10091 during the evening of April 22nd 2013.

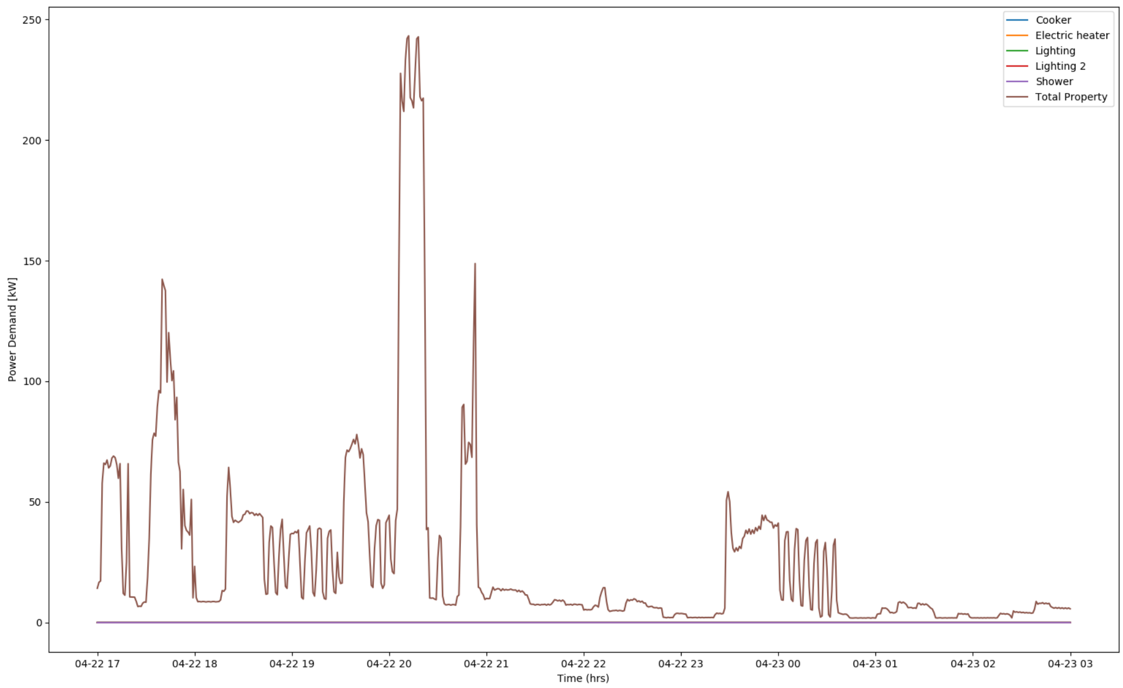

Figure A.5: Power demand for household 10105 during the evening of April 22nd 2013.

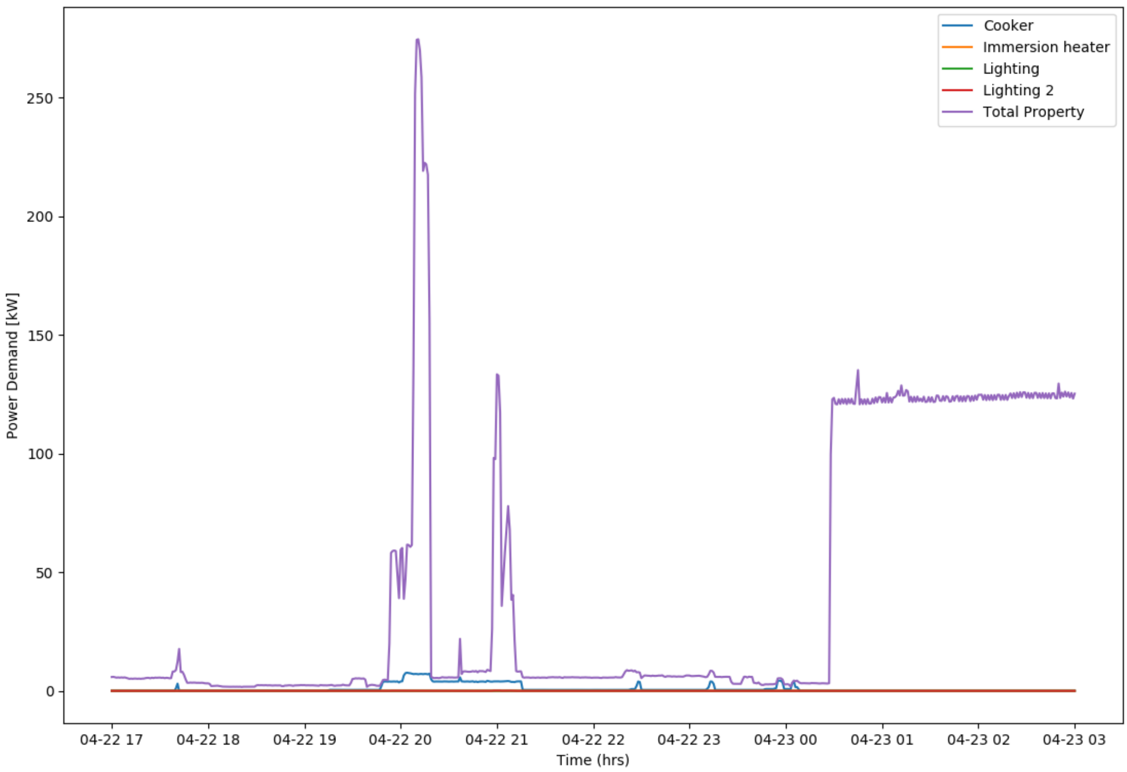

Figure A.6: Power demand for household 10156 during the evening of April 22nd 2013.

