NILM: Energy Monitoring, Modelling, and Disaggregation

PhD Thesis

David Meek Murray

A thesis presented for the degree of
Doctor of Philosophy

Department of Electronic and Electrical Engineering
University of Strathclyde
United Kingdom
02/06/2023

Declaration

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree. The copyright of this thesis belongs to the author under the terms of the United Kingdom Copyright Acts as qualified by University of Strathclyde Regulation 3.50. Due acknowledgement must always be made of the use of any material contained in, or derived from, this thesis.

Abstract

Smart meters have begun to replace traditional energy meters across the world. These meters will be able to give near real-time energy usage to consumers, industry and suppliers, the benefits of this are marketed as; easier energy management, saving money and reducing emissions. However by themselves smart meters are unlikely to be able to achieve this as only a small sub-set of users are likely to remain engaged with their smart meter long term. Non-intrusive Load Monitoring (NILM) aims to provide a method to explain to consumers in more detail about their energy usage without need for their input or attention, be it explaining which appliances are causing high energy consumption within the home or in an industrial setting, explaining appliance/machinery usage to maximise scheduling with time of use tariffs. We demonstrate the steps and methodology to produce meaningful and explainable results which could as part of an energy suppliers service to provide enhanced billing information, similar to some credit card statements, showing a breakdown of appliance usage and statistics. This thesis provides steps from data collection to results and visualisation as part of a complete NILM workload. We demonstrate data management, pre-processing, appliance modelling for analysis of individual appliances, neural network model creation and evaluation for both commercial and residential premises, the need for transfer learning to work at scale, and the explainability of the networks and results, necessary to provide accurate information and ensure customers can understand any errors they might see from a NILM system.

Contents

List of Figures

List of Tables

Acronyms

CNN Convolutional Neural Network

DNN Deep Neural Network

GRU Gated Recurrent Unit

LCA Life Cycle Assessment

NILM Non-intrusive Load Monitoring

SMETS Smart Metering Equipment Technical Specification

UK United Kingdom

Preface

Initially the concept of load monitoring was inherently an invasive one, with a need to individually monitor each appliance with it's own hardware. This was challenged when Hart [54] proposed the idea of monitoring the entire building and identifying appliance signatures from the aggregate power signal, described as NILM. As data has become more more widely available with many more open source datasets, there has been renewed interest in non-invasive load monitoring. This was accelerated further by the huge improvements in Graphical Processing Units (GPUs) which meant that training large deep neural networks became significantly easier and with this a huge increase in the application and development of neural networks. The use of neural networks being applied for the process of NILM was cemented with Jack Kelly's paper [77].

Acknowledgements

I'd like to thank my supervisors Lina and Vladimir for putting up with my inability to write a thesis. Georgia Elafoudi, Jing Liao, and Srdjan Lulic for their support. And finally my parents for their belief in me.

Chapter 1

Introduction

The 'Smart Home' aims to fully connect the household with appliances and the house with the wider world. The definition of 'smart' is loose covering anything from smart fridges to thermostats, each one providing its own subset of 'smart' features. The one unifying part of the smart revolution is data collection, processing and feedback. Starting in the mid-90's the Internet-of-Things has been gaining pace rapidly, connecting nearly every device via WiFi, Bluetooth, and other protocols. Ideally devices could provide their own energy monitoring features, however the IoT market focused more on small convenience based devices like remote power switching and convenience applications and is only now introducing energy monitoring features [101].

A key component of the ideal smart home is the 'smart meter', an electricity (or gas) meter which is able to communicate with firstly the energy supplier and secondly to other services and devices. The United Kingdom (UK) version of a smart meter is defined in the Smart Metering Equipment Technical Specification (SMETS) and will be the benchmark in this thesis for smart meter capability. [114] 'Dumb' meters which rely on user reported values are usable only as a means of correcting billing after the fact and require manual confirmation by suppliers to confirm consumer readings periodically, requiring a on-site visit, one of their most intrusive features.

Smart meters which report hourly or monthly readings directly to suppliers can be used initially for correct billing but additionally for a host of other metrics not previously available, such as accurate forecasting for bill estimation with data also being made available to infrastructure providers [60]. The UK implementation would require users to opt in to have their data monitored at higher rates, and likely would require a separate hardware/software as the meter's as specified are designed firstly for very low frequency readings to suppliers (see SMETS2 [114]). Once these higher level processes are in place smart meters can allow for even more in depth energy management, other finer detail such as usage patterns and demand management can be implemented and made available to consumers.

On a larger scale, smart meters and the foundation of the smart grid. Being able to track consumer demand enables a host of tools and features to energy providers and grid operators. Demand management is a vital tool for developing countries and where there are deficits in power generation. Smart meters allow for dynamic pricing which is beneficial as the customers who can benefit the most are generally considered fuel poor. To enable these more advanced features the data rate has to be higher than monthly, smart meters generally will have the ability to record the households aggregate at a rate higher than 15 minutes, down to 1 second [114]. This higher data rate allows external programs and hardware to analyse the signal and provide feedback. At these higher data rates comes the ability to extract individual appliance consumption, a process known as NILM, where the aggregate is broken down to its constituent appliances. This allows more granular feedback to the consumer, estimating individual appliance load estimation without the need for additional sub-metering, real efficiency values via data driven life cycle appliance modelling, informing flexibility in appliance use for more effective demand management, and other applications whereby appliance-level consumption is needed, including anomaly detection.

This thesis looks at the challenges and solutions to intrusive and non-intrusive load monitoring, appliance modelling and load prediction. Firstly looking at how data collection and preparation is handled and the challenges associated with scaling systems assuming massive uptake, secondly the application of NILM to detect appliance usage, thirdly appliance modelling, which has many wide ranging uses from assisting with prediction to commercial studies and consumption prediction focusing on the needs of utilities and micro-grids when it comes to supply and demand management. Finally it looks at the explainability of these systems, so that business to consumers can gain insight and help to improve their understanding.

Now that smart meter data is being gathered in many countries, applications can now be created such as NILM (see [152] and [132] and references therein), consumption forecasting [38], demand response [81], and load scheduling & energy saving feedback [131]. Currently, there are many smart meter datasets with individual appliance monitoring openly accessible from different countries, providing domestic consumption patterns for geographically specific locations, UK [78, 107], Switzerland [10], India [8], Germany [143], China [87]. The models proposed in this thesis can facilitate the usage of these existing datasets to study sustainability in geographically spread households, e.g., by quantifying energy wastage and usage.

1.1 Research Motivation and Aims

The motivation of this research was the exploration of the potential of smart meters and the practical non-billing applications that they could enable. The question about how to monitor and store data securely, what can be analysed and what data can be presented meaningfully to consumers or power companies. The benefits of smart meters are only just beginning to be fully recognised by power companies and attempts to centralise the data by the UK government are still to be defined. Currently smart meters (in the UK) are limited to sampling rates below 1Hz, in the range of minutes [114].

Due to this the low cost of additional sensing to enable investigation into higher frequency NILM (1 - 10 second range) helps to generate analysis and modelling which is far more beneficial and promotes the uptake of NILM and it's associated applications. This leads to the following questions:

R.Q.1: What data best encapsulates the average UK household and how do consumers interact with the introduction of smart meter enabled technologies?

R.Q.2: What insights can the analysis of consumption data enable in the context of real world appliance usage for both consumers and suppliers for energy savings?

R.Q.3: How can real world consumer data be used to generate consumption models based on limited operational knowledge, and how can this be leveraged by industry for better environmental impact reporting?

R.Q.4: How can neural networks be used to create fast and accurate and generalised NILM models for residential and commercial properties, what advantages do different approaches have and can they be used across the world to reduce the need for additional datasets and monitoring?

R.Q.5: How can the outputs of NILM be explained? Neural networks are generally considered black boxes, and explaining their operation and output to less technical parties is of importance due to the legal requirement when used in commerical applications that could affect a user either through billing or automated decisions.

Chapter 3 elaborates answering R.Q.1:, where the REFIT dataset is introduced and explains its creation and curation as well as its many benefits in comparison with others datasets. Chapter 4 looks to answer R.Q.2: by investigating the data collected from the REFIT dataset and apply it to real world scenarios, in this case investigating participants change of appliance in an attempt to be more energy efficient. R.Q.3: was answered in collaboration with Nestec S.A. (trading as Nestlé) and is discussed in Chapter 5. Following the work done as part of R.Q.2:, Nestlé reached out to propose an industrial application of the work that would help them with product development, specifically Life Cycle Assessment (LCA) which look at the carbon footprint of appliance usage as part of a large process e.g. food preparation. In Chapter 6 covers both R.Q.4: & R.Q.5:, looking both at the implementation of a Gated Recurrent Unit (GRU) network with a comparative Convolutional Neural Network (CNN) network both trained identically to produce outputs for consumption and on/off values. This early work on a neural network NILM implementation is then followed by work on answering R.Q.5: how to understand the output of NILM networks. The intermediate layers of a network contain the encoded values of the input, extracting these from the trained network layers can help to explain the networks 'thought' process. The extraction of these values is then plotted in a heatmap to better explain the network process.

1.2 Contribution of Thesis

This research is based around the complete NILM process; from data collection, processing, dissemination, and process explainability.

In summary the main contributions of this thesis are as follows:

The collection and curation of one of the most extensive open source residential power datasets in the EU, with a wide variety of occupancy and appliance types. Described in Chapter 3.

Analysis of the REFIT dataset, time of use analysis, consumption patterns, and comparison across houses throughout seasons. Described in Chapter 4.

Proposal of data driven modelling methodology, taking into account real usage and user bias. The modelling was designed to improve on old lab based models which required a large number of variables and were designed specifically for use in LCA, the proposed method greatly simplifies and improves upon previous models, described in Chapter 5.

Proposing a neural network model which addressed both classification and regression for NILM. Incorporating both metrics in the same model helped to reduce complexity, and model training was done with no synthetic data. Demonstration of model transferability across distinct datasets namely REDD (USA) and REFIT (UK). Previously NILM results had only shown transferability across houses within the same dataset. Finally proposal of graphical plots capable of showing sequence to point network attention using a sliding occlusion window to aide explainability. Contribution described in Chapter 6.

1.3 Organisation of Thesis

Firstly Chapter 2 contains a full literature review across the different topics (data collection, life cycle assessments, and neural networks) covered in this research. Chapter 3 reviews the creation and curation of the REFIT dataset and how it provided a unique combination of features making it extremely valuable to the NILM community. Providing the methodology, post-processing steps taken and the data availability. The dataset was also made accessible to meet FAIR guidelines and the post processing resulted in two datasets: raw and cleaned. Chapter 4 introduces the requirement to keep appliance modelling up to date and the ability of smart meters to provide data which could be used to better understand the usage and energy efficiently of large populations, including the ability to predict future usage patterns. Chapter 5 further investigates appliance modelling and its relevance to industry groups via the life cycle assessment, a requirement for many products where the state-of-the-art has not been updated or requires lab based work to calculate. We aim to simplify these older models to simple variables which produce more accurate results. Chapter 6 looks at non-intrusive load monitoring conducted via the use of deep learning. Firstly looking at model transferability between highly distinct datasets and secondly making use of attention via occlusion to aide in explainability for non technical users.

1.4 Publications

Journals

1. David Murray, Lina Stankovic, Vladimir Stankovic, and Namy Daniela Espinoza-Orias. 2018. Appliance electrical consumption modelling at scale using smart meter data. Journal of Cleaner Production, Volume 187, 2018, Pages 237-249,

2. David Murray, Lina Stankovic, and Vladimir Stankovic. 2017. An electrical load measurements dataset of United Kingdom households from a two-year longitudinal study. Sci Data 4, 160122 (2017).

3. David Murray, Jing Liao, Lina Stankovic, and Vladimir Stankovic. Understanding usage patterns of electric kettle and energy saving potential. Applied Energy, Volume 171, 2016, Pages 231-242.

Conference Proceedings

1. David Murray, Lina Stankovic, and Vladimir Stankovic. 2021. Transparent AI: explainability of deep learning based load disaggregation. In Proceedings of the 8th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation (BuildSys '21). Association for Computing Machinery, New York, NY, USA, 268-271.

2. David Murray, Lina Stankovic, and Vladimir Stankovic. 2020. Explainable NILM Networks. In Proceedings of the 5th International Workshop on Non-Intrusive Load Monitoring (NILM'20). Association for Computing Machinery, New York, NY, USA, 64-69.

3. David Murray, Lina Stankovic, Vladimir Stankovic, Srdjan Lulic, and Srdjan Sladojevic. 2019. Transferability of neural networks approaches for low-rate energy disaggregation. International Conference on Acoustics, Speech, and Signal Processing, Brighton, United Kingdom.

4. David Murray, Jing Liao, Lina Stankovic, and Vladimir Stankovic. 2015. How to make efficient use of kettles: understanding usage patterns. Proceedings of the 8th International Conference on Energy Efficiency in Domestic Appliances and Lighting: EEDAL'15 27 Aug 2015, p. 1-13.

5. David Murray, Jing Liao, Lina Stankovic, and Vladimir Stankovic. 2015. A data management platform for personalised real-time energy feedback. Proceedings of the 8th International Conference on Energy Efficiency in Domestic Appliances and Lighting: EEDAL'15

6. Tom Hargreaves, Richard Hauxwell-Baldwin, Mike Coleman, Charlie Wilson, Lina Stankovic, Vladimir Stankovic, David Murray, Jing Liao, Tom Kane, Steven Firth, and Tarek Hassan. 2015. Smart homes, control and energy management: How do smart home technologies influence control over energy use and domestic life? European Council for an Energy Efficient Economy (ECEEE) 2015 Summer Study on Energy Efficiency

7. Tom Kane, Val Mitchell, Jing Liao, Lina Stankovic, Steven Firth, Stuart Cockbill, Charlie Wilson, David Murray, Farid Fouchal, Andrew May, Vanda Dimitriou, Vladimir Stankovic, and Tarek Hassan.

2015. Supporting Retrofit Decisions using Smart Meter Data: A Multi-Disciplinary Approach. European Council for an Energy Efficient Economy (ECEEE) 2015 Summer Study on Energy Efficiency

8. David Murray, Bochao Zhao, Georgia Elafoudi, Jing Liao, Lina Stankovic, and Vladimir Stankovic. 2014. Combined network coding and paillier homomorphic encryption for ensuring consumer data privacy in smart grid networks. Algebra, Codes and Networks 2014 - Bordeaux, France.

1.5 Author's Contribution to Publications

Journals

1. Appliance model research, development and testing of proposed generalised appliance consumption models for use in LCA's, paper writing, industry specific knowledge provided by N.D. Espinoza-Orias, supervisory input (paper writing) from Lina Stankovic and Vladimir Stankovic. Referenced in Chapter 6.

2. Dataset management, processing, curation, and distribution. Technical write up of curation process and background. Supervisory input (administration and paper writing) from Lina Stankovic and Vladimir Stankovic. Referenced in Chapter 3.

3. Appliance consumption research, development and testing of proposed model and paper writing, with supervisory input (background research, paper writing) from other three authors (Jing Liao, Lina Stankovic and Vladimir Stankovic). Referenced in Chapter 4

Conference Proceedings

1. Research and design of machine learning model capable of producing encoded time series heatmaps for explainability, paper writing. Super-

visory input (paper writing, background research) from Lina Stankovic and Vladimir Stankovic. Referenced in Chapter 6.

2. Research and design of attention layer within previously created model, paper writing. Supervisory input (paper writing) from Lina Stankovic and Vladimir Stankovic. Referenced in Chapter 6.

3. Research, design, and testing of proposed GRU model, and paper writing. Second author contributed CNN model design. Equal contribution with second author of dataset processing and analysis. Supervisory input (paper writing) from Lina Stankovic, and Vladimir Stankovic. Background research input from Srdjan Sladojevic. Referenced in Chapter 6.

4. Appliance consumption research, development and testing of proposed model and paper writing, with supervisory input (background research, paper writing) from other three authors (Jing Liao, Lina Stankovic and Vladimir Stankovic). Referenced in Chapter 4

5. Dataset curation and analysis, paper writing. Additional data analysis from second author Jing Liao. Supervisory input (administration and paper writing) from Lina Stankovic and Vladimir Stankovic. Referenced in Chapter 3.

6. Supporting author providing data collection and post processing advice for appliance activity.

7. Supporting author providing data collection and post processing advice for household consumption.

8. Design and implementation of a encryption model for smart meter data anonymisation, paper writing. Secondary background and advice by authors; Bochao Zhao, Georgia Elafoudi, and Jing Liao. Supervisory input (background research and paper writing).

Chapter 2

Literature Review

Background

The idea of the 'smart home' has featured in literature since the 1950's in these early days the technology was far from what we think of today. The current smart home in commercial literature focuses on the idea that appliances will be handle everything for the occupant, setting the temperature, controlling blinds, lights, music, hoovering, etc. All of which are designed to reduce the need for interaction from the occupants via routines, voice commands, mobile apps. In literature the focus is on the background systems and methodologies that makes this possible via, data collection, real time monitoring, sensing and the backend data processing and machine learning.

NILM is the concept of monitoring only household aggregate and from the consumption patterns estimating appliance energy usage. This idea was first introduced by George Hart, in his paper titled "Nonintrusive Appliance Load Monitoring" [54]. This thesis explores multiple aspects related to NILM from data collection & management, to visualisation and end user engagement. The literature review is broken down chronologically and reflects the content and focus of each chapter, firstly collection, through to modelling, prediction, NILM, and finally visualisation and explainability.

Data Collection

Until smart meters and smart devices are present in all homes, which can self report electrical consumption, there is a need for commercial or bespoke hardware and software. This is crucial to aiding in the creation of NILM algorithms, as with all algorithms, data is needed. Data collection is however a difficult and long process, which requires willing participants, and importantly reliable ones. Participation can also result in a bias towards certain groups within society, those who are technologically inclined, interested in research, and those willing to put up with some disruption to their lives. There is also a issue that even if smart meters are present in the majority of homes, getting access to this data, at frequencies that can be meaningfully used for research may prove to be difficult if not impossible due to data protection laws as well as data access from suppliers or governments.

If no smart meter is available, a monitoring device is required, many of which have propitiatory software and hardware requirements. The market is also very volatile as it is targeted at an audience which craves accuracy and ease of use. Many of the original offerings quickly became obsolete or lacked any sort of continued support or a sustainable business model, which made the choosing of a meter for a data collection project all the more important. For example Jack Kelly avoided using a commercial offering instead opting to build his own meter to create his dataset UK-DALE [78]. More recently metering offerings have inhibited the ability to extract raw data, in most cases hiding this behind an app or requiring technical knowledge to extract it from a webpage or via packet inspection e.g. TP-Link Smart Home App Kasa [139] reports values but they have to be extracted via Python.

A number of open source datasets exist currently with new ones being added, in some cases new datasets are continuations of a previous dataset but with new data covering a longer period of time, or the same houses but recorded with a new sampling rate or additional appliances/features as technology becomes available. More recently synthetic datasets have started to be made available, these are either curated collections of data created by a model or the model itself which is either trained or can be trained to mimic household usage given training data. Synthetic datasets or generators which can successfully mimic real houses are ideal for research purposes as they remove the largest obstacle in many projects which is the recruitment and retention of participants, as the creation of datasets for non-intrusive modelling is inherently intrusive. [82]

Table 3.1 contains a number of datasets and highlights a number of their features, the complexity of the dataset generally affects key areas such as frequency, data type, and duration. The higher frequency typically the shorter the duration for example.

The frequency of the data collected will greatly affect the type of deliverables a dataset can achieve, hourly would for example only suit usage patterns and predicted consumption, whereas high frequency e.g. 1Hz would support investigation into specific appliances usage. See [67] for one of the most up to date reviews of NILM datasets, describing 42 publicly available real and synthetic datasets. In addition to synthetic datasets, the models which can create these datasets are also being made available and important step in being able to understand and reproduce any results created using these datasets, the CREST model by Richardson [124] is one such example.

Appliance Modelling

Appliance load modelling involves estimating for a given appliance, the consumption based on a number of known factors. Initially there is a requirement to gather data from a range of appliances, this includes power consumption and specific settings. Knowing the setting is the most labour intensive part of recording data. It requires either the research team to run the appliance on every setting or for the household owner to note the setting used, the second being far more prone to incorrect data entry or apathy. Appliance modelling can be used in a number of ways, by manufacturers to better estimate usage patterns, consumption and commonly used features. By regulators to better understand how to regulate the market, and by power generator/distribution entities to better estimate the expected load.

The food industry is one of the world's largest contributors to carbon emissions, due to energy consumption throughout the food life cycle. In this thesis I focus on the residential consumption phase of the food LCA, i.e., energy consumption during home cooking. The LCA covers every aspect of energy consumption from the creation of raw ingredients to transport, etc. Specifically, while much effort has been placed on improving appliance energy efficiency, appliance models used in various applications, including the food LCA, are not updated regularly. This process is hindered by the fact that the cooking appliance models are either very cumbersome, requiring knowledge of parameters which are difficult to obtain or dependent on manufacturers' data which do not always reflect variable cooking behaviour of the general public. In [105] we proposed a methodology for generating accurate appliance models from energy consumption data, obtained by smart meters that are becoming widely available worldwide, without detailed knowledge of additional parameters such as food being prepared, mass of food, etc. Furthermore, the proposed models, due to the nature of smart meter data, are built incorporating actual usage patterns reflecting specific cooking practice. We validate our results from large, geographically spread energy datasets and demonstrate, as a case study, the impact of up-to-date models in the consumption phase of food LCA.

The food industry worldwide accounts for a significant fraction of carbon emissions. For example, in the United Kingdom (UK) alone, the food industry is responsible for about 14% of the energy consumption of the entire industry sector, equivalent to 7 million tonnes of carbon emissions per year [121]. The food LCA estimates material and energy input and output at all stages of the food product's life cycle − from acquisition of raw materials, production, processing, and packaging, to consumer use, and waste/recycling.

For many food products, including ready meals, domestic cooking takes up a significant fraction of the total energy consumption of the product's LCA [41, 130, 138] Although the energy consumption of domestic cooking in developed countries has decreased by 31% from 1991 to 2008 [138], mainly due to improved energy efficiency of cooking appliances, cooking and beverage preparation in domestic settings still consume a substantial amount of energy, or approximately 7 MegaJoule(MJ)/Kilogram(kg) [39, 53]. For example, according to [25], in the United States, domestic cooking accounts for 8-16% (equivalent to 6.9×10^8 Giga Joule per year) of the total national annual energy consumption [58]. Similarly, the report on Energy Consumption and Efficiency Trends in the European Union (EU) [13] estimates the energy consumption for ovens and hobs in private households in the EU-27 to be approximately 60 Tera-Watt hour (TWh) and states that for cooking appliances, there is still potential for energy savings. The situation is worse in many developing countries, where cooking consumes up to 90% of the overall residential energy consumption [1, 31], and it is mainly based on non-renewable energy.

Jungbluth [70] provides an early inventory of data for cooking at home using various appliances, so that food LCA studies can accurately quantify the food preparation stage. Sensitivity analysis of efficiency (ratio of energy input to energy output) shows that the type of vessels used, type of electrical appliance used (grill, oven, microwave oven, cast iron plate, induction stove), and the preparation method, have a large influence, resulting in the conclusion that there are large differences between efficiencies. Lakshmi et al. [86] tested different methods to cook rice in a microwave with different power levels, including the influence of soaking the rice prior to cooking. It was concluded that an electric rice-cooker is more energy efficient despite a longer cooking time compared to the microwave, but microwave cooking is as energy efficient as using a pressure cooker. A similar study on energy efficiency of cooking rice is presented in Das et al. [30], but only an electric rice cooker and a pressure cooker are tested.

Zufia and Arana [155] carried out the LCA of an industrially cooked dish, namely cooked tuna with tomato sauce, to assess the environmental impact of the production and distribution. However, the electricity consumed at home by a microwave for heating the ready meal is not considered. Calderón et al. [18] model the microwave consumption using its power rating, and focus on LCA of a canned ready meal, a stew product based on cooked pulses and pork meat cuts, highlighting subsystems with the highest environmental loads, concluding that 11% of the total energy consumption in the product life cycle is attributed to the domestic level. More recently, in [17], environmental burdens of the same dish at four production scales is considered, including canned food (reheated at home), the restaurant (cooked and heated in a traditional way and served), and the home-made dish with electricity consumption at household level estimated from electrical appliance specifications.

Oberascher et al. [113] highlight the variability of behaviour of consumers in using electrical appliances at the domestic level and provides empirical data on electricity consumption for a number of cooking processes, including heating water, baking potatoes and boiling eggs, concluding that the energy consumption of the microwave for heating water is lower than stove with a pot with and without a lid, but higher than kettle. Experiments conducted by Vattenfall [142] show that using the most energy efficient appliance in the kitchen for a specific cooking job is undoubtedly an effective way of lowering energy consumption. However, Fechner shows differences of up to 50% in energy consumption when six chefs all cooked the same meal with the same equipment (cited as per [146]), which agrees with the findings of DeMerchant [32]. Kemna expects no further energy-saving potential possible for electric ovens from 2010 to 2020 from a purely technical design point of view; however, with a change to more sustainable consumer behaviour, the additional potential for energy savings is expected to be about 10% [79].

The literature survey above demonstrates a clear need to capture the effect of different cooking styles, and consequently variable appliance usage, and explore the impact of the consumer phase, particularly at domestic level, in food LCA studies. This requires wider studies that can only be enabled via accurate and *scalable* models for computing the energy aspect of end-user cooking at home, and not relying on appliance manufacturer's specifications to estimate the consumption.

To this end, in [130], general models for quantifying energy consumption related to the food preparation in private households are proposed, including frying, boiling, oven roasting and microwave cooking. This is achieved via appliance load modelling, through exhaustive tests in laboratory conditions using different appliances and cooking settings. However, the models generated require a large number of variables to be known, such as type of food, its mass, the evaporation mass of food, which is not possible at scale. Similarly, the modelling of the combined oven in a restaurant setting, proposed recently in [16], also includes a large number of parameters which are not approachable at scale. Industrial scale applications already have a number of studies conducted, which model industrial versions of household appliances, such as modelling the cooking properties of industrial bread ovens [116]. These, however, are very specific to the industry in question and therefore non transferable methodologies are applied.

This work is inspired by some of the work done over a decade ago [130], and the appliance modelling performed, albeit in a different context, e.g., [99] where lines of best fit are applied for each setting of washing machines and [47] that use online survey, in-home study and laboratory experiments to assess the energy consumption of refrigerators. The state-of-the-art microwave (MW) energy consumption model in [130] requires knowledge of the food being cooked as well as water content, which is infeasible to collect at scale. A similar microwave model is used in [86]. We note that other studies, such as [17, 18, 86, 113], rely on heuristic measurements in laboratory conditions, or focus on preparation of particular dishes, or rely on power rating of the appliances and cooking recipes with the risky, and sometimes wrong assumption, that they will be followed correctly. The model for an oven, validated in [130] using 23-59 litre ovens and data supplied by The Swedish Consumer Agency (`http://www.konsumentverket.se`) which had oven volumes ranging between 18-65 litres, requires a large number of variables to be known to estimate consumption again showing that at scale outside of a laboratory these models are not applicable.

This work proposes a methodology for generating *general appliance models* in a scalable manner to quantify energy consumption related to the usage of cooking appliances at domestic level. Specifically, our research hypothesis is that only smart meter data, comprising active power measurements with at least 60 second sampling in a similar format to the UK SMETS version 2 [114], is sufficient for building accurate major cooking appliance models. To prove our hypothesis, we draw upon load profiling, appliance mining and user activities assessment methodologies related to energy demand literature [104, 132]. For example, in [104], smart meter data is used to assess energy efficiency and sustainability of electric kettle usage. In [132], energy demand of different energy consuming domestic activities such as cooking and laundering is quantified. Load profiles of 11 major domestic appliances is studied in [120] showing that the majority of cooking load profiles are single state even when there may be large variation during the operation of the appliance.

While the proposed methodology is generic to all cooking appliances, for ease of understanding, the methodology is illustrated on two electric cooking appliances with wide ownership in most countries, namely the electric oven and microwave. According to [64], almost 70% of households with ovens in England have electric ovens and just under 30% have gas ovens. Thus, in our study we chose to focus only on electrical appliances as they have a market dominance and the eventual move away from fossil fuels should increase the ownership of electrical appliances.

The appliance models proposed have a two-fold advantage over existing models: (i) models reflect current, more energy efficient and wider capacity appliances and (ii) models are not reliant on parameters that are difficult to estimate by non-specialists in large field studies. The proposed models are validated with existing literature, as well as on large electrical measurements datasets where electrical consumption of the two appliances in question was recorded. Furthermore, in order to demonstrate applicability and impact on food LCA, a case study is provided where figures for the microwave oven and electric oven are calculated for frozen ready meals and compared with recent LCA for the same food [129].

Non-intrusive Load Prediction

NILM has been researched for over 30 years [54] and has become an active area of research again recently due to ambitious energy efficiency goals, smart homes/buildings, and large-scale smart metering deployment programmes worldwide.

