VENICE

WP2: Modelling Approach & Findings

July 2022

Issue: 1

Prepared for: Western Power Distribution

SYSTEMS ● ENGINEERING ● TECHNOLOGY

VENICE

WP2: Modelling Approach & Findings

Client: Western Power Distribution

Client Ref.:

Date: July 2022

Classification: FRAZER-NASH IN CONFIDENCE

Project No.:	010837	**Compiled By:**	Zoe Hodgins, Katie Chippendale and Greg Nelson
Document No.:	52726R	**Verified By:**	Megan Smith, Neil Headings, Rob Wadsworth and Mark Ashworth
Issue No.:	1	**Approved By:**	Mark Joyce
		Signed:	

Distribution

Copy	Recipient	Organisation
1	Stuart Fowler	Western Power Distribution
2	File	Frazer-Nash Consultancy Limited
3		
4		
5		

COPYRIGHT

The Copyright in this work is vested in Frazer-Nash Consultancy Limited. Reproduction in whole or in part or use for tendering or manufacturing purposes is prohibited except under an agreement with or with the written consent of Frazer-Nash Consultancy Limited and then only on the condition that this notice is included in any such reproduction.

DISCLAIMER

This document (including any annexes, appendices and associated computer files) is issued in confidence solely for the purpose for which it is supplied. Frazer-Nash Consultancy Limited accepts no liability for loss or damage suffered by any party resulting from use of this document, other than as provided for by the contract under which this document has been produced.

Originating Office: FRAZER-NASH CONSULTANCY LIMITED
Stonebridge House, Dorking Business Park, Dorking, Surrey, RH4 1HJ
Tel: +44 (0)1306 885050

Executive Summary

Distribution Network Operators (DNOs) offer support to vulnerable households in the form of ongoing assistance and advise, and the provision of priority services during an interruption to supply. As DNOs do not own the relationship with customers, it is difficult to identify these households and provide them with this assistance. Frazer-Nash were tasked with determining whether it was possible to identify a vulnerable household from its smart meter data.

A household is considered vulnerable for a variety of reasons and each of these can represent themselves differently in smart meter data. Work completed prior to this report identified a selection of behavioural characteristics that may be present in a household's smart meter data if they have a specific vulnerability. Three separate models were developed, with each one identifying a different type of behavioural characteristic in smart meter data that may be indicators of vulnerability. There are:

1. Appliance Disaggregation and Prediction: This model is used to identify vulnerabilities related to appliance usage.
2. Cohort Comparison: This model is used to identify vulnerabilities related to the overall level of usage a specific household has.
3. Overall Changes in Usage: This model is used to identify vulnerabilities related to a significant change in its occupants' behaviour.

Each model was developed and tested in isolation using different, open source, datasets. This document details the methods undertaken to complete each of the three models in the Modelling Approach section of this report. Whilst the open source datasets were used to develop each model, none have been validated using real households with known vulnerabilities. This is due to the restrictions in place which limit smart meter data access to DNOs and the public. During the project, we tried to obtain smart meter data from vulnerable households through direct communication with consumers, through energy support companies and directly from suppliers; however, none of these proved fruitful. Due to this lack of validation and realistic training data, currently, none of these models are ready to be deployed or used by DNOs. The limitations to these models are detailed in the Conclusions section of this report.

The appliance disaggregation prediction model concluded that the simplistic test cases proved the model could determine how a household was using its appliances. This was done by determining the likelihood the power drawn in a 30-minute window was from each combination of appliances and multiplying this by the probability each appliance is used in that specific 30-minute window. This probability is iterated upon to become specific for each household and appliance. The key model limitations are driven by high uncertainty in the power drawn by an appliance, as expected. These conclusions show that the approach has shown potential, but further development should be undertaken to reduce the uncertainties. Once this is completed, the model should be tested with medical appliances.

The cohort comparison model was developed using a proxy for household characteristics and average usage data. The results showed that some of the variation in a household's average usage could be attributed to the characteristics. To develop the model further, more testing must be undertaken with real average household usage data for homes with known characteristics.

The overall changes in usage model concluded that step changes in usage could be identified and quantified. Here, data from the UK Power Network's Low Carbon London project (SmartMeter Energy Consumption Data in London Households, 2014), was used. This model was tested by artificially adding vulnerabilities into the smart meter data, and therefore if this model were to be developed further, real household data would have to be obtained. This is mainly because the magnitude and consistency of the expected changes in vulnerable customers is currently unknown.

To conclude, three models were developed to identify different behavioural characteristics that may be an indicator of vulnerability from a household's smart meter data. All three models showed that the behavioural characteristic was detectable using the data sources available. The next step would be to test each model using usage from known households, however, this has not been possible to date due to lack of access to smart meter data for normal and vulnerable households. If this information were obtained, and the model's capability assessed, they could be combined and deployed as a novel and efficient means to detect household vulnerability, to aid the DNOs discharging their responsibility.

Contents

1 Introduction

1.1 Industry Landscape

1.1.1 Ofgem Vulnerability Strategy

The Government's target for the United Kingdom (UK) to reach net zero by 2050 has accelerated the electrification of heat and transport in the UK. As a result, over the next few years, digitalisation, decarbonation and decentralisation are likely to radically change business models, creating new costs, benefits and capability challenges for consumers. When these changes happen, the energy networks have a responsibility to make sure the most vulnerable are adequately protected in this future market.

The energy regulator, Ofgem, defines a household as vulnerable if (Ofgem, October 2019):

> *A consumer's personal circumstances and characteristics combine with aspects of the market to create situations where they are significantly less able than a typical domestic consumer to protect or represent their interests, or they are significantly more likely than a typical domestic consume to suffer detriment or that detriment is likely to be more substantial.*

This broad definition covers a wide range of situations that a household could have, including financial, heating and insulation difficulties, as well as physical or mental impairment. Supporting and protecting consumers in vulnerable situations is a key priority for Ofgem. They have generated a Consumer Vulnerability Strategy (Ofgem, October 2019) with five areas of focus to drive strong improvements for consumers in vulnerable situations:

1. Improving identification of vulnerability and smart use of data,
2. Supporting those struggling with their bills,
3. Driving significant improvements in customer service for vulnerable groups,
4. Encouraging positive and inclusive innovations,
5. Working with partners to tackle issues that cut across multiple sectors.

These areas of focus have guided the scope of innovation projects in the energy networks, and as a result, Project VENICE is being undertaken by Western Power Distribution (WPD) to respond to focus areas 1 and 2.

1.1.2 Smart Meter Roll-out

Smart meters are an essential digital upgrade to our energy system. Smart meters work by measuring the electrical current flow and voltage at regular intervals and adding this up to calculate the power used in a half-hour period. The uptake of smart meters across the UK will provide the foundation for a more sophisticated, green, and consumer-friendly energy system.

The government has set targets for energy suppliers to install smart meters in every home in Great Britain by 2024 (Western Power Distribution, July 2021). By the end of 2021, 26.1 million smart meters had been installed in domestic properties across Great Britain, accounting for 50% of all domestic meters (BEIS, March 2022). The total roll-out is estimated to cost £10.9 billon.

The Smart Meter Data Communications Centre (Smart DCC) provide the communications infrastructure that handles smart meter data and is responsible for all access to smart meter data. The energy suppliers are responsible for offering and installing smart meters to all homes. Ofgem are responsible for regulating the Smart DCC and energy suppliers during the rollout and ongoing support for smart meter use in the home.

Missing from this collection of smart meter roll-out roles and responsibilities are the DNOs. DNOs are only licensed to access certain extracts of smart meter data via the Smart DCC, which must be anonymised or aggregated for data privacy. This means that currently DNOs cannot access the smart meter data for a specific household.

1.1.3 DNO Responsibility to Vulnerable Consumers

One way that Ofgem ensures vulnerable consumers are supported is by ensuring all Distribution Network Operators (DNOs) offer support to vulnerable households. This is in the form of ongoing assistance and advise, and the provision of priority services during an interruption to supply.

It is the responsibility of each DNO to keep an up-to date Priority Services Register (PSR) of the vulnerable households in their region. This includes the household's address, inhabitants, and the reason the household is considered vulnerable. There are many vulnerabilities that are considered by the DNOs, some of which are households that have inhabitants that use medical equipment, who have mobility issues, are of pensionable age or who have any cognitive impairment.

WPD's PSR has approximately 1.9 million customers registered, and it is an ongoing difficulty for all DNOs to maintain their PSRs. The customer services teams in each DNO reply on consumers self-identifying as vulnerable and informing their DNOs if there is a change in their circumstances. Due to the high internal cost to DNOs to maintain the PSR, and Ofgem's Consumer Vulnerability Strategy, there is a drive to use more data driven methods to ensure the PSR remains accurate and that vulnerable households receive the support they require.

1.1.4 Smart Meters and Vulnerability

One of Ofgem's key focuses for vulnerability is to improve the identification of vulnerability and smart use of data. Smart meters can play a key role in achieving this aim as they provide regular monitoring of a household's usage. Research recently published by 2020Health (J. Paxman, M. Jones, Dr E. Constanza and J. Mannin, November 2020), explores the possibilities for harnessing smart meter data in health and care monitoring systems. The report concludes that:

> *The opportunity for smart meter data analysis in health and care monitoring is in fact unprecedented, since never before has there then a government-driven roll-out of communications hardware into people's homes. If used as health and care monitoring technology, the smart meter could soon become a virtually ubiquitous tele-health-care solution.*

2020Health presents the table below which demonstrates the potential monitoring opportunities using smart meter data, primarily targeting single occupant households.