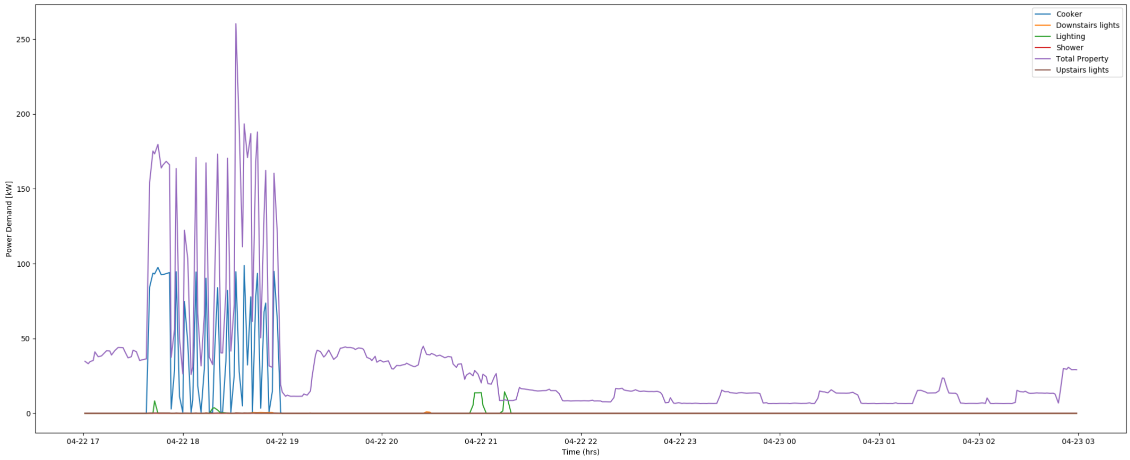

Figure A.7: Power demand for household 10157 during the evening of April 22nd 2013.

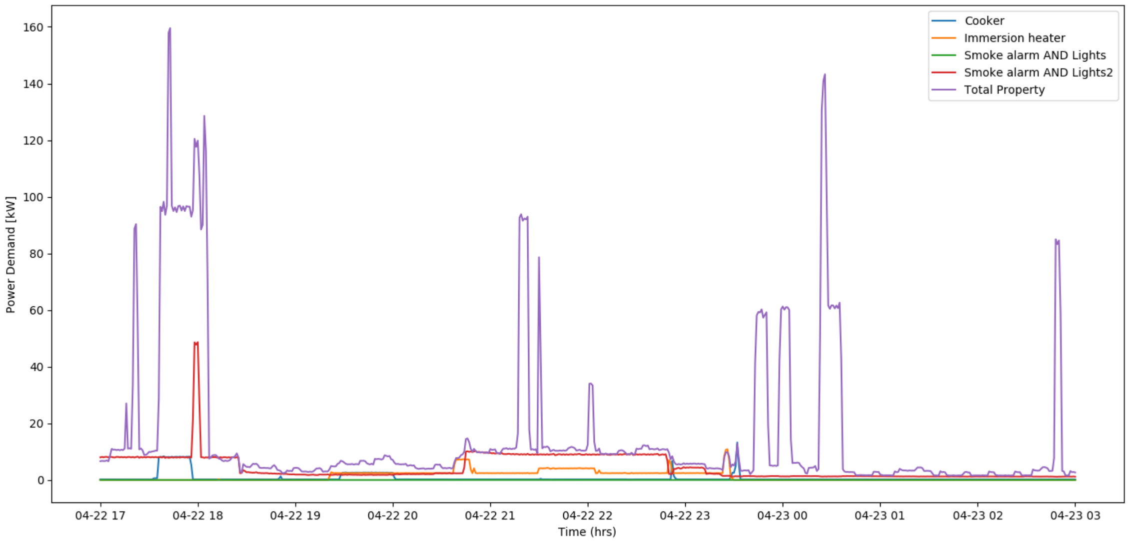

Figure A.8: Power demand for household 10181 during the evening of April 22nd 2013.

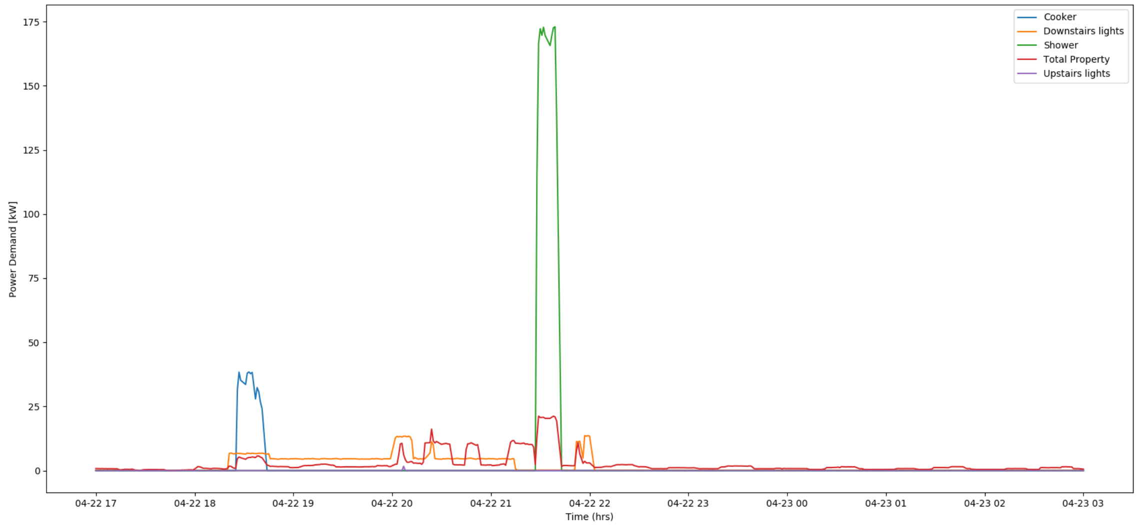

Figure A.9: Power demand for household 10252 during the evening of April 22nd 2013.