Different approaches have been proposed for NILM, using various signal processing and machine learning techniques (reviews can be found in [148, 154]). Approaches proposed include include Hidden Markov Models (HMM)-based methods and their variants (see, e.g., [51, 117]), signal processing methods, such as dynamic time warping [22, 43, 89], single-channel source separation [84], graph signal processing [55,152], decision trees [89], support vector machines with K-means [3], genetic algorithms [42, 150] and neural networks [126].

The recent increase in the availability of load data [78, 84, 107] for model training, has ignited *data-driven* approaches, such as Deep Neural Network (DNN) using both CNN, and recurrent neural network (RNN) architectures [7, 35, 77, 80]. Currently, DNN-based NILM relies on creating a new network for each house and each appliance. Some approaches use a single model to predict the output of several appliances, although these are more recent developments due to the complexity of training. With the availability of a sufficient and good training datasets, these networks perform well as they are highly targeted, but if NILM is to become widespread and scalable, networks will need to be trained on a wide range of electrical load signatures. As such, the challenge is to design a single network to accurately disaggregate any appliance across multiple 'unseen' houses, i.e., houses not present in the training dataset a.k.a transfer learning. Transfer learning is extremely difficult due to the vast amount of different make and models of appliance, this is made even more challenging by changes to appliance operation to meet new any new energy efficiency standards that might come into law, or a new technology will fundamentally change how the appliance operates completely changing it's consumption signature. The introduction of heat pump tumble dryers is an example of this. [11]

Though the previous DNN-based approaches demonstrated competitive results [7,35,46,56,77,80,149] , they do not fully exploit the DNN potential. Indeed, the approaches of [77] and [35] are limited by generation of synthetic activations, which do not necessarily capture 'noise' well, here defined as unknown simultaneous appliance use, usually present in the dataset. In Maucnh's papers [94,95], an long short-term memory (LSTM) & DNN-HMM approach was used to rebuild the appliance signal but due to the difference in aggregate and sub-metered sampling rates in the REDD dataset, synthetic data was used exclusively in both papers, created by summing all sub-meters to create a synthetic 'aggregate'; this limits the amount of noise as appliances not sub-metered would be excluded. [80] uses real "noisy" dataset, but requires thousands of epochs to generate accurate results, which is not a feasible approach for online disaggregation, while the architecture of [7] contains large number (i.e., 44) layers designed only for identification of appliance state, without generating disaggregation or load consumption estimations.

Auto-encoders have become one of the go-to network designs when investigating NILM, the ability to rebuild the appliance signal is significant both in state detection and more accurate load prediction. Auto-encoders can contain a number of the layers previously mentioned but generally follow a specific hour glass shape rather than specific layer types. There have also been differences in approach when it comes to output. Sequence-to-Sequence, where the appliance signal is recreated in it's entirety from an input window, and Sequence-to-Point where only a single output value is generated per window. These two approaches have different benefits, S2S means that less processing would need to take place across a dataset but depending on the step value could mean that full appliance activations would not be fully captured (dependent on the appliance and the window length). Sequence to point on the other hand moves across the appliance activation and although the result may fluctuate more as the model is making more predictions, with post-processing it is possible to better construct an appliance than if a sequence to sequence did not correctly predict.

Another issue with NILM literature is that there is many ways of assessing the quality of the results. Generally the most commonly accepted methods are by the use of MAE, however due to the nature of some appliances the usage over a long period may only be less than a percent of the recorded time. [52] Even a kettle which is used multiple times a day would only be on for around 1-5 minutes at a time. This means that MAE becomes a poor indicator of quality, if the model were only to predict zero the average error would be very close to 0. To combat this a number of NILM specific metrics have been devised, in the following paper they discuss the merits and issues associated with these metrics [118], this paper also makes use of [98] which uses a number of the metrics and ranks them in order of of usefulness, with Match Rate scoring the best, even then there is a caveat that the metric has to be used in a way which makes it meaningful between heavy and light users, and across different time scales.

Metric	Description	Equation Energy
RE	Relative error [98] indicates quality of energy estimation relative to ground-truth.	$\sum_{i=1}^{N} \frac{E_i - \hat{E}_i}{\hat{E}_i}$
RMSE	Root mean square error [9,61,98]. The RMSE reports based on how spread-out errors are, and is the most commonly used as it reports in the same unit as the data.	$1 - \sqrt{\frac{1}{N}\sum_{i=1}^{N}(E_i - \hat{E}_i)^2}\bar{E}$
AE	Average error [61,98] indicates if estimation is above or below the expected.	$\frac{1}{N}\sum_{i=1}^{N} \Delta E_i$
r^2	R-Squared [61,96,97,98]	$1 - \sum_{i=1}^{N} \frac{(E_i - \hat{E}_i)^2}{(E_i - \bar{E}_i)^2}$
EE	Energy Error [9,98] the ratio of the absolute difference between estimated and true energy, and the total amount of true energy.	$\frac{\sum_{i=1}^{N} \lvert \hat{E}_i - E_i \rvert}{\sum_{i=1}^{N} E_i}$
MR	Considered the best NILM metric by [98], varying between 1 (strong match) and tending towards 0 for a poor match)	$\frac{\sum_{i=1}^{N} min\{E_i, \hat{E}_i\}}{\sum_{i=1}^{N} max\{E_i, \hat{E}_i\}}$
Fraction of Energy Explained	[50,54]	$\frac{\sum_{i=1}^{N} \hat{E}_i}{\sum_{i=1}^{N} E_i}$
E_t	Metered energy at time interval t	
$\bar{E}$	Average metered energy over the dataset	
$\hat{E}_t$	Estimated energy at instant t	
ΔE_t	Error between NILM and metered data at instant t	

Table 2.1
Methods used to determine the ability of NILM algorithms energy estimation capabilities. A subset of the full table which can be found in Pereira [118].

The quality of NILM metrics however does not represent the quality of the NILM model. In many cases the test data will be part of the open dataset but the authors will not specify the time period used or quantify the complexity of the time period used, an example might be a very strong test score but the data only contains couple of very different appliance signatures. So a good result may be obtained on very clean data and score very well on the metrics but this would not be representative of data during a busier time of the day where the same model may produce poor results. This is fundamental to explaining a 'good' NILM model and acknowledging false positives are important as this is the main issue that affects consumer trust, more than false negatives where the expectation is that the result will improve.

Recently attempts to improve the description of data has been made by Klemenjak [83]. In this paper a metric which describes the noisiness of the data is described as noise-to-aggregate ratio (NAR). This is an important metric to be used in conjunction with others on standardised test sets. The ability of a network to score well in a noisy test environment will help to show it's ability to generalise and therefore perform better in transfer learning and widespread deployment.

To help show the objective performance of a given NILM model it would be worth using specific benchmark time periods where there are a number of known appliances, a realistic amount of noise and so on. This would then be made available as a dataset, with which all models are compared against. Currently papers where comparisons are made using the same model designs but are unlikely to be trained in the same way as the original authors did, unless they made their model weights available which very few have.

This thesis looks at the application of sequence to point models and their transferability across datasets, making use of both normal and synthetic datasets. I describe the challenges of multi-state appliances, and the difficulties with universal NILM models which are capable of detecting multiple appliances (similar to image classifiers which can handle 10's to 1000's of classes [85]).

Chapter 3

Data Collection & Curation

3.1 Introduction

This chapter looks at the challenges of collecting smart home data from households which do not have smart metering installed and dealing with issues such as erroneous & missing data which can occur with off the shelf solutions (specifically if they rely on external servers e.g. TP-Link energy monitoring smart plugs [139]). As discussed in the literature review, a number of datasets have been created with focus on 1 characteristic over others, such as extremely high sampling frequency, but sacrificing duration e.g. a number of hours/days, or scale, e.g. containing only a single house. This can be seen in Table 3.1.

The approach taken to create the REFIT dataset means that observation frequency, duration and sample size are all well balanced resulting in a dataset which contains significantly more information than others, remaining generally useful rather than specifically useful. Households involved are also varied, giving an overview of middle class English households over a two year period. This gives the REFIT dataset an advantage over other datasets when training neural networks, the range of appliances and frequency of use, along with challenging and noisy conditions means that networks can be

Dataset	Location	Duration, Year	No. Houses	Energy Sensors	Data Recorded	Readings Freq.
ACS-F1 [49]	CHE	2*1 hour sessions, 2013	N/A	100 App. (10 types)	V, I, f, P, Q, Φ	10 secs
ACS-F2 [125]	CHE	2*1 hour sessions, 2013	N/A	225 App. (15 types)	V, I, f, P, Q, Φ	10 secs
AMPds [92]	CAN	1 year, 2012	1	21 App.	V, I, f, pf, P, Q, S, E	1 min
AMPds2 [91]	CAN	2 years, 2012	1	21 App.	V, I, f, pf, P, Q, S, E	1 min
BLUED [4]	USA	8 days, 2011	1	Agg.	V, I	12 kHz
DRED [141]	NED	6 months, 2015	1	Agg., 12 App.	P	1 Hz
ECO [10]	CHE	8 months, 2012	6	Agg., 6-10 App.	V, I, P, Q, Φ	1 Hz
GREEND [103]	AUT, ITA	1 year, 2013	9	9 App.	P	1 Hz
HES [153]	GBR	1 month (255 houses), 1 year (26 houses), 2010	251	Agg., 1-10 Sub., 13-51 App.	P	10 min
iAWE [8]	IND	73 days, 2013	1	Agg., 10 App.	Agg. V, I, f, P, Q, S, E, Φ App. V, I, f, P, S, E, Φ	1 Hz
IHEPCDS [57]	FRA	4 years, 2006	1	Agg., 3 Sub.	Agg. P, Q Sub. E	1 min
REDD [84]	USA	3-19 days, 2011	6	Agg., 9-24 App.	Agg. V, P App. P	Agg. 15 kHz App. 3 secs
REFIT	GBR	2 years, 2013	20	Agg., 9 App.	P	8 secs
Smart* [6]	USA	3 months, 2012	3	House A. Agg., 26 Sub., 55 App. House B, C. Agg., 21 Sub.	Agg. V, f, P, S Sub. V, f, P, S App. P	Agg. 1 Hz Sub. 1 Hz App. 2.5 secs
Tracebase [123]	DEU	1883 days, 2012 onwards	15	158 App. (43 types)	P	1 Hz
UK-DALE [78]	GBR	655 days, 2012	5	Agg., 5-54 App.	Agg. V, I App. P	Agg. 16 kHz App. 6 secs
AMBAL [15]	Synth Res		N/A	14 App.	App. P	App. 1 Hz
SmartSim [20]	Synth Res	7 days	N/A	25 App.	App. P	App. 1 Hz
SHED [59]	Synth Com	14 days	N/A	66 App.	App. P	App. 0.033 Hz
SynD [82]	Synth Res	180 days	N/A	21 App.	App. P	App. 5 Hz

Table 3.1
Household Power/Appliance Open Access Datasets
Agg. = Aggregate, App. = Appliance, Sub. = Power circuit, e.g., the fuse which all appliances in a single room are connected to.
Types is in relation to appliance groups in situations where only appliances were monitored.
Active Power (P), Reactive Power (Q), Apparent Power (S), Energy (E), Frequency (f), Power Factor (pf), Phase Angle (Φ), Voltage (V) and Current (I).
ACS-F1 and ACS-F2 datasets contain appliance signatures obtained in a laboratory setting instead of real homes.

trained to generalise well, meaning that they can provide meaningful results across a range of households and appliances. The time scale included also means that data can take into account the change of seasons, and therefore seasonal usage patterns that other datasets may not capture. Through appliance identification, and demonstration of clean continuous data we show that the REFIT dataset is a useful resource for researchers looking to model energy appliance energy consumption or usage patterns. Additionally the REFIT makes for a challenging benchmark dataset for both academic and commercial entities due to the large number of appliances contained in each household [83]. The content of this chapter is taken from the paper on 'An electrical load measurements dataset of United Kingdom households from a two-year longitudinal study' [107] which as of writing has been cited over 160 times.

3.2 Methodology

The following section contains the steps needed to create a dataset similar to REFIT. Focusing on the reasoning behind what was included and the methods used to present the finalised dataset.

3.2.1 Selection

The first step in the creation of any real world dataset is recruitment. The participants in the REFIT study were recruited via email and leaflet drops. In total, 57 households replied with basic information about their household. [73] Final selection was based on the householder's familiarity with information and communication technology (ICT) as well as a mix of household occupancy, including retired couples, working couples and families with children ranging from infants to young adults. Some houses were excluded for a number of reasons, mainly related to connectivity, such as utility meters being underground meaning that signal acquisition would be difficult, or absence of a broadband connection [72]. Occupancy and physical characteristics of each house relevant to electricity consumption is shown in Table 3.2. Ethics approval was granted by Ethics Approvals (Human Participants), Sub-committee, Research Office, Loughborough University and all participants gave informed consent and understood how their data would be used. This selection process does mean that there is an inherent bias towards middle class households due to the exclusion criteria described in [73], as well as as those likely to respond to this type of recruitment are people who would regularly seek out new technology and engage with it are likely to be tech savvy. The physical requirement of accessible metering also excluded smaller terraced houses typical in the study area, meaning that larger households with higher incomes and a larger number of appliances, specifically technology and pet related appliances would be included which added to the amount of background noise. [73]

House	Occupancy	Dwelling Age	No. of Appliances	Dwelling Type	Size
1	2	1975-1980	35	Detached	4 bed
2	4	-	15	Semi-detached	3 bed
3	2	1988	27	Detached	3 bed
4	2	1850-1899	33	Detached	4 bed
5	4	1878	44	Mid-terrace	4 bed
6	2	2005	49	Detached	4 bed
7	4	1965-1974	25	Detached	3 bed
8	2	1966	35	Detached	2 bed
9	2	1919-1944	24	Detached	3 bed
10	4	1919-1944	31	Detached	3 bed
11	1	1945-1964	25	Detached	3 bed
12	3	1991-1995	26	Detached	3 bed
13	4	post 2002	28	Detached	4 bed
15	1	1965-1974	19	Semi-detached	3 bed
16	6	1981-1990	48	Detached	5 bed
17	3	mid 60s	22	Detached	3 bed
18	2	1965-1974	34	Detached	3 bed
19	4	1945-1964	26	Semi-detached	3 bed
20	2	1965-1974	39	Detached	3 bed
21	4	1981-1990	23	Detached	3 bed

Table 3.2
Additional information about the houses involved in the study. Occupancy column shows the number of people living in the house during the monitoring period. Number of Appliances shows the total number of electrical appliances in the house based on the conducted house survey. Size is given as number of bedrooms as this is a more common representation of dwelling size in the UK.

In each house, nine appliances were selected to be monitored via individual plug meters. In some cases an extension cord, increased the number of appliances that were recorded on a single plug meter. Appliance selection was motivated by the completed Household Electricity Survey (HES) [153], a large study conducted by the UK's Department of Environment and Climate Change (DECC). Since the HES study focused on collecting a large amount of data about consumer attitudes towards energy consuming practices and energy demand, the study prioritised appliances with relatively high electrical consumption and/or frequent use to be monitored.

With respect to the monitoring priorities from HES (see Appendix II of the HES report [153]), the main appliances from the energy demand point of view are cold appliances (refrigerators, freezers, fridge-freezers, etc.), cooking appliances (microwaves, kettles, etc.), ICT (computers, screens, printers, consoles, etc.), utility room appliances (washing machines, tumble dryers, dishwashers), while low priority items include mobile phone chargers, hair straighteners and similar small items which may not be used regularly or moved frequently.

In the REFIT study, this HES prioritisation list was used as a guide when selecting appliances to monitor, unless study participants explicitly requested monitoring unusual appliances, such as a vivarium or pond pump. Table 3.3 lists all appliances monitored in each house. Column 4 in Table 3.2 shows the total number of electrical appliances in each house according to the house survey obtained at the beginning of the study. Note that all REFIT study houses used gas central heating and hot water systems as primary source of fuel and there were no other HVAC systems present (although some houses made use of mobile space heaters during winter months), in the UK this is highly typical but does mean that the dataset lacks a number of key consuming appliances that would be found in the rest of the world such as air conditioning, immersion water heaters, and heat pumps [2]. Several houses did have solar panel installations, which provided instantaneous power in which excess was sold back to the power grid. At the time of the study domestic battery storage was uncommon.

	House Number																				
Appliance	1	2	3	4	5	6	7	8	9	10	11	12	13	15	16	17	18	19	20	21	Total
Television	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗		2✗	2✗	✗	✗	✗	✗	✗	✗	✗	21
Hi-Fi		✗							✗		✗							✗			4
Fridge-Freezer		✗	✗	✗	✗				✗	✗	✗	✗		✗	2✗	✗	✗			✗	14
Fridge	✗			✗			✗	✗			✗						✗	✗	✗		7
Freezer	2✗		✗	✗		✗	2✗	✗		2✗						✗	✗	✗	✗		13
Microwave		✗	✗	✗	✗	✗		✗	✗	✗	✗	✗	✗	✗		✗	✗	✗	✗		16
Cooker Hood		✗																			1
Kettle		✗	✗	✗	✗	✗	✗	✗	✗		✗	✗	✗	✗		✗		✗	✗	✗	16
Toaster		✗	✗		✗	✗	✗	✗				✗		✗							8
Misc Kitchen										2✗								✗		✗	4
Washing Machine	✗	✗	✗	2✗	✗	✗	✗	✗	✗	✗	✗		✗	✗	✗	✗	✗	✗	✗	✗	20
Washer Dryer									✗								✗				2
Tumble Dryer	✗		✗		✗		✗	✗					✗	✗		✗			✗	✗	10
Dishwasher	✗	✗	✗		✗	✗	✗		✗	✗	✗		✗	✗	✗		✗		✗	✗	15
Computer	✗			✗	✗	2✗		✗			✗		✗	✗	✗	✗	✗		✗		13
Router											✗										1
Electric Heater	✗								✗						2✗						4
Lamp																		✗			1
Misc															✗	✗				2✗	4

Table 3.3
Monitored appliances in each house organised as shown in HES Appendix II [153]. The total number of appliances of the same type monitored are shown in the final column. Small/unique appliances are grouped into 'Misc' or 'Misc Kitchen' as they may only appear in one house.
Misc. Appliances include: House 21 - Pond Pump & Vivarium, House 16 - Dehumidifier, House 17 - Bedroom Plug.
Misc. Kitchen: House 10 - Mixer & Blender, House 19 - Bread Maker, House 21 - Mixer.

3.2.2 Monitoring Set-up

Figure 3.1 shows the schematic of the data collection platform. To ensure reliability, scalability and performance, all equipment used in the REFIT study was commercial off-the-shelf hardware available for purchase at the beginning of the study. Off-the-shelf hardware had the benefit of already being tested, with support available for issues, and importantly an accessible API with which the team was able to access the data without action being required by the participants.

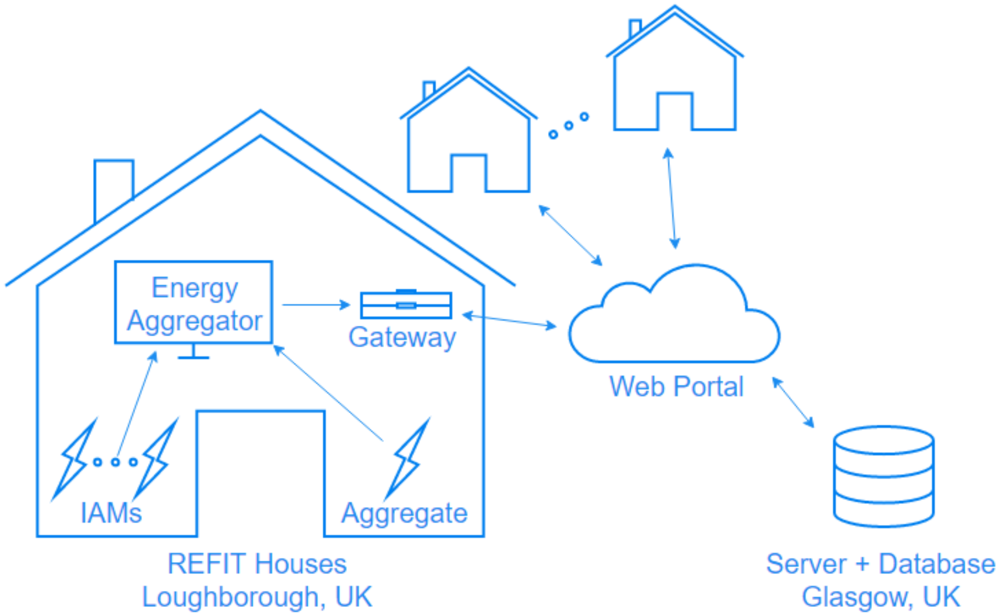


Fig. 3.1. REFIT remote real-time data collection schematic.

Energy sensors (nine in each house) wirelessly sent readings every 8 seconds to an energy aggregator, which was connected to a communications gateway. The gateway, connected to the broadband router, forwarded readings to a web portal. From the web portal, the data was requested by a server at Strathclyde and stored in a MySQL database.

The overall platform is designed to be as similar as possible to a typical smart meter installation [114] in terms of data collection and in-home presence. Indeed, the used aggregator comes with an IHD that displays usage information and basic historical statistics similar to what will be available after smart meter roll-out [114]. However, we note that individual appliance monitors (IAMs) will not be part of the smart meter installation [114] and it will be up to the utility supplier if one is provided or not, but they are helpful to correlate use times and power usage, design, model, test and validate analytical approaches.

In the following, we describe each component of the monitoring platform.

Household Aggregate

The most important measurement in each house is the household aggregate energy consumption as this will imitate what smart meters will be able to provide. The household aggregate was measured by a CurrentCost transmitter (specifications: [26]), which contains a single phase current clamp and a transmission module which wirelessly transmits readings every 8 seconds using Radio Frequency (RF) 433MHz to the energy aggregator. These were installed in all of the houses within the study in a single phase configuration which is the most common in households within the United Kingdom. The aggregator used was a CurrentCost EnviR module that also contains an IHD, this was hidden from the study participants for a period of time, to study the effectiveness of the information the IHD provided.

CurrentCost monitoring equipment has been used successfully in many previous trials, e.g., in trials [23,66,74,78,122,136,137], which is why it was chosen over other options available on the market around the time the study started. It should be noted that the sensor (a split-core current transformer) does not measure mains voltage, thus there may be variation in the calculated active power. The manufacturer did not give any details with regards to the internal workings of their sensors, however testing quantifies their relative error of around 6% [78]. The value of 6% aligns with the UK's declared voltage tolerance which is -6% to +10%.

Fig. 3.2. Current Cost in home display & hub

Of the 20 households recruited, six homes had solar panels installed. In three cases rewiring was done to remove the effect of solar panel generation (Houses 1, 6 & 7). In the other three (Houses 3, 11 & 21) rewiring was not possible and the aggregate of these houses was recorded as is with solar interference as the sensor used to measure aggregate energy consumption was unable to distinguish the direction of power flow, solar panels appeared as additional power consumption resulting in a bell-curve-shaped power consumption increase during the day with significant noise due to weather changes, such as clouds passing overhead.

Individual Appliance Monitors (IAM)

Each house was supplied with 9 CurrentCost IAMs, which is the maximum number supported by the associated EnviR module without the likelihood of causing data loss from transmission collisions. Each IAM provided the power consumption (in Watts) for each appliance which was connected, at a sampling rate of 8 seconds. All IAM readings were collected via the EnviR aggregator which was then connected to the communications gateway.

Similarly to the household aggregate, the IAMs only monitor the current and not the voltage which means there may be a variation in the supply voltage which introduces an error in the reading. By default, all of the installed load monitoring devices had the voltage pre-set to 240V, suitable for the UK where mains voltage is rated at 230V +10% to -6%.

IAMs were only capable of broadcasting their readings, which results in the readings not being synchronised with the aggregate readings (discussed in this GitHub [76], the type of plug used was the 'CC_TX'). The timestamp assigned to data was the UNIX timestamp when the data was received at the Strathclyde server. Since a data request grabs data from the aggregator, which is the aggregate and last broadcasts from all the IAMs, the timestamp is the sampling time of the aggregate reading. Thus, all of the IAMs received at timestamp n+1 will be lagging the aggregate reading by up to the time since the last sample (n), that is, the offset between IAM I and Aggregate A readings will always satisfy n < offset < n+1, where n and n+1 are two consecutive sampling time of aggregate reading. See Subsections Code Availability and Known Issues. Figure 3.3 shows a time representation of the readings from the Current Cost system. Note that each sampling time period is of 8sec length.

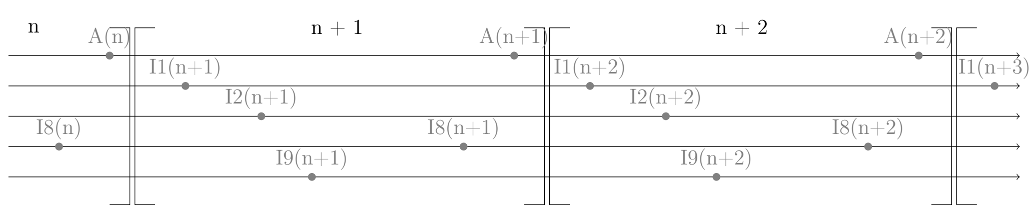


Fig. 3.3. Each circle represents a reading taken. Each line represents a sensor, A being the aggregate and I1 representing IAM1 and so on. The numbers in the bracket represent sample number.

The appliances monitored were recorded during initial installation and households were advised not to unplug or move monitors during the monitoring period without notifying the REFIT project team. Any changes in the appliance monitored by an IAM that occurred during the trial are recorded in detail in the ReadMe file included with the dataset. An example of this would be House 10, which moved IAM 2 from a Freezer to a Toaster on 25/06/2014.

3.2.3 Energy Aggregator

The EnviR aggregator with an IHD came bundled with the CurrentCost transmitter used for measuring the household aggregate. The EnviR [27] ties all of the CurrentCost devices together acting an an energy aggregator. Its display provides information about all of the CurrentCost devices which were installed with a simple interface using buttons as navigation. Together with the transmitter for aggregate measurements this pairing best represented the combination of smart meter and IHD that would be supplied as part of the UK roll out (the supply on an IHD was supplier dependant [33]). The EnviR communicated via a USB cable to the communications gateway allowing data to be recorded remotely.

Communications Gateway

The communications gateway used in the REFIT project for electrical measurements was the Vera3 smart home controller [145]. All sensors reported data wirelessly to the EnviR which then forwarded information to Vera3 using a USB connection. Finally, Vera3 sent the data to the cloud, which was an on-line dashboard available via the Vera Control (formally MiCasaVerde [44]) on-line portal [144]. Vera3 has a number of interfaces to enable additional monitoring with the following technologies: WiFi, USB, LAN, and Z-Wave. In the REFIT study, Vera3 was connected to a home broadband router via its LAN interface, to EnviR via the USB interface, and to additional sensors (measuring temperature, humidity, light intensity) via the Z-Wave interface.

Web Portal & Data Collection

Data collected through the communications gateway was available on the web portal via an application programming interface (API). The API requests are available remotely and all REFIT houses were linked to a single web portal account with a user account for each household (so that household could benefit from basic energy feedback available via the web portal) as well as an administration account. The list of available requests can be found at [28, 29]. Responses are given by default in the JavaScript Object Notation (JSON) formatting language.

The simplest set-up was to make a call for each sensor in a house individually. However, there are several issues with this configuration: (1) the number of requests being sent from the same account ID will be over 200, i.e., every 8 seconds (20 houses $\times$ 10 requests per house), (2) if the CurrentCost is reset or connections to IAMs lost, different ID numbers could be assigned to these IAMs. To eliminate this possibility our python scripts requested only sensor values which had changed and used the sensor's universally unique identifier, this was a feature of the CurrentCost API. This method was more robust as only 20 API calls were made every 8 seconds. Furthermore, this reduced bandwidth and storage requirements but does introduce the synchronisation issue as mentioned in section 3.2.2.

Server & Database

Requests for new data were issued to the web portal and the replies were recorded on the MySQL server hosted at the University of Strathclyde, Glasgow UK, with the following specification:

DELL PowerEdge R320 with an Intel® Xeon® E5-2407 and 16 gigabytes of RAM, running on Linux Debian V8.3 with MySQL 5.5.47.

Checking the connectivity of houses was done via a web page which displayed time passed from the last insert for each house by the Readings.py script. Any home which had not updated recently would show a large time difference and that home was then be contacted to check if the any of the in-home kit (aggregate sensor, IAMs, aggregator, gateway) had been inadvertently moved or unplugged. A similar page was constructed which showed all IAMs.

Initially code had been written in the Perl programming language, which was subsequently replaced with Python code due to Python's readability and versatility.

3.2.4 Code Availability

The code used to collect and check data and monitor the collection process is available at `https:\github.com/David-Murray/REFIT`. The code runs using Python 3 on a Debian server. The time-based job scheduler CRON, available on most Unix distributions, is used to run some segments of code at particular time intervals. The following python scripts are available:

HouseUpdater.py : This python script was responsible for keeping the information about the houses monitored up to date including the server address which API should be made to; this was run hourly via CRON.

SensorUpdater.py : This python script kept the list of sensors within the houses up to date; this would also record when the sensor was last checked which helped to show any sensors that were no longer available.