Table 1: Potential monitoring opportunities using smart meter data, primarily targeting single occupant households from 2020Health (J. Paxman, M. Jones, Dr E. Constanza and J. Mannin, November 2020).

Context	Service using smart energy data	Relevance
Informal and formal care	Monitoring of vulnerable people	Frailty, Learning Disabilities, Detection of early state neurological disease
Health and social care	Monitoring of long-term conditions progression	Alzheimer's Dementia, Parkinson's, Multiple sclerosis, COPD
Health care	Post-operative or post-discharge monitoring	Stroke, Heart Failure, Hip/ knee surgery, Vascular surgery.
Health care	Impact monitoring of health intervention	Sleep medication, CBT, SSRI antidepressants, Physical therapy

1.1.5 Combined Strategy

The health and care monitoring opportunities listed in Table 1 align almost directly with the household vulnerabilities that DNOs consider. Therefore, it is thought that DNOs could use smart meters to determine if a household may be considered vulnerable. If possible, a model could be developed to analyse smart meter data from anonymised households. If a house is identified, DNOs could request access to the personal information from the Smart DCC, with permission from Ofgem. This would benefit the consumer by ensuring they receive the support they require. It would also benefit DNOs by reducing the burden of maintaining the PSR and it would ensure the full social benefit of the smart meter roll-out is realised.

1.2 Project VENICE

Project VENICE (Vulnerability and Energy Networks Identification and Consumption Evaluation) is a network innovation project funded through Ofgem's Network Innovation Allowance, supported by Western Power Distribution (WPD). The project started in August 2021 and is using a three-pronged approach to support vulnerable customers in the future.

The first work package focuses on a net zero Community, at Wadebridge in Cornwall. Led by Wadebridge Renewable Energy Network (WREN), a community energy group, the project will establish how net zero is likely to impact on fuel poverty, as growing numbers of people switch to electricized transport and heat. WREN will then look at how it can work with WPD to support vulnerable customers through this transition, finding ways for them to participate in the decarbonisation of the energy system to benefit the community and achieve net zero equality. This will include investigating different commercial models to establish which ones work best for consumers.

The second strand, led by Frontier Economics, is looking at changes in electricity use during the pandemic and how likely these changes are to continue. For instance, the shift to home working, and whether this will have an impact on customers in vulnerable situations. This could be invaluable to all DNOs for their business planning.

Finally, the third stand is undertaken by us, Frazer-Nash Consultancy, and is presented in this report. Our work package explores the concepts discussed in Section 1.1; whether smart meter data can be used to identify vulnerable households.

1.3 Report Structure

This report details the analysis undertaken to conclude whether smart meters could be used to identify vulnerable households. This report builds on findings from a research project undertaken by Frazer-Nash Consultancy, (Lily Darling, Elsie Roberts and Tom Saunders, March 2022), which identified features of vulnerability that may be recognisable in a household's smart meter data.

The remainder of this report is structured as follows:

- Section 2 details the modelling approach.
- Sections 3, 4, 5 and 6 detail the method, results and conclusions from each of the different modelling techniques used.
- Section 7 gives a summary of the conclusions and recommendations following this work.

2 Modelling Approach

The aim of this workstream is to determine if vulnerabilities can be identified from smart meter data. The research presented in (Lily Darling, Elsie Roberts and Tom Saunders, March 2022) and discussed in Section 1.2, found that many of the vulnerabilities considered by DNOs could be identified from smart meter data. The links between behavioural characteristics and vulnerability are complex, hence some behavioural characteristics may be indicators of multiple vulnerabilities, whilst some characteristics may be indicators of both a vulnerability and a normal behaviour.

Some behavioural characteristics which might be identifiable, and their potential link to a vulnerability, are:

a. The use of medical equipment in the home,

b. A household using its appliances more frequently than anticipated due to dementia[1],

c. A household using more electricity than anticipated in the winter due to poor insulation,

d. A household using consistently less usage than anticipated due to payment concerns,

e. A household increasing its day-time usage due to unemployment[2],

f. A household increasing its night-time usage due to mental health difficulties,

g. A household using less electricity at the end of the month due to payment concerns.

These features are referred to as the behavioural characteristics of consumers in a household, and the developed model will aim to identify these in smart meter data. As these behavioural characteristics are not exclusive to vulnerability, the identification of these within smart meter data should not be used in isolation. When the model predictions are combined with other knowledge of the household, customer service teams within the energy networks can conclude whether the household may be vulnerable, and then contact the household and target their services.

Looking at the features listed above, the types of behavioural characteristics that are linked to vulnerabilities can be separated into three categories:

- Appliance related characteristics (this covers examples *a* and *b* above),
- Household usage levels different to its cohort (this covers examples *c* and *d* above),
- Overall changes in behaviour (this covers examples *e*, *f*, and *g* above).

Therefore, three separate models have been developed to specifically identify each of these characteristics. The remainder of this section gives an overview of each model, and the following sections in this report detail the approach and prediction ability for each of these three models.

[1] Dementia could lead to other behavioural characteristics, such as a sudden change in appliance usage, but the listed behaviour is the most identifiable characteristic for this vulnerability. As mentioned, the models should be used in conjunction with household data to ensure correct interpretation of results.
[2] As an example, unemployment may not be discernible from someone who has started working from home. This is where using household data alongside analysis from the models is paramount.

2.1 Appliance Disaggregation and Prediction

The aim of this model is to determine how a household is using its appliances and compare this to how a vulnerable household is thought to use their appliances. This is completed by first determining how appliances are used in a household, which are referred to as the appliance usage statistics. These are used to disaggregate smart meter data into the likelihood that each individual appliance is used in each 30-minute window in 24 hours. The model does this for each day and uses the behaviour to determine how a household uses each appliance. The development of this model is separated into two:

- Section 3 details the method and results for determining household appliance usage statistics.
- Section 4 details the development of the appliance disaggregation and prediction model using the usage statistics.

Details of the datasets used for the development and testing of this model are given in Section 3. The conclusions, limitations and future developments for this model are given in Section 4.

2.2 Cohort Comparison

The aim of the model is to determine whether a household is using statistically more or less electricity than anticipated, given its cohort's usage. A cohort is defined through the household characteristics, for example: the location, size, EPC rating and number of inhabitants. This is done by developing a model that uses the known household features to predict the anticipated usage and then comparing this to the known usage from the smart meter data. Details of the method, datasets used, testing results and conclusions of this model is given in Section 5.

2.3 Overall Changes in Usage

The aim of this model is to identify changes in a household's usage pattern. This is done by first identifying a change in usage behaviour, and then attributing this change to a statistically significant increase or decrease in usage in each 30-minute window through the day. Details of the method, datasets used, testing results and conclusions of this model is given in Section 6.

2.4 Training, Testing and Validation Datasets

Each of the three models have different training, testing and validation dataset requirements. A summary of the requirements for each of the three models is given in Table 2 below. The entries with an * indicate datasets which were not obtained for this work package due to difficulties obtaining smart meter data.

Initially, we tried obtaining smart meter data from households with known vulnerabilities as validation data for the cohort analysis and overall changes in usage models. We set out to do this through surveying the public to gain access to their smart meter data. We set up GDPR sharing confirmations and procedures so that we could gather smart meter data with the customers permission. We then asked Frazer-Nash staff, friends and family, and customers of WREN through the inclusion of the survey in a newsletter that WREN publishes. We also asked Severn Wye Energy to send our survey out to their customers, but this did not take place due to GDPR concerns and limitations from Severn Wye. Unfortunately, the results from this survey were not extensive enough for us to use as validation data. There were not enough cases where people gave us permission to access their smart-meter data either from lack of awareness or lack of desire. Out of the responses we did receive, we were lacking in cases where people identified as vulnerable.

Table 2: Dataset requirements for training, testing, and validating each of the three models. Entries with a * indicate datasets that were not obtained for this analysis.

Model	Training	Testing	Validation
Appliance Disaggregation and Prediction	High resolution (<30 second) appliance monitoring for multiple households and appliances.		30-minutely household usage for households with irregular appliance usage behaviours and medical appliances. *
Cohort Analysis	Average household usage with known characteristics. A proxy could be used for initial model generation.	Average usage for real households with known characteristics. *	Average usage for households with known vulnerabilities. *
Overall Changes in Usage	Real smart meter data for many houses. No information of the house is required.		Smart meter data for houses with a change in circumstance at a known time. *

It was decided that the training and testing datasets that were obtained for each model[3] were sufficient to develop the models and conclude whether they could detect each behavioural characteristic. However, the models must be tested with household data with known vulnerabilities before they can be utilised. For example, the increase in evening usage for households can be estimated by us and included in training data to see if the change can be detected. This will likely vary between different household types and therefore is an assumption we have not validated. These validation datasets would be sourced as follow-on work if the models were found to perform suitably.