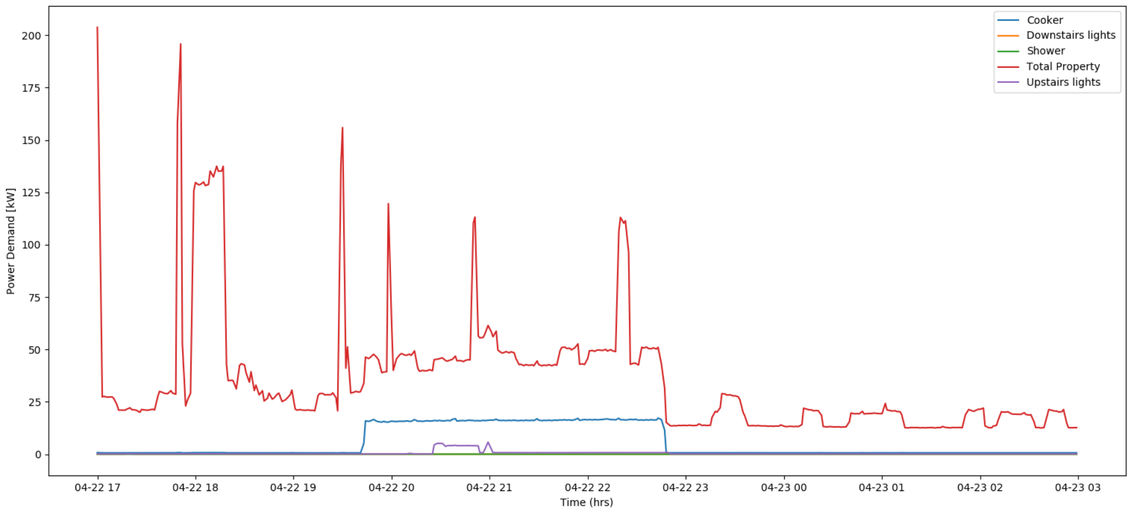

Figure A.10: Power demand for household 10257 during the evening of April 22nd 2013.

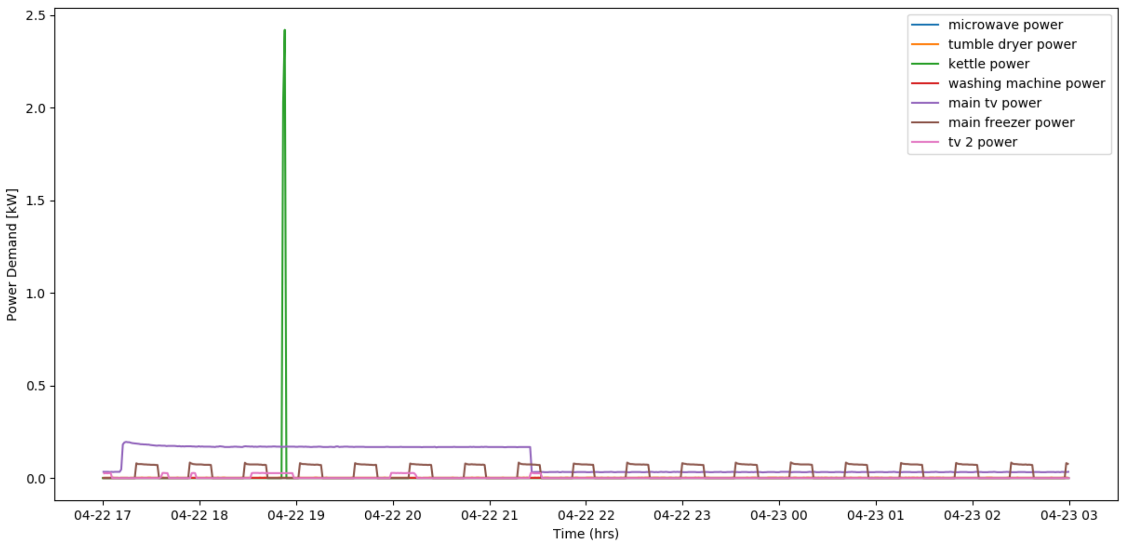

Figure A.11: Power demand for household 10028 during the evening of April 22nd 2013.