ReadingsTaskMaster.py : A python script which generated child processes for each house (see Readings.py); once each house's script was running it would check for new houses that had been added/changed or had come back on-line and would restart their Readings.py script if required.

Readings.py : A python script that would run continuously querying and inserting sensor values into the database every 8 seconds. The reading time was determined by the time at which the server received the response from the API call. As IAMs were only able to broadcast their readings every 8 seconds they will not be synchronised with the reading of the aggregate. This means that the time associated with a record may have IAM readings up to 7 seconds old.

3.2.5 Known Issues

- CurrentCost IAMs: Occasionally, IAMs reported readings much greater than the maximum load for standard household appliances, i.e., 4000 Watts (W), due to sensor malfunctioning. These readings were removed from the raw data.

- CurrentCost IAMs: Reporting time synchronisation - although data was recorded at set intervals for all devices, the time between when IAMs reported a reading will not be in synchronisation to the current second and therefore may show a mismatch to the aggregate e.g. the aggregate reading was less than an individual appliances reading.

- Houses' 3, 11 & 21 aggregate readings are affected by solar panel generation as re-wiring was not possible.

- In some cases the step change in values will differ between IAM and Aggregate, as the CurrentCost system did not monitor other variables to adjust for voltage or phase angle, this is caused by inductive and capacitive loads. The difference should be accounted for with the knowledge of the monitoring errors from both Aggregate and IAM sensors.

3.3 Data Records

The REFIT Electrical Load Measurements' Dataset is available in the form of CSV files. Each house has one associated CSV file containing all aggregate and IAM measurements for the entire monitoring period.

The data has been cleaned by correcting the date/time due to British Summer Time (BST), removing IAM spikes of greater than 4000 W, and forward filling gaps of less than 2 minutes with previous values or, if the gap is larger than 2 minutes, filling with zeros, and moving data streams where appliances had been switched between plugs so there is a continuous record of each appliance (see Algorithm 1).

Data: $Time$, $Power$
Result: Forward fill NaN values of time gaps less than $<$2 minutes
Start;
for $n \leftarrow 2$ **to** $length(Power)$ **do**
  if *time(n)-time(n-1)* $<$ *120 seconds* **then**
    $power(n) = power(n-1)$;
  else
    $power(n) = 0$;
  end
end

Algorithm 1: Pseudocode for removing Not-a-Number (NaN) values in data.

Datetime is in the format Year(YYYY)-Month(MM)-Day(DD) hours(HH):minutes(mm):seconds(SS). The CSV files have the following columns:

- DateTime [YYYY-MM-DD HH:mm:SS]
- UNIX Timestamp
- Aggregate [W]

- IAMs 1-9 [W].

A ReadMe TXT file is also included to provide additional information about the dataset. This includes a list of the appliances (including make & model where known) attached to each IAM as they were set-up by the REFIT team and subsequent changes that were discovered via visits to households, by being informed that an appliance had been removed/replaced or by visual inspection followed by querying the household. The format of the ReadMe file is the following:

- Introduction to the dataset
- Licensing information
- Naming scheme
- File formats
- Appliance list per house including changes made during the monitoring period and make and model where known.

The data availability for all REFIT houses can be seen in Figure 3.4. The gaps indicate periods when the data was unavailable. The vertical right edge of Figure 3.4 shows uptime per house, calculated by summing the time between gaps that are greater than one hour, and normalising this by the total monitoring duration for each house. The average uptime across all houses was 88%, with House 2 having the lowest at 76% and House 18 the highest at 94%.

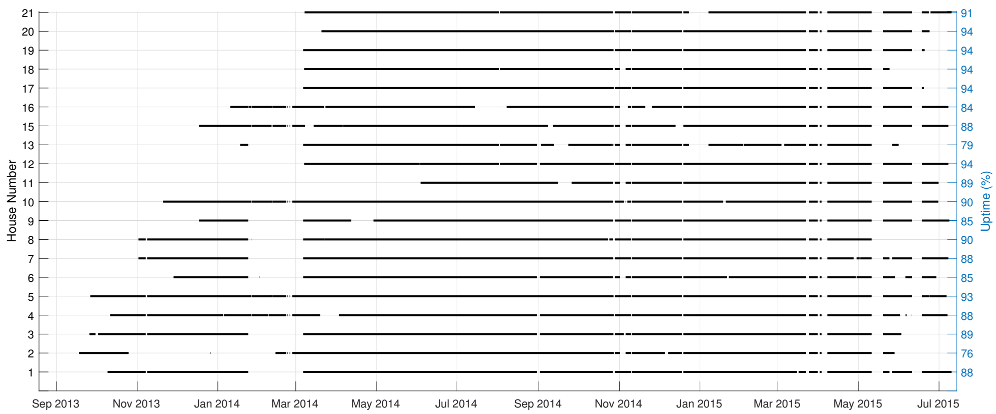


Fig. 3.4. REFIT Data Availability. Gaps in the line represents an area where data was unavailable for more than a quarter of a day.

The raw dataset is available on University of Strathclyde's PURE data repository at Raw:
`http://dx.doi.org/10.15129/31da3ece-f902-4e95-a093-e0a9536983c4` (Data Citation 1). The cleaned dataset, where IAMs which had appliances swapped between them, have been correctly stitched together to create a continuous data stream, is also available at Cleaned:
`http://dx.doi.org/10.15129/9ab14b0e-19ac-4279-938f-27f643078cec` (Data Citation 2). for those wishing to analyse cleaned, labelled electrical measurements immediately.

3.4 Technical Validation

Over the course of the study there were 119,495,879 timestamped readings taken from all houses combined, with each timestamp referring to 9 appliances and an aggregate per house - leading to 1,194,958,790 readings in total. Of these, 6.4% were Not a Number (NaN) values, which represent an unchanging reading or the IAM failing to respond to requests. NaN values are still available in the Raw Data version of the dataset. In the Cleaned Data version, a notes column has been added per house to indicate when the sum of recorded IAM readings is larger than the aggregate for the corresponding sample. These are described in the ReadMe file supplied with the dataset.

All the IAM streams have been visually checked to assess the validity of the signatures which are recorded. In all cases the ReadMe file associated with the dataset accurately reflects the known appliances which were plugged in. In some cases additional signatures may appear as households have removed an appliance for a very short period of time, e.g., replacing a toaster for an infrequently used kitchen appliance.

The quality of some appliance readings is affected by the location or interference from other devices. This is more notable on appliances further from the energy aggregator as well as IAM plugs located behind devices such as washing machines and other white goods. This was detected via visual inspection against the same appliance during a similar regular usage. For example, washing machines during a spin cycle are characterised by frequent changes in power; in some cases, the power will remain static (originally NaN values which have been forward filled) for a long period of time due to a connection loss that caused a lack of updates.

Note, however, that infrequently used devices will have many NaN values only due to not being used for large periods of time, e.g., electric heaters which are typically not used during summer but left plugged in.

All IAMs exhibited erroneous spikes, some more frequently than others. There was however no correlation between appliance monitored and the number of errors which occurred. It should be noted that these errors represent less than 0.004% of total IAM readings and that across all houses there was an average of only 215 errors per IAM (over the entire 2-year monitoring period).

Previously we mentioned that IAMs did not report values synchronously, as manifested by a lead with respect to the aggregate. Indeed, most IAMs should lead the aggregate by 2 or 3 readings at most. As shown in Figure 3.5, typically, this issue will not affect the analysis as appliance switching on and off events can clearly be observed in the aggregate readings with a delay of up to 1-2 readings.

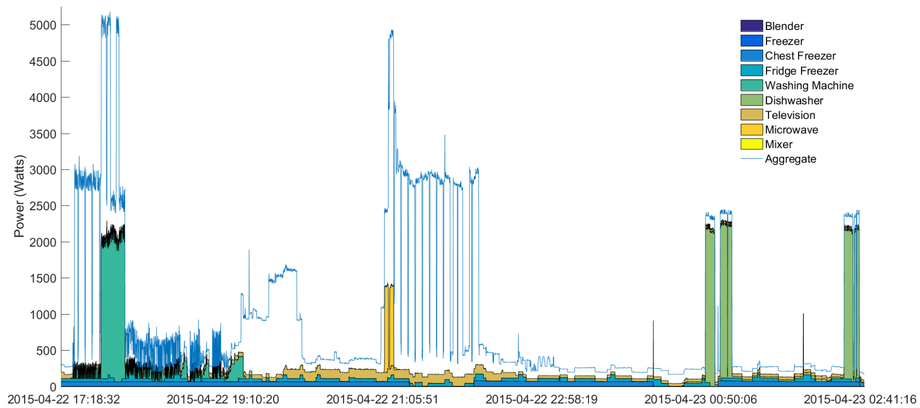


Fig. 3.5. Power demand for House 10 during the evening of April 22$^{\mathrm{nd}}$ 2015. The gap between the stacked IAM plot and the aggregate represents the power consumed by other appliances not monitored by IAMs.

Meter readings were taken from several houses during installation and at subsequent visits by the REFIT project team. A comparison between the reading taken from the utility meter and the measured power by our platform is shown in Table 3.4. In some cases the monitored values may be higher than expected due to spikes which occurred in the aggregate. Also, as readings were only taken once every 8 seconds it is possible that the estimated consumption which is based on a reading multiplied by time difference to the next reading was under or over estimated. We found that the houses without solar interference (from solar panel installs which caused issue with the aggregate sensor) had generally less than 12% difference to the utility meter estimated consumption.

House	Dates	Metered [kWh]	Monitored [kWh]
8	29/09/2014 - 15/10/2014	226	240
8	15/10/2014 - 27/01/2015	1785	1810
8	27/01/2015 - 05/03/2015	657	695
10	15/10/2014 - 24/03/2015	2799	2676
13	04/10/2014 - 26/11/2014	640	583
17	13/11/2014 - 02/12/2014	178	192
18	15/10/2014 - 18/11/2014	333	323
18	18/11/2014 - 16/12/2014	316	345
19	02/12/2014 - 11/12/2014	79	73

Table 3.4
Recorded Meter Consumption. Metered represents the difference in readings between the two dates which was recorded by the utility installed meter for the house. Monitored is the value calculated using the recorded data from the REFIT study.

Table 3.5 shows the % of total household consumption captured by sub-metering. It can be seen that up to 55% of consumed energy can be attributed to appliances directly monitored via appliance plugs. We note that the fact that some large consumers, such as the electric ovens and electric showers were not monitored, resulted in a relatively low % of energy consumption captured by plugs in some houses.

	House Number																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	15	16	17	18	19	20	21
% Captured by sub-metering	35	33	N/A	48	48	40	46	22	38	43	N/A	36	48	36	37	45	55	35	46	N/A

Table 3.5
The amount of power captured by the IAM plugs compared to the total consumption over the monitoring period. N/A value in shown for houses with solar panels.

In studies aimed at conducting NILM or appliance modelling, it is important to capture a large number of appliance uses. We have analysed the entire IAM dataset to record the number of uses for different appliance types. To estimate the number of uses per appliance, edge detection was used to help build up a pattern of usage.

In Table 3.6 we show the number of use events captured for 15 types of appliances. Note that in some cases appliances may have been monitored but used rarely. In the table, Number of Uses represents a start and end event recorded for an appliance; in the case of fridges this is a cooling cycle, e.g., from the motor starting till motor winds down to a stop. For some appliances, this number may not represent the total number of uses recorded by the REFIT study as the edge detection used was not accurate enough to classify uses in appliances where multiple devices were monitored on the same IAM, which was sometimes the case with ICT equipment (a computer, printer and monitor will be connected to the same IAM using a power strip) and Television site (Television, DVD Player, TV set-up-box were monitored together). Number of Appliances shows the total number of appliances across all REFIT households. The Pond Pump and Vivarium stayed on almost constantly, with a fixed load, throughout the study and are therefore classified as continuous use.

Appliance Type	# of Uses	Consumption (kWh)	# of Appliances
Fridge Freezers	121,752	20020	13
Fridge	53,163	1310	7
Freezers	133,967	5486	13
Washing Machines	6865	3994	21
Dishwashers	4250	6827	14
Tumble Dryers	2372	4210	10
Kettle	40,092	3298	16
Microwave	12,946	1208	16
Toaster	5364	257	9
ICT Equipment	4176	3104	13
Television Site	11274	5995	21
Electric Heater	503	1023	4
Bread Maker	206	56	1
Pond Pump	Continuous	282	1
Vivarium	Continuous	208	1

Table 3.6
The amount of data collected per appliance type across all the REFIT households.

3.5 Summary

This chapter describes the creation and management of the REFIT dataset. It contained a combination of features previously not made available in public datasets.

The use of IAM's made it ideal for NILM applications, as well as appliance modelling. The time period covered allowed for seasonal trends to be calculated and implementation of new regimes effects to be monitored. At the time of publication the REFIT dataset was considered the largest and best resolution datasets publicly available. It provided data for a large number of follow up papers in a number of different fields. The data from the REFIT dataset is used heavily within the following chapters and provides an invaluable base for this thesis.

Chapter 4

Visualisation & Understanding Patterns of Use

4.1 Introduction

Appliance modelling is a key activity enabled by NILM datasets. Appliance usage as defined in energy labels is static and based on general assumptions around usage, for example a washing machine would be used 3 times a week, and used on a setting that consumes a certain amount of kWh. This gives a reasonable guide when comparing two similar appliances against each other, but does not take into account personal usage. Personal usage may change the experience of the appliance and making use of specific features may either increase of decrease consumption massively. Therefore understanding appliance usage is an important part of appliance modelling, and helps understand how the user interacts. The content of this chapter is taken from the paper on 'Understanding usage patterns of electric kettle and energy saving potential' [104].

Many empirical studies on consumer attitudes and interactions with energy consuming appliances have been reported [12, 119, 133]. Interestingly, despite the fact that the humble but ubiquitous and widely used kettle has a non-negligible influence on electricity demand [68], modelling and forecasting methods to understand and predict demand, as well as calculating energy-wasteful usage, have not been analysed in detail so far for this appliance. Earlier studies on kettle efficiency are led by energy charities or government where the emphasis is on assessing over-boil, minimum water volumes, and daily/monthly/annual costs based on average estimates [140].

The kettle is one of the most (inefficiently) used appliances in the UK as well as the appliance with the highest rates of ownership (according to UK's Department for Environment, Food and Rural Affairs' 2006 report [121] 97% of UK households own a kettle). Indeed, more than nine in ten people (95%) use the kettle every day, with 40% doing this five times a day or more. In a survey of 86,000 homes in the UK, by the Energy Saving Trust [140], it was found that three-quarters of British households admit to overfilling their kettle when boiling water and are subsequently wasting GBP68 million each year.

Though a relatively low consumer when compared to an electric heater or washing machine, the kettle can significantly influence electricity generation and power distribution network, mainly due to the so-called TV pick-up effect, that manifests itself through significant and synchronised usage of appliances during TV programme breaks. This is especially a problem in the UK where individual programmes often attract a massive audience, and householders use commercial breaks for boiling water, using microwaves, opening the refrigerator door, etc. Hence, understanding domestic routines related to kettle usage are important for demand response measures.

Kettles are not currently subject to any efficiency labelling guidelines; this means that the consumer may not understand that a lower power rated kettle will take much longer to boil than a high rated one, but would consume the same amount of energy. The time to boil might help encourage energy efficient behaviour as the occupant is aware that the kettle will take much longer to boil if it is overfilled. Eco kettles have also become available, featuring insulated housing which can help reduce consumption by maintaining a higher water temperature between usage (The UK brand Vektra is currently the most well known insulated kettle). However, 86% of people do not choose kettles based on their features, but on looks to match a kitchen design/already owned products [100].

To the best of our knowledge, there has been no in-depth study which accurately measures, in a longitudinal study, kettle consumption and analyses patterns of consumption, personalised costs based on household composition and overboiling. This may be due to the challenge of measuring fill water volumes (since it is impractical to measure and record water volume for every kettle use), as well as individual electricity consumption in order to carry out an in-depth analysis.

This overcomes the above problem by measuring the individual kettle consumption (kWh) and estimating the water volume from this measurement using mathematical modelling. In particular, using measurements with different kettle types, a mathematical model is built that relates the water volume of a kettle, its consumed power and water temperature. We verify the proposed models using four kettles: two standard and two smart kettles (where temperature and keep warm modes are available).

In order to demonstrate how we can effectively use load data, which could be obtained via disaggregation from smart meters or individual appliance monitors (widely available on the market) or from smart appliances themselves, we launched a field trial comprising 14 UK houses, monitored over a period of two years as part of the REFIT study. The timestamped kettle power consumption was collected via a plug monitor that measures active power every 6-8 sec [106]. The monitored houses are of different occupancy and age groups (e.g., retirees, working couples, families with children and single occupants), some energy conscious and others not.

Equipped with the collected power consumption information and time of use information, together with the kettle model, we revealed households behaviour in terms of water overboiling and energy wasting. Additionally, time-use analysis enabled us to capture established routines and usage synchronicity across the monitored households as well as seasonal effects. Finally, having observed that kettle usage patterns follow fairly distinct routines, we study demand predictability (by adapting recognised prediction tools, namely Support Vector Machine (SVM) [24] and Adaptive Neuro Fuzzy Inference System (ANFIS) [69]), and show that kettle energy consumption can accurately be predicted on a daily and weekly basis.

In summary, the key contributions of this chapter are:

- Generic mathematical model that relates water volume, consumed power, and water temperature, obtained heuristically using three different kettle types, that improves previously proposed models.

- Understanding kettle use through the lens of household composition, time of use, seasonal effects and overboiling.

- In depth analysis of kettle usage patterns in terms of established routines and synchronicity.

- Identifying households, routines w.r.t kettle usage, we propose a kettle energy demand prediction model that accurately predicts future kettle energy consumption.

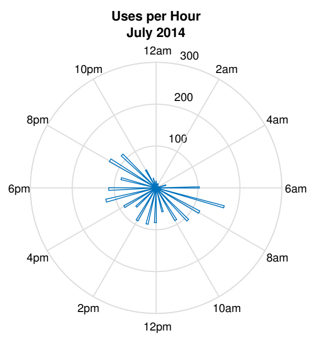
4.2 Usage Analysis

Presented is the observations from monitoring kettle usage over a period of two years in 14 UK households as part of the REFIT project (Chapter 3). The houses in the study were selected to achieve a good cross section of middle-high income British households, from the Loughborough area. Each of the 20 houses monitored had up to 9 appliances, aggregate, gas and environmental data monitored. We chose a subset of 14 households, since their kettles were directly monitored by IAMs, recording consumption data in Watts (W) at a 6-8 second intervals. See [106] for more details about energy monitoring platform. Each house has been given a unique ID by which it will be referred to throughout e.g. (House 1, House 2, ...). This data is available publicly at DOI [10.15129/31da3ece-f902-4e95-a093-e0a9536983c4].

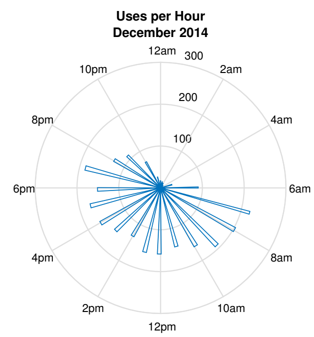
Based on the collected energy consumption data, we look at both time of use information and energy consumption data trying to identify distinct usage patterns, household routines and possibly synchronicity between households.

4.2.1 Time of Use

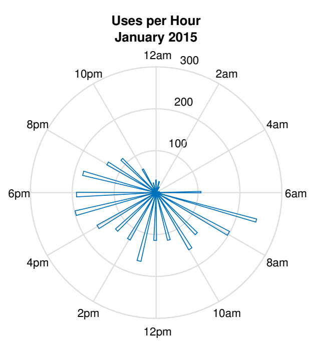
Fig. 4.1 shows kettle usage, presented as rose charts, aggregated from 14 monitored houses, for four different months across the years 2014-2015. In each plot, the bins represent the number of times kettles were used in the houses during the specified one-hour period. Across the year the difference in usage can be seen: Autumn/Winter months have a higher usage, as expected, due to colder weather. Peaks at 7am and 5pm are prominent in most months, signalling the pre-departure and post-arrival usage around the average UK workday. December (Fig. 4.1b) has less prominent peaks, which could be due to a number of factors, time-off work (i.e., holiday season), guests and additional cooking activities.



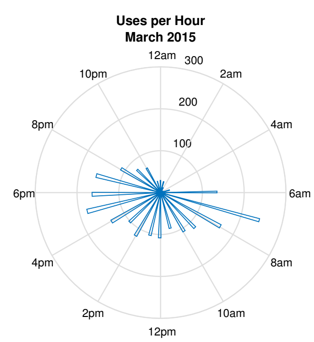
(a) July 2014



(b) December 2014



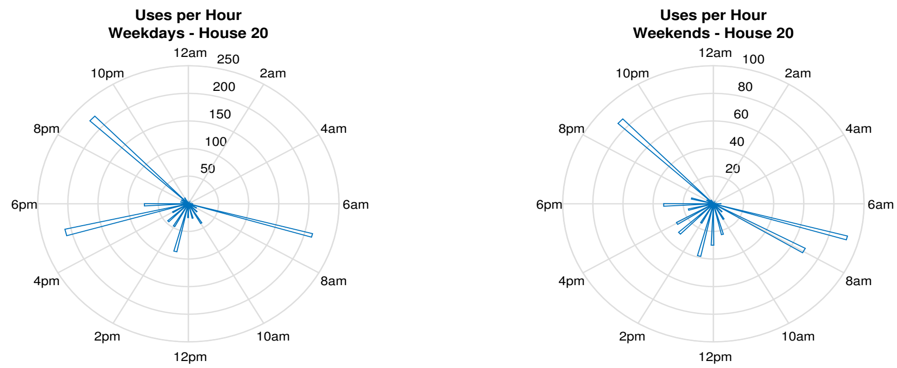
(c) January 2015



(d) March 2015

Fig. 4.1. Total number of uses per hour per month in all 14 monitored houses.

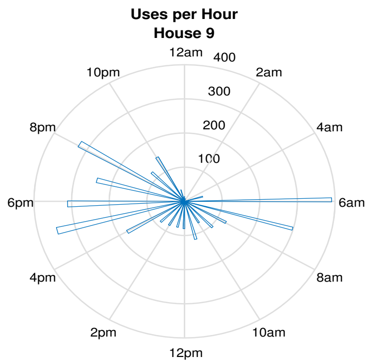
We aim to show that, within a house, patterns of kettle use are maintained throughout the year, embedded as the household's steady routines, where only the number of kettle uses differ − less in summer months and more in winter. Looking at data across the households, there are two main usage patterns that can be observed: working and retiree households. While both types of household occupants will likely have morning and evening peaks in usage, the difference can be observed during normal work hours.



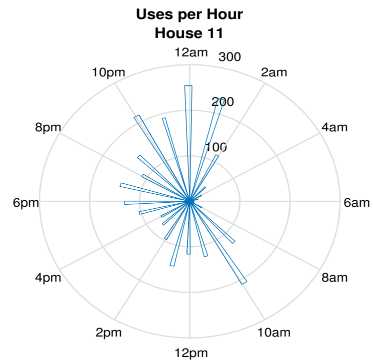
(a) Weekday (b) Weekend

Fig. 4.2. Total number of kettle uses per hour in House 20 over a continuous period of 11 months.

Fig. 4.2 shows results for one particular house (House 20) for the 11-month period July 2014 - June 2015 broken down as workday and weekend. Weekday usage has four well defined peaks: 7am, 1pm, 5pm and 9pm. This follows the schedule of common office hours in the UK and can help determine occupancy levels within the house; in this case, the peak at 1pm can be attributed to the young adult living at home returning from university for lunch. The weekend has 3 significant peaks: at 7am, 8am and 9pm. This new morning pattern shows that the three occupants have a different routine over the weekend. The night time use at 9pm suggests there is a clearly defined sleeping pattern as uses after 9pm are almost non-existent. This latter observation is useful for social scientists to understand activities in the home.



(a) Working Household



(b) Retiree Household

Fig. 4.3. Total number of kettle uses per hour in Houses 9 (Dec13-Jun15) and 11, (Jun14-Jun15).

Fig. 4.3 shows temporal usage patterns for Houses 9 and 11. House 9 has a more regular pattern in relation to today,s general working lifestyle - a large peak at 6/7am signifies the waking time of at least one occupant and a secondary peak at 4/5pm signifies their return home after work. This is expected as the occupants of House 9 are a couple with no children; hence, usage throughout the day is at a minimum and can be accounted for by weekend and holiday periods.

House 11 has the usage pattern of a “night owl” (Fig. 4.3b). Indeed, usage is unusually low during the most expected hours of usage (7-9am); instead, the usage usually starts at midday and continues over the afternoon till late night with a final spike at midnight dropping of as 2am passes. The low usage throughout the day, suggests that the occupants are at home and do not have any ingrained daily schedule or work commitments. This more unusual pattern can therefore be attributed to there being a single retired occupant in the property who has a nocturnal sleeping pattern. This is confirmed by discussions with the householder.

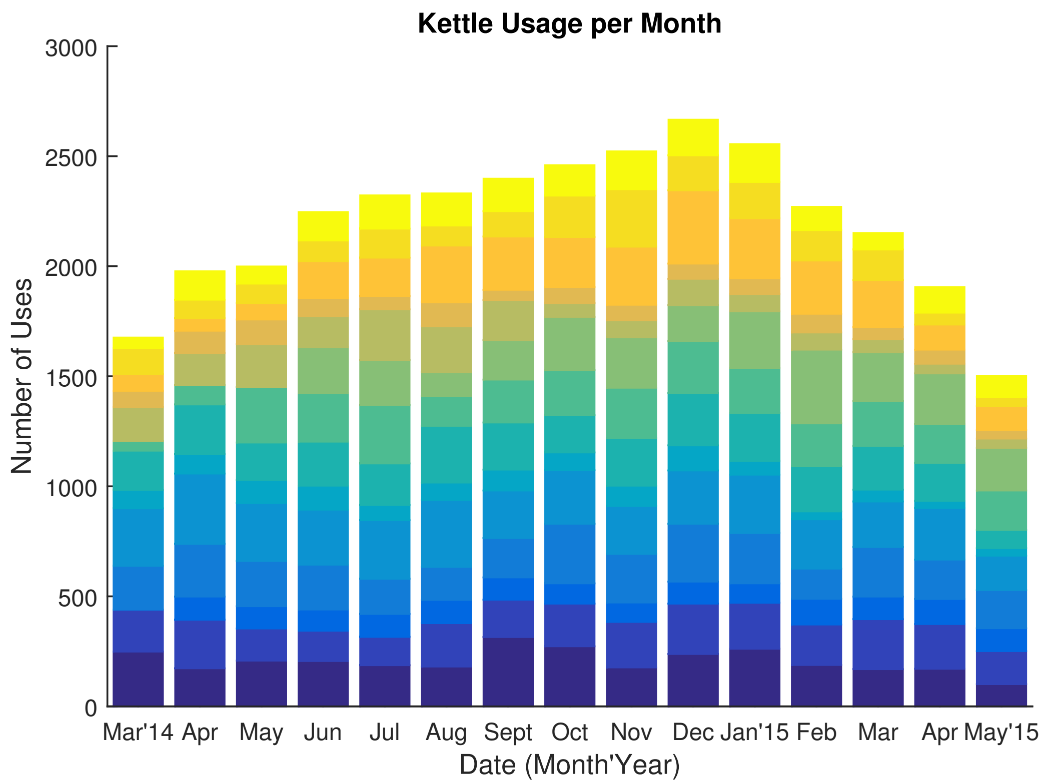


Fig. 4.4. Seasonal kettle usage for all 14 houses. The result for each house is represented with a unique colour. Variations within a house and across all houses is evident. As expected, kettle usage increases during winter months.

The effect of seasons on kettle usage can be visualised in Fig. 4.4. During the study period the expected trend of increased usage as winter approaches has proved correct with an upward trend from July to December. The slight decrease in August is attributed to a number of households going on holiday.

4.2.2 Electricity Consumption

A major factor in consumption is the occupancy of the household. Table 4.1 shows the kettle consumption for each house over the month of December 2014. It can be seen that consumption as well as kWh per use, varies significantly even in households with a similar occupation. It can be seen that kWh per use also varies. Noticeably, some houses have a much higher kWh per use than others. For example, Houses 9 and 12 are close to 0.1 kWh per use (a relatively high value) requiring a deeper investigation into their usage habits, to identify why they fill the kettle significantly more than other households.

Table 4.1
The number of occupants, total electrical consumption, and kettle electrical consumption for all 14 monitored houses.

House	Occupancy	kWh Total	kWh Kettle	Total Monthly Cost [GBP]	kWh per Single Kettle Use	% of aggregate use
3	2	621.45	14.83	1.96	0.062	2%
2	4(2)	471.17	18.28	2.41	0.072	4%
4	2R	270.59	6.87	0.90	0.068	3%
5	4(2)	676.58	19.41	2.56	0.073	3%
6	2	324.74	15.04	1.98	0.060	5%
7	4(2)	514.88	8.69	1.14	0.075	2%
8	2R	571.73	16.09	2.12	0.067	3%
9	2	537.10	23.24	3.07	0.098	4%
11	1	152.51	12.02	1.58	0.072	8%
12	3	305.78	19.07	2.52	0.097	6%
13	4(2)	317.26	6.09	0.80	0.088	2%
17	3(1)	324.57	21.01	2.77	0.062	6%
19	4(2)	216.38	9.00	1.19	0.057	4%
20	3	291.55	11.65	1.54	0.067	4%

4(2) means there are 4 occupants including 2 minors. R refers to retired occupants. The consumption results are given for December 2014. Total monthly cost assumes 0.13GBP per 1kWh[1].