2.5 Limitations of Modelling Approach

Each model analyses a single household and therefore does not consider global events which may impact all households, for example the start of the pandemic. This could be added at a future date by utilising the Low Carbon London dataset and identifying global events which impacted usage during that period. It should be noted however, that all the models are currently developed to consider changes over a long period, so it is not overly sensitive to holidays and similar small changes. Therefore, the impact of isolated events, for example an additional bank holiday, would not significantly impact the results.

In addition, the models do not directly account for changes in local weather and seasonal changes, but it is indirectly included by design. For example, if a household is sensitive to weather changes, when reviewing their usage pattern, they would be more sporadic in their usage than a household which is less impacted by weather.

[3] Detailed in Sections 4, 5, and 6 for the appliance disaggregation and prediction, cohort analysis and overall changes in usage, respectively.

3 Usage Statistics

To produce the appliance disaggregation and prediction model, we must gather information on how different appliances are used: when each appliance is used during the day, how long the appliance draws power for, and how much power that appliance draws when in use. These are referred to as the appliance usage statistics, and with this information, we can build a model to detect the appliance from smart meter data, detailed in Section 4.

We have calculated the usage statistics for household appliances by utilising two open source, high resolution datasets, that recorded individual appliance usage for various houses in the UK for up to two years. Using these, we developed an algorithm to identify when an appliance was used across all the houses. These usages formed the usage statistics for each appliance.

This Section is structured as follows:

- Section 3.1: Description and preparation of high resolution datasets,
- Section 3.2: Identification of appliance power consumption profile types,
- Section 3.3: Identification of the appliance power consumption profiles,
- Section 3.4: Development of appliance usage detection algorithm,
- Section 3.5: Determination of appliance usage statistics.

3.1 Dataset Description and Preparation

Two open source high resolution datasets were used to extract real appliance usage powers: UKEDC and REFIT. These datasets recorded the power drawn by various appliances in different houses over a period of at least a year. A significant amount of processing was required to ensure the datasets were in a suitable format for the analysis.

3.1.1 UKEDC

The UKEDC dataset comes from the UK Domestic Appliance Level Electricity Disaggregated appliance power data published by UK Energy Research Centre Energy Data Centre in 2015 (UK Energy Research Centre Energy Data Centre, 2017). The dataset contains data from five different houses in the London area, with data taken between 2012 and 2015, but only recorded for one to two years for each house. Each house in the UKEDC dataset contains individual appliance power monitoring and a household aggregate, for different appliances for each house. The power drawn by each appliance was recorded with IAMs (Individual Appliance Monitors), which provided data for approximately every 6 seconds.

The UKEDC data also contains information about the households from which power was recorded. This is very useful for looking at how certain scenarios might affect how a household might use their appliances, such as number of occupants or how often the occupants are in the house. The information about each house is given in Table 3.

Table 3: Additional information about each household in the UKEDC dataset.

House	Building Type	Construction	Number of Occupants	Energy Improvements	Notes
1	End of Terrace	1905	2 adults, 2 children. One child born in 2011, the other in 2014	Solar thermal, loft insulation, solid wall insulation, double glazing	3 bedrooms, 1 bathroom.
2	End of terrace	1900	2 adults	Cavity wall insulation, double glazing	1 adult at work all day, the other is sometimes at home.
3	N/A	N/A	N/A	N/A	N/A
4	Mid-terrace	1935	1 adult, 1 pensioner	Loft insulation, double glazing	House fully owned
5	Flat	2009	2 adults	N/A	Has a communal boiler

3.1.2 REFIT

This dataset comes from the project entitled Personalised Retrofit Decision Support Tools for UK Homes using Smart Home Technology (REFIT) which was supported by the Engineering and Physical Sciences Research Council and three UK Universities (David Murray et al., 2015). The dataset includes data from 20 different households from the Loughborough area over the period 2013 to 2015. The nine highest power drawing appliances were monitored for each household, at approximately every 30 seconds. There is no information about the houses recorded in REFIT, such as in Table 3 for the UKEDC dataset.

3.1.3 Pre-processing the Datasets

Both datasets are ideal for determining household appliance usage statistics because they contain the power drawn for approximately 125 different appliances for 25 different households, giving a substantial dataset to work with. There are, however, a total of 334 appliances across the different households and the data was found to be at inconsistent time intervals. We therefore have undertaken to pre-processing steps: resampling and appliance categorisation.

3.1.3.1 Resampling

The UKEDC datasets nominally had timesteps of every six seconds, but upon inspection, the time ranges from two to ten seconds. The REFIT datasets were similar with timesteps being between every 10 to 25 seconds. To make the analysis simpler and more interpretable, we resampled the datasets to contain self-consistent steps:

- UKEDC was resampled to 10 second intervals,
- REFIT was resampled to 30 second intervals.

3.1.3.2 Appliance Categorisation

Across the 25 houses, there are a total of 334 appliances monitored. Some of the appliance names vary between houses (for example, *vacuum* vs *hoover*), and some houses have multiple of the same appliance and are named accordingly (for example, *fridge1* and *fridge2)*. To account for this variation, all 334 appliances were grouped into 24 categories, which were then used for the analysis. These categories are shown in Table 4 along with the number of houses in each dataset that contained an appliance from each category.

The miscellaneous category contains the appliances which are not considered in the analysis, for example: office fan, dehumidifier, and pond pump. We also discounted 12 appliances from across six houses due to either: inconsistent monitoring or unknown appliance. For example, one household had the washing machine, microwave and bread-maker recorded as one appliance. The categories marked with an * in Table 4 represents the appliances that draw high enough power to be recognisable from 30-minute averaged usage, and therefore the ones the modelling focuses on.

For the remainder of this report, when an appliance is referenced or discussed, it is referring to the categories detailed here.

Table 4: Appliance categories used, with the number of UKEDC and REFIT houses that contained at least one appliance within the category.

Category Name	Number of UKEDC houses	Number of REFIT houses
Kettle*	4	15
Microwave*	3	16
Toaster*	3	9
Dishwasher*	3	15
Vacuum*	2	0
Fridge	2	7
Freezer	1	10
Fridge-Freezer	1	15
Electric hob	1	0
Tumble dryer*	0	10
Oven*	3	0
Electric heater*	1	3
Washing machine*	3	19
Tv	2	19
Iron	2	0
Computer	4	12
Boiler	2	0
Heat pump	1	0

Continuous	1	0
Small kitchen appliances	3	4
Small bedroom appliances	2	1
Small lounge appliances	4	4
Lighting	1	1
Miscellaneous	4	3

3.2 Appliance Power Consumption Profile Types

To detect when an appliance is in use from the power it has drawn, we started by exploring the different power consumption profiles that appliances produce.

We researched ways in which appliance disaggregation from smart meter data had been undertaken in academia to help aid our decision in which methods we were going to utilise. During this research, we found that an appliance can have four different types of power usage profile whilst they are being used. These are:

- Type I: have two states of operation, either on or off. For example, a toaster or a kettle.
- Type II: multi-state appliances with a finite number of operating states greater than one. This switching pattern can be repeatable. For example, a washing machine or dishwasher.
- Type III: continuously varying devices due to a variable power draw, with no fixed number of states. For example, a dimmable lamp.
- Type IV: in use constantly, consuming energy at a constant rate. These include permanent consumer devices. For example, a Wi-Fi router.

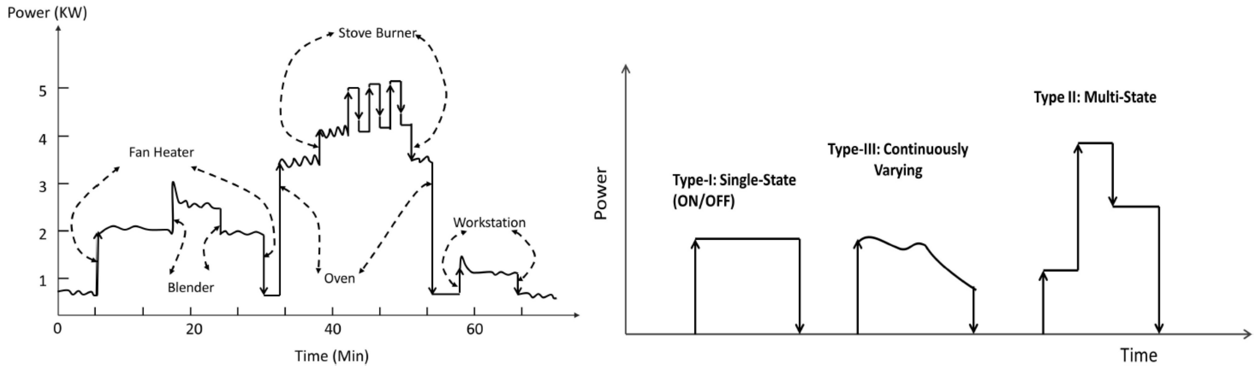

Figure 1: Plots showing the different types of power draw an appliance might have during a usage (David Murray et al., 2015).

An example of what these usages might look like is shown in Figure 1. By determining which power consumption profile Type each appliance has, and what each of the known power states are, we can determine how much power it would draw when turned on. However, if an appliance is too variable, or does not draw a high amount of power, a model will not be able to identify it.