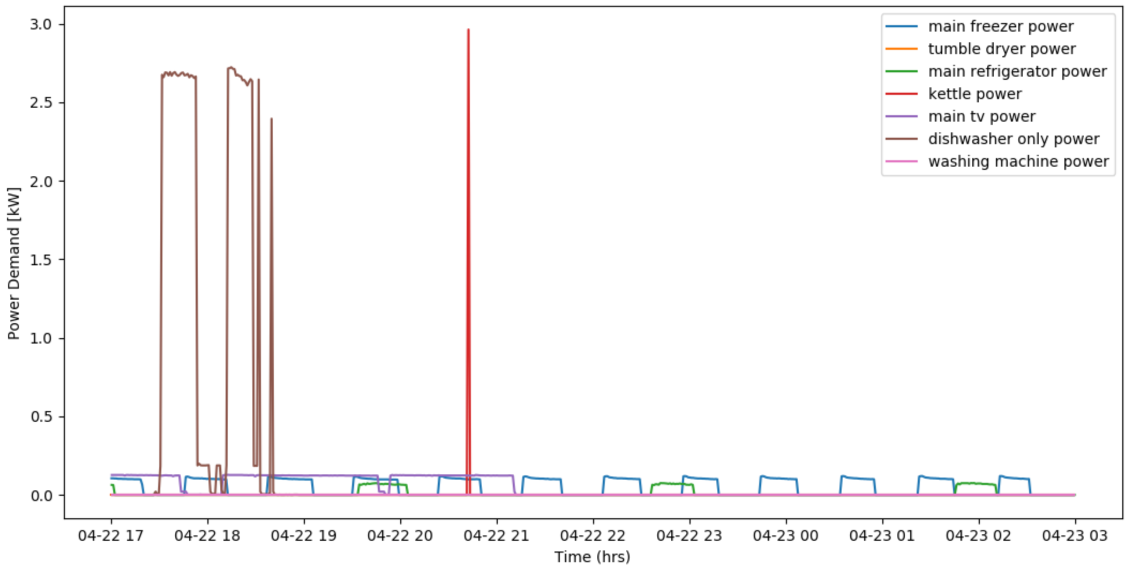

Figure A.12: Power demand for household 10116 during the evening of April 22nd 2013.

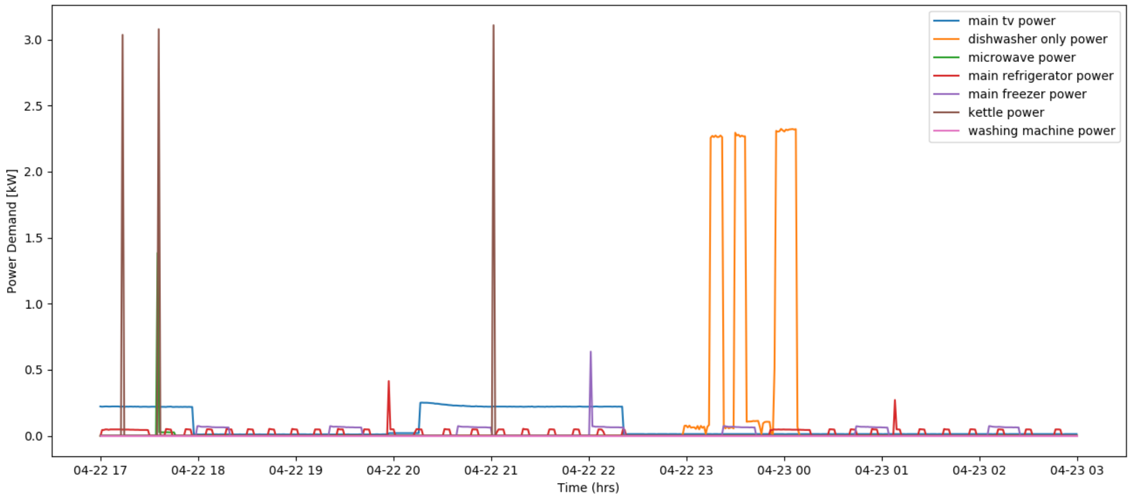

Figure A.13: Power demand for household 10073 during the evening of April 22nd 2013.

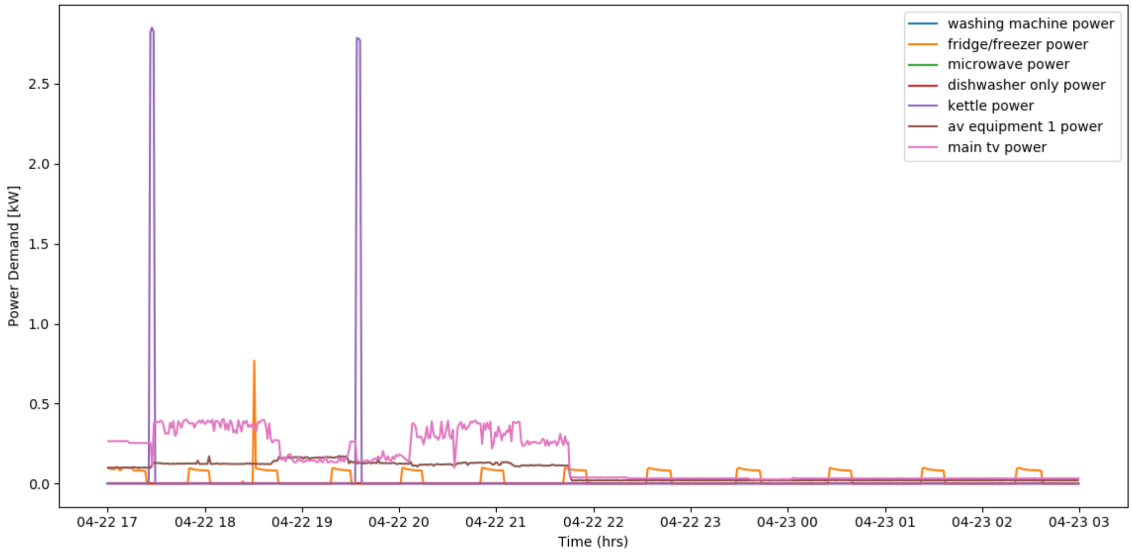

Figure A.14: Power demand for household 45617 during the evening of April 22nd 2013.