[1]Standard tariff for Loughborough area `https://www.scottishpower.co.uk/tariff-information.process?execution=e1s3`

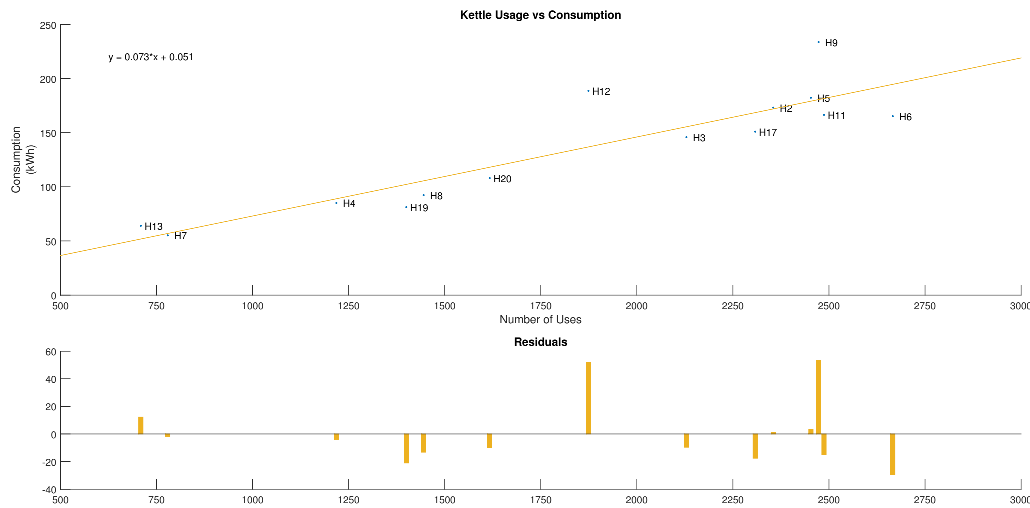
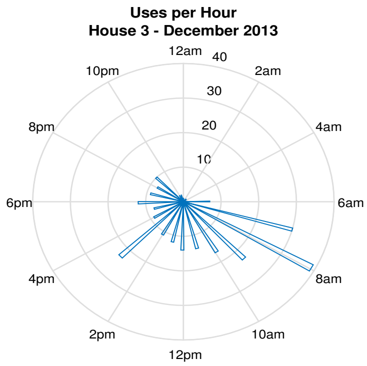


Fig. 4.5. Household consumption plotted against the number of kettle uses together with the fitted line. The bottom figure shows the difference to the best fit line.

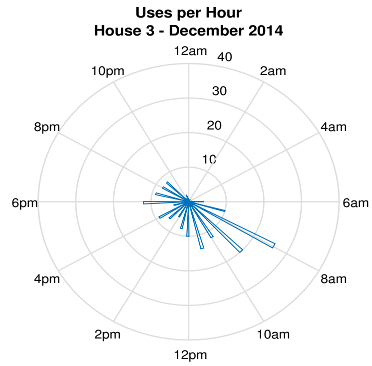
Fig. 4.5 shows the energy consumption of each household plotted against the number of times the kettle was switched on i.e., the number of uses. It can be seen that a clear linear trend is apparent, where 10 houses fall below this fitted line and 3 are above it. The houses that fall below the line of the best fit are prime candidates to help understand efficient kettle usage, to help save both water and power. On the other hand, Houses 9 and 12 far exceed the mean consumption for their number of uses and investigating these houses may help show where improvements can be made. With respect to the Energy Saving Trust's findings [140] we could expect 3-4 houses of the 14 not to overfill the kettle; it can be seen that, House6 and House19 studied appear to have kettle usage habits which do not excessively overfill.

4.2.3 Standard vs. Eco Kettle

Household 3 changed its kettle during the study (Winter 2013 - Spring 2015) introducing a vacuum(eco) kettle as an energy saving measure. The vacuum kettle keeps water hot for longer, thus potentially reducing the number of uses. We can therefore look at their usage before and after the change to show the advantage/disadvantage of this new kettle.



(a) Standard kettle



(b) Eco vacuum kettle

Fig. 4.6. Standard and eco kettle usage in House 3 over the month of December 2013 and 2014, respectively.

The consumption pattern between the standard and vacuum kettle can be seen in Fig. 4.6. The standard kettle usage was much higher at 7am compared to the vacuum kettle. This can be attributed to the fact the vacuum kettle will be used once with a large amount of water and retain that heat throughout the hour. This pattern can also be seen at 3pm in 2013 when there were uses in the following two hours. In 2014, the following two hours have significantly less usage; the hour immediately after has a much smaller, and the following hour slightly more, possibly attributed to reheats where the water is not considered hot enough for the drink being prepared.

Table 4.2
Kettle usage in House 3 in December 2013 (using standard kettle) and December 2014 (using eco kettle).

Year	Uses	Consumption (Kilo Joules)	kWh	Cost (13.52p/kWh)
2013–December [standard]	241	63,253	17.57	2.38
2014–December [eco]	199	45,075	12.52	1.69

From Table 4.2, it can be seen that the eco-kettle has a significantly fewer number of uses and therefore the associated cost has been reduced by close to

GBP0.70 in the comparative months of December. Over the period of a year this could mean a possibility for saving around GBP8.00. This represents close to a 50% saving based on the figure found on CarbonFootprint.com. This helps demonstrate that a desire to become more eco-friendly is possible by making little changes to appliances. The initial cost of the eco kettle, however, is around GBP80, therefore there is a significant period of time before the kettle will be cost effective.

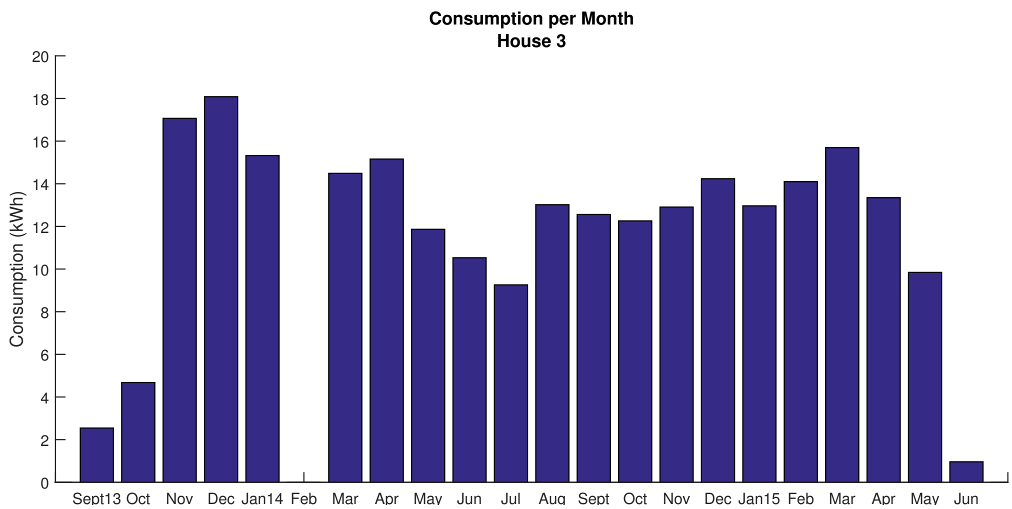


Fig. 4.7. Consumption due to kettle in House 3. Note, no data was available in Feb'13.

The vacuum kettle was introduced in April 2014 and was removed in January 2015. From Fig. 4.7, it can be seen that after its introduction, consumption reduces. However, this can also be attributed to the seasonal flux in usage as the weather is hotter. August 2014 onwards shows an almost stable trend in usage with a peak in December'14. In late January'15 the kettle is replaced with the previous (standard) kettle that was in use. It can be seen that February-March'15 have higher peaks; however, this then returns to lower than previous levels in April-May'15.

4.2.4 Feedback

The residents of House 3 were given a breakdown of usage, along with textual explanation of the findings. A survey was completed prior to delivery of the consumption breakdown to assess the residents thoughts. The survey revealed a number of traits about the household. The residents were committed to being eco-friendly and were positive about buying other products aimed at consumption reduction. They believed that they had changed their habits significantly as they actively incorporate the vacuum kettle into their routine.

As shown above, this can be seen in the comparisons made, in both usage and water consumption which led to a more economical usage style. They also made a note of the fact that they try to avoid re-heating water and this is backed by the fact that only 7% of their kettle usage is within a 5-minute window of a previous usage. Shortly after the time period included in this work, the household stopped using this vacuum kettle due to a fault which once fixed never made it back into daily usage. Interestingly, this was not due to any effects on performance, but due to the noise the kettle made, which was annoying to the occupants. The feedback, however, was well received and the residents believed that this would be of benefit, and expressed that a monthly breakdown of appliance usage would be beneficial toward supporting their efforts towards being eco-friendly.

4.3 Energy waste prediction via modelling

In this section we describe the proposed mathematical modelling method used to predict water volume based on measured consumed power and use it to estimate energy waste due to overfilling and re-boiling the kettle.

4.3.1 Mathematical Modelling

The objectives of mathematical modelling are: (i) to determine whether there exists one generic model or equation that can estimate the water volume of a standard or smart kettle using consumed power data only with high accuracy, (ii) assess its relative accuracy compared to the 'specific heat' model described in [111], where the kettle is treated as a classic heating problem, and (iii) determine whether separate models for standard and smart kettles yield higher relative estimation accuracy.

In [111], volume, temperature and consumed power are related as:

$$W = \frac{\alpha}{\beta * \Delta T} \tag{4.1}$$

where W denotes water volume (in millilitres), α is consumed power in kiloJoules (kJ), β is the specific heat capacity of water (4.19 kJ/kg.°C) and ΔT the change in temperature (degrees Celsius).

We perform experiments using four non-faulty kettles, namely 2 standard kettles and 2 smart kettles, measuring the following parameters: consumed power in kWh, water volume, starting water temperature, finishing water temperature ($< 100\,°C$). A standard kettle is defined as a kettle that boils water to 100 °C with no additional 'boil' temperatures and no 'keep warm' or additional functionalities. A smart kettle would included additional heating temperatures 70 - 100 °C and/or a keep warm functionality.

280 experiments were carried out, 140 with standard kettles and 140 with smart kettles. 5/7 of the data was used for training the model and the remaining 2/7 used for validation of the model. As a result, three kettle models were developed using surface fitting with the data obtained from the experiments, namely: (1) generic kettle model, combing smart kettle and standard kettle together, (2) standard kettle model, built with standard kettle data only, (3) smart kettle model, built with smart kettle experimental data only.

Due to the simplicity of a kettle driven by a heating element, and the fact that the boil time is nearly linear with respect to the volume of water, we assume initially that a linear function would be suitable to model the relationship between water volume and consumed power, taking starting temperature into account. Fig. 4.8 shows that this assumption generally holds true, for both the standard and smart kettles, as well as for the model combining both standard and smart kettles.

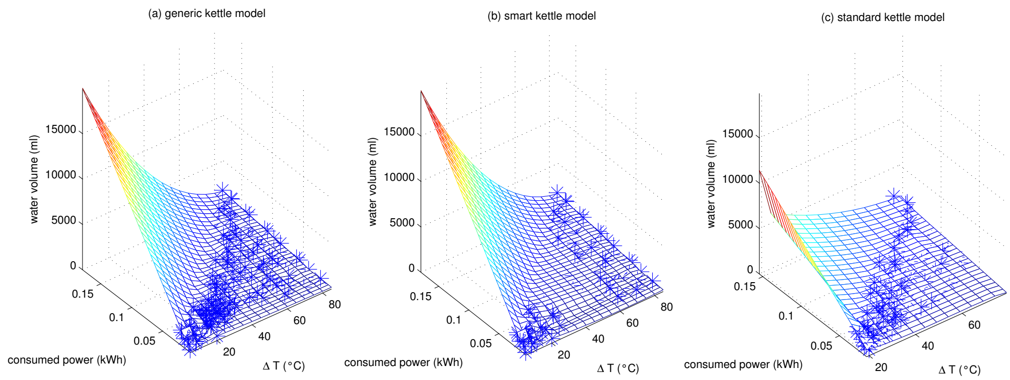


Fig. 4.8. Consumption comparison of the three kettle models using linear interpolation method.

Fig. 4.8 presents the results of linear modelling for all three kettle models, showing the relationship between water volume in millilitres (mL), consumed power in kWh and ΔT in degrees Celsius. From the relative curvature of the plane, one can see that the three models are not identical. We therefore assess best fit using the following classic linear methods: polynomial linear, locally weighted linear regression, and the linear interpolation methods. The accuracy of the models is assessed by the root-mean square error (RMSE) between the actual water volume and the estimated water volume, across 80 experiments for the generic kettle model and across 40 experiments for the standard and smart kettles, respectively. The results are shown in Table 4.3.

Table 4.3
RMSE in millilitre for the three kettle models and using Eq. (4.1).

	Linear Interpolation	Polynomial	Locally Weighted	Eq.(4.1)
Generic	120.79	**94.79**	147.67	173.24
Standard	217.32	**53.69**	153.85	136.75
Smart	74.40	**72.64**	133.78	203.28

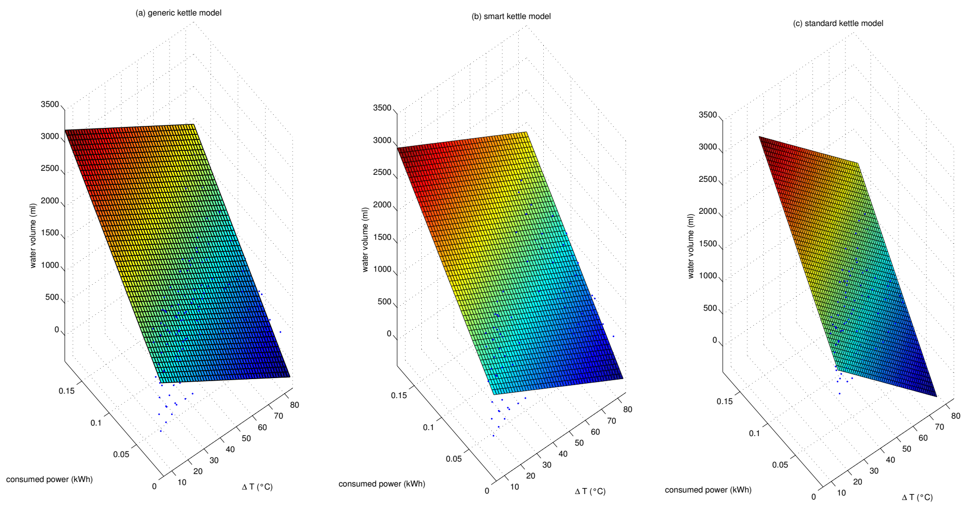


Fig. 4.9. Consumption comparison of the three kettle models using the polynomial linear interpolation method. The true measurements are shown as blue dots.

Note from the table that the RMSE of Eq.4.1 [111] is consistently higher than all other linear models evaluated. The polynomial linear interpolation method provides the lowest RMSE. As expected, the general kettle model performs worse. Thus, the generic model should be used only if estimating the water volume of a kettle, whose type is unknown.

The polynomial linear interpolation method works best for all three proposed models. It is defined as:

$$W(L) = p_0 - p_1 \Delta T + (p_2 * P), \tag{4.2}$$

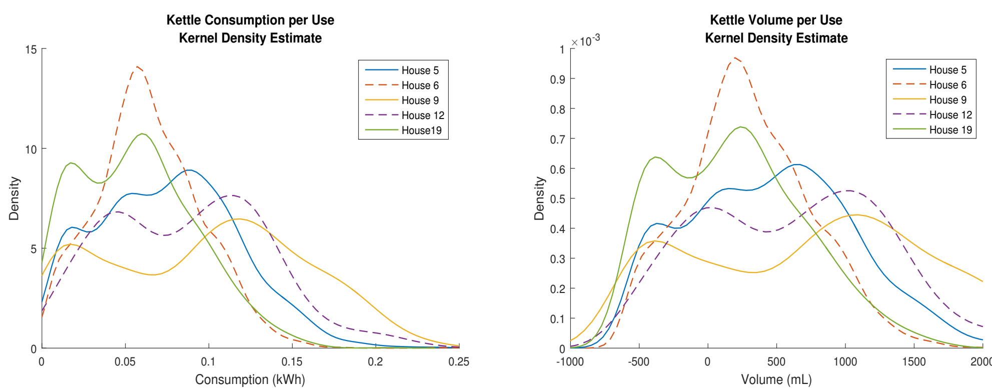
where the value of the coefficients p_0, p_1 and p_2 are shown in Table 4.4, P is consumed power in kWh, W is water volume in litres and ΔT is the change in temperature. Fig. 4.9 shows the obtained results: the surface model good fit with the true measurements.

Table 4.4
Kettle models coefficients for the polynomial linear interpolation given by (4.2).

	Generic	Standard	Smart
p_0	1025	1244	905
p_1	15.95	22.09	14.01
p_2	12.34	14.54	11.82

4.3.2 Energy wasted due to overfilling and reboiling

Equipped with the previous model, we can estimate the amount of wasted energy due to overboiling and re-boiling the kettle using only the collected power measurements. All houses in the study had standard kettles at the time readings were taken, apart from House 3, as discussed in Section 4.2.3.



(a) Energy consumption per use (b) Water volume per use

Fig. 4.10. Probability density function of kettle consumption and water volume per use for 5 houses. The negative volume scale represents re-boils.

Fig. 4.10 shows statistical distribution of energy consumption and water volume estimated in the kettle per use for 5 houses. Obviously, different households have different preferred levels of water per boil, and consequently, consume differently per single kettle use. It can be seen from Fig. 4.10a that each house has its own distinct fill patterns. Indeed, Houses 6 and 19 have a similar water-filling pattern: a very narrow boil range between 0.05-0.075 kWh; House 19 has a higher percentage of re-boils (hence, the second 0.025 kWh peak). The other end of the spectrum is House 12, peaking at more than 0.1 kWh with a wide bell curve on each side. This suggests that this house is the least efficient house in the survey group. House 9 has two peaks of relative magnitude, 0.05-0.075kWh and 0.125 kWh, respectively. This wide range of consumption per usage suggests that the kettle is filled with little thought as to the purpose. House 5 is between the two extrema − no significant peak, but the probability after 0.1kWh falls of at an equivalent rate to Houses 6 and 19 suggesting a slightly more economical usage with comparison to Houses 9 and 12.

It can be seen from Fig. 4.10b that House 6, which has been shown to be an economical kettle user, has a much lower number of uses where the water level has been above 1 Litre and the usage peaks at just under 0.5 Litres. Similarly, House 9 has been shown to be one of the less economical users (Table 4.1). Two distinct peaks are visible, one at -500 and 1250 with a slow tail off. The negative volume scale represents re-boils, i.e., when the occupant uses the kettle before the water has cooled to room temperature. House 19 which has the highest reboil peak has 37% of all recorded uses estimated as re-boils consuming 15.14kWh. House 12 which has a much larger number of uses has the lowest peak where 22% of uses are re-boils accounting for 10.77kWh of consumption.

From Fig. 4.10, we can estimate how much energy could be saved assuming that the household cuts down on overfilling their kettle, as well as re-boils and assume a minimum of 500mL (many kettles minimum fill) and a range of 250-300mL per additional occupant for each household. As an example, for House 9 which has 2 occupants, working from the assumption that a usage is minimum 500mL, for two people ideal water volume will be around 550mL. Over the entire study period, House 9 had a recorded 3220 uses of which 1978 were above 1000mL. This accounts for a consumption of 271.31kWh of a total 312.36kWh. Based on their kettle performance, the average kWh cost for 525-575mL is 0.08kWh. If all of the uses above 550mL were reduced to 550mL a saving of 110.57kWh could have been made over the 18 months monitored for House 9, or 73.71kWh per year. Similarly House 12 which has 3 occupants should be around 825mL. House 12 has 1874 recorded usages with 790 of those usages being greater than 825mL.

This accounts for 105.54kWh of a total 163.92kWh recorded over the study period. Reducing these overboils to the maximum 825mL could result in a saving of around, 26.23kWh or 20.98kWh per year.

Table 4.5
Energy savings in GBP that could be made by the reduction of filling levels. Consumption is given in kWh. Second column represent the number of months the house was monitored. Volume refers to the optimal volume estimated based on the number of occupants. Consumption above the volume is the energy consumption due to filling the kettle over the optimal volume level. Savings denotes the total savings in GBP assuming 0.13GBP per 1kWh, if kettle would have always been filled upto the optimal volume level.

House	Months Recorded	Kettle Consumption (kWh)	Volume(mL)	Consumption Above Volume (kWh)	Savings per Year (kWh)	Savings per Year (GBP)
2	20	255.32	825	126.76	15.32	1.99
3	20	251.16	550	171.06	28.85	3.75
4	20	135.86	550	45.02	6.29	0.82
5	21	314.66	825	148.85	17.32	2.25
6	19	273.60	550	122.75	16.67	2.17
7	20	109.84	825	42.21	5.17	0.67
8	18	245.68	550	171.83	23.41	3.04
9	18	312.36	550	271.31	73.71	9.58
11	12	182.02	500	83.78	29.99	3.90
12	15	163.92	825	105.54	20.98	2.73
13	16	103.24	825	62.32	7.37	0.96
17	15	183.63	550	98.98	16.99	2.21
19	15	108.27	825	26.64	3.56	0.46
20	15	136.11	825	19.66	1.64	0.21

Table 4.5 shows each household's total electricity consumption for the kettle across the entire survey period, the volume that would be expected for the number of occupants (assuming that occupants under 18 consume half the fill volume of an adult) and the potential savings. It can be seen that most households' consumption above this level accounts for, in many cases, more than half of the total consumption. Savings per year are calculated using a mean consumption value for the volume range that has been estimated for each household, and then all occurrences above the volume limit have been set to this calculated kWh average. The resulting value is the difference between the consumption above volume limit and the consumption that could be achieved if filling to this limit; this value is then calculated into yearly savings based on the number of months the house was monitored.

4.4 Demand Prediction

As shown in Section 4.2, kettle usage patterns are part of established domestic daily routines (e.g., a high likeliness of usage early in the morning), and hence it is natural to assume that they can be accurately analytically predicted. Presented in this section are the findings on kettle use and energy demand predictability additionally predicting kettle demand, has not been studied before.

The ability to predict individual appliance demand could potentially help improve overall household's demand prediction. Moreover, appliance demand prediction is useful in time-use studies to understand routines and practices in the home. To predict usage patterns we look at two different methods, namely Support Vector Machines (SVM) and adaptive network-based fuzzy inference system (ANFIS) established previously for predicting power demand. As one of the most powerful binary classifiers, SVM is suited to predicting kettle uses (kettle is used vs. kettle is not used). On the other hand, ANFIS is more suited to predicting load demand due to kettle.

4.4.1 SVM-based prediction

SVM has been used for load prediction in numerous papers [36, 71, 88]. SVMs are simple supervised learning mechanisms used as a way of classifying information that can easily be adapted for use as a time series predictor.

We develop an SVM-based predictor for predicting if kettle will be used within a given hour in the future. Let $\mathbf{X} = \Big\{x(t-n), \cdots, x(t)\Big\}$, be a vector of binary variables indicating kettle usage within n consecutive hours in the past, where $x(t-i)$ takes value 1 is the kettle was used i hours ago, or 0 otherwise. Then, the task is to predict the value of $x(t+1)$, that is, if the kettle will be used in the next hour. To do that, we input to the SVM training module, the vector

$$input = \Big\{x(t-26),\ \cdots\ x(t-22),\ x(t-2),\ \cdots,\ x(t)\ , weekDay\Big\}.$$

$$target = \Big\{x(t+1)\Big\}.$$

Note that, we use the three previous hours from the current time t (where t contains the previous hours consumption) for prediction as well as five hours during the previous day. Binary variable $weekDay$ is set to 1 if it is a weekday and 0 otherwise, so that only weekdays are used for predicting usage over a weekday and vice versa.

We took two approaches to SVM classification, running a binary classifier for each hour and one across all hours. Table 4.6 shows the predictions for the period 2015-01-Jan to 2015-01-Jul trained with the data from 2014-12-Dec to 2014-31-Dec using data from all hours in one classifier which uses the Iterative Single Data Algorithm (ISDA) solver [75] (MATLAB implementation). The number of hours where the kettle is not used will always far exceed those where it is, therefore, there are very few positive predictions using this method.

Table 4.6
SVM performance using a single SVM classifier. TP (True Positive), FP (False Positive), TN (True Negative), FN (False Negative), Correct = TP+TN.

House	Correct	TP	FP	TN	FN	Correct (%)
2	117	1	2	116	26	80
3	108	0	0	108	37	74
4	129	0	0	129	16	88
5	118	2	2	116	25	81
6	104	4	7	100	34	71
7	134	0	0	134	11	92
8	117	9	13	108	15	80
9	111	0	3	111	31	76
11	94	14	18	80	33	64
12	110	1	10	109	25	75
13	145	0	0	145	0	100
17	110	0	0	110	35	75
19	121	0	0	121	24	83
20	117	6	3	111	25	80

Secondly, we look at the effectiveness of training 24 SVM's, one for each hour of the day. Table 4.7 shows the results of ISDA solver using the radial basis function (RBF) kernel which produces the best results.

Table 4.7
SVM performance using a hourly SVM classifiers.

House	Correct	TP	FP	TN	FN	Correct (%)
2	105	16	29	89	11	72
3	104	9	13	95	28	71
4	132	6	3	126	10	91
5	91	10	37	81	17	62
6	113	16	10	97	22	77
7	127	1	8	126	10	87
8	105	5	21	100	19	72
9	108	14	20	94	17	74
11	98	0	0	98	47	67
12	109	8	18	101	18	75
13	145	0	0	145	0	100
17	112	10	8	102	25	77
19	119	0	2	119	24	82
20	114	12	12	102	19	78

It can be seen that SVM prediction heavily favours “Not Used" predictions over “Used" due to the much higher number of occurrences of zeros, this is common with many appliances in NILM, where for the vast majority of time they will be in an off or in a standby state. However, in hours of peak usage, e.g., mornings between 6am-9am “Used" will almost always be predicted. Balancing data can also be problematic, either reducing the amount of data available considerably or making models more prone to predicting one appliance as another resulting in many more false positives.

4.4.2 ANFIS-based prediction

The adaptive network-based fuzzy inference system (ANFIS) is a type of artificial neural network that is based on the Takagi-Sugeno fuzzy inference system [69], suitable for time-series prediction. Previous papers have focused at load prediction such as [62] which looks at hourly prediction of large scale power system load. Ying [147] looks at a number of prediction algorithms including ANFIS for regional load prediction on a yearly basis for regional load in Taiwan concluding that ANFIS is the most effective of the models trialled. In Ozturk [115], ANFIS is applied to predict residential customer load profiles for a number of appliances use two models per appliance one for time of use and another for operating duration.

ANFIS-Hourly Prediction

Using ANFIS, we investigate kettle consumption prediction on a daily and hourly basis. Similarly to the SVM case, we define $\mathbf{Y} = \Big[y(1), \cdots, y(t)\Big]$, as a vector of random variables indicating energy consumption due to kettle usage within consecutive hours. Then, to predict the value of $y(t+1)$ we built an ANFIS model taking as input $\Big[y(t-25),\quad y(t-24),\quad y(t-23),\quad y(t-1),\quad y(t)\Big]$, which was heuristically found to lead to the highest prediction accuracy.

One can see that the ANFIS model is built using the previous hour and the same time the previous day since these usage patterns influence the following hour the most. Built-in MATLAB Fuzzy C-means clustering [14] was used to generate a fuzzy interference system, which provided better results than grid partitioning. Clustering was done with the Sugeno-Type Fuzzy Inference method [135].