Therefore, we concluded that:

- Type I appliances could be considered in our model if they draw a high enough power when in use. These are referred to as ‘single step’ appliances for the continuation of the report.
- Type II appliances could be considered in our model if at least one of the power steps draws a high enough power when in use. These are referred to as ‘multi step’ appliances for the continuation of the report.
- We have excluded all Type III appliances from this analysis as we would not be able to determine the start and end of a usage within a 30-minute period. Our judgement is that only a small number of household appliances are Type III and therefore their exclusion is unlikely to affect the conclusions of this work.
- We have excluded the identification of all Type IV appliances from the analysis. We have accounted for the power drawn by all these appliances by considering a constant baseload power, more detail is given in Section 4.2.1.

3.3 Appliance Consumption Profiles

With the appliances reduced to 24 categories, the next step is to review the power drawn by each of the categories to determine whether they were a single or multi-step appliance (Section 3.2) and whether they used a high enough power to be identifiable in smart meter data.

The 24 categories we down selected to final nine that the analysis was run for: kettle, microwave, toaster, dishwasher, vacuum, tumble dryer, oven, electric heater and washing machine. This selection was based upon initial inspections into the power drawn by the appliances, and all low power appliances were excluded. For example, the fridge is continuously a low level of power and increases when the door is opened. But as the fridge is only opened for a matter of seconds, this usage will not be identifiable in smart meter data. Additionally, all categories with different appliance types, for example ‘small bedroom appliances’, were discounted.

To review the power drawn by the selected categories, we extracted the power individually drawn by each appliance, in each category, and plotted the Gaussian kernel-density estimate distribution. This gives the power most likely drawn by that appliance when in use, and the variation in power drawn by the appliance across different households. For example, different makes and models of the same appliance are likely to draw different power when on. By considering the power for all appliances in each category, we can get a greater understanding for this variability and ensure it is accounted for in the analysis.

This was completed for each of the nine selected categories, and the results from the kettle and dishwasher are presented here for discussion. The Gaussian kernel-density estimations for the remaining seven categories can be found in the Annex. For each category, the estimations have been calculated for all powers greater than 10W, to avoid a large spike in probability when the appliance is idle (plugged in but not in use).

Appliance Power Consumption Profile: Kettle

Figure 2 shows the power probability densities for all kettles monitored across the households, 19 in total. The densities show that the kettle is a single step appliance as there is only one main peak in the probability density for each house; the small ‘bumps’ lower than 1000W are likely due to the kettle approaching and leaving full power. This can be confirmed by reviewing a single kettle usage, as shown in Figure 3, where the power drawn is clearly only from one step.

By then looking at location of the peaks, Figure 2 shows that there is significant variation in the power drawn by a kettle depending on the household. This implies that different makes or models of kettles, and potentially different ages, affect the power drawn. This variation must be captured in the analysis when attempting to identify a kettle.

Appliance Power Consumption Profile: Dishwasher

Figure 4 shows the power probability densities for all dishwashers monitored across the households, 18 in total. The dishwasher shows considerably more peaky behaviour, and there are two clusters of peaks. This shows that the dishwasher is a double step appliance, and the power drawn by one step is around 100W and the second step draws a much higher power of between 1500W and 3000W, depending on the specific appliance. When a single dishwasher usage is isolated, as shown in Figure 5, the different powers drawn between two distinct power steps are observable.

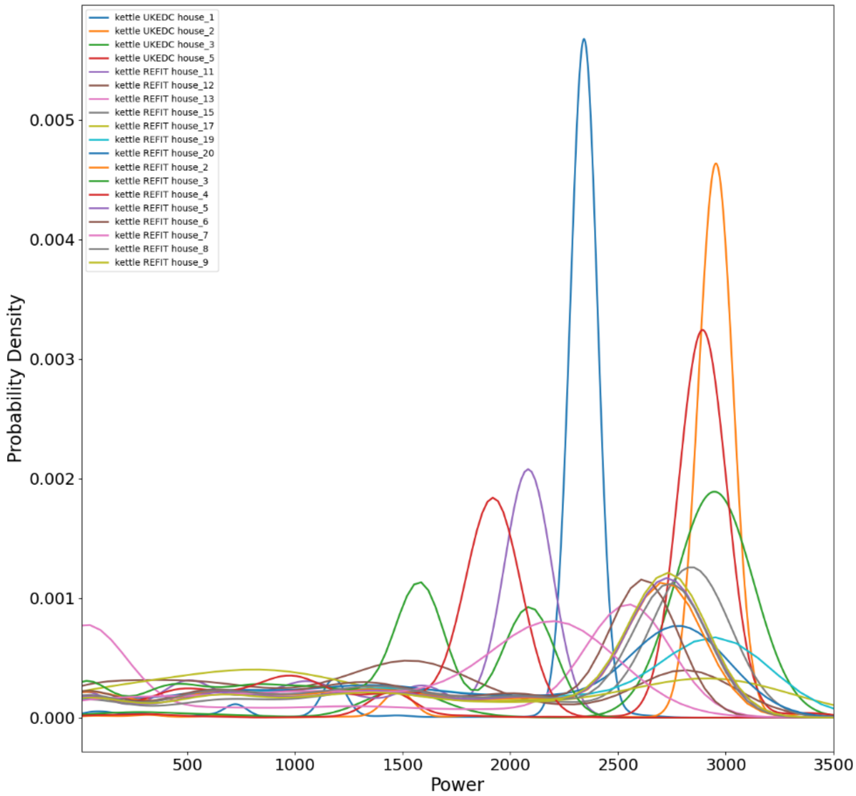

Figure 2: Gaussian kernel-density estimates of the power drawn by each appliance within the kettle category.

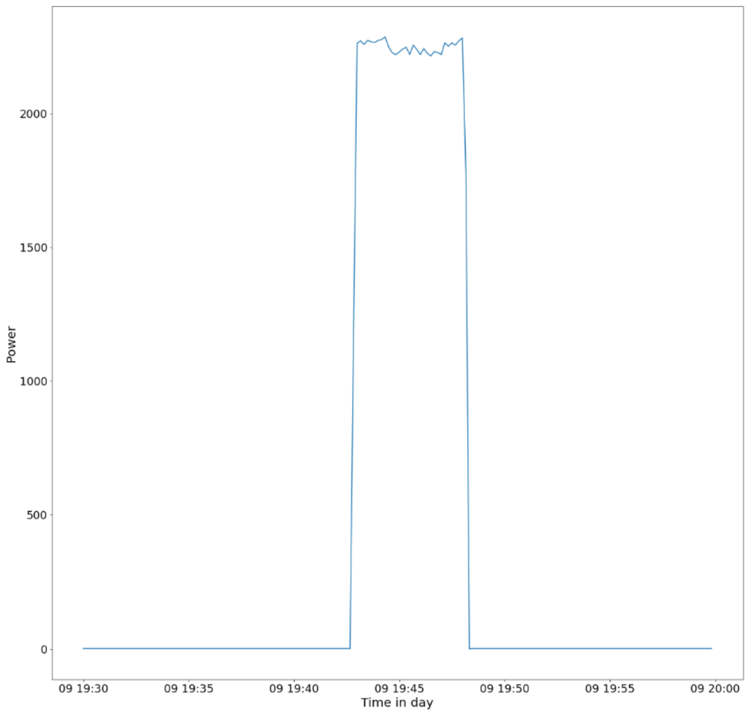

Figure 3: A randomly selected usage from the power drawn by the kettle from the UKEDC House 1 dataset.

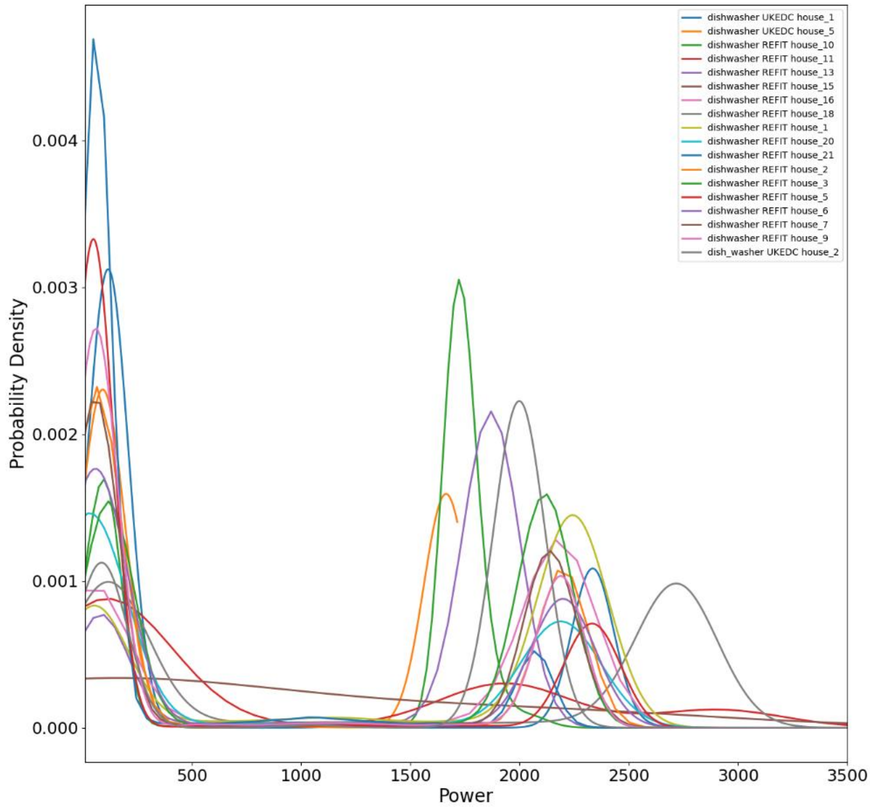

Figure 4: Gaussian kernel-density estimates of the power drawn by each appliance within the dishwasher category.