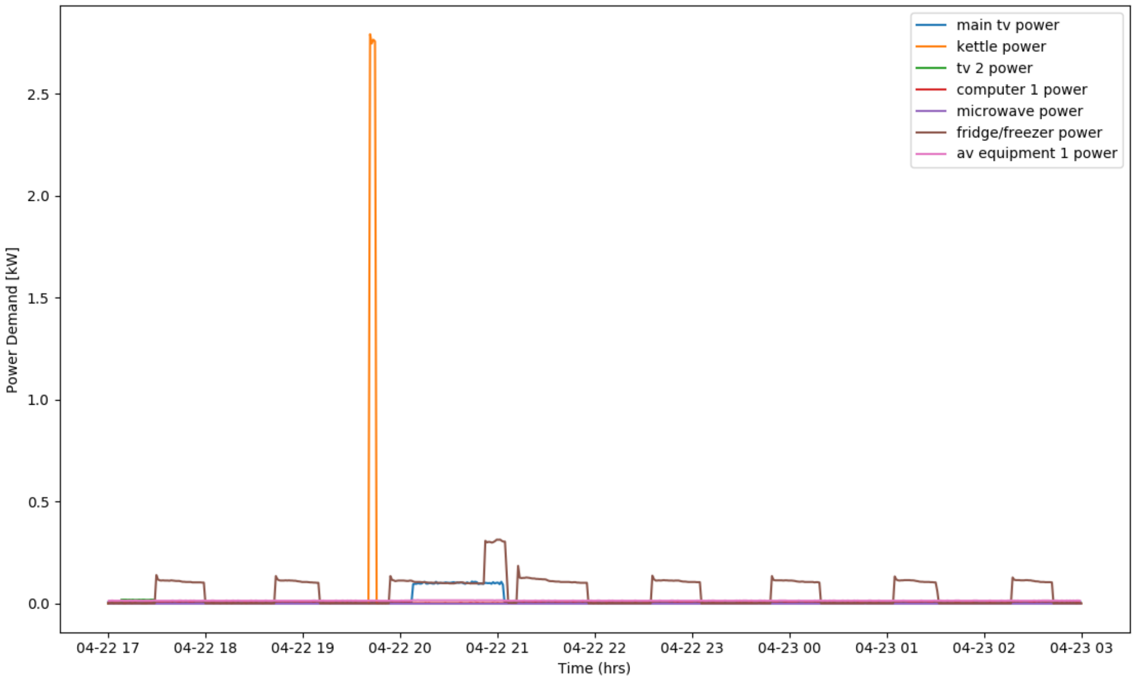

Figure A.15: Power demand for household 10129 during the evening of April 22nd 2013.

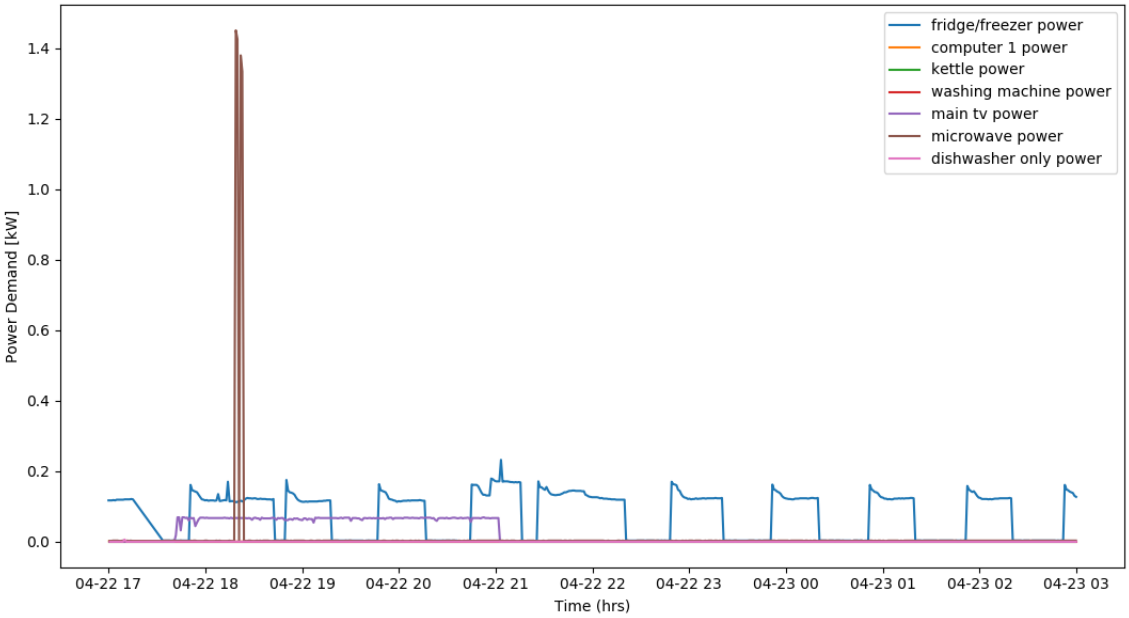

Figure A.16: Power demand for household 10085 during the evening of April 22nd 2013.