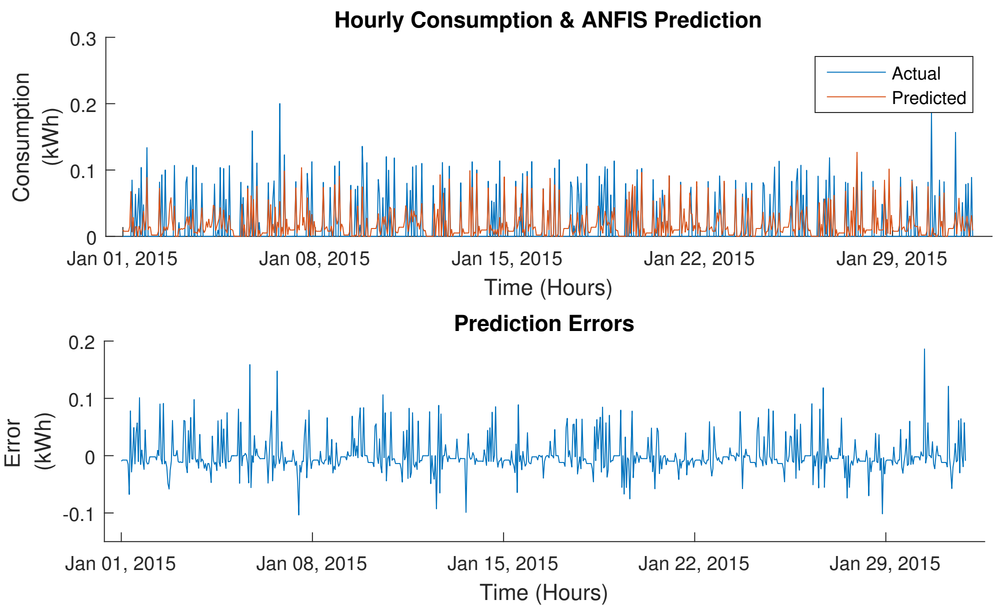


Fig. 4.11. ANFIS-based prediction for House 20. Error is $\varepsilon = y - \hat{y}$, e.g. over prediction is negative

Fig. 4.11 shows the obtained results for House 20. In some cases, there is a large error close to 0.2kWh predicted in a single hour. Negative predictions are set to zero as they effectively predict no usage. Fig. 4.12 shows the number of uses against the RMSE for each hour. It can be seen that 7am has a significantly lower RMSE due to a high number of predictable uses in the morning; additionally, 8am is likely being affected by the the previous hour as additional usage is based on a number of factors such as an occupant being late to rise, time of year or wanting another drink. Other examples are that 1pm has a high number of uses, however it also has a relatively high RMSE due to occasional uses at lunch time.

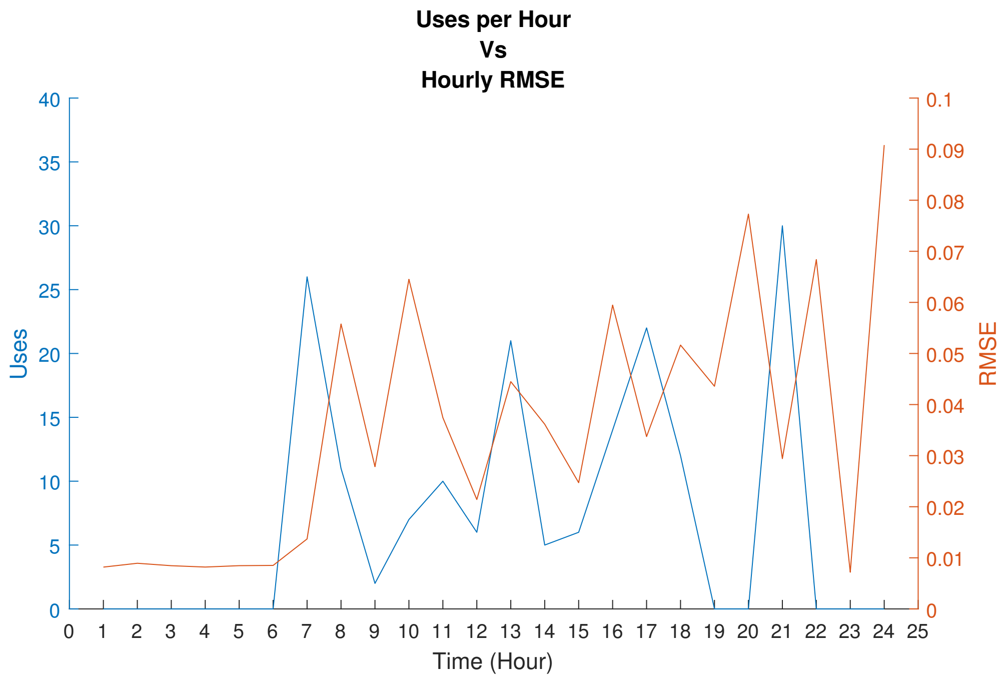


Fig. 4.12. Uses during hour (Blue) and average prediction RMSE per hour (Red) over the month of January

ANFIS-daily Prediction

Next, we investigate the accuracy of a daily prediction model. This model uses an input vector of: $\left[isweekend(t), y(t-7), y(t-6), y(t-2), y(t-1), y(t)\right]$ to predict $y(t+1)$, where $y(t-i)$ denotes kettle consumption i days ago. Note that to predict kettle consumption in kWh for the next day, we use the consumption of the current day, one, two, six and seven days ago to capture established weekly routines. The results for House 20 can be seen in Fig. 4.13.

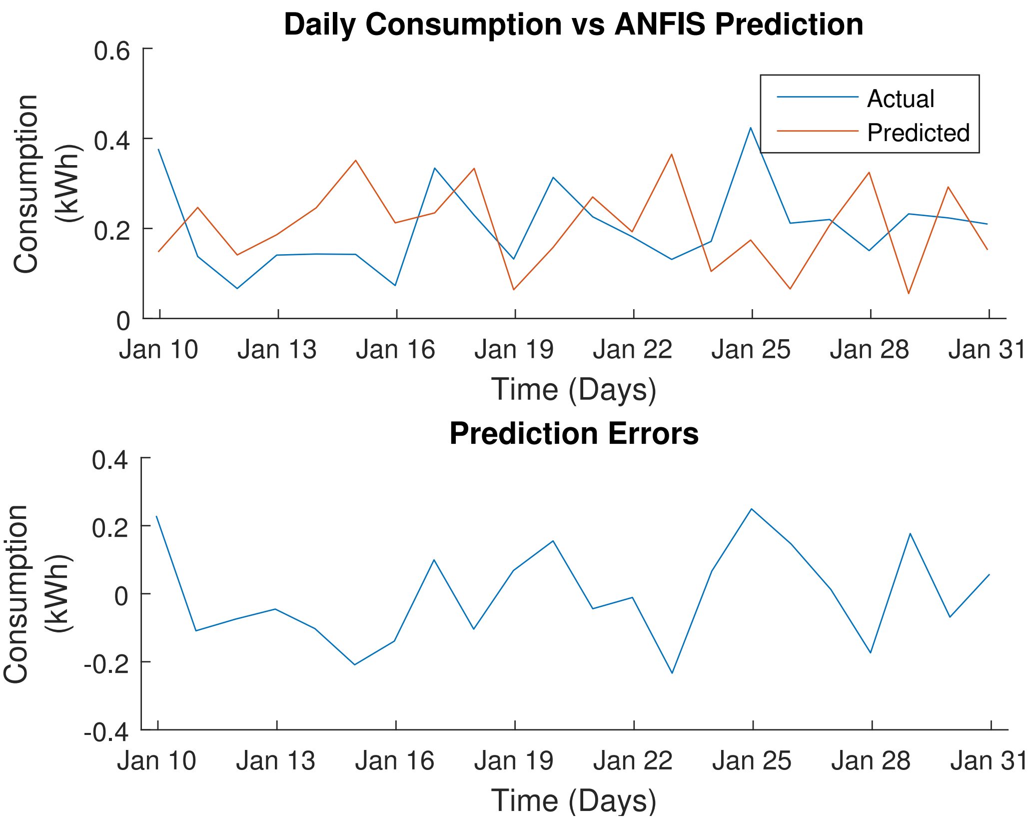


Fig. 4.13. Daily consumption comparison for House 20 over January 2015.

Table 4.8
RMSE of different training months

Tested \ Trained	Jan	Feb	Mar	Apr	May
Jan		0.13	0.07	0.12	0.15
Feb	0.00		0.06	0.06	0.00
Mar	0.05	0.02		0.01	0.14
Apr	0.13	0.01	0.13		0.13
May	0.04	0.10	0.04	0.06	

Table 4.8 shows the results for the same house over 5 months. It can be seen from the table that some months better represent a households general pattern with

comparison to others although this can be greatly affected by external factors, e.g. holidays. The results show that even within a single house multiple models may be required to account for variations in occupant behaviour based on larger changes such as the seasons.

Table 4.9 shows RMSE for the hourly and daily models for the 14 houses in the study, obtained using the ANFIS method. Prediction accuracy varies per household. Some households have very low error rates around work periods; however, higher errors tend to occur during unexpected hours, e.g., 4pm just before the expected return from work.

Table 4.9
Prediction RMSE for kettle energy consumption obtained using the proposed ANFIS-based prediction.

House	Hourly (kWh)	Daily (kWh)
2	0.054	0.0294
3	0.042	0.0114
4	0.025	0.0025
5	0.050	0.0558
6	0.050	0.1537
7	0.029	0.1158
8	0.043	0.0180
9	0.058	0.0783
11	0.055	0.1982
12	0.020	0.0662
13	0.040	0.0450
17	0.069	0.3681
19	0.043	0.1709
20	0.032	0.0230

4.5 Summary

This chapter presents a number of approaches to understanding, modelling and forecasting kettle usage in households purely from individual kettle load data. Time of use analysis shows well defined patterns of use with respect to weekdays during standard “office hours", pattern variation depending on type of occupancy and general daily schedule, and seasonal variation.

Our analysis shows that kettle usage patterns are semi-regular with clear peak times (morning, evening around dinner) and sporadic usage otherwise during the day. Usage patterns are correlated to working patterns, family size, and age group: working couples will likely have no or only few uses between the hours of 9am and 5pm, while retired couples would have more sporadic usage of kettle. This pattern motivated us to use prediction tools, such as SVM and ANFIS, to forecast kettle demand hourly and weekly with accurate results.

Additionally, we show quantitatively, in-line with previous studies, that a significant percentage of households do overfill their kettle. Our study aimed to improve on previous which had consisted of self reported questionnaires, due to the technical challenges of in depth measurement. However, a bigger factor is reheating water soon after it has boiled. In these cases households that appear not to overfill, based on the number of occupants, waste energy on reheating or reboiling.

The habit of reheating is prevalent across many of the households, which can be addressed by informing them that their habit of refilling/reheating is detrimental to being economical. This process of reheating could also be contributed to a lack of communication or forgetfulness: if a person is unaware the kettle has been boiled recently or has left it for a period of time there is a tendency to reheat. Analysis of the kettle usage data suggests that most people perform a reheat soon after it has boiled, in some cases less than 5 minutes after boiling. Options to improve on this could be as simple as audio queue’s from the kettle similar to a microwave, or more in depth such as disallowing reboiling above a certain temperature threshold.

Chapter 5

Appliance Modelling

5.1 Introduction

As a follow up to the previous chapter (and the reception of the associated paper [104], cited over 50 times), modelling of additional new appliances using new real world data was seen as advantageous. Nestec S.A. (commercially known as Nestlé) reached out to us to help update their appliance models making use of the data collection and procedures we had created prior. Large companies are required to complete a LCA for any new product they want to bring to market. The LCA is the complete product story, taking into account the gathering of raw materials, their transport, product manufacture, product usage and finally disposal. Companies rely on models to estimate the consumption of every part of the process, the models can range from highly complex (See Eq. 5.1) to very simplistic e.g. trucks consume x tonnes of carbon per y distance. Therefore developing models which are accurate but simple to implement provide great value, these up to date real world models allow companies to improve their current assessments and enable regulators to better understand real world usage and consumption when creating energy labelling schemes or devising testing procedures [68]. The content of this chapter is taken from the paper on ‘Appliance electrical consumption modelling at scale using smart meter data’ [110].

5.2 Methodology

The aim is to provide models which can be easily applied to numerous existing smart meter energy datasets (see [107] for partial list) or used in large longitudinal consumer and energy studies, where information may be lacking with regards to specialised knowledge such as cooking settings, food temperature, appliance makes and models. The research hypothesis is that by using only smart meter energy data we can build accurate energy consumption models of major cooking appliances. To prove this hypothesis we first conducted a small field study to collect data for building mathematical energy consumption models, and then, validate the developed models using state-of-the-art models that use many parameters, difficult to collect in practice.

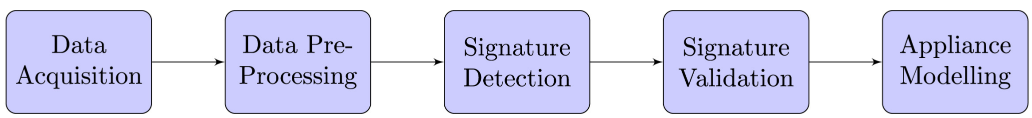


Fig. 5.1. Proposed methodology.

The methodology, from data acquisition to mathematical modelling, is summarised in Fig. 5.1. The input required is either total household energy consumption, which can be obtained directly from a household smart energy meter, or energy consumption of the appliance of interest, that can be measured using commercial plug meters [107]. Note that collected energy measurements are not readily usable for appliance modelling, requiring (1) *Data pre-processing* to filter out outliers, erroneous samples and noise, synchronise samples, and fill-in missing values; (2) *Signature detection* via edge detection or other NILM techniques [152] to isolate individual appliance power loads from the total household consumption (unless plug energy monitors are used at appliance level); (3) to acquire a relatively 'clean' or usable set of load signatures, *signature validation* is performed by using expert knowledge or an appliance database, to confirm that the signature belongs to a particular appliance of interest and remove signatures that were wrongly detected by NILM. The applied *appliance modelling* procedure is similar to that of [104].

5.2.1 Data Acquisition

Electrical consumption data for building the proposed appliance models, was collected during a field trial. Four households were recruited and monitored using a combination of Raspberry Pi (https://www.raspberrypi.org/) and the off-the-shelf Smappee (http://www.smappee.com/uk/home) real-time energy monitor. Measurements were taken at 1 second intervals, for up to 120 days. At the beginning of the study, all four households filled an Appliance Usage Survey containing information about appliance ownership and general patterns of appliance usage, again for the purposes of signature validation. The survey confirmed that all houses owned a microwave and all but one was likely to use it daily; ovens are most likely to be used in the afternoon and evenings.

In addition, for the purposes of validation, time diaries were also kept by household participants, including recording the start time of each appliance use along with settings of the appliance in use. Recorded start times in time diaries are in general accurate to around 1-2 minutes of the actual usage time. It was found that the recording of the recipe or type of food prepared using an electrical appliance did not affect the energy consumption in a significant way and thus did not affect the appliance consumption models.

5.2.2 Data Pre-processing

The first step is making sure the measurements are suitable for use by filtering out outliers and erroneous measurements, including spikes, following a similar methodology to [107].

The second step is synchronisation of readings. The raw data did not have a uniform sampling rate and therefore isolating appliance signatures becomes more complex when varying time frequencies are involved. Re-sampling timestamped data to one second intervals helps to achieve this uniformity; each sample is rounded to the nearest second (either up or down, to remove the millisecond component). Once this is completed, forward filling using the previous sample fills any gaps which have arisen. Forward filling however should be limited (2 seconds = 0.0016kWh at 3000W) to avoid skewing electrical signatures significantly. These methods are generic regardless of the data source.

5.2.3 Signature Detection and Validation

Smart meters measure only the household's total energy consumption. To use smart meter data for appliance load modelling, it is necessary to isolate appliance usage, which is commonly done via NILM [152]. For this study, we develop a simple, supervised, edge-detection based method given in Algorithm 2.

The input to the algorithm are smart meter active power readings collected at time t, pow_t in Watts (W), which are used to calculate $edges$, as $\Delta pow_t = pow_t - pow_{t-1}$, defined as difference in power value between sequential timestamps. An edge could be either a rising or a falling edge: a *RisingEdge* is a positive change in power (e.g., when an appliance is switched on or goes to a high consuming state) and a *FallingEdge* is a negative change in power (e.g., an appliance going to a low consuming state or is switched off).

The appliance metadata are data gathered from time diaries and individual appliance monitoring (IAM); these include instantaneous maximum and minimum observed power draw for the appliance of interest (pow_{min} and pow_{max}) and time durations (dur_{min} and dur_{max}), and the number of states. Since the power draw of the appliances of interest oscillates between high power state (e.g., heating state) and low power state (maintaining the heat), to avoid picking up low power state as switching off the appliance (that is, as a *FallingEdge*), we also estimate *switching_ time* as the time between an appliance going to a low power state (where it may appear to be off) before returning to a high powered state.

After all candidate *RisingEdge* and *FallingEdge* are detected, each *RisingEdge* is matched with the closest in time *FallingEdge* if the time duration between these two edges is within the acceptable limits, that is, between dur_{min} and dur_{max}, minimum and maximum appliance operation time duration observed from the Appliance metadata. Next, by comparing the time difference between the rising and falling edges with *switching_ time*, a check is performed to ensure that the *FallingEdge* is not due to appliance transitioning to a low power state.

Input : Data(Time(t), Power(pow_t)), Appliance Metadata
Load Appliance Metadata:
Minimum & Maximum Time Duration ($dur_{min}; dur_{max}$) ;
Minimum & Maximum Power ($pow_{min}; pow_{max}$);
switching_time;
Output: Time Duration t_{secs} & Energy Consumption E_{app}
RisingEdge = {}, FallingEdge = {}, Edges = {}, Signature = {} ;
for *(t, pow_t) in Data* **do**
    Calculate: $\Delta pow_t = pow_t - pow_{t-1}$;
    Store: RisingEdge$\{(t, |\Delta pow_t|)\}$ **IF** $pow_{min} \leq \Delta pow_t \leq pow_{max}$;
    Store: FallingEdge$\{(t, |\Delta pow_t|)\}$ **IF** -$pow_{max} \leq \Delta pow_t \leq$ -pow_{min};
end
i=0;
for t_{RE} *in RisingEdge* **do**
    for t_{FE} *in FallingEdge* **do**
        if $dur_{min} \leq t_{FE} - t_{RE} \leq dur_{max}$ **then**
            Store: Edges$\{(t_{RE}, \Delta pow_{RE})$, $(t_{FE}, \Delta pow_{FE})\}$;
            Calculate:
            $E_{app_i} = (\Delta pow_{RE} + \Delta pow_{FE}) \times (t_{FE} - t_{RE})/(2 \times 10^6)$;
            $i = i + 1$;
        end
    end
end
for *(t_{RE}, t_{FE}) in Edges* **do**
    if $|t_{FE} - t_{RE+1}| \leq switching\_time$ **then**
        $(t_{FE}, \Delta pow_{FE}) = (t_{FE+1}, \Delta pow_{FE+1})$;
        Calculate: $E_{app_{FE}} = E_{app_{FE}} + E_{app_{FE+1}}$;
        if $t_{FE+1} - t_{RE} \leq dur_{max}$ **then**
            Store: Signature$\{(t_{FE}, \Delta pow_{FE}, t_{RE}, \Delta pow_{RE}, E_{app_{FE}})\}$;
        end
    end
end
for *(t_{FE}, t_{RE}) in Signature* **do**
    Calculate Duration as: $t_{secs}[sec]$=t_{FE}-t_{RE};
end

Algorithm 2: Signature Detection Pseudocode. The appliance values, max and minimum powers etc, were optimised empirically. A table of these is available within the code respository. The Matlab code is available at `https://github.com/David-Murray/Matlab-Appliance_EdgeDetection`.

The proposed edge detection algorithm is limited by multiple simultaneous appliance uses and appliances with similar-valued consumption and duration. More sophisticated NILM methods can be used to provide higher disaggregation accuracy [56, 78, 152].

Algorithm 2 is run separately for each appliance of interest within the same houses smart meter data, and it returns the time duration and energy consumption of each detected appliance-of-interest use. To ensure that none of these detected uses comes from another appliance with similar load, or with multiple appliances being switched on/off at the same time, for each paired output Rising/Falling edge, we check with time diary or IAM measurements, if available, or validate it against known appliance signatures, e.g. Altrabalsi's signature dataset [3], to check if they fall within a valid range of values (e.g., similar consumption, duration, time of day). This is done to ensure that incorrectly disaggregated signatures are not used for the modelling stage.

All correct, labelled appliance signatures, namely the duration $t_{secs}[sec]$ and energy consumption ($E_{app}[MJ]$) per use, are then fed to the Appliance Modelling stage in Section 5.3 where *app* refers to the appliance being modelled, e.g., MiW for microwave.

5.3 Appliance Modelling

In this section, we perform curve fitting on the processed and cleaned energy measurements, namely time duration and energy consumption, to construct generalised models of appliance consumption based only on information gathered from smart energy meters as described in Section 5.2 and are widely available in numerous existing smart meter datasets that have been made public (see Table 1 in [107]). Note that other parameters such as temperature, food weight etc., which are difficult to gather during a large-scale longitudinal energy and consumer studies, are excluded. Resulting mathematical models for the microwave and the oven are presented next.

5.3.1 Existing Energy Consumption Models

We review the existing models for microwave and electric ovens for quantifying energy consumption related to food preparation in the domestic sector. Both appliances have some functionality overlap, but use completely different heating methods.

The state-of-the-art microwave (MiW) energy consumption model [130] requires knowledge of the food being cooked as well as water content, infeasible to collect at scale. It is given by:

$$E_{MiW}[MJ] = (m_{food} \times T_{elev} \times c_p + (m_{evap} \times e_{ew}))/e_{totmw}, \tag{5.1}$$

where the following parameters are required:

$$\begin{aligned}
m_{food} &= \text{mass of product (in grams [g])} \\
t_{elev} &= \text{difference in food temperature (in Celsius degrees } [^oC]) \\
c_p &= \text{heat capacity of the food product } [MJ/(g{\times}^oC)] \\
m_{evap} &= \text{mass of water evaporated [g]} \\
e_{ew} &= 2.26{\times}10^{-3} \text{ [MJ / g water evaporated]} \\
e_{totmw} &= e_{support} \times e_{trans} \times e_{magn} \times e_{mwcoup} \\
e_{support} &= 0.95 \text{ (efficiency for fan \& lamp \& controls)} \\
e_{trans} &= 0.86 \text{ (efficiency for transformation)} \\
e_{magn} &= 0.73 \text{ (efficiency for the magnetron)} \\
e_{mwcoup} &= 0.57 + 3.8{\times}10^{-4} \times m_{food} \text{ (for } 200 < m_{food} < 1000\text{g)} \\
e_{mwcoup} &= 0.95\text{(for } m_{food} > 1000\text{g).}
\end{aligned}$$

A similar microwave model is used in [86]. We note that other studies, briefly reviewed in the Introduction, such as [17,18,86,113], rely on heuristic measurements in laboratory conditions, or focus on preparation of particular dishes, or rely on power rating of the appliances and cooking recipes with the risky, and sometimes wrong assumption, that they will be followed.

The model for oven, validated in [130] using 23-59 litre ovens and data supplied by The Swedish Consumer Agency (`http://www.konsumentverket.se`) which had oven volumes ranging between 18-65 litres, is given by:

$$\begin{aligned} E_{OVEN}(MJ) = {} & E_{hu} \times V \times T + E_{mt} \times V \times T \times t \\ & + E_{hp} \times m_{tot} \times \Delta T + 2.26{\times}10^{-3} \times m_{wevap} \\ & + 3.34{\times}10^{-4} \times m_{frozen}. \end{aligned} \tag{5.2}$$

The model requires the following input parameters:

$$\begin{aligned} E_{hu} &= 2.0 * 10^{-4} \\ &\quad \text{Energy for heating one litre of oven volume. } [(\text{MJ}/(\text{litre}{\times}^{o}C))] \\ V &= \text{volume of the oven in [litres]} \\ T &= \text{temperature the oven is heated to } (-20^{o}C \text{ start temp})[^{o}C] \\ E_{mt} &= 4.2 * 10^{-6} \\ &\quad \text{Energy for maintaining a certain oven temperature in} \\ &\quad \text{one litre for one minute } [MJ/(litre{\times}minutes)] \\ t &= \text{time for cooking (excluding preheating) [minutes]} \\ m_{tot} &= \text{mass of product [g]} \\ m_{wevap} &= \text{mass of water evaporated [g]} \\ m_{frozen} &= \text{mass of the product if frozen [g]} \\ e_{hp} &= \text{heat capacity of the food product } [\text{MJ}/(\text{kg}{\times}^{o}C)]. \end{aligned}$$

Obviously, many of these parameters cannot be easily collected during a field study, since collecting some of these parameters requires information on devices not widely available, such as scales, thermometer, heat capacity of foods etc. Furthermore, the collection is often timely, cumbersome, expensive and cannot be expected to be conducted at scale by untrained volunteers in field studies. Thus, to the best of our knowledge, scalable appliance consumption models, suitable to estimate accurate energy consumption that captures varied cooking styles in a longitudinal study, is missing.

5.3.2 Microwave Consumption Model

The microwave's power consumption over time is pulsed as microwaves are typically run at either 100% power, or variations on this at 20% intervals (20, 40, 60, 80%). To develop a model, the time duration (t_{secs}) and energy consumption (E_{MiW}) values were obtained using the Signature Detection algorithm (i.e., Algorithm 2) as explained in Section 5.2.3. The other two parameters used in the model are: $power_{watts}$, that is, the rated magnetron power of the microwave, obtained from the appliance manual or information labels of the respective microwaves, which is fixed for each microwave; and $setting_{percentage}$, which denotes the power setting of the microwave (e.g., 60% of $power_{watts}$), as set by the consumer per run and, for the purpose of model development, was obtained from time diaries.

Following curve fitting on the output of Section 5.2.3, which isolated 584 valid microwave signatures, the following mathematical linear model for the consumed energy is obtained:

$$E_{MiW}(MJ) = (0.0010899 \times t_{secs}) + (power_{watts} \times 5.8681 \times 10^{-6}). \tag{5.3}$$

It should be noted that the above microwave model is limited by the availability of consumption data, and thus it is inadvertently biased towards the tested microwaves, which ranged between 700W and 900W. Reduced power band settings, e.g., 80% of 700W (560W), will be estimated well while, 60% of 700W will be over estimated. To reduce errors of predicting lower powered microwaves and, as such give a better overall range, 700W uses were weighted appropriately to help improve model predictions. This resulted in a more accurate quadratic microwave model, given in equation below:

$$
\begin{aligned}
E_{MiW}(MJ)^{Quadratic} = {} & (0.0015893 \times power_{watts}) \\
& + (-1.4823 \times setting_{percentage}) \\
& + (0.0001032 \times t_{secs}) \\
& + (0.0020663 \times (power_{watts} \times setting_{percentage})) \\
& + (-9.9386e - 08 \times (power_{watts} \times t_{secs})) \\
& + (0.0012324 \times (setting_{percentage} \times t_{secs})) \\
& + (-2.237e - 06 \times (power^2_{watts})) \\
& + (0.018915 \times (setting^2_{percentage})) \\
& + (-5.1134e - 07 \times (t^2_{secs})).
\end{aligned} \tag{5.4}
$$

Compared to the previous linear model, this quadratic model incorporates the power percentage setting ($setting_{percentage}$) of the microwave as a variable (e.g., 60%, 80%). Note that, in contrast to previous models (see Section 5.3.1), the model predicts the amount of consumed energy E_{MiW} based purely on the duration t_{secs}, power rating $power_{Watts}$ (which is a fixed parameter for each microwave, e.g., 900, 800W, etc.), and $setting_{percentage}$ of the microwave, which can be easily recorded. Alternatively, the model can predict microwave setting, based on the measured E_{MiW} and t_{secs}, that is, purely from the Algorithm 2's output.

We use Normalized Root Mean Square Error (NRMSE) as a measure of the error. NRMSE was chosen as large errors are more heavily weighted than Mean Absolute Error (MAE); this is desirable, as a lower NRMSE shows that the model is less prone to inaccurate estimations and therefore when performing LCA or yearly consumption estimates the model is proven to be a reliable guide. The NRMSE captures the error in the models predicted consumption against the actual measured consumption, and was calculated as:

$$
NRMSE = \frac{1}{\frac{1}{N}\sum_{i=1}^{N} E_{app_{measured_i}}} \sqrt{\frac{\sum_{i=1}^{N}(E_{app_{model_i}} - E_{app_{measured_i}})^2}{N}} \tag{5.5}
$$

where $E_{app_{model_i}}$ and $E_{app_{measured_i}}$ denote the energy consumption estimated by the model and measured, during the i-th use, respectively, of the appliance in question (microwave or oven), and N is the number of appliance uses detected in the dataset.

Firstly the proposed models is validated visually by comparing the predicted energy consumption values from the model with actual energy consumption values, as shown in Figure 5.2. Six distinct power settings were used in the time diaries gathered, resulting in $power_{watts} \times setting_{percentage}$ consumption bands of 900W, 800W, 700W, 560W, 420W, and 280W of which, there were 28, 6, 148, 185, 208, 9 uses, respectively. The normalised root mean squared error (NRMSE), calculated using Eq(5.5), for these 6 settings is 0.44, 0.15, 0.002, 0.04, 0.05, and 0.04, respectively.

It can be seen from Figure 5.2, that the Predictions from the quadratic model show a very good fit with the actual measurements for all power settings, including the 900W setting that shows the highest NRMSE. Therefore, we conclude that the NRMSE between actual measurements and predicted values from the model of the microwave of around 0.5 is acceptable.