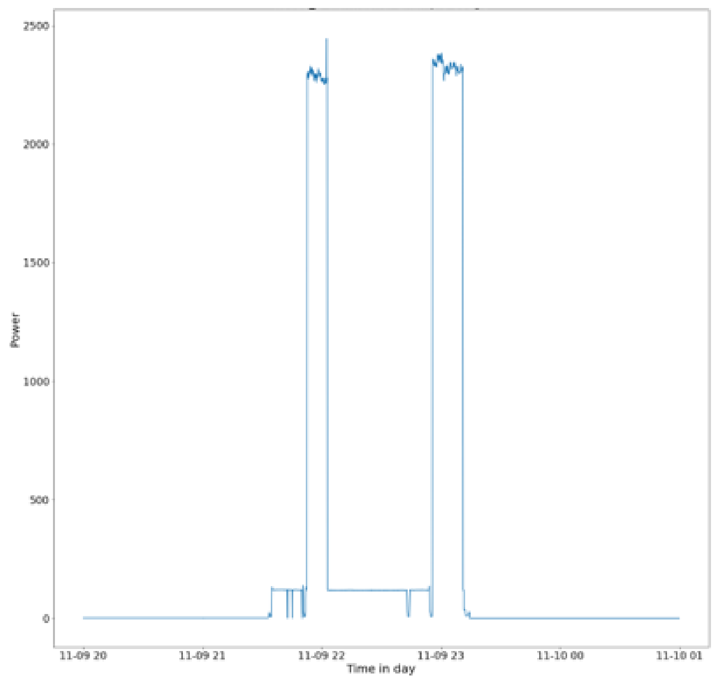

Figure 5: A randomly selected usage from the power drawn by the dishwasher from the UKEDC House 1 dataset. The time axis is displayed as the date and then the hour, for example, 11:09 20 represents 11th September at 8pm.

3.4 Usage Detection Algorithm

Once the characteristics of power consumption for the down-selected appliance categories were established, we developed an algorithm to identify when each appliance was in use and extract the usage details. The algorithm was developed for both single-step and multi-step appliances (Type I and Type II) and it determines:

- The usage start time,
- The average power drawn at each step,
- The time spent at each power step,
- The total usage length.

The following sections detail how the algorithm was developed, it's key functionality and the resultant usage statistics.

3.4.1 Algorithm Inputs

The appliance specific usage inputs that the algorithm requires are:

- N Steps: number of distinct power steps used by the appliance.
- Power limits: minimum and maximum power limits for each step in the usage.
- Base Power: power limit below which the appliance is considered not in use.
- Grace period: the length of time the power can be below the base power before it is considered not in use. This ensures temporary drops in power are not mistaken for a new usage.
- Minimum usage period: The minimum duration for the appliance to be in a power step to be considered 'in use' or 'in step'. This ensures temporary spikes in power are not mistaken for a new usage.

These inputs were defined by observing clusters of peaks in the appliance Gaussian kernel-density estimate distributions, detailed in Section 3.3, and extracting the range of peak values. This ensures the input power values are applicable for all the appliances from all the households.

3.4.2 Algorithm Logic

The usage detection algorithm works by determining when an appliance is 'in use' and 'in step'. A single step appliance will only ever in 'in use' or 'not in use'. A multi-step appliance will be 'in use' and 'in step X', where X represents the step number (range from 1 to N). It does this by:

1. Iterating over all recorded powers for each appliance.
2. Determining if the appliance is in use by checking if the power is above the specified base power.
3. For any multi-step appliances (where the number of steps is greater than 1), determine which step it is in, based on the power limits provided. By using these ranges, any noise in the usage is either ignored or accounted for.
4. Consider the appliance as 'in use' or 'in step' if the power is held within the correct range for the minimum usage period. This allows for any noise that might look like a usage to be excluded, if it is less than the minimum time that an appliance will be on for.
5. For any multi-step appliances, if the power moves to a new step's power range for longer than the grace period, then consider the appliance now in the new step.
6. If the power falls below the base power for longer than the grace period, then appliance is usage finishes.

7. The algorithm was run for the nine categories, to extract all the usages for each appliance across all the houses.

For each usage, the following information was extracted:

- Usage start time,
- Length of total usage,
- Number of power steps in the usage (a double step appliance could have four steps if it cycles between each step multiple times),
- Average power of each step within the usage,
- Length and start time of each step within the usage.

This usage data was then used to determine the usage statistics for each appliance category.

3.5 Presentation of Usage Statistics

3.5.1 Method

We can represent the appliance usages as a set of usage statistics for each category. These are used to identify trends in usage for each appliance. Three sets of information are extracted and presented for the usages for each category across all houses. These are:

- Ratio of usage count over maximum usage count. This is the number of times the appliance was turned on (the usage start time) in each 30-minute window, divided by the maximum number of time that appliance was turned on across all of the 30-mintue windows. This is so the relative probabilities can be compared across the appliances.
- The average power drawn for each step when in use.
- The total usage length.

3.5.2 Multi-step Appliances

For multi-step appliances, the usage detection algorithm identifies the average power when at each level, and how long the appliance is at that level. This accounts for the variation in power drawn by these appliances when on different cycles. For example, the largest power draw for a washing machine is heating the water at the start, but the temperature it is heated to depends on the exact cycle. Therefore, this first high-power step will have varying length, depending on the cycle, which results in a distribution. If generating the usage statistics for a multi-step appliance, we determine the distributions of average power drawn and total usage length, for each power step.

This results in many distributions and complex relationship between each distribution. A decision was therefore made to exclude all multi-step appliances from the analysis for the appliance prediction (Section 4), due to the extra complexity required. The multi-step appliances could be added into future analysis if the method is proven to work.

This resulted in five appliance categories that could be used for the appliance prediction: kettle, electric heater, toaster, microwave, and vacuum.

© FNC 2022

3.5.3 Limitations

The limitations of the usage statistics are:

1. The usage statistics are only extracted for single-step appliances (Section 3.5.2). We judge this as an appropriate initial analysis and the exclusion of multi-step appliances is unlikely to affect our conclusions.
2. The top and bottom 5% of values for the usage length and average usage power have been removed for the figures. This is because over many usages, there are some that were clearly erroneous.
3. The REFIT houses only have usage at 30 second intervals, and therefore the usage lengths appear ‘peaky’.

3.5.4 Results

The resultant usage statistics for the kettle are presented in Figure 6 and discussed below. The usage statistics for the other single step appliances (electric heater, vacuum, toaster, and microwave) are presented in the Annex.

The top graph shows that the kettle is used most in the morning and is also used throughout the day, and peaking again in the evening, when people tend to prepare dinner. There are not many usages in the early hours of the morning. This is in line with expectations from inherent knowledge of when the kettle is used.

The middle graph shows the distribution of the average power drawn by the kettle, which varies significantly. These peaks directly align to the power peaks observed in the Gaussian kernel-density estimation, shown in Figure 2 and therefore, we can conclude that this variation is due to the different makes and models of kettles between the houses. This conclusion is important as it means that: when the kettle is in use, it could draw a significantly different power depending on what kettle type the house has. This in turn adds a large amount of uncertainty to the power a kettle could draw in a 30-minute period. This must be accounted for in the method used to determine whether a kettle is used from smart meter data, detailed in Section 4.

The bottom graph in Figure 6 shows the distribution of the length of time the appliance was used for. The histogram of samples is peaky due to the different usage lengths (see Limitation 3). Nevertheless, the distributions show that the kettle is most likely to be used for durations up to two minutes, which is realistic.