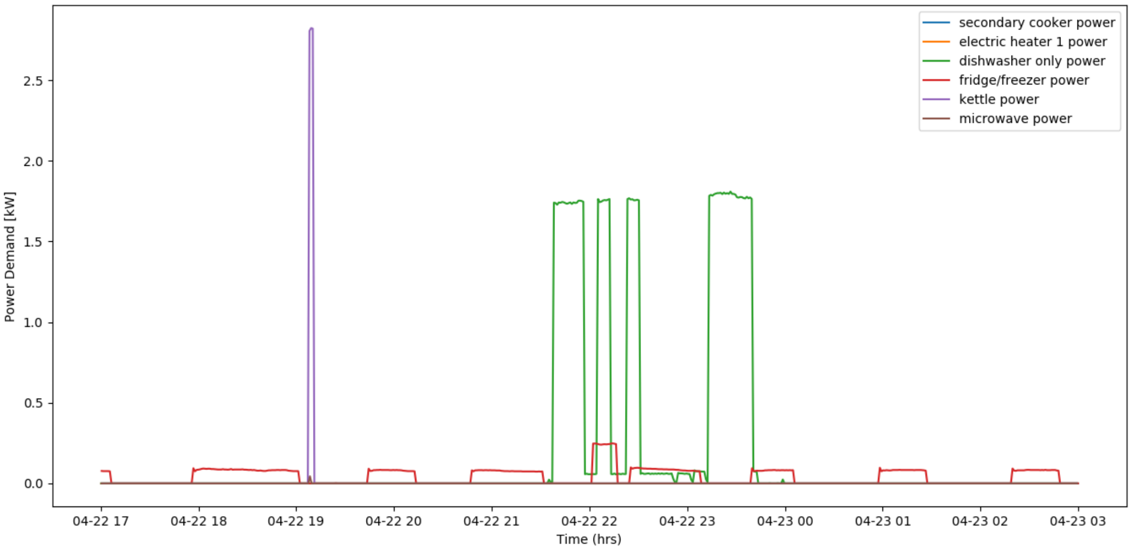

Figure A.17: Power demand for household 10218 during the evening of April 22nd 2013.

Appendix B

Channels in datasets

B.1 Dataset: *tc2amicro*

Identifier	Channels
1	Bathroom AND Kitchen heating
2	Bathroom heater
3	Bathroom lights
4	Bedroom heater
5	Bedroom lights
6	Cellar lights
7	Cooker
8	Cooker2
9	Cooker3
10	Downstairs lights
11	Downstairs lights2
12	E7 meter
13	Electric heater
14	Electric heater2
15	Electric heater3
16	Electric heater4
17	Freezer

Continuation of Table	
Identifier	Channels
18	Fridge
19	Garage lights
20	Ground floor Lights
21	Hall lights
22	heat pump power consumption
23	heat pump power consumption2
24	Immersion heater
25	Immersion heater2
26	Kitchen AND Front room lights
27	Kitchen Lights
28	Landing lights
29	Lighting
30	Lighting 2
31	Lighting 3
32	Lighting 4
33	Lounge AND Bedroom heating
34	Lounge lights
35	Outside lights
36	Oven
37	Shower
38	Shower2
39	Smoke alarm AND Lights
40	Smoke alarm AND Lights2
41	Total Property
42	Upstairs lights
43	Upstairs lights2
44	Washing Machine

B.2 Dataset: *tc2apassiv*

Identifier	Channels
1	solar power
2	heat pump power consumption
3	dishwasher only power
4	fridge/freezer power
5	kettle power
6	main tv power
7	tumble dryer power
8	washing machine power
9	secondary freezer power
10	main freezer power
11	main refrigerator power
12	secondary cooker power
13	whole home power import
14	microwave power
15	electric heater 1 power
16	whitegoods 1 power
17	computer 1 power
18	set top box 1 power
19	downstairs light 1 power
20	electric heater 3 power
21	av equipment 1 power
22	tv 2 power
23	hob power
24	washer dryer power
25	secondary refrigerator power
26	main cooker/oven power
27	av equipment 2 power

Continuation of Table	
Identifier	Channels
28	whitegoods 4 power
29	wine cooler power
30	water heater power
31	appliance 1 power consumption
32	air conditioner 1 power
33	small appliances 1 power
34	small appliances 2 power
35	electric heater 2 power
36	appliance 2 power consumption

Appendix C

CLNR Data Pre-processing

C.1 Channels Categorisation

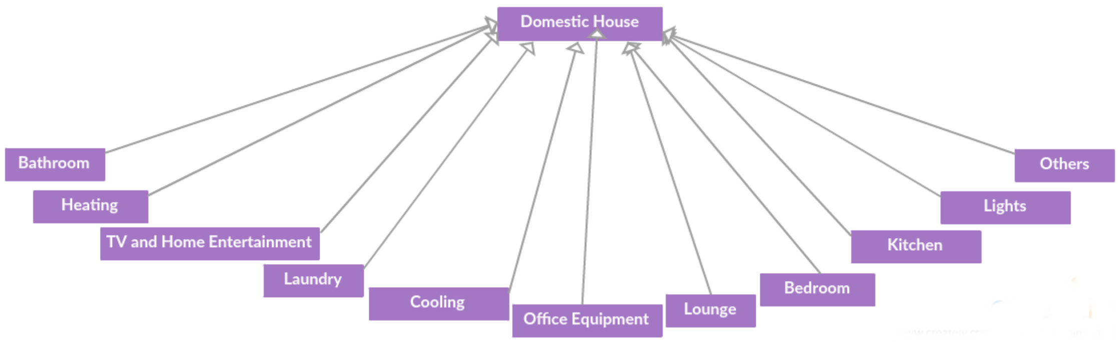

Figure C.1: Main channels in abstraction tree.