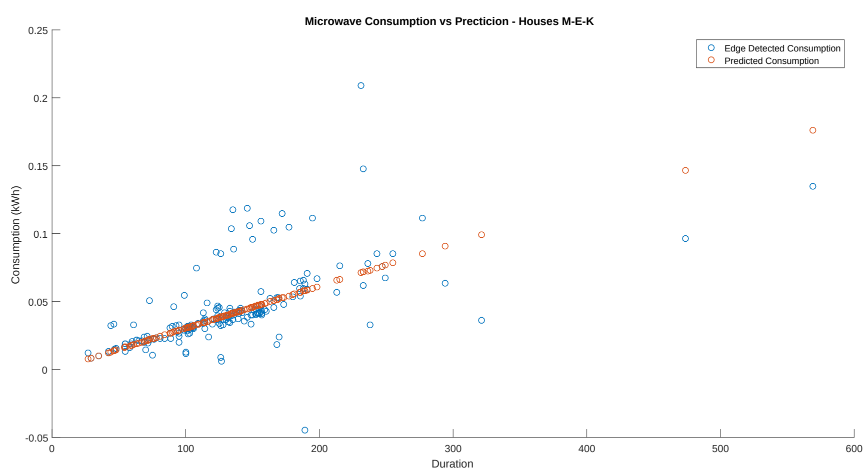


Fig. 5.2. Edge detected microwave consumption (actual) vs predicted values using the proposed quadratic model Eq. 5.4.

Furthermore, Table 5.1 shows how the model output compares to that of the state of the art [130], described in Section 5.3.1. As in [130], the ratio, measured as the predicted energy usage from the model divided by the measured energy usage, is used for assessing the accuracy, i.e., the results closer to 1 indicate better prediction. The better performing method is highlighted in green.

Microwave Power	Duration (mins/secs)	Son. Model Prediction	Proposed Model	Proposed Model (Quad)	Son. Measured	Ratio Son.	Ratio Proposed	Ratio Proposed (Quad)
750	3/5	0.23	0.21	0.23	0.22	1.05	0.95	1.09
750	3/5	0.20	0.21	0.23	0.22	0.91	0.95	1.09
800	2/53	0.25	0.19	0.23	0.22	1.14	0.86	1.05
700	3/18	0.25	0.22	0.23	0.22	1.14	1	1.05
750	4/30	0.37	0.30	0.32	0.32	1.16	0.94	1.00
750	4/30	0.28	0.30	0.32	0.32	0.88	0.94	1.00
800	3/0	0.28	0.20	0.23	0.22	1.27	0.90	1.09
700	4/0	0.28	0.27	0.27	0.25	1.12	1.08	1.08

Table 5.1
Comparison of the model prediction of Sonesson [130] (abbreviated to Son.) and the proposed linear Eq. 5.3 & quadratic model Eq. 5.4 using the values recorded by [130] given in [MJ].

We can see that the proposed quadratic model is more accurate than the proposed linear model and benchmark model of [130]. This confirms that knowledge of the food being prepared does not impact the consumption phase of the microwave in a significant manner.

In [18], it claimed that an 800W rated microwave at 100% for approx 8 minutes consumes 0.1kWh (0.36 MJ). Using the linear model the same microwave would be expected to consume 0.1394kWh (0.5 MJ), and the quadratic model would be consuming 0.1919kWh (0.69 MJ).

5.3.3 Oven Consumption Model

Oven-baking is still one of the most used cooking practices in Europe, with electric ovens present in most homes [16, 130]. The oven's used in this work are electric fan ovens. The oven operates in two stages: (i) preheating/restoring heat, when the oven typically draws a constant power load, and (ii) maintaining heat. The set temperature, type and amount of food being baked, affects only the time the oven will be in each of the two states. Within this section, to build a model, the time duration (t_{secs}) and energy consumption (E_{Oven}) were obtained from the edge detection algorithm (i.e., Algorithm 2) as described in Section 5.2.3; the set temperature ($temp_{Oven}$) of the oven has been obtained from time diaries, or if this is not available, estimated based on the initial preheating duration, which is close to linear for each oven.

Most cooking recipes require the oven to be pre-heated to a set temperature. During this phase, the oven door is usually kept shut meaning that the time to reach temperature will be linear with the assumption that the oven is in good working order and condition. The cooking phase may consist of door openings as food is introduced and later checked or taken out. When the door is opened the oven will lose temperature and require another heating stage to return to the set temperature. Different oven settings produce different signatures; however, they still retain the general heating/maintenance cycle with adjusted timings. In the models we assume that the door is not opened frequently or unnecessarily and that only one dish is added for the cooking stage. Note that the model is limited by the available oven data, that is, we expect that the model will not accurately represent very large ($\geq$ 80 litres) or very small ovens ($\leq$ 50 litres).

In developing a scalable model for the oven load, we start from the Sonesson model given by Eq.(5.2) and attempt to remove dependency of the model on parameters that are difficult to acquire.

First, using the data collected from smart meters (see Subsection 3.1) we estimate two static parameters of Eq.(5.2) - E_{hu} and E_{mt}. As the values reported in [130] were set back in 2003, the obtained estimates differ from Eq.(5.2) due to the newer and more energy efficient oven designs. This situation further motivates the proposal for appliance modelling methodologies that can be updated regularly, without significant effort and cost.

In the field study, the 58-litre oven is the oldest and worst performing from an energy point of view. The average energy values given by Eq.(5.2) for $V = 58$ litre oven, for energy needed for heating and maintaining the heat, respectively, are $E_{hu} = 2.0{\times}10^{-4}$ $[MJ/(litre{\times}^oC)]$ and $E_{mt} = 4.3{\times}10^{-6}$ $[MJ/(litre{\times}minutes)]$. The experiments show that this value should be lowered to around $1.1055{\times}10^{-4}[MJ/(litre{\times}^oC)]$ for E_{hu}. The recommendation is to increase slightly the E_{mt} value from Eq.(5.2) to $1.7288{\times}10^{-5}[MJ/(litre{\times}minutes)]$ due to the increase in the size of oven cavities. Table 5.2 summarises the mean and variance of the values obtained from the models from the detailed field study of two ovens, where the combined average is the recommended value for E_{hu} and E_{mt}.

Volume (Litre)	E_{hu}		E_{mt}	
	Mean	Variance	Mean	Variance
58	1.2905×10^{-04}	1.6243×10^{-10}	2.1575×10^{-05}	2.3426×10^{-11}
74	8.0806×10^{-05}	1.6753×10^{-10}	1.0398×10^{-05}	1.0288×10^{-11}
Combined Avg.	1.1055×10^{-04}		1.7288×10^{-05}	

Table 5.2
Experimentally obtained values for the energy needed for heating and maintaining the heat, for the two ovens in the study from Eq. 5.2

Table 5.2 is an update to the model of [130] described in Section 5.3.1; however the large number of variables that need to be known hamper using the model in large-scale studies. Instead, we propose the following scalable model, if the oven volume is unknown:

$$E_{Oven}[MJ] = (0.0037372\times temp_{Oven}[^oC]) + (0.0011084\times t_{secs}[s]). \tag{5.6}$$

The difference in the volume of the oven causes variations resulting in an NRMSE between predicted and actual values of 0.25, as calculated by Eq.(5.5). We conclude that volume is important for estimating power consumption, especially with the current large variation in oven volumes on the market. A revised model that incorporates time duration, set temperature and oven volume is given by the following equation:

$$\begin{aligned} E_{Oven}[MJ] = & \\ & (-0.064859\times V[Litre]) \\ & + (0.028626\times temp_{Oven}[^oC]) \\ & + (0.00094777\times t_{secs}[sec]). \end{aligned} \tag{5.7}$$

With the addition of the Volume (V) component, Eq. (5.7) has a reduced NRMSE of 0.19 when compared to Eq.(5.6). Fig. 5.3 shows the validation results for the energy consumption vs. operation time for the two ovens, this time showing a much better agreement between the predicted values from the revised model from Eq.(5.7) and the actual measured values. Therefore, we define an acceptable value for NRMSE between actual measurements and predicted values from the model of the oven to be around 0.3.

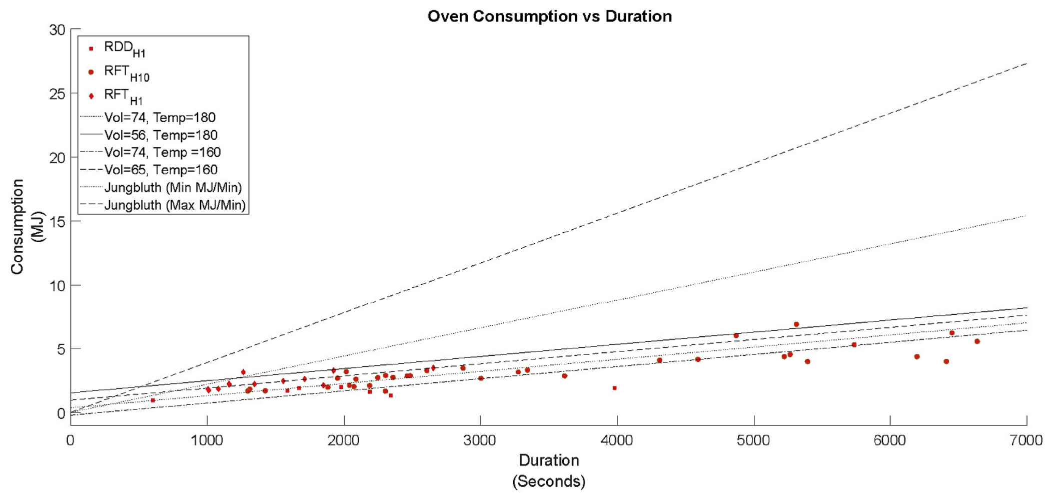


Fig. 5.3. Predicted oven model traces against actual oven consumption.

5.4 Summary

The models presented in this chapter show an advance and improvement over models widely used by industry. These models require significantly less information to get a reasonable consumption estimate for given parameters, additionally they by design include user specific habits as the data is taken from real world usage rather than ideal lab conditions. The methodology means that updating the models is also less intensive than previous state of the art, and would be easy to carry out especially by a large company or industry partner. The comparison against the currently used examples shows the need for this work, and for continuous refreshing of the models given the previous were updated around 20 years ago and do not take into account massive improvements in design, manufacturing, and energy efficiency.

Chapter 6

Load Disaggregation

6.1 Introduction

The content of this chapter is taken from the paper on 'Transferability of neural networks approaches for low-rate energy disaggregation' [109] and 'Transparent AI: explainability of deep learning based load disaggregation' [108].

Energy disaggregation of appliances using NILM represents a set of signal and information processing methods used for appliance-level information extraction out of a meter's total or aggregate load. Large-scale deployments of smart meters worldwide and the availability of large amounts of data, motivates the shift from traditional source separation and Hidden Markov Model-based NILM towards data-driven NILM methods. Furthermore, we address the potential for scalable NILM roll-out by tackling disaggregation complexity as well as disaggregation on houses which have not been 'seen' before by the network, e.g., during training. We focus on low rate NILM (with active power meter measurements sampled between 1-60 seconds), this represents the best case scenario in terms of performance against processing/data cost (high frequency data (>10kHz PLAID dataset [45] or UK-DALE [78]) provides minimal improvements in performance [93]) and present two different neural network architectures, one, based on convolutional neural network, and another based on gated recurrent unit, both of which classify the state and estimate the average power consumption of targeted appliances. Our proposed designs are driven by the need to have a well-trained generalised network which would be able to produce accurate results on a house that is not present in the training set, i.e., transferability. Performance results of the designed networks show excellent generalization ability and improvement compared to the state of the art.

Table 2 in [63] summarises the state-of the-art DNN-based NILM methods. Though prior work considered transferability across houses within the same dataset (e.g., [77] [35]), few at the time had considered cross dataset evaluation, [65] was one exception (using curve fitting and DBSCAN to generate a generic model for each appliance), i.e., transferability across datasets. Transferability has become much more common, and the UK-DALE & REFIT dataset are the two most popular test sets. Transferability is particularly challenging due to the large variation in sampling rates, appliances, usage patterns, climate, age (different energy labels) and electrical specifications (e.g., voltage, phase) across datasets. Cross-dataset transferability is very much needed in order to be able to use the developed models at scale and for practical commercial usage.

The main contributions are:

(a) showing that a single neural network can be trained to accurately target *at once* both NILM problems (which have been addressed separately or unevenly so far), that is, to identify occurrences AND estimate the contribution to the total load of a specific appliance. Our approach addresses these problems *inseparably* with flow of information from the classification part of the network to the load estimation part. This is in contrast to previous work that focused on binary classification of appliance state (ex. [7,35,80]) or estimation of appliance load mainly (ex. [77,149]).

(b) The proposed architectures are designed to facilitate successful transfer learning between very distinct datasets.

(c) Our proposed networks represent a significant reduction in complexity (the number of trainable parameters) compared to previous approaches [7,35,77,80,149], even though our proposed networks are tested on arguably more challenging real datasets.

(d) We do not make use of synthetic data and perform both training and testing on *balanced* data to avoid the issue of bias due to lack of appliance activations, which is a feature of many NILM datasets.

In order to demonstrate transferability, three datasets are used, namely UK REFIT [107] and UK-DALE [78], which are known to have similar appliances, as well as the US based REDD [84], whose appliances are different in terms of electrical signatures compared to UK appliances. The difference in voltage between UK and US affects how appliances operate, specifically appliances with a heating element in the UK are usually 2000-3000 Watts in power, where a similar US product would be around 1500W.

6.2 Proposed Network Architectures

/

We introduce two networks, both of which are suited to processing temporal data: (1) a GRU architecture, as shown in Figure 6.1, and (2) a CNN architecture, as shown in Figure 6.2. Both architectures remain purposely simple with a two-branch layout, with the side branch considering state prediction and feeding it back to the main branch to assist with consumption estimation.

It is worth noting that prior work has generally focused on either state or consumption estimation, using a single-branch network [7, 80, 149], or attempting to rebuild the signal hence generating both state and consumption as an output [77] [35] [46] [56]. In the latter, an autoencoder network is used where the network takes in an aggregate window and attempts to rebuild the target appliance signal only; these network types require a large amount of labelled data and generally make use of synthetic data. In addition, each of our networks differs from the literature, by training on fewer epochs or by having many less trainable parameters.

The GRU is a variant of the LSTM unit, especially designed for time series data to handle the vanishing gradient problem of networks. As such, they are designed, as LSTM, to 'remember' patterns within data, but are more computationally efficient. GRUs have fewer parameters and thus may train faster or need less data to generalize. Therefore, a GRU is more suited to online learning and processing than the LSTM unit. The specific variation used in this section is the original version, proposed in [21], using an NVidia CUDA Deep Neural Network library (CuDNN) accelerated version and implemented in Keras (CuDNNGRU). The GRU network

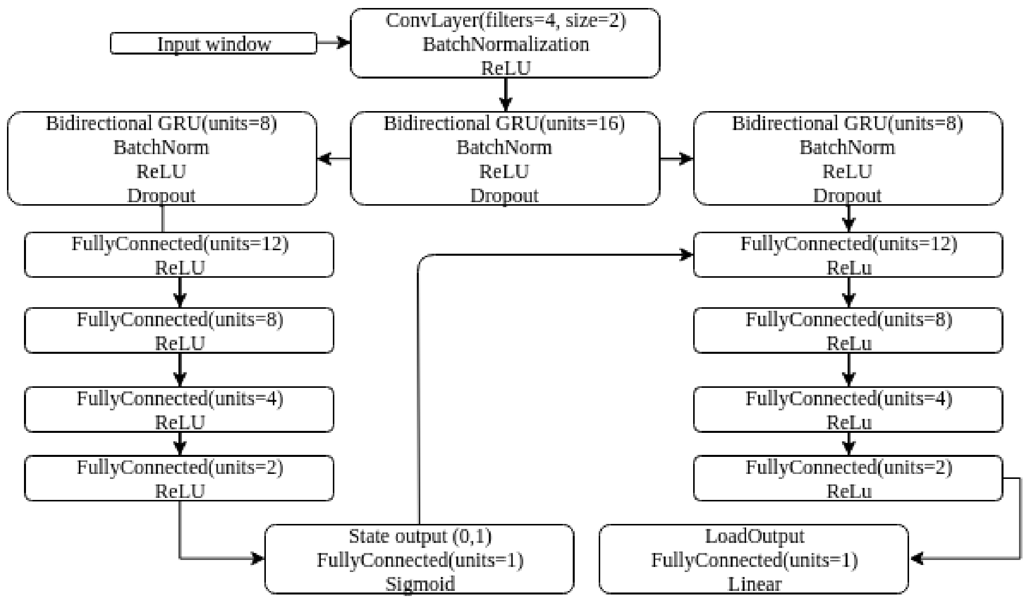


Fig. 6.1. Proposed GRU Network Architecture.

contains 4,861 parameters, out of which 4,757 are trainable and 104 non-trainable, i.e., hyper-parameters.

The proposed CNN consists of Conv1D (Keras) layers. 1D convolutional layers takes sub-samples (kernel_size) of the input window and steps through the time series building up a feature map for the input, this is done a number of times based on the number of filters the layer has (denoted by 'units' in Fig. 6.2) see Fig.6.3. The more of these layers that are connected the deeper and more complex the representations can be under the correct conditions. The CNN network contains 28,696,641 parameters, out of which 28,696,385 are trainable, and 256 non-trainable, hyper-parameters.

In both proposed networks, we make use of the ReLU function [112] as the network activation. This activation is monotonic and half rectified, that is, any negative values are assigned to zero. This has the advantage of not generating vanishing gradients, exploding gradients or saturation. Vanishing gradients are caused when the backpropagation algorithm calculates smaller and smaller gradients, as it works back through the network (output to input), as such the weights of the

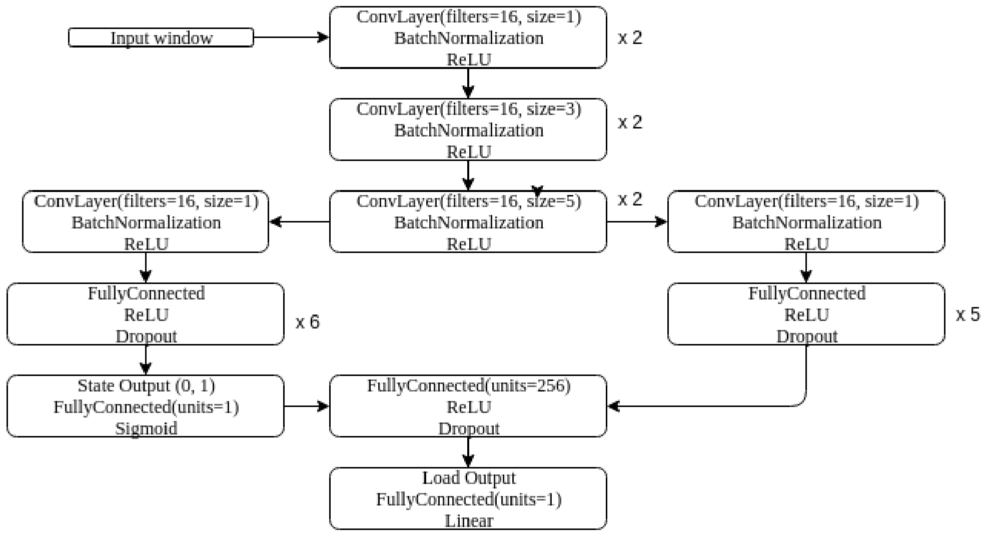


Fig. 6.2. CNN Network Architecture proposed by Srdjan Lulic for comparison.

higher layers (closer to the input) are never updated. Exploding gradients are the opposite, where the gradients become larger resulting in large weight updates causing the network to become unstable. However, ReLU activations can cause dead neurons; we therefore use drop out to help mitigate the effect of dead neurons which may have been generated during training. Dead neurons refer to neurons where their weights become 0 and can cause training to stall. Drop out can help mitigate the effect of dead neurons; by artificially severing the connections between neurons and the lower layers of the network (from input to output), these neurons are no longer considered during training.

Both proposed networks also use sigmoid activations for the state prediction and linear activations for the power estimation. The sigmoid function is used as it only outputs between 0 and 1, thus ideal for the probability that the appliance is on or off; in our networks, we assume a value greater than 0.5 to be on and anything below to be off. Linear activations can be any value and therefore are the best when estimating power. Both networks are implemented using the TensorFlow wrapper library Keras using Python3.

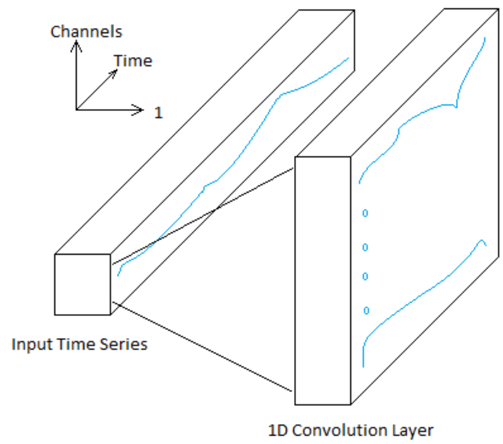


Fig. 6.3. 1D Convolution

6.3 Training of Proposed Networks

We train the proposed networks using both REDD and REFIT datasets, both containing sub-metered data. Note that sampling rates in these two datasets are different. To account for this, we pre-processed all data down to 1 second (using forward filling), then back to uniform 8 second intervals. Data was standardised by subtracting the mean, then dividing by the standard deviation across each window.

We train on a number of houses, except House 2, in both REDD and REFIT datasets, for the entire duration of the respective datasets. Testing is then performed on unseen House 2 in REDD and House 2 in REFIT, as well as UK-DALE House 1. The latter was used as it was monitored for the longest period of time. Details of houses used for training each appliance model are shown in Table 6.1.

An example of a typical day within each of the datasets is shown in Figure 6.4. It can be seen that the aggregate of the REDD dataset typically has very few appliance activations and a low noise level. On the other hand, the REFIT and UK-DALE datasets are similar in complexity with both having multiple large appliance activations with a complex low consumption noise level at below 500 watts. [82]

Four models are trained, one for each target appliance: dishwasher (DW), refrigerator (FR), microwave (MiW) and washing machine (WM). As each appliance has a different duty cycle, windows were chosen to capture a significant portion of

Table 6.1
Appliances and Houses Used

Appliance	REDD Houses	REFIT Houses	Window Size (samples)	On State (Watts)
MiW	1, 2, 3	2, 6, 8, 17	90 (12 mins)	> 100
DW	1, 2, 3, 4	2, 3, 6, 9	300 (40 mins)	> 25
FR	1, 2, 3, 6	2, 5, 9, 15, 21	800 (1.78 hours)	> 80
WM		2, 3, 10, 11, 17	300 (40 mins)	> 25

a single activation, shown in Table 6.1 along with the watt thresholds, obtained using training data, and are used to decide if the appliance is deemed to be on, i.e., if the threshold was exceeded.

From the paper [91] the noise-to-aggregate ratio (NAR) of a house is the amount of consumption which is not attributed to individual appliance monitors. A NAR of 25% would mean that 25% of the aggregate consumption is not sub-metered. This can be applied on an appliance by appliance basis to highlight the complexity of disaggregation. A high NAR would be more difficult due to a large amount of unknown consumption, while a lower ratio would show that there is fewer other appliances within the household. Fig. 6.5 shows the NAR ratio for the appliances most commonly used to show disaggregation results.

Input data was balanced to avoid a training bias within the networks, by limiting the majority class to that of the minority class (in nearly every appliance the minority class (appliance state) is off, apart from possibly refrigerator). Limiting the majority class was done by selecting samples at random until the classes were balanced. Validation data was then generated from randomly sampling from 10% of balanced training data. Each network was trained to 10 epochs with early stopping monitoring “Validation Loss"; if this failed to improve after 2 epochs the best performing network weights were used. Both networks used binary cross entropy as the loss function for state classification, for consumption the CNN uses mean square error (MSE) and the GRU logcosh. The CNN uses the stochastic gradient descent (SGD) optimiser and the GRU uses RMSprop.

Four performance metrics are used, F1-score (state prediction), Accuracy, Root MSE (RMSE) & Mean Absolute Error (MAE) (consumption estimation), which frequently appear in literature:

$$F1 = \frac{2 \cdot precision \cdot recall}{precision + recall} \tag{6.1}$$

$$Accuracy = (1 - \frac{\sum_{n=1}^{\infty} |e_t|}{2 * \sum_{n=1}^{\infty} true}) * 100 \ [\%] \tag{6.2}$$

$$RMSE = \frac{\sqrt{\Sigma_{i=1}^{n} \left(e_t\right)^2}}{n} [\text{ Watts}], \tag{6.3}$$

$$MAE = \frac{1}{n} \sum_{t=1}^{n} |e_t| [\text{ Watts}], \tag{6.4}$$

where n is the number of samples and

$$\begin{gathered}
precision = \tfrac{\text{True Positives}}{\text{True Positives+False Positives}}, \\
recall = \tfrac{\text{True Positives}}{\text{True Positives+False Negatives}}, \\
e_t = \text{predicted load} - \text{actual load}, \\
true = \text{actual load}.
\end{gathered}$$

The testing data was also balanced to avoid artificially improving scores; that is, in NILM datasets there is a higher likelihood that an appliance will be in an off state than it will be on (fridges and freezer being the exception). For example, a microwave may only be used once or twice per day or around 0.14% of a day. Therefore with unbalanced testing data, a network that only predicts the microwave in the off state will score well assuming that the microwave is used infrequently. Therefore, balancing the test data clearly shows the network is working well if it has an F1-score above 0.5.

Before assessing transferability across datasets, we establish baseline performance by training and testing on the same dataset. Tables 6.2 and 6.3 show the results of testing of each network on unseen House 2 from within the same dataset, i.e., electrical load measurements from Houses 2 of REDD and REFIT datasets were not used at all for training. The tables show that GRU tends to perform appliance state prediction marginally better (as shown by F1-score), while

Appliance	F1-Score CNN	F1-Score GRU	Accuracy [%] CNN	Accuracy [%] GRU	RMSE [W] CNN	RMSE [W] GRU	MAE [W] CNN	MAE [W] GRU
Microwave	0.95	0.95	**76.4%**	55.7%	**165.73**	252.17	**68.02**	127.79
Dishwasher	0.71	**0.74**	71.4%	**76.3%**	185.72	**136.79**	119.35	**98.90**
Refrigerator	0.67	0.67	**83.5%**	53.9%	**16.17**	31.15	**10.14**	28.31

Table 6.2
Testing on “unseen" House 2, after training the networks on all other REDD houses.

Appliance	F1-Score CNN	F1-Score GRU	Accuracy [%] CNN	Accuracy [%] GRU	RMSE [W] CNN	RMSE [W] GRU	MAE [W] CNN	MAE [W] GRU
Microwave	0.82	**0.87**	**68.7%**	65.6%	**88.75**	107.57	**35.49**	39.08
Dishwasher	0.82	0.82	82.9%	**84.8%**	**200.98**	211.78	82.74	**73.53**
Refrigerator	**0.93**	0.85	**76.9%**	64.1%	**14.77**	23.94	**8.56**	13.30
Washing Mac	0.79	**0.86**	**71.8%**	68.9%	**176.22**	190.05	**71.99**	79.33

Table 6.3
Testing on “unseen" REFIT House 2, after training the networks on all other REFIT houses.

CNN performs slightly better for appliance consumption (as shown by Accuracy, RMSE and MAE). However, overall, both networks perform in a similar manner and demonstrate very good performance when training and testing on unseen houses on the same dataset, when compared against a modern federated learning approach [151]. We thus show that the proposed methodology transfers well for unseen houses from within the same dataset.

6.4 Results

In this section, we demonstrate our networks' ability to transfer across datasets. This real-world test shows the ability of the network to handle completely unknown appliances, duty cycles and consumption - see, for example, Figure 6.6.

We first present the results when the models are trained using only REFIT houses (as per Table6.1), and tested on House 2 from the REDD dataset. This is shown in Table 6.4. Compared to Table 6.2, we can observe a drop in performance for MiW and DW, due to a difference in make/models of appliances between UK houses and the US house. Similar conclusions can be made from Table 6.5, where we show results when the models are trained using only REDD houses and tested on one REFIT house.

Note that in Table 6.5, the accuracy of Fridge is missing due to the window size selection; that is, with this window size, in the REDD dataset, there is always a fridge that is on, which means transferability between REDD to REFIT is biased to predicting the fridge always being on. This can be seen in Fig. 6.6, where the REDD fridge has a considerably smaller duty cycle than in the REFIT and UK-DALE datasets. This can be remedied by choosing a smaller window size; however in real-world applications this would only become apparent after testing, and multiple fridge networks may have to be generated.

Table 6.6 shows the results of training on REFIT houses and testing on unseen UK-DALE House 1. The UK-DALE dataset is similar to the REFIT dataset as it is also UK based, therefore has similar appliance types. This is reflected in the scoring metrics, as it has minimal performance drop compared to Table 6.3.

When comparing state prediction and consumption estimation performance of the proposed CNN and GRU networks across all results, we observe that they both perform similarly.