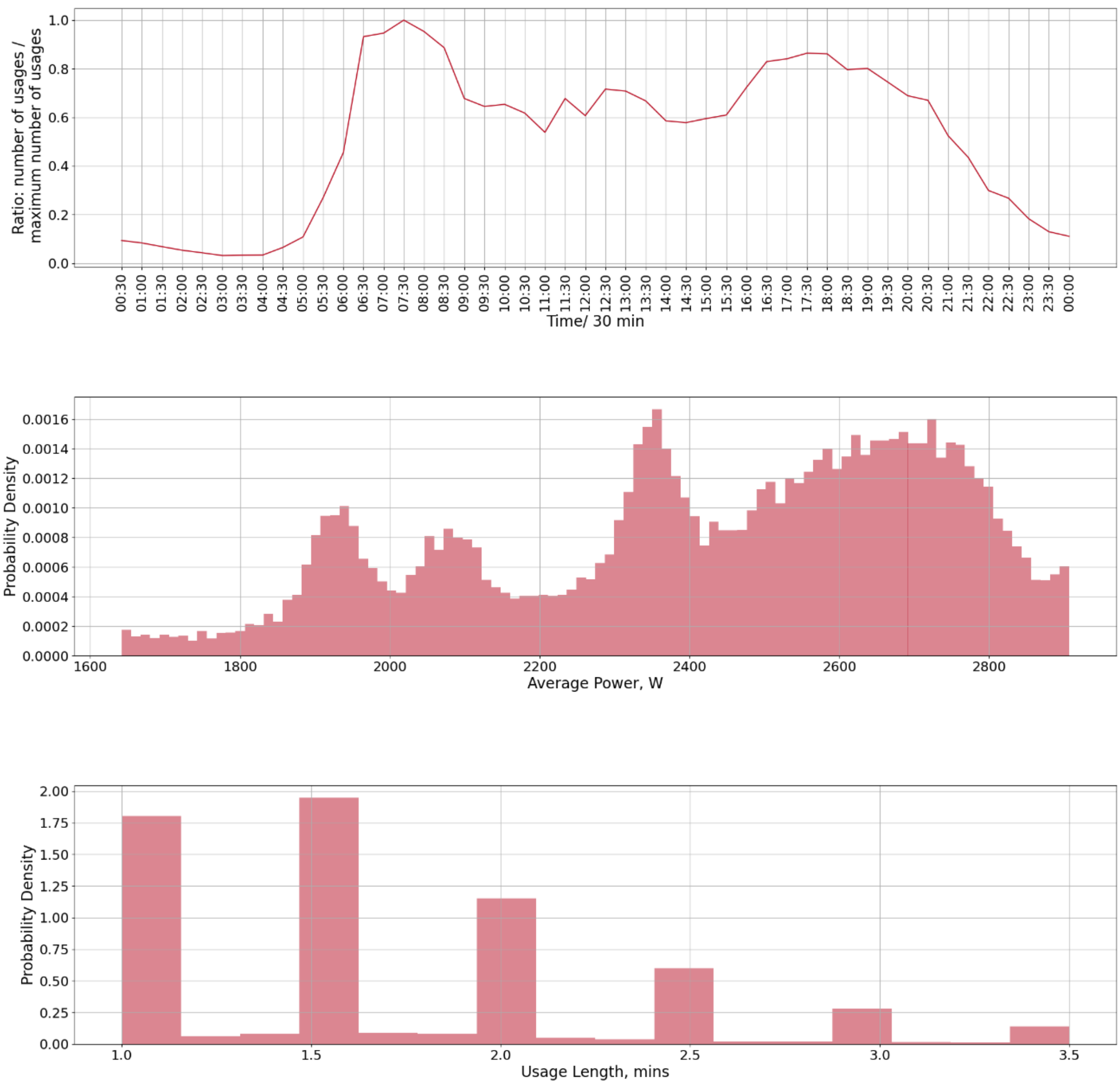

Figure 6: Usage statistics extracted for the kettle across all datasets.

3.6 Conclusions

To predict when an appliance is used in smart meter data, we must first know how appliances are used. This has been done by extracting details of appliance usages from two sets of open-source appliance monitoring data.

The two datasets monitored the power drawn by many appliances across multiple years (all pre-pandemic) for 25 different houses in the London and Loughborough regions. There was a total of 334 appliances monitored in total. This was reduced to a more manageable selection of appliances by categorising them and determining their appliance power consumption profiles. For each category, we determined:

- The power consumption profile type: single step, discrete multi-step, variable steps or continuous. Only single and multi-step appliances are considered in this analysis.
- If the appliance draws enough power to be recognisable in smart meter data, which is averaged over 30-minute periods.

Following this down selection, the usage details for appliances in nine categories were calculated. These usage statistics were extracted using a bespoke algorithm developed, to determine the average power and usage length for single and multi-step appliances.

From these usages, the usage statistics were extracted for single step appliances only. The double step appliances were excluded for the remainder of the analysis due to the complex relationship between the power of each step and the number of steps in each usage. The double step appliances can be included in future analysis is the method is successful. The usage statistics were therefore determined for:

- Kettle,
- Toaster,
- Microwave,
- Vacuum,
- Electric heater.

These usage statistics enable us to build a model to identify whether one of these appliances was used in smart meter data, and ultimately, determine how a household uses their appliances. The method and results from this analysis are given in Section 4.

4 Appliance Disaggregation & Prediction

The aim of this model is to determine how a household is using its appliances and compare this to how a vulnerable household is thought to use their appliances. To do this, for each appliance in the household, we calculate a probability that on any given day the appliance has been turned on at a specific time, which is referred to as the Time of Day prior (ToD prior). With this, we can determine if an appliance is being used more frequently than anticipated, or at different times of the day. This method was developed for household appliances that we had obtained usage statistics for, discussed in Section 0.

This section details:

- Findings from existing research into appliance disaggregation,
- The method implemented to calculate a household and appliance specific ToD prior,
- The method and results of testing this model on repeat data,
- The method and results for this model using real household usage,
- The recommended next steps for model development.

4.1 Existing Research

Appliance disaggregation is a large area of research in academia. An initial literature review found many research papers presenting analysis that uses machine learning techniques to identify appliance usages in aggregated usage data. Notably, (Hana Altrabalsi et al., December 2014) concluded that:

> *Despite increased research efforts, non-intrusive appliance load monitoring (NALM) techniques that can disaggregate power loads at low sampling rates are still not accurate and/or practical enough, requiring substantial customer input and long training periods.*

The research papers concluded that the greatest difficulty in the machine learning methods for appliance disaggregation is the large training set requirement for the model to learn each specific power consumption profile. These then become obsolete when a different make or model is used, or a new, unknown, appliance is in the aggregated load. As a result, the research concludes that appliance disaggregation at a resolution less than six minutely has not been used successfully.

Given these conclusions, we changed our modelling approach to directly include the uncertainty in an appliance power consumption profile in the model, and thus shifted our approach to be probabilistic. This ensures that the level of confidence which can be placed in results is assessed and that the model will be extensible if proved successful with the simplistic test cases presented here.

4.2 Method

An overview of the method used to disaggregate appliances from 30-minutely aggregated load, and then to calculate the ToD for each appliance for a household is shown in Figure 7. The remained of this section is structured as follows:

- Section 4.2.1 details the generation of the matrix likelihoods.
- Section 4.2.2 details the generation of the initial ToD priors.
- Section 4.2.3 details the generation of an iterative ToD prior, resulting in a specific ToD prior for a household and appliance.

1. Obtain 30-minute usage data for a household to run the analysis for. This is selected from the datasets detailed in Section 3.1.
2. Determine all the combinations of appliances that could be used in each 30-minute period for that household. These are the matrix combinations, discussed in Section 4.2.1.1.
3. Calculate the probability that each appliance combination happened, for each 30-minute period, for one day of the household data. These are the matrix likelihoods, discussed in Section 4.2.1.3.
4. Determine the initial ToD prior for each appliance. This is representative of the inherent knowledge “the kettle is turned on at 8am”. The calculation is detailed in Section 4.2.2.
5. Refine the ToD prior based on the observed usage of the previous day. The calculation is detailed in Section 4.2.3. Repeat steps 4 and 5 for each day.
6. Conclude the final ToD prior for each appliance in the household. The testing and results are discussed in Section 4.3 and 4.4, respectively.

Figure 7: Overview of the appliance disaggregation and prediction method used.

4.2.1 Matrix Likelihood Probabilities

4.2.1.1 Matrix Combinations

For a given half an hour, the observed energy usage is a combination of the power drawn by all the appliances that have been turned on in that 30-minute window, plus the power drawn by all appliances that are still on from being turned on in the previous 30-minute window, and so on. This can be represented using a binary matrix where the horizontal axis represents the appliance, and the vertical axis represents whether the appliance was turned on in the current window, in the previous window etc. Therefore, if you consider all combinations of *m* appliances, for *n* 30-minute windows, there will be a total of 2^{m*n} combinations. These binary matrices are referred to at the matrix combinations, and are used to calculate the matrix likelihoods. A representation of a matrix combination is presented in Figure 8.

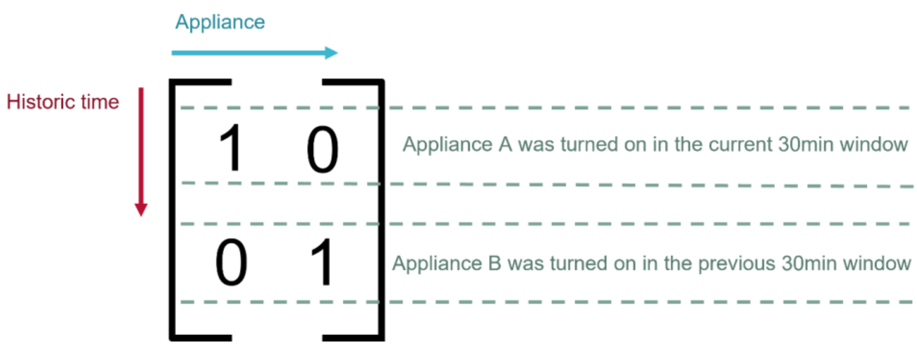

Figure 8: Representation of matrix combinations.

4.2.1.2 Matrix Combination Power Distributions

The matrix likelihood is calculated from the matrix combination by determining the distribution of power that could be drawn for each element in the matrix combination and combining this into a total matrix combination distribution of power drawn. This method accounts for the uncertainty in usage by an appliance in a 30-minute window and is calculated by extracting the appliance power distribution from the usage statistics.