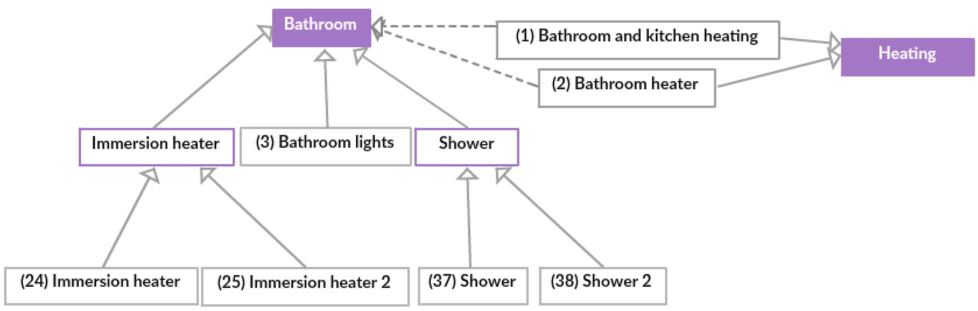

Figure C.2: Nodes under Bathroom Category.

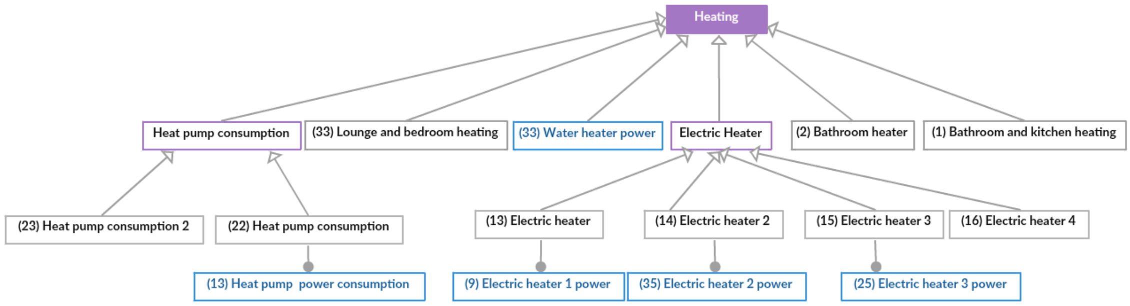

Figure C.3: Nodes under Heating Category.

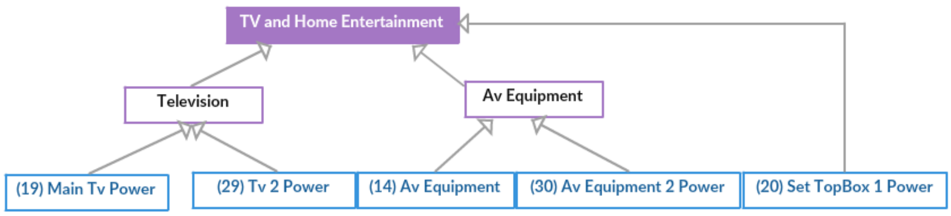

Figure C.4: Nodes under TV and Home Entertainment Category.

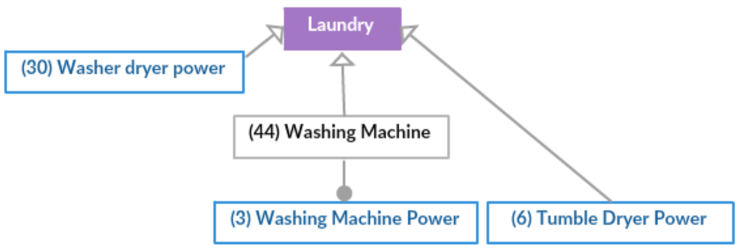

Figure C.5: Nodes under Laundry Category.

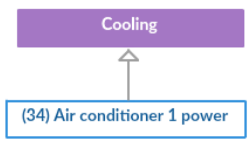

Figure C.6: Nodes under Cooling Category.

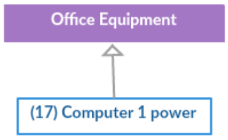

Figure C.7: Nodes under Office Equipment Category.

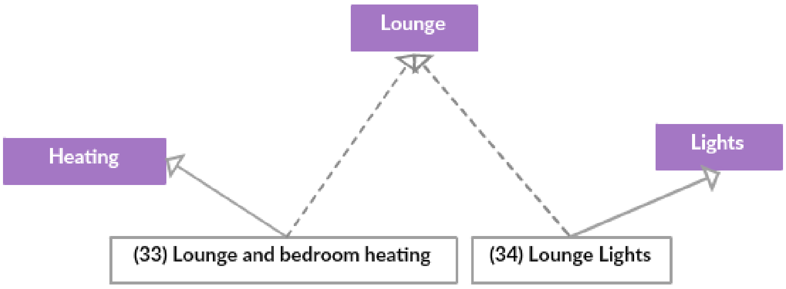

Figure C.8: Nodes under Lounge Category.