Though the metrics used are similar to those in the NILM literature, we cannot directly compare our consumption estimation results with the literature because the network outputs are different, additionally there are no commonly agreed benchmark data, so testing data selection even with in the same house is also not directly comparable. In [77] for example, the network output generates a single value which is then stitched together to recreate the original appliance window without the aggregate; as such the MAE value is the error at each individual point in time, not the error of the estimated consumption over the entire window as in our work. However, as an indication of classification performance, [77] achieves F1 scores of 0.26 for MiW, 0.74 for DW and 0.87 for FR when training on UK-DALE and testing on an unseen house also in the UK-DALE dataset. Our cross-dataset results in Table 6.6 show superior F1 performance for MiW and FR. Comparing results for House 2 REDD, i.e., Tables 6.2 and 6.4, our best F1 scores show similar results

	F1-Score		Accuracy [%]		RMSE [W]		MAE [W]	
Appliance	CNN	GRU	CNN	GRU	CNN	GRU	CNN	GRU
Microwave	0.41	**0.49**	**64.1%**	54.7%	120.01	**103.71**	90.96	**39.26**
Dishwasher	0.44	**0.57**	**50.2%**	39.6%	305.22	**284.34**	**183.27**	222.34
Refrigerator	**1.00**	0.98	**76.0%**	65.5%	**44.44**	59.92	**38.42**	55.11

Table 6.4
Training on REFIT houses only and testing on unseen House 2 from REDD.

	F1-Score		Accuracy		RMSE [W]		MAE [W]	
Appliance	CNN	GRU	CNN	GRU	CNN	GRU	CNN	GRU
Microwave	0.70	**0.78**	47.9%	**50.8%**	114.89	**100.17**	59.20	**55.82**
Dishwasher	**0.80**	0.62	**62.8%**	54.0%	431.61	**386.91**	**179.83**	222.43
Refrigerator	**0.67**	**0.67**	–	–	68.97	**56.57**	63.73	**53.37**

Table 6.5
Training on REDD houses only and testing on unseen House 2 from REFIT.

	F1-Score		Accuracy		RMSE [W]		MAE [W]	
Appliance	CNN	GRU	CNN	GRU	CNN	GRU	CNN	GRU
Microwave	**0.79**	0.7	**77.3%**	65.11%	**66.96**	144.46	**41.30**	63.70
Dishwasher	0.21	**0.46**	44.1%	**52.09%**	**43.09**	44.62	29.08	**24.97**
Refrigerator	**1**	0.69	**82.0%**	73.08%	**14.38**	19.56	**11.15**	16.69

Table 6.6
Training on REFIT houses only and testing on unseen UK-DALE House 1.

with respect to Nascimento [35] best scores for MiW (0.95), better for FR (1 vs 0.94) but slightly worse for DW (0.74 vs 0.82).

6.5 Neural Network Explainability

Recent years have seen significant research in defining, at a high level, how inference models can be designed to be explainable to end-users, in the case of NILM consumers or companies which have bought a NILM model to enhance their customer offering. Explainability leads to trust in data-driven AI systems ensuring that complex machine learning (ML) models underpinning these systems are understandable to the end user and decisions or recommendations are transparent. Despite a large number of publications from different disciplines, including many tutorial and feature articles [37, 48, 102], most explainable AI implementations focus on technology designs, e.g., for the purpose of removing the bugs in the code or improving the models [19], while other potential users of the technology are neglected. Furthermore, the bulk of the literature tends to focus on explainability of image processing [127] and natural language processing, while raw time-series sensor signals processing, e.g. energy measurements, is almost non-existent.

We focus on the explainability of deep-learning based NILM [54] of electrical smart meter data that provides feedback to householders or building managers about energy consumption of individual appliances [152] [34]. We demonstrate, using the popular sequence2point deep learning NILM architecture [34], how heatmaps can be used to explain NILM outputs.

We refer to a model being interpretable if it is possible to mathematically predict its output, and interpretability as the ability to support user comprehension of the model decision making process and predictions. Making ML models interpretable has become a hot topic of research, industry and policy makers [5, 37, 90]. We refer to explainability as the ability to explain the underlying model and its reasoning with accurate and user comprehensible explanations. Explainability is essential when assessing effects of biases in the data, degrees of fairness and other ethical implications of research, since the methods need to be replicated and tested in a new environment (using different, potentially biased dataset), and its decisions need to be mathematically tractable [37].

There have been only few attempts to explain time-series data models [128], where it is challenging to relate decisions to raw signals, and hence explanations have mainly been related to quantifying the importance of each feature; however, with deep learning models that take raw signals and integrate the feature engineering steps, this is often impossible. Similarly, there have been no attempts to explain NILM specifically besides [108], which targeted tech developers by visualising trained network weights at the early layers.

NILM or load disaggregation refers to estimating individual appliance load contributing to the metered household aggregate energy consumption without sub-metering. NILM can be seen as a single-channel source separation problem, where individual appliance load needs to be estimated from the total aggregate load, i.e., $p(t) = \sum_j p_j(t) + n(t)$, where $p(t)$ denotes the total power measured at time instant t, $p_j(t)$ is the power consumption of the j-th load, and $n(t)$ is the noise that includes measurement noise and unknown loads. NILM has been widely tackled in the literature as an event detection problem, assigning an appliance to a given aggregate load. More recently, NILM is an event estimation or regression problem, where a model is built based on labelled appliance load data in training, then estimates load directly from the aggregate testing data. Numerous approaches for NILM have been used previously, and a review can be found in [152]. To illustrate explainability tools for NILM, we use a sequence2point network of D'Incecco [34] that is a widely used for benchmarking deep learning based NILM work. We note that the approaches presented apply to other architectures also. The architecture of [34, 149], is a novel sequence2point approach for NILM, based on CNN that extracts meaningful latent features with appliance transfer learning and cross-domain transfer learning. A sliding window of the aggregate is mapped into a single middle value point of the targeted appliance, this way predicting the appliance consumption value for each sample in time. [34] presents results showing excellent performance of the proposed approach for a range of appliances on three datasets.

6.6 Explainability of NILM

We illustrate how NILM deep learning models can be interpreted and explained using the washing machine, considered a challenging appliance to disaggregate, due to multiple consumption states, with power values similar to numerous other appliances. To explain how the model makes decisions, we occlude (null values) part of the raw input and slide the occlusion window across the data. This highlights the features learned by the model, occluding a non-important section should result in little to no change in the networks output but occluding an important section e.g. the change in appliance state, we can highlight what is important and this can then explain why appliances may be missed, for example if another appliance was running at the same time causing the state change of the target appliance to be ‘hidden’. For each window position, we estimate the model’s singular point output. This is used to generate a heat map as shown in Figure 6.7 (bottom). This methodology makes no changes to the network’s internals unlike methods such as Attention Networks which require the addition new methods/layers.

In Figure 6.7 (top) we show an example of the input aggregate signal, target signal (washing machine), and predicted (non-occluded) signal. The horizontal axis shows sample number and vertical, consumed power. In this case we show a true positive result on the ECO dataset [10], the model being trained on the REFIT dataset [107]. The occlusion window blocks 50 consecutive samples and is stepped across the input window from index 0 to 549. This is then used to generate the heat map in Figure 6.7 (bottom). For a fixed sample point (horizontal axis), vertically, the values in the map correspond to the network output for different starting positions of the occluding window (from 0 to 549).

The heat map should be read diagonally to keep the occluded window stationary as the network target moves along the x-axis. The occluding window starting at x=0, y=0, would move diagonally down and left to stay at the same point as each window is passed to the network, due to the network targeting the centre point of the window. The heat map is aligned with the top plot along the x-axis to better indicate where the target point is, with the colour representing consumption estimation at a given point. The horizontal bar across the centre of the heat map represents where the centre point of the input sequence window is occluded, e.g., samples 249 to 299. When this occurs the network struggles to predict, as the input sample corresponding to target sample is null. Importantly, this leaves the model vulnerable should errors occur around the window centre, and makes a case for explaining how data is filled/processed to end users.

The ellipses in the heat map represent three key features. Ellipse 1 shows what we consider the main feature of the washing machine, the heating element turn on, around sample 705700; when this is occluded (top plot area 1) the predicted load drops significantly. Shown in Ellipse 2, the lowest load estimate occurs when this feature is fully occluded highlighting its importance, it represents the last stage of the washing machine cycle, the draining cycle (shaded area 2 on the top plot), when occluded the network looses confidence in it’s estimation massively reducing the heating element wattage estimation. The third ellipse highlights the spin cycle. In this case, the occluded section, which is not fully captured by ellipse 3 (and corresponding top plot shaded area), shows distinct band of blue which begins shortly after the washing machine heating element has turned off, but not immediately, indicating that the model is expecting a certain duration of spin cycle.

In Figure 6.8, we show an example that illustrates the limitation of the trained model to handle overlapping activations, where a number of appliances are used simultaneously. In this example, another appliance usage occurs at the end of the washing machine heating cycle (sample 575500). When occluding this area, we expect a truer estimate of the previous load. Indeed, the result is a much higher network prediction, shown by ellipse 1 in the heat map. Ellipse 2 shows the importance of the draining cycle in order to detect washing machine uses. If this segment is even partially occluded, the estimated consumption drops to near 0. Additionally, there is another appliance which overlaps this feature (Sample 575700, end of area 2 on the top plot); this, along with the second overlapping appliance, helps us to explain why the network likely missed this activation. Finally, occlusion 3 (in the top plot) corresponds to the false detection of the spin cycle that in Figure 6.8 has a number of unknown appliance uses causing network confusion. Ellipse 3 in the heat map plot shows a false positive occurring if the end of the second appliance is occluded, e.g., the network thinks that a second heating cycle is in progress. This false positive (along the y axis of the heatmap (moving the occlusion window) the network does not detect this activation unless the state change of the unknown appliance is occluded) shows the trained network can be confused by similar consumption made up of multiple appliances overlapping.

Heat maps provide a model agnostic way to visually interpret time series results, working with both sequence-to-sequence and sequence-to-point style networks. Depending on the complexity of the input and target signals, the number of learned features will become apparent when occluding the input signal. Depending on the size of the target signal, the size of the occluding signal can highlight features, and shrinking the occluding window can show what the model considers the most impactful features. This methodology could also be used to discover adversarial examples in which outputs are vastly influenced by the change of a single point in the input window e.g. a single value affected by noise could cause the network to fail to predict consumption correctly [134]. The visualisation of stacked plots, allows those not familiar with the field to understand features which are considered important, and could help to create a “stress testing" set of tricky examples to be used for trained model benchmarking.

Figures 6.7 and 6.8 also report the MAE, SAE and NDE performance measures for these particular uses as defined in [34]. The values of all three metrics are lower for Figure 6.7 compared to those of Figure 6.8, which is expected since the former is a TP sample whilst the latter is a FN. However, these measures do not clearly have a range of values that are comparable.

While interpretability explains the decisions made by the model, it is often focused on technical explanations (e.g. feature importance without visualisation) and not understandable by the end user, e.g., a householder trying to understand their appliance consumption estimate in regards to their electricity bill. In Dwivedi [40] a large number of methods are discussed along with their pitfalls. Thus we provide explainability by attempting to explain the measures in relation to the top plot in Figure 6.8 which is data they can understand (their consumption), with the heatmap. Clearly, the predicted consumption of the appliance is under-estimated compared to the actual. The MAE is the only metric that captures this wide difference in reconstructing the signal, compared to the other two metrics but does not necessarily explain the underestimation, which would not provide a realistic consumption to the end-user trying to understand the real consumption of their appliance. Over the entire dataset, however, MAE is less explainable as the MAE value becomes lower due to the fact that appliances spend a significant period turned off. Therefore, we wish to highlight that these metrics, commonly used for evaluating the performance of deep learning approaches in the recent NILM literature [118], are not truly explainable since they are not necessarily intuitive.

6.7 Summary

We address one of the biggest NILM challenges that is yet to be demonstrated and hence limiting commercial take-up: scalability. This is reflected in performance vs complexity trade-off of NILM solutions and the ability to disaggregate appliance loads, which have previously not been seen (or trained) by the NILM solution, i.e., transferability. Driven by the increasing availability of smart meter data, we thus design and propose two data-driven deep learning based architectures that perform appliance state prediction and classification estimation inseparably, and can generalize well across datasets. We show the ability of our trained CNN- and GRU-based networks to accurately predict state and consumption across 3 publicly available datasets, commonly used in the literature. We show that our proposed trained networks have the ability to transfer well across datasets with minimal performance drop, compared to the baseline when we train and test on the same dataset, albeit on an unseen household within the same dataset. Both GRU- and CNN-based networks show similar performance but the GRU-based network has fewer trainable parameters and is thus less complex than the CNN-based network. Secondly, explainability is key to NILM becoming more widely used in customer facing products. The ability to understand why a model is reacting a certain way to data, firstly helps to reassure a consumer about the validity of the model and secondly allows for a more targeted investigation into either improving training or adding additional post processing. We propose heatmaps as the best solution allowing for the visualisation of many sequence to point traces within a small diagram and to aide explainability requiring little explanation to consumers, where the model output data can be seen in line with understandable household consumption data without the need for explanation of the models internal layer structure.

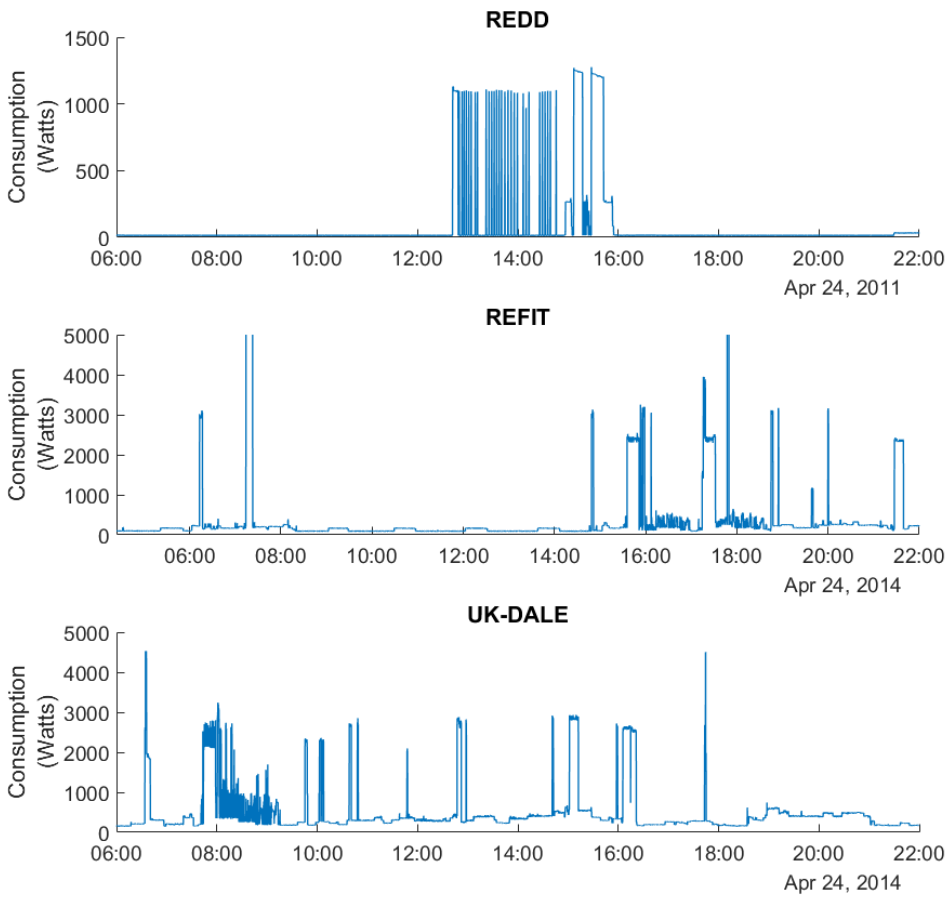


Fig. 6.4. Aggregate load measurements for a typical day for Houses 2 in REFIT and REDD datasets and House 1 UK-DALE, showing relatively higher noise levels for UK REFIT and UK-DALE houses.

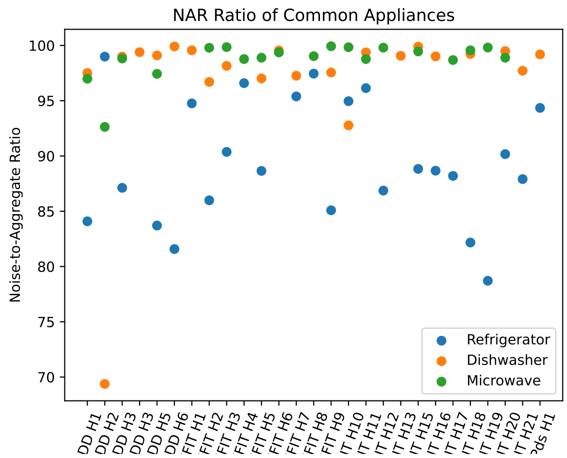


Fig. 6.5. Aggregate load measurements for a typical day for Houses 2 in REFIT and REDD datasets and House 1 UK-DALE, showing relatively higher noise levels for UK REFIT and UK-DALE houses.

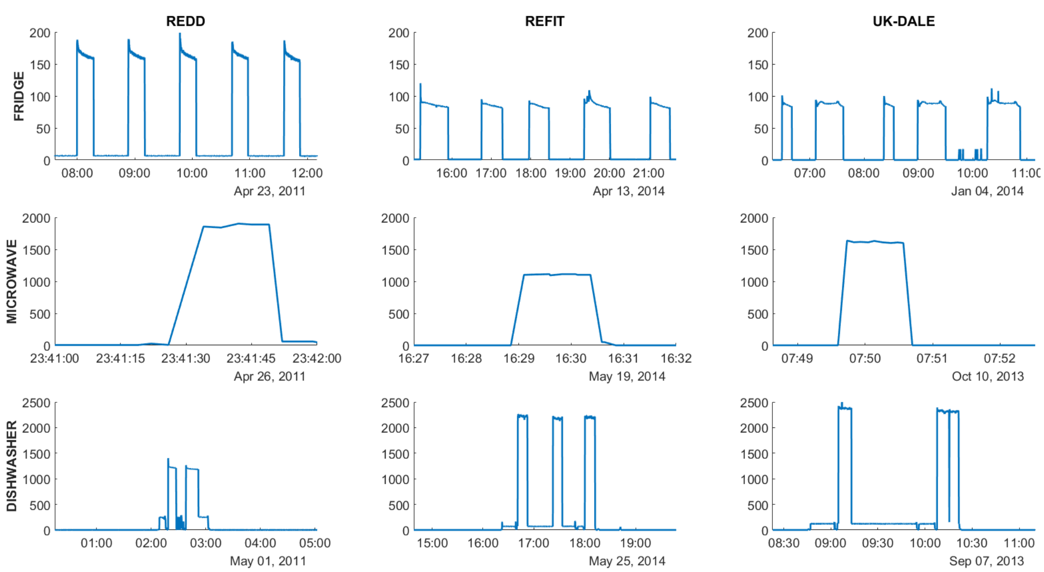


Fig. 6.6. Typical appliance signatures for MiW, DW and FR across REDD, REFIT and UK-DALE datasets.

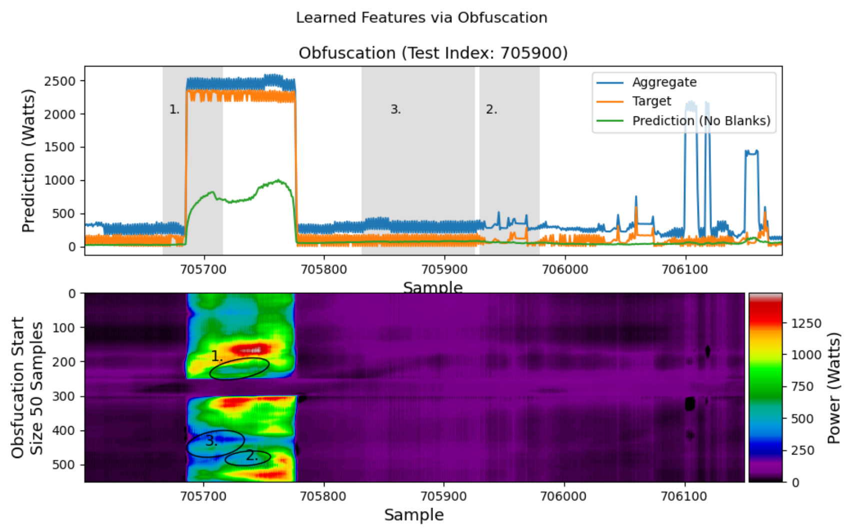


Fig. 6.7. The heat map generated for the Washing Machine in the ECO dataset, house 1. The model is trained using the entire REFIT dataset. The obtained performance measures for this sample are: MAE:292.73, SAE:0.62, NDE:0.45. The occlusion area numbers (top) and ellipse numbers (bottom) are explained more in the text.

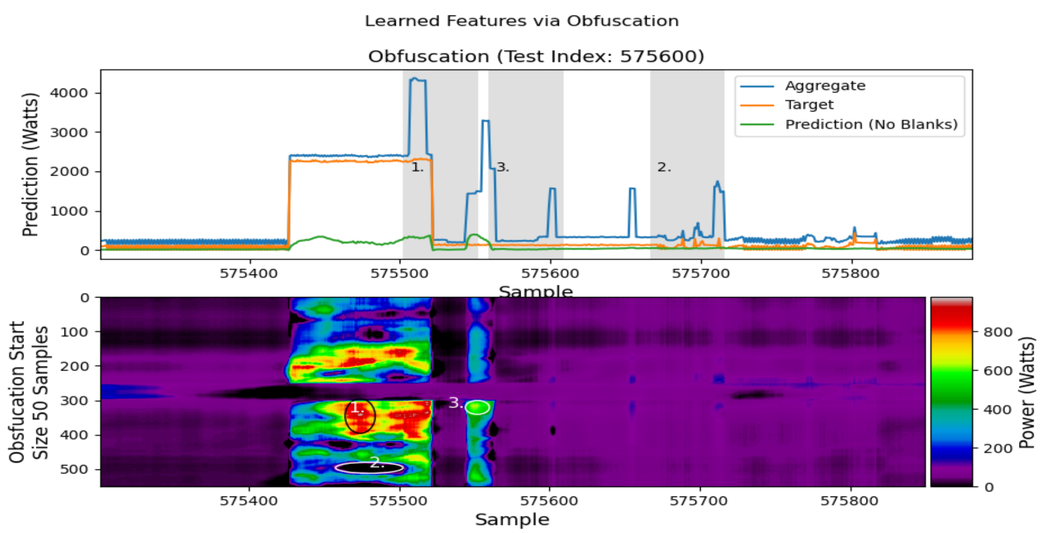


Fig. 6.8. The heat map generated for the Washing Machine in the ECO dataset, house 1. The model is trained using the entire REFIT dataset. The obtained measures for this sample are: MAE:383.18, SAE:0.83, NDE:0.79. The occlusion area numbers (top) and ellipse numbers (bottom) are explained more in the text.

Chapter 7

Conclusion

Summary

NILM is a field which is extremely challenging, no two houses will be the same, nor will the usage of two identical appliances by different users.

In chapter 1, the REFIT dataset is described. The REFIT dataset was very well received and the associated Nature Scientific Data journal paper has over 200 citations [107]. It combines a number of the most desirable features in a NILM dataset, namely duration, sampling frequency and scale. It provides the framework on which all the work in this thesis is based on.

In chapter 2, making use of the REFIT dataset a focus on individual appliances was considered, at the time there was little longitudinal work on appliances. The 2 year study period of the REFIT dataset enabled this, and allowed for seasonal trends to be explored as well as changes in attitude and appliances. This enabled the in depth analysis of the real world benefits of changing to a more energy efficient appliance using real world data rather than lab based efficiency claims.

In chapter 3, expanding on the previous work, and working with industry, the modelling of appliance consumption for use in Life Cycle Analysis reports proved to be an extremely beneficial to Nestec S.A.. The industry standards were based on old lab data and making use of the REFIT dataset, and some targeted appliance monitoring, updated models were created. These new models provided better results than the previous state of the art while requiring significantly less variables and which take into account real appliance usage.

Finally in chapter 4, NILM is introduced. Making use of neural networks a number of methods were proposed making use of CNN and GRU layers, these were demonstrated to show transferability to similar datasets, and were compared against the latest papers at the time. Secondly a newer network is presented, this network has a sequence to point architecture. This work provided a novel way of explaining the output of a NILM network in a visual way, by displaying the network input (time series) and extracting the features of importance within the input using occluded windows and mapping them into a heatmap. The proposed method allows the user to identify at what point the network begins to correctly identify the appliance within the sliding window. This is extremely useful for understanding window size choice and curation of training sets, with NILM larger training sets are typically not useful as the network either specialises or reverts to extreme generalisation (0 watts for nearly every appliance apart from cold appliances).

Future Work

Future NILM work would involve more rigorous benchmarking. Currently nearly every paper suffers from a number of issues (also guilty): Metrics are usually not suitable for long term data, for example over a year predicting a kettle at 0 watts would give a great result for a number of metrics as the kettle is going to be off for 98% of the time. Data selection is usually skewed to provide better results, either few activations which are correctly identified or low noise areas of data which would be considered easier. Multiple appliance output, every model is looked at individually while this makes sense academically, NILM is a highly commerical problem and misidentifying one appliance as another is a major issue, requiring explainability or being able to decide which model is correct.

A new direction would be a generalised network which was capable of outputting labelling for any number of appliances, similar to larger image based networks capable of multi-labelling or more recently the massive improvements in large language models such as ChatGPT. This would then be expanded to include regression values. Finally NILM could be combined with time of use and a more in depth user understanding to help plan appliance usage to best take advantage of either stored power from solar batteries or from time of use based electricity tariffs, this is already prevalent in South Africa where load shedding (sections of the country endure rolling blackouts due to power demand above generation capacity) is already common and is likely to become more prevalent as global warming, ageing infrastructure and fossil fuel scarcity becomes more widespread.

Bibliography

[1] International Energy Agency. World energy outlook. Technical report, International Energy Agency, 2006.

[2] International Energy Agency. Heat pumps. Webpage, 2021. `https://www.iea.org/reports/heat-pumps`.

[3] Hana Altrabalsi, Vladimir Stankovic, Jing Liao, and Lina Stankovic. Low-complexity energy disaggregation using appliance load modelling. *AIMS Energy*, 4:1–21, 2016.

[4] Kyle Anderson, Adrian Filip Ocneanu, Diego Benitez, Derrick Carlson, Anthony Rowe, Mario Berges, and Mario Bergés. Blued: A fully labeled public dataset for event-based non-intrusive load monitoring research. In *Proceedings of the 2nd KDD Workshop on Data Mining Applications in Sustainability*, pages 1 – 5, 2012.

[5] Alejandro Barredo Arrieta, Natalia Díaz-Rodríguez, Javier Del Ser, Adrien Bennetot, Siham Tabik, Alberto Barbado, Salvador Garcia, Sergio Gil-Lopez, Daniel Molina, Richard Benjamins, Raja Chatila, and Francisco Herrera. Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai. *Information Fusion*, 58:82–115, 6 2020.

[6] Sean Barker, Aditya Mishra, David Irwin, Emmanuel Cecchet, Prashant Shenoy, and Jeannie Albrecht. Smart*: An open data set and tools for enabling research in sustainable homes. In *SustKDD'12: Workshop on Data Mining Applications in Sustainability*, 2012.

[7] Karim Said Barsim and Bin Yang. On the feasibility of generic deep disaggregation for single-load extraction. *arXiv preprint*, 2 2018.

[8] Nipun Batra, Manoj Gulati, Amarjeet Singh, and Mani B. Srivastava. It's different: Insights into home energy consumption in india. In *Proceedings of the 5th ACM Workshop on Embedded Systems For Energy-Efficient Buildings*, BuildSys'13, pages 1–8, New York, NY, USA, 2013. Association for Computing Machinery.

[9] Nipun Batra, Jack Kelly, Oliver Parson, Haimonti Dutta, William Knottenbelt, Alex Rogers, Amarjeet Singh, and Mani Srivastava. Nilmtk: An open source toolkit for non-instusive load monitoring. In *Proceedings of the 5th international conference on Future energy systems - e-Energy '14*, pages 265–276. ACM Press, 2014.

[10] Christian Beckel, Wilhelm Kleiminger, Romano Cicchetti, Thorsten Staake, and Silvia Santini. The eco data set and the performance of non-intrusive load monitoring algorithms. In *Proceedings of the 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings*, BuildSys '14, page 80–89, New York, NY, USA, 2014. Association for Computing Machinery.

[11] Peder Bengtsson, Jonas Berghel, and Roger Renström. Performance study of a closed-type heat pump tumble dryer using a simulation model and an experimental set-up. *Drying Technology*, 32(8):891–901, 2014.

[12] Petra Berkholz, Rainer Stamminger, Gabi Wnuk, Jeremy Owens, and Simone Bernarde. Manual dishwashing habits: an empirical analysis of uk consumers. *International Journal of Consumer Studies*, 34:235–242, 3 2010.

[13] Paolo Bertoldi, Javier López Lorente, and Nicola Labanca. Energy consumption and energy efficiency trends in the eu-28 2000-2014. Technical report, Joint Research Centre, 2016.

[14] James C. Bezdek. *Pattern Recognition with Fuzzy Objective Function Algorithms.* Kluwer Academic Publishers, 1981.

[15] Nadezda Buneeva and Andreas Reinhardt. Ambal: Realistic load signature generation for load disaggregation performance evaluation. In *2017 IEEE International Conference on Smart Grid Communications (SmartGridComm)*, pages 443–448, 2017.