Power Drawn by the Baseload

This analysis can only consider appliances that draw a significant amount of power and therefore not all appliances can be explicitly modelled. Any appliance that draws a consistent, low power, is included in a baseload, modelled as a truncated normal distribution[4]. We sample this normal distribution thousands of times to build up a set of baseload powers that may be observed in a 30-minute window. For example, if all appliances are modelled as off (i.e. the matrix combination is all zeros), then you would expect to observe a power in a 30-minute window of this baseload.

For this analysis, the baseload is considered to always be on, and the power drawn is a normal distribution centred at 10Wh with a standard deviation of 30Wh, shown in Figure 9. The red line shows the probability density of the sampled values: it is not exactly aligned to a normal distribution due to the random nature of sampling from a truncated normal. The value of the baseload can be increased or decreased depending on the household.

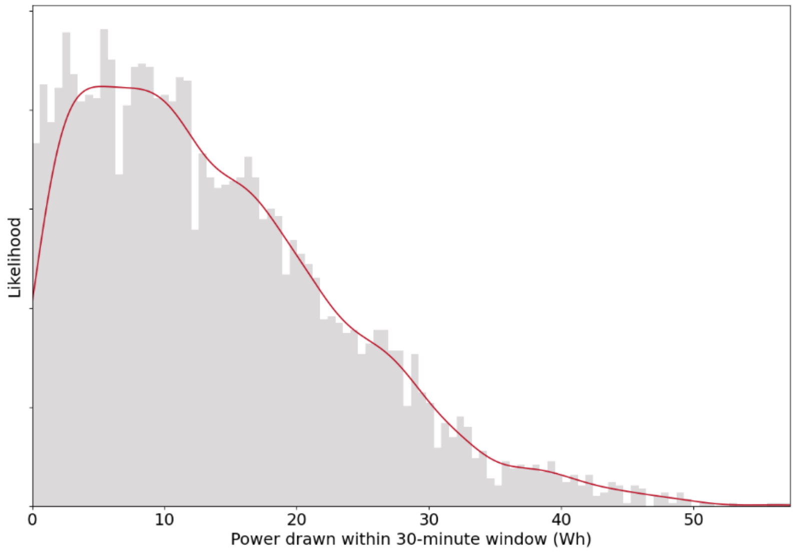

Figure 9: Distribution of the sampled power drawn by the baseload over a 30-minute period.

Power Drawn by an Appliance

To determine the variation in the power drawn by an appliance within a 30-minute window, we must calculate how much power it *could* draw in a 30-minute window. There are three uncertainties that must be accounted for: when in the window is the appliance turned on, how much power the appliance draws when it's turned on and how long it's turned on for within the 30-minute window. To do this, we use the appliance usage statistics, discussed in Section 3.5. From these, we have the distribution of power drawn and time used for all appliances that are recorded in those datasets. For example, we have the power drawn by the kettle at 10 second intervals from the UKEDC houses, and at 30 seconds from the REFIT houses.

We randomly sample from the distribution of powers and times to create a set of sampled appliance usages, and assign each usage a start time within the 30-minute window. For example, you may sample that the kettle is used 5

[4] The distribution is truncated to ensure there are no negative powers.

minutes into the window for 3 minutes and draws an average power of 2000W over that period[5]. Over the 30-minute window this contributes to the total energy drawn of 100 Wh. If the appliance is used for longer than is remaining within the window, the power drawn is only calculated for the time it is on in that window. This sampling is repeated for thousands of usages to build up a distribution of energy that the appliance could draw. The same method is applied for creating the power drawn distributions for the previous 30-minute window, and so on.

Figure 10 shows the histogram and probability density for 5,000 samples of the kettle usage in a 30-minute window, plus the samples from the baseload over that time (see Figure 9). The peak of the probability distribution represents usages of the kettle entirely contained within the 30-minute period. The standard deviation of the peak, and the tail to the right represents the variation in power drawn by the kettle (see Figure 2 for the variation in kettle power). The left of the peak, the very low power values, represents when the kettle is only used for a small amount of time within the window (plus the baseload).

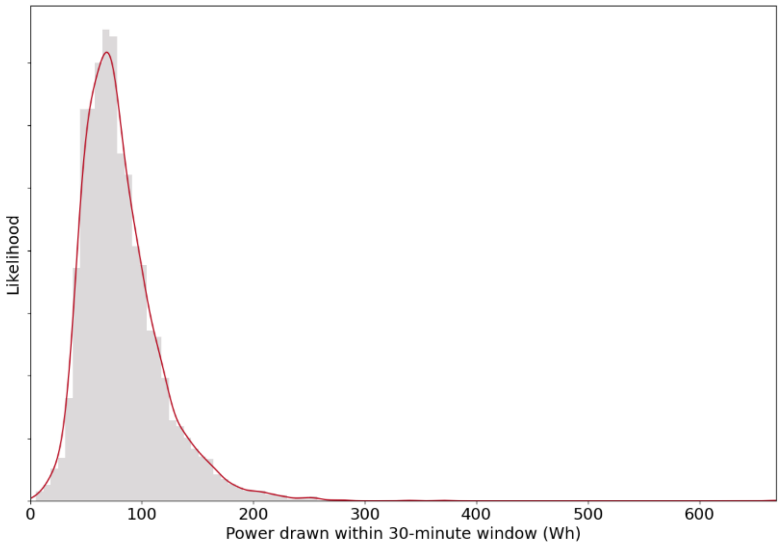

Figure 10: Distribution of power drawn by the kettle in a 30-minute window.

This method ensures that all uncertainty about when the appliance is used within the 30-minute window is captured: the power drawn in that window may be low but that does not mean a high-powered appliance was not turned on.

Combining Samples

The samples for the power drawn by the constant baseload and by each appliance for each 30-minute window, are combined to give the resultant samples of power for each matrix combination. For example, using the matrix combination represented in Figure 8, the samples of power drawn from the baseload, appliance A being turned on in the current 30-minute window and the power from appliance B being turned on in the previous 30-minute window, are all combined to give a resultant distribution of power that would be drawn in the 30-minute window if that matrix combination had occurred.

4.2.1.3 Matrix Likelihood

The matrix likelihood is the probability that each matrix combination results in the observed 30-minute usage. The equation below represents this for the i^{th} combination. The matrix likelihood is calculated by taking the Gaussian kernel-density estimate (KDE) from the sampled appliance powers. The problem is constructed so that the sum of all

[5] Note: only single step appliances have been used for this analysis.

the matrix combination probabilities (the matrix likelihoods) is equal to one, as all possible occurrences are considered, where N is the number of matrix combinations[6].

$$matrix\ likelihood_i = P_{KDE}(matrix\ combination_i)$$

$$\sum_i^N matrix\ likelihood = 1$$

Figure 11 shows the resultant Gaussian kernel-density estimates for matrix combinations when considering two appliances- kettle and vacuum- for a single 30-minute window[7]. The lines are coloured by the matrix combination where:

[0, 0]	Power observed is from the baseload only.
[0, 1]	Power observed is from the baseload and the vacuum being turned on in the 30-minute window.
[1, 0]	Power observed is from the baseload and the kettle being turned on in the 30-minute window.
[1, 1]	Power observed is from the baseload and the kettle and vacuum being turned on in the 30-minute window.

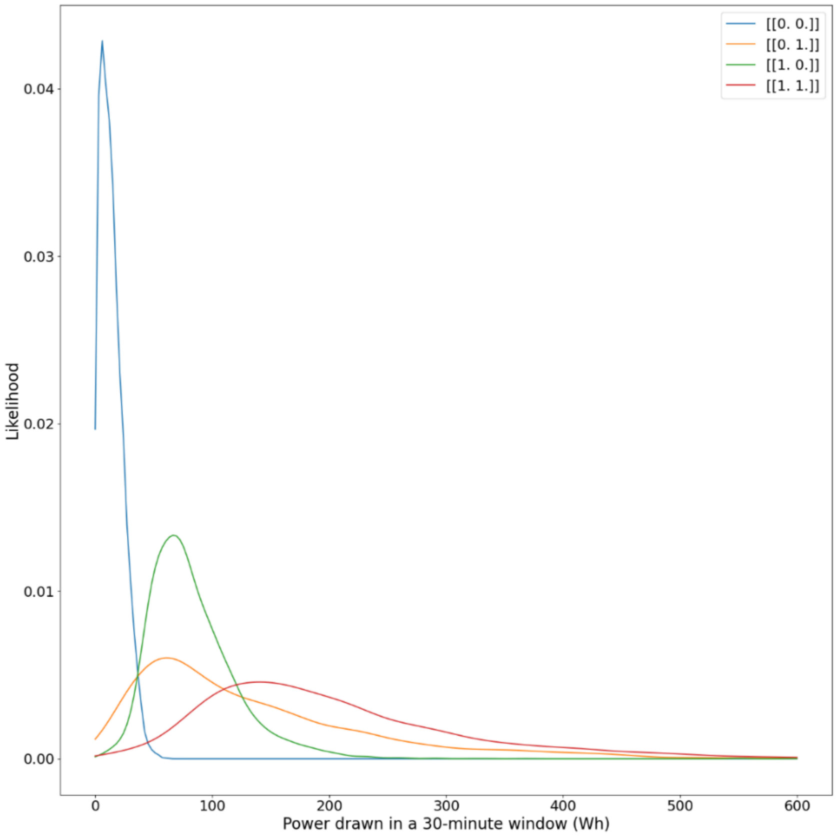

Figure 11: Gaussian kernel-density estimate of the matrix combinations when considering the kettle and vacuum for a single 30-minute window.