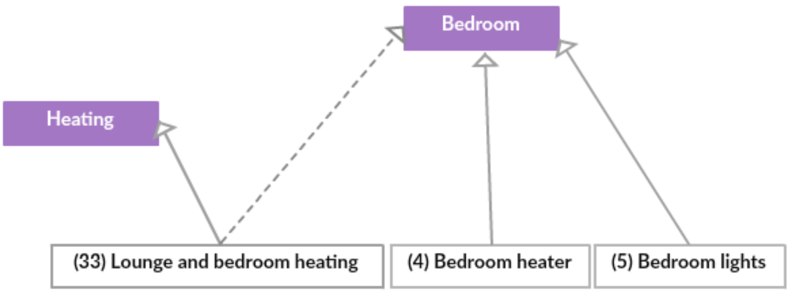

Figure C.9: Nodes under Bedroom Category.

C.2 Explanation of abstraction tree diagram for channels

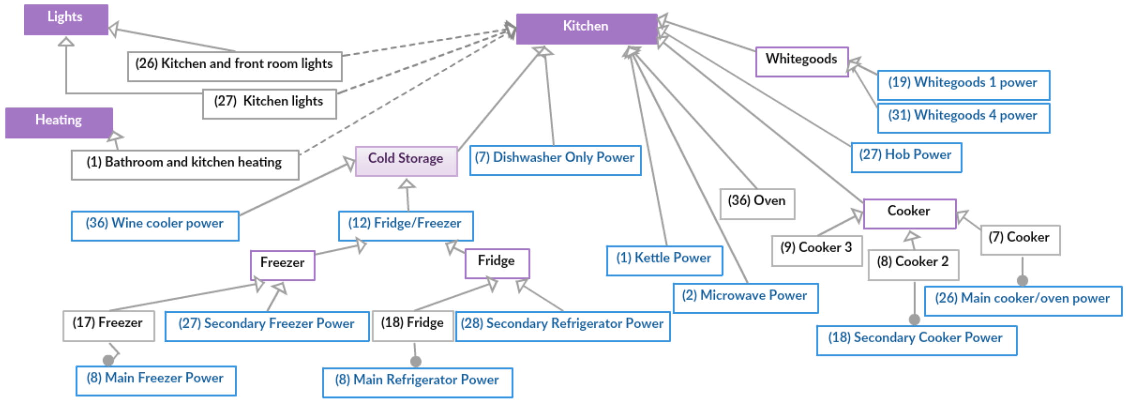

Figure C.10: Nodes under Kitchen Category.

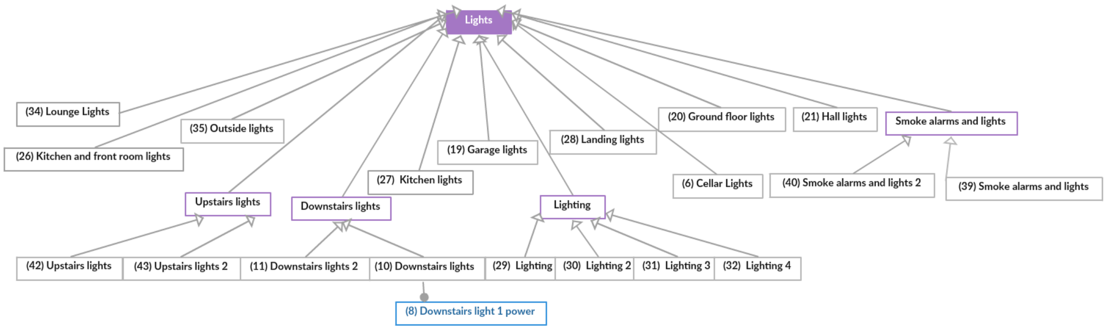

Figure C.11: Nodes under Lights Category.

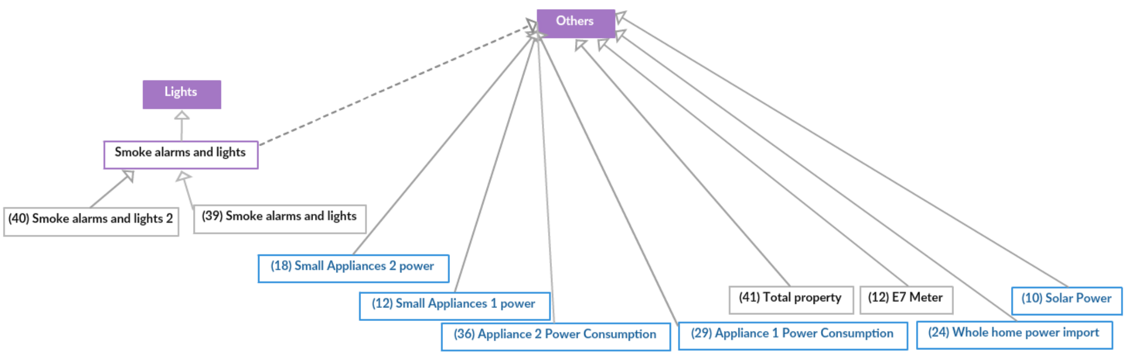

Figure C.12: Nodes under Others Category.

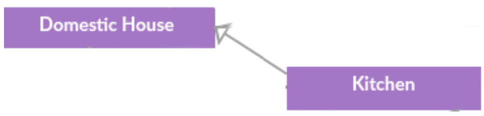

Figure C.13: Parent node and child.

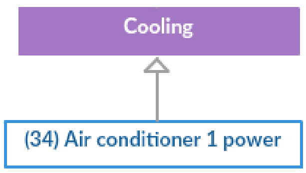

Figure C.14: Air conditioner has the node *Cooling* as a parent node.