[16] Fabio Burlon. Energy efficiency of combined ovens. *Energy Procedia*, 82:986–993, 12 2015.

[17] Luis Alberto Calderón, Mónica Herrero, Adriana Laca, and Mario Díaz. Environmental impact of a traditional cooked dish at four different manufacturing scales: from ready meal industry and catering company to traditional restaurant and homemade. *The International Journal of Life Cycle Assessment*, 23:811–823, 4 2018.

[18] Luis Alberto Calderón, Loreto Iglesias, Adriana Laca, Mónica Herrero, and Mario Díaz. The utility of life cycle assessment in the ready meal food industry. *Resources, Conservation and Recycling*, 54:1196–1207, 10 2010.

[19] Diogo V. Carvalho, Eduardo M. Pereira, and Jaime S. Cardoso. Machine learning interpretability: A survey on methods and metrics. *Electronics*, 8:832, 7 2019.

[20] Dong Chen, David Irwin, and Prashant Shenoy. Smartsim: A device-accurate smart home simulator for energy analytics. In *2016 IEEE International Conference on Smart Grid Communications (SmartGridComm)*, pages 686–692, 2016.

[21] Kyunghyun Cho, Bart van Merrienboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using rnn encoder-decoder for statistical machine translation. *arXiv preprint*, 6 2014.

[22] Andrea Cominola, Matteo Giuliani, Dario Piga, Andrea Castelletti, and Andrea Emilio Rizzoli. A hybrid signature-based iterative disaggregation algorithm for non-intrusive load monitoring. *Applied Energy*, 185:331–344, 2017.

[23] Gianpiero Costantino and Fabio Martinelli. Privacy-preserving energy-reading for smart meter. In *Inclusive Smart Cities and e-Health*, pages 165–177, Cham, 2015. Springer International Publishing.

[24] Nello Cristianini and John Shawe-Taylor. *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. Cambridge University Press, 3 2000.

[25] Amanda D Cuellar and Michael E Webber. An updated estimate for energy use in u.s. food production and policy implications. In *Proceedings of the ASME 2010 4th International Conference on Energy Sustainability*, pages 1–10. ASME, 2010.

[26] CurrentCost. Currentcost transmitter. Webpage, 2016. `http://www.currentcost.com/product-transmitter.html`.

[27] CurrentCost. Envir. Webpage, 2016. `http://www.currentcost.com/product-envir.html`.

[28] CurrentCost. Luup requests. Webpage, 2016. `http://wiki.micasaverde.com/index.php/Luup_Requests`.

[29] CurrentCost. Ui notes. Webpage, 2016. `http://wiki.micasaverde.com/index.php/UI_Notes`.

[30] Tribeni Das, R. Subramanian, A. Chakkaravarthi, Vasudeva Singh, S.Z. Ali, and P.K. Bordoloi. Energy conservation in domestic rice cooking. *Journal of Food Engineering*, 75:156–166, 7 2006.

[31] Dilip K. De, Nathaniel Muwa Shawhatsu, Narendra Nath De, and Mathias Ikechukwu Ajaeroh. Energy-efficient cooking methods. *Energy Efficiency*, 6:163–175, 2 2013.

[32] Elizabeth Ann Demerchant. *User's Influence on Energy Consumption with Cooking Systems Using Electricity*. PhD thesis, Virginia Polytechnic Institute and State University, 1997.

[33] Energy & Industry Strategy Department of Buisiness. Smart metering implementation programme policy conclusion following energy suppliers' trails of alternatives to in-home displays. Technical report, Department of Buisiness, Energy & Industry Strategy, 2019.

[34] Michele D'Incecco, Stefano Squartini, and Mingjun Zhong. Transfer learning for non-intrusive load monitoring. *IEEE Transactions on Smart Grid*, 11:1419–1429, 3 2020.

[35] Pedro Paulo Marques do Nascimento. *Applications of Deep Learning Techniques on NILM.* PhD thesis, COPPE - Universidade Federal do Rio de Janeiro, 2016.

[36] Bing Dong, Cheng Cao, and Siew Eang Lee. Applying support vector machines to predict building energy consumption in tropical region. *Energy and Buildings*, 37:545–553, 5 2005.

[37] Finale Doshi-Velez and Been Kim. Towards a rigorous science of interpretable machine learning. *arXiv*, 2 2017.

[38] Pei Du, Ju'e Guo, Shaolong Sun, Shouyang Wang, and Jing Wu. A novel two-stage seasonal grey model for residential electricity consumption forecasting. *Energy*, 258:124664, 2022.

[39] Chris E Dutilh and Klaas J Kramer. Energy consumption in the food chain. *AMBIO: A Journal of the Human Environment*, 2000.

[40] Rudresh Dwivedi, Devam Dave, Het Naik, Smiti Singhal, Rana Omer, Pankesh Patel, Bin Qian, Zhenyu Wen, Tejal Shah, Graham Morgan, and Rajiv Ranjan. Explainable ai (xai): Core ideas, techniques, and solutions. *ACM Comput. Surv.*, 55(9), jan 2023.

[41] G. Edwards-Jones, K. Plassmann, E.H. York, B. Hounsome, D.L. Jones, and L. Milà i Canals. Vulnerability of exporting nations to the development of a carbon label in the united kingdom. *Environmental Science & Policy*, 12(4):479–490, 2009. Special Issue: Food Security and Environmental Change.

[42] Dominik Egarter, Anita Sobe, and Wilfried Elmenreich. Evolving non-intrusive load monitoring. *Lecture Notes in Computer Science*, pages 182–191, 2013.

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

[44] Engadget. Mi casa verde vera review: Home automation, simplified. Webpage, 2010. `https://www.engadget.com/2010-12-17-mi-casa-verde-vera-review-home-automation-simplified.html`.

[45] Jingkun Gao, Suman Giri, Emre Can Kara, and Mario Bergés. Plaid: a public dataset of high-resoultion electrical appliance measurements for load identification research. In *Proceedings of the 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings - BuildSys '14*, pages 198–199. ACM Press, 2014.

[46] Felan Carlo C. Garcia, Christine May C. Creayla, and Erees Queen B. Macabebe. Development of an intelligent system for smart home energy disaggregation using stacked denoising autoencoders. *Procedia Computer Science*, 105:248–255, 2017.

[47] Jasmin Geppert. *Modelling of domestic refrigerators' energy consumption under real life conditions in Europe*. PhD thesis, University of Bonn, 2011.

[48] Leilani H. Gilpin, David Bau, Ben Z. Yuan, Ayesha Bajwa, Michael Specter, and Lalana Kagal. Explaining explanations: An overview of interpretability of machine learning. *arXiv*, 2018.

[49] Christophe Gisler, Antonio Ridi, Damien Zufferey, Omar Abou Khaled, and Jean Hennebert. Appliance consumption signature database and recognition test protocols. In *2013 8th International Workshop on Systems, Signal Processing and their Applications (WoSSPA)*, pages 336–341. IEEE, 5 2013.

[50] Mario E. Bergés Gonzalez. *A framework for enabling energy-aware facilities through minimally-intrusive approaches*. PhD thesis, Carnegie Mellon University, 2010.

[51] Hugo Gonçalves, Adrian Ocneanu, and Mario Bergés. Unsupervised disaggregation of appliances using aggregated consumption data. In *1st KDD Workshop on Data Mining Applications in Sustainability (SustKDD)*, 2011.

[52] Stephen Haben, Jonathan Ward, Danica Vukadinovic Greetham, Colin Singleton, and Peter Grindrod. A new error measure for forecasts of household-level, high resolution electrical energy consumption. *International Journal of Forecasting*, 30:246–256, 4 2014.

[53] Tiffany J. Hager and Ruben Morawicki. Energy consumption during cooking in the residential sector of developed nations: A review. *Food Policy*, 40:54–63, 6 2013.

[54] George William Hart. Nonintrusive appliance load monitoring. *Proceedings of the IEEE*, 80:1870–1891, 1992.

[55] Kanghang He, Lina Stankovic, Jing Liao, and Vladimir Stankovic. Non-intrusive load disaggregation using graph signal processing. *IEEE Transactions on Smart Grid*, 9:1739–1747, 5 2018.

[56] Wan He and Ying Chai. An empirical study on energy disaggregation via deep learning. In *Proceedings of the 2016 2nd International Conference on Artificial Intelligence and Industrial Engineering (AIIE 2016)*, pages 338–342. Atlantis Press, 2016/11.

[57] Georges Hebrail and Alice Barard. Individual household electric power consumption data set (ihepcds). UCI Machine Learning Repository, 2006. `https://archive.ics.uci.edu/ml/datasets/Individual+household+electric+power+consumption`.

[58] Martin C Heller and Gregory A Keoleian. Life cycle-based sustainability indicators for assessment of the u.s. food system. Technical report, Cen-

ter for Sustainable Systems, School of Natural Resources and Environment, University of Michigan Ann, 2000.

[59] Simon Henriet, Umut Şimşekli, Benoit Fuentes, and Gaël Richard. A generative model for non-intrusive load monitoring in commercial buildings. *Energy and Buildings*, 177:268–278, 10 2018.

[60] Mike Hewitt and Stuart Coleman. Unlocking smart meter data to accelerate the energy revolution, July 2021. https://www.smartdcc.co.uk/news-events/unlocking-smart-meter-data-to-accelerate-the-energy-revolution/.

[61] Chris Holmes. Non-intrusive load monitoring (nilms) research activity: End-use energy efficiency & demand response. Presentation, 03 2014.

[62] Rahmat-Allah Hooshmand, Habib Amooshahi, and Moein Parastegari. A hybrid intelligent algorithm based short-term load forecasting approach. *International Journal of Electrical Power & Energy Systems*, 45(1):313–324, 2013.

[63] Patrick Huber, Alberto Calatroni, Andreas Rumsch, and Andrew Paice. Review on deep neural networks applied to low-frequency nilm. *Energies*, 14:2390, 4 2021.

[64] Jack Hulme, Adele Beaumont, and Claire Summers. Energy follow-up survey 2011. Technical report, Department of Energy and Climate Change, 2013.

[65] Bontor Humala, Akshay S.N. Uttama Nambi, and Venkatesha R. Prasad. Universalnilm: A semi-supervised energy disaggregation framework using general appliance models. In *Proceedings of the Ninth International Conference on Future Energy Systems - e-Energy '18*, pages 223–229. ACM Press, 2018.

[66] Tim D Hunt, Dileep Rajendran, Mark Nikora, Susan Bennett, and Andy Fendall. A minimally intrusive monitoring system that utilizes electricity consumption as a proxy for wellbeing. *Journal of Applied Computing and Information Technology*, 18, 2014.

[67] Hafiz Khurram Iqbal, Farhan Hassan Malik, Aoun Muhammad, Muhammad Ali Qureshi, Muhammad Nawaz Abbasi, and Abdul Rehman Chishti. A critical review of state-of-the-art non-intrusive load monitoring datasets. *Electric Power Systems Research*, 192:106921, 3 2021.

[68] Fraunhofer ISI and VITO. Ecodesign and energy labelling preparatory study on electric kettles. EU Project, 2020. https://energy.ec.europa.eu/studies/preparatory-studies/ecodesign-and-energy-labelling-preparatory-study-electric-kettles_en.

[69] J.-S.R. Jyh-Shing Roger Jang. Anfis: adaptive-network-based fuzzy inference system. *IEEE Transactions on Systems, Man and Cybernetics*, 23:665–685, 1993.

[70] Niels Jungbluth. Life-cycle-assessment for stoves and ovens. Technical report, Umweltnatur & Umweltsozialwissenschaften, 1997.

[71] Wang Junping, Chen Quanshi, and Cao Binggang. Support vector machine based battery model for electric vehicles. *Energy Conversion and Management*, 47:858–864, 5 2006.

[72] T. Kane, S.K. Firth, T.M. Hassan, and V. Dimitriou. Heating behaviour in english homes: An assessment of indirect calculation methods. *Energy and Buildings*, 148:89–105, 2017.

[73] Tom Kane, Val Mitchell, Jing Liao, Lina Stankovic, Steven Firth, Stuart Cockbill, Charlie Wilson, David Murray, Farid Fouchal, Andrew May, Vanda Dimitriou, Vladimir Stankovic, and Tarek Hassan. Supporting retrofit decisions using smart meter data: A multi-disciplinary approach. In *ECEEE 2015 Summer Study on Energy Efficiency*, pages 1–14, 2015.

[74] Aqeel H. Kazmi, Michael J. O'Grady, and Gregory M. P. O'Hare. *Non-intrusive Identification of Electrical Appliances*, pages 190–195. Springer International Publishing, Cham, 2013.

[75] V. Kecman, T.-M. Huang, and M. Vogt. *Iterative Single Data Algorithm for Training Kernel Machines from Huge Data Sets: Theory and Performance*, pages 255–274. Springer, Berlin, Heidelberg, 4 2005.

[76] Jack Kelly. Rfm edf ecomanager wiki. Webpage, 2013. `https://github.com/JackKelly/rfm_edf_ecomanager/wiki`.

[77] Jack Kelly and William Knottenbelt. Neural nilm: Deep neural networks applied to energy disaggregation. In *Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments*, BuildSys '15, pages 55–64, New York, NY, USA, 2015. Association for Computing Machinery.

[78] Jack Kelly and William Knottenbelt. The uk-dale dataset, domestic appliance-level electricity demand and whole-house demand from five uk homes. *Scientific Data*, 2:150007, 3 2015.

[79] R Kemna and R. Stamminger. Energy consumption of domestic appliances in european households - results from stock model calculations. In *Proceedings of the 5th International Conference EEDAL'09 16-18 June, Berlin, Germany. volume 3*. E3P, 2004.

[80] Jihyun Kim, Thi-Thu-Huong Le, and Howon Kim. Nonintrusive load monitoring based on advanced deep learning and novel signature. *Computational Intelligence and Neuroscience*, 2017:1–22, 2017.

[81] J.G. Kirkerud, N.O. Nagel, and T.F. Bolkesjø. The role of demand response in the future renewable northern european energy system. *Energy*, 235:121336, 2021.

[82] Christoph Klemenjak, Christoph Kovatsch, Manuel Herold, and Wilfried Elmenreich. A synthetic energy dataset for non-intrusive load monitoring in households. *Scientific Data*, 7:108, 12 2020.

[83] Christoph Klemenjak, Stephen Makonin, and Wilfried Elmenreich. Towards comparability in non-intrusive load monitoring: On data and performance

evaluation. In *2020 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, pages 1–5. IEEE, 2 2020.

[84] J. Zico Kolter and Matthew J. Johnson. Redd: A public data set for energy disaggregation research. In *SustKDD workshop on Data Mining Applications in Sustainability*, pages 1–6, 2011.

[85] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. In *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1*, NIPS'12, page 1097–1105, Red Hook, NY, USA, 2012. Curran Associates Inc.

[86] S. Lakshmi, A. Chakkaravarthi, R. Subramanian, and Vasudeva Singh. Energy consumption in microwave cooking of rice and its comparison with other domestic appliances. *Journal of Food Engineering*, 78:715–722, 1 2007.

[87] Jiangfeng Li, Dandan Li, Xin Zeng, Vladimir Stankovic, Lina Stankovic, and Qingjiang Shi. Energy disaggregation data set in a non-residential room in shanghai, china (sc-ednrr). *Dataset*, 2022.

[88] Qiong Li, Qinglin Meng, Jiejin Cai, Hiroshi Yoshino, and Akashi Mochida. Predicting hourly cooling load in the building: A comparison of support vector machine and different artificial neural networks. *Energy Conversion and Management*, 50:90–96, 1 2009.

[89] Jing Liao, Georgia Elafoudi, Lina Stankovic, and Vladimir Stankovic. Non-intrusive appliance load monitoring using low-resolution smart meter data. In *2014 IEEE International Conference on Smart Grid Communications (SmartGridComm)*, pages 535–540. IEEE, 11 2014.

[90] Zachary C. Lipton. The mythos of model interpretability. *Queue*, 16:31–57, 6 2018.

[91] Stephen Makonin, Bradley Ellert, Ivan V. Bajić, and Fred Popowich. Electricity, water, and natural gas consumption of a residential house in canada from 2012 to 2014. *Scientific Data*, 3:160037, 6 2016.

[92] Stephen Makonin, Fred Popowich, Lyn Bartram, Bob Gill, and Ivan V. Bajic. Ampds: A public dataset for load disaggregation and eco-feedback research. In *2013 IEEE Electrical Power & Energy Conference*, pages 1–6. IEEE, 8 2013.

[93] Franco Marchesoni-Acland, Camilo Mariño, Elías Masquil, Pablo Masaferro, and Alicia Fernández. End-to-end nilm system using high frequency data and neural networks. arXiv, 2020.

[94] Lukas Mauch and Bin Yang. A new approach for supervised power disaggregation by using a deep recurrent lstm network. In *2015 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pages 63–67. IEEE, 12 2015.

[95] Lukas Mauch and Bin Yang. A novel dnn-hmm-based approach for extracting single loads from aggregate power signals. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2384–2388. IEEE, 3 2016.

[96] E. Mayhorm, R. Butner, M. Baechler, G. Sullivan, and H. Hao. Characteristics and performance of existing load disaggregation technologies. In *(Technical Report) Pacific Northwest National Laboratory*, 2015.

[97] E. T. Mayhorn, G. P. Sullivan, T. Fu, and J. M. Petersen. Non-intrusive load monitoring laboratory-based test protocols. In *(Technical Report) Pacific Northwest National Laboratory*, 2017.

[98] Ebony T Mayhorn, Joseph Petersen, Ryan S Butner, and Erica M Johnson. Load disaggregation technologies: Real world and laboratory performance. In *2016 ACEEE Summer Study on Energy Efficiency in Buildings*, 2016.

[99] A. Milani, C. Camarda, and L. Savoldi. A simplified model for the electrical energy consumption of washing machines. *Journal of Building Engineering*, 2:69–76, 6 2015.

[100] Mintel. The consumer. Technical report, Mintel, 2004.

[101] Vittorio Miori and Dario Russo. Domotic evolution towards the iot. In *2014 28th International Conference on Advanced Information Networking and Applications Workshops*, pages 809–814, 2014.

[102] Sina Mohseni, Niloofar Zarei, and Eric D. Ragan. A multidisciplinary survey and framework for design and evaluation of explainable ai systems. *ACM Trans. Interact. Intell. Syst.*, 11(3-4), sep 2021.

[103] Andrea Monacchi, Dominik Egarter, Wilfried Elmenreich, Salvatore D'Alessandro, and Andrea M. Tonello. Greend: An energy consumption dataset of households in italy and austria. In *2014 IEEE International Conference on Smart Grid Communications (SmartGridComm)*, pages 511–516. IEEE, 11 2014.

[104] D. M. Murray, Jing Liao, L. Stankovic, and Vladimir Stankovic. Understanding usage patterns of electric kettle and energy saving potential. *Applied Energy*, 171:231–242, 2016.

[105] David Murray, Jing Liao, Lina Stankovic, and Vladimir Stankovic. How to make efficient use of kettles : Understanding usage patterns. In *8th Int. Conf. Energy Efficiency in Domestic Appliances and Lighting, EEDAL*, 2015.

[106] David Murray, Jing Liao, Lina Stankovic, Vladimir Stankovic, Charlie Wilson, Michael Coleman, and Tom Kane. A data management platform for personalised real-time energy feedback. In *8th Int. Conf. Energy Efficiency in Domestic Appliances and Lighting, EEDAL*, pages 1–15, 2015.

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

[108] David Murray, Lina Stankovic, and Vladimir Stankovic. Explainable nilm networks. In *NILM 2020 - Proceedings of the 5th International Workshop on Non-Intrusive Load Monitoring*, 2020.

[109] David Murray, Lina Stankovic, Vladimir Stankovic, Srdjan Lulic, and Srdjan Sladojevic. Transferability of neural network approaches for low-rate energy

disaggregation. In *International Conference on Acoustics, Speech, and Signal Processing*, 2019.

[110] D.M. Murray, L. Stankovic, V. Stankovic, and N.D. Espinoza-Orias. Appliance electrical consumption modelling at scale using smart meter data. *Journal of Cleaner Production*, 187, 2018.

[111] R L Myers. *The Basics of Physics*. Greenwood Press, 2006.

[112] Vinod Nair and Geoffrey E Hinton. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th International Conference on International Conference on Machine Learning*, pages 807–814. Omnipress, 2010.

[113] Claudia Oberascher, Rainer Stamminger, and Christiane Pakula. Energy efficiency in daily food preparation. *International Journal of Consumer Studies*, 35:201–211, 3 2011.

[114] Department of Energy and Climate Change. Smart metering implementation programme smart metering equipment technical specifications version 2. Technical report, Department of Energy and Climate Change, 2013.

[115] Yusuf Ozturk, Datchanamoorthy Senthilkumar, Sunil Kumar, and Gordon Lee. An intelligent home energy management system to improve demand response. *IEEE Transactions on Smart Grid*, 4:694–701, 6 2013.

[116] Davide Papasidero, Sauro Pierucci, and Flavio Manenti. Energy optimization of bread baking process undergoing quality constraints. *Energy*, 116:1417–1422, 12 2016.

[117] Oliver Parson, Siddhartha Ghosh, Mark Weal, and Alex Rogers. An unsupervised training method for non-intrusive appliance load monitoring. *Artificial Intelligence*, 217:1–19, 12 2014.

[118] Lucas Pereira and Nuno Nunes. Performance evaluation in non-intrusive load monitoring: Datasets, metrics, and tools—a review. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8, 11 2018.

[119] James Pierce, Diane J. Schiano, and Eric Paulos. Home , habits , and energy : Examining domestic interactions and energy consumption. *Proceedings of the 28th international conference on Human factors in computing systems - CHI '10*, pages 1985–1994, 2010.

[120] Manisa Pipattanasomporn, Murat Kuzlu, Saifur Rahman, and Yonael Teklu. Load profiles of selected major household appliances and their demand response opportunities. *IEEE Transactions on Smart Grid*, 5:742–750, 3 2014.

[121] Defra Market Transformation Program and Briefing Note. Bnck06 : Kettle trends. Technical report, DEFRA, 2006.

[122] Yatindra Kumar Ramgolam, Shanmuga Veerapen, Anshu Prakash Murdan, Vishwamitra Oree, and Krishnaraj Madhavjee Sunjiv Soyjaudah. Apparent and standby energy consumption in mauritian middle-income households. In *2014 IEEE International Energy Conference (ENERGYCON)*, pages 1121–1128. IEEE, 5 2014.

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

[124] Ian Richardson and Murray Thomson. Integrated simulation of photovoltaic micro-generation and domestic electricity demand: a one-minute resolution open-source model. *Proceedings of the Institution of Mechanical Engineers, Part A: Journal of Power and Energy*, 227(1):73–81, 2013.

[125] Antonio Ridi, Christophe Gisler, and Jean Hennebert. A survey on intrusive load monitoring for appliance recognition. In *2014 22nd International Conference on Pattern Recognition*, pages 3702–3707. IEEE, 8 2014.

[126] Antonio G Ruzzelli, C. Nicolas, Anothony Schoofs, and Gregory M P O'Hare. Real-time recognition and profiling of appliances through a single electricity sensor. In *2010 7th Annual IEEE Communications Society Conference on*

Sensor, Mesh and Ad Hoc Communications and Networks (SECON), pages 1–9. IEEE, 6 2010.

[127] Wojciech Samek, Thomas Wiegand, and Klaus-Robert Müller. Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models. *ITU Journal: ICT Discoveries - Special Issue 1 - The Impact of Artificial Intelligence (AI) on Communication Networks and Services*, 1:1–10, 10 2017.

[128] Udo Schlegel, Hiba Arnout, Mennatallah El-Assady, Daniela Oelke, and Daniel A. Keim. Towards a rigorous evaluation of xai methods on time series. In *Proceedings - 2019 International Conference on Computer Vision Workshop, ICCVW 2019*, 2019.

[129] Ximena C. Schmidt Rivera, Namy Espinoza Orias, and Adisa Azapagic. Life cycle environmental impacts of convenience food: Comparison of ready and home-made meals. *Journal of Cleaner Production*, 73:294–309, 2014. Towards eco-efficient agriculture and food systems: Selected papers from the Life Cycle Assessment (LCA) Food Conference, 2012, in Saint Malo, France.

[130] Ulf Sonesson, Hans Janestad, and Birgitta Raaholt. Energy for preparation and storing of food - models for calculation of energy use for cooking and cold storage in households. Technical report, Foundation for Strategic Environmental Research, 2003.

[131] Ameena Sorour, Meghdad Fazeli, Mohammad Monfared, Ashraf A Fahmy, Justin R. Searle, and Richard P. Lewis. Forecast-based energy management for domestic pv-battery systems: A U.K. case study. *IEEE Access*, 9:58953–58965, 2021.

[132] L. Stankovic, V. Stankovic, J. Liao, and C. Wilson. Measuring the energy intensity of domestic activities from smart meter data. *Applied Energy*, 183:1565–1580, 12 2016.

[133] Bruce Stephen, Stuart Galloway, and Graeme Burt. Self-learning load characteristic models for smart appliances. *IEEE Transactions on Smart Grid*, 5:2432–2439, 9 2014.

[134] Jiawei Su, Danilo Vasconcellos Vargas, and Kouichi Sakurai. One pixel attack for fooling deep neural networks. *IEEE Transactions on Evolutionary Computation*, 23:828–841, 10 2019.

[135] Michio Sugeno. *Industrial Applications of Fuzzy Control*. Elsevier Science Inc., 1985.

[136] Guoming Tang, Kui Wu, Jingsheng Lei, and Jiuyang Tang. Plug and play! a simple, universal model for energy disaggregation, 2014. arXiv:1404.1884.

[137] Guoming Tang, Kui Wu, Jingsheng Lei, and Jiuyang Tang. A simple model-driven approach to energy disaggregation. In *2014 IEEE International Conference on Smart Grid Communications (SmartGridComm)*, pages 566–571. IEEE, 11 2014.

[138] Christoph Thim. *Technology Guide*, chapter Domestic appliances, pages 458–461. Springer Berlin Heidelberg, 2009.

[139] TP-Link. Kasa smart home app. Webpage, 2023. `https://www.tp-link.com/uk/smarthome/`.

[140] Energy Saving Trust, Energy Saving Trust, and Energy Saving Trust. At home with water. Technical report, Energy Saving Trust, 2013. http://www.energysavingtrust.org.uk/reports/home-water.

[141] Akshay S.N. Uttama Nambi, Antonio Reyes Lua, and Venkatesha R. Prasad. Loced: Location-aware energy disaggregation framework. In *Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments*, BuildSys '15, pages 45–54, New York, NY, USA, 2015. Association for Computing Machinery.

[142] Vattenfall. Energie sparen macht spass und spart geld. Technical report, Vattenfall, 2008.

[143] Apostolos Vavouris, Lina Stankovic, and Vladimir Stankovic. Smart meter electricity of a household in germany with electric vehicle charging annotation. Dataset, 2022. `https://pureportal.strath.ac.uk/en/datasets/smart-meter-electricity-of-a-household-in-germany-with-electric-v`.

[144] Vera. Mios control panel. Webpage, 2016. `https://cp.mios.com/login.php`.

[145] Vera. Vera3. Webpage, 2016. `http://getvera.com/controllers/vera3/`.

[146] G. Wood and M. Newborough. Dynamic energy-consumption indicators for domestic appliances: environment, behaviour and design. *Energy and Buildings*, 35(8):821–841, 2003.

[147] Li-Chih Ying and Mei-Chiu Pan. Using adaptive network based fuzzy inference system to forecast regional electricity loads. *Energy Conversion and Management*, 49(2):205–211, 2008.

[148] Michael Zeifman and Kurt Roth. Nonintrusive appliance load monitoring: Review and outlook. *IEEE Transactions on Consumer Electronics*, 57:76–84, 2 2011.

[149] Chaoyun Zhang, Mingjun Zhong, Zongzuo Wang, Nigel Goddard, and Charles Sutton. Sequence-to-point learning with neural networks for non-intrusive load monitoring. *The Thirty-Second AAAI Conference on Artificial Intelligence*, 12 2018.

[150] Guanchen Zhang, Gary Wang, Hassan Farhangi, and Ali Palizban. Residential electric load disaggregation for low-frequency utility applications. In *2015 IEEE Power & Energy Society General Meeting*, pages 1–5. IEEE, 7 2015.

[151] Yu Zhang, Guoming Tang, Qianyi Huang, Yi Wang, Xudong Wang, and Jiadong Lou. Fednilm: Applying federated learning to nilm applications at the edge. arXiv, 2021.

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

[153] Jean-Paul Zimmermann, Matt Evans, Tim Lineham, Jonathan Griggs, Griggs Surveys, Les Harding, Nicola King, and Penelope Roberts. Household electricity survey: A study of domestic electrical product usage. *Intertek*, page 600, 2012.

[154] Ahmed Zoha, Alexander Gluhak, Muhammad Imran, and Sutharshan Rajasegarar. Non-intrusive load monitoring approaches for disaggregated energy sensing: A survey. *Sensors*, 12:16838–16866, 12 2012.

[155] J. Zufia and L. Arana. Life cycle assessment to eco-design food products: industrial cooked dish case study. *Journal of Cleaner Production*, 16(17):1915–1921, 2008.