[6] N = $2^{(m*n)}$, where m = number of appliances, n = number of 30-minute windows considered.
[7] Only one 30-minute window is presented here as the kettle and vacuum are both short usage appliances.

The higher the line on the y-axis, the more likely it is that the matrix combination results in the corresponding usage. For example, if the observed power drawn in the 30-minute window is 80Wh, the most likely occurrence is [1, 0] (kettle used), but there is also some possibility the power is made from [0, 1] (vacuum used) or [1, 1] (both used). The combination [1, 1] has some probability it makes the 80Wh, but this is quite low. This is likely because: to make 80Wh, both the kettle and vacuum would have to both have been used for a short period in the 30-minute window, or those specific appliances draw a lower amount of power than a general one when used (a left-hand peak in Figure 2).

These Gaussian kernel-density estimations are used to determine the matrix likelihood combinations, which are the probability that each matrix combination (Section 4.2.1.1) occurred in the 30-minute window. The matrix likelihoods are calculated for each matrix combination and for each 30-minute window throughout a day.

4.2.1.4 Informed Matrix Likelihood

The matrix likelihoods are used to determine the most likely combination of appliance usage for each 30-minute window. This calculation can be improved by including the prior knowledge of when an appliance is more or less likely to be used. Therefore, the informed matrix likelihoods are determined by multiplying the calculated matrix likelihoods for a 30-minute window (t), by the ToD prior for that window (calculation of the ToD prior is given in Sections 4.2.2 and 4.2.3):

$$informed\ matrix\ likelihood_t = matrix\ likelihood_t \times ToD\ prior_t$$

The informed matrix likelihoods are normalised to sum to one for each 30-minute window (t), because all possible combinations are included in the matrix likelihoods.

The informed matrix likelihood is the final prediction of which combinations of appliances is most likely to make up the observed power drawn through a day.

4.2.2 Initial Time of Day Prior

The ToD prior is a representation of the inherent knowledge that we all have that an appliance is more likely to be used at a specific time of day. The initial ToD priors are derived from the usage statistics, discussed in Section 3.5.

For all the appliance usages across all the households, the number of times the appliance was used in each 30-minute window through a day was counted and normalised by the total number of days the appliance could have been used (the total number of days available for each household). This returns the probability an appliance was used in each 30-minute window for each house. For example, if one house has a year of appliance monitoring data and the kettle was used 100 times in total between 07:00 and 07:30, the probability that house uses the kettle in the 13th window (07:00 to 07:30) is 100 / 365. This is repeated for each house that contains the appliance, and a set of probabilities is obtained. A normal distribution is fit to these values, which is referred to as the q_{dist}. This distribution is normalised so that the area underneath sums to one.

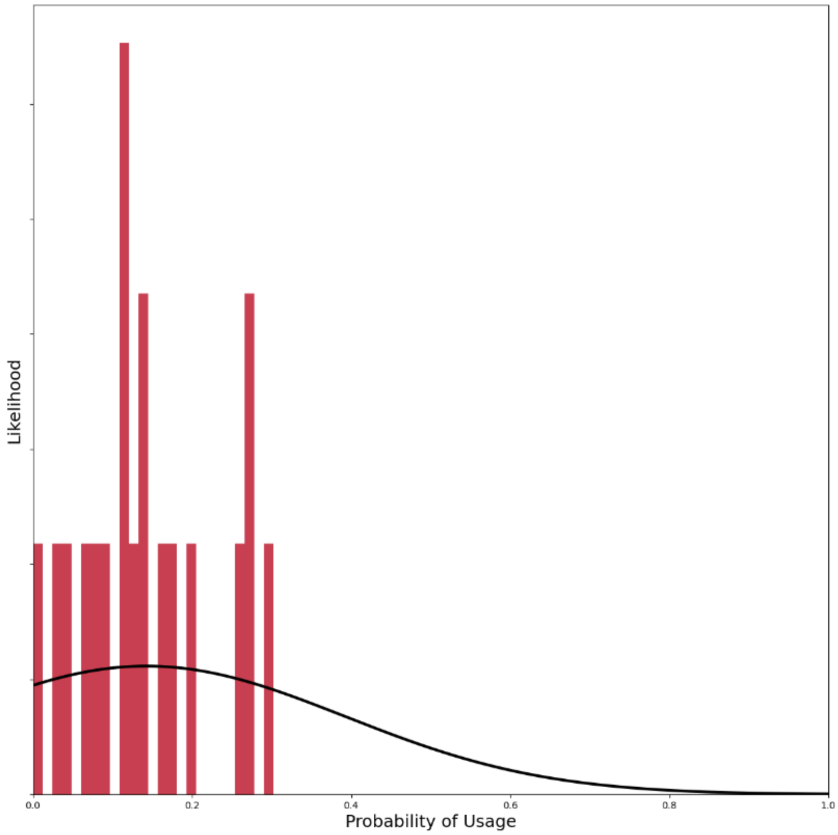

Figure 12: Samples of the number of usages across all houses in the REFIT and UKEDC datasets of the kettle in the 35th 30-minute window of the day (17:30 to 18:00), alongside the resultant q_{dist}.

Figure 12 shows the samples and resultant q_{dist} for the 35th 30-minute window (17:30 to 18:00) for all the houses that used the kettle in that 30-minute window of the day. There is a relatively high number of houses that used the kettle in the 35th window (19 out of 25 houses), but the probability is it used is centred around 0.2, which is quite low. This is because the households had appliance monitoring for many months and even if the appliance was used once a week, this still equates to a small overall probability is it used. This low number of usages results in a very uncertain q_{dist}- the normal distribution does not appear to be a very good fit to the household usage probabilities (the histogram). When developing this analysis, different q_{dist} fit types were tested, but we found that the normal distribution was the most suitable due to the high uncertainty from the low number of usages for some appliances. For example, when considering appliances with fewer houses that used than the kettle, if a more specific statistical fit was used, it tended to over-fit to the very small amount of data. Additionally, due to the low number of samples for some appliances, normal distribution was forced to have a standard deviation of at least 0.25, to ensure no over-fitting was occurring.

The initial ToD prior was calculated from the resultant q_{dist} as the expected value, which is calculated using:

$$TOD\ prior_t \ = \ E[X_t] = \int_0^1 x q_t(x) dx$$

Where $E[X_t]$ is the expected value for the t^{th} 30-minute window for a day, where $t = 0$ is the window 12:00 to 12:30. $q_t(x)$ is the q_{dist} probability distribution the t^{th} 30-minute window for a day and x is the integration variable. The expected value from the distribution shown in Figure 12 is calculated to be 0.26, which is slightly to the right of the most likely value (the peak).

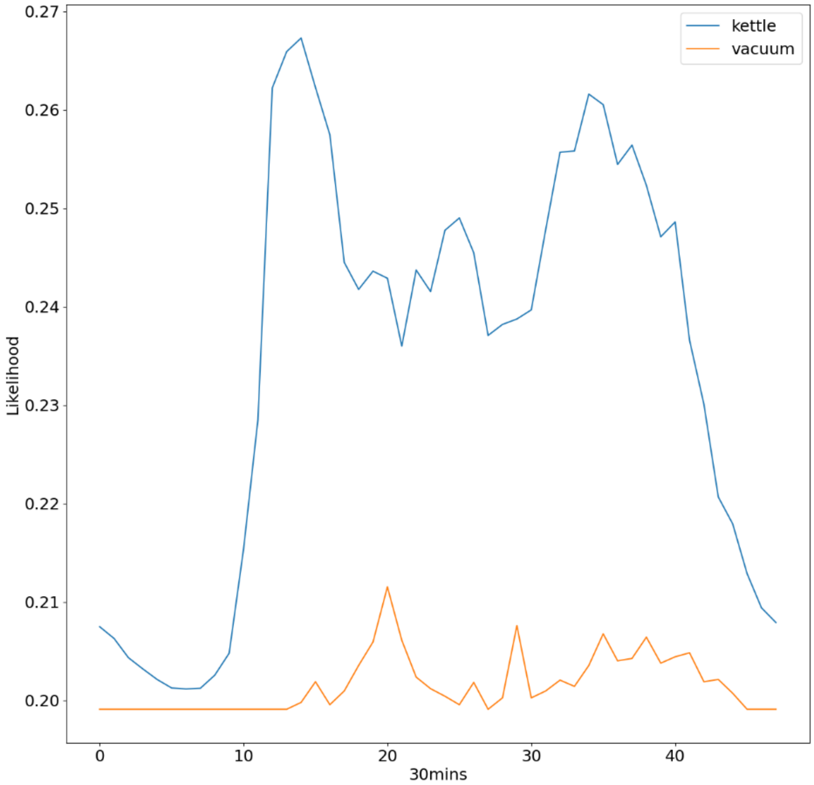

Figure 13: Initial ToD prior for the kettle and vacuum.

Figure 13 shows the result of this calculation for the kettle and vacuum, for each 30-minute window through the day (from 0 to 47)- these are the initial ToD priors. The initial ToD prior for the kettle at the 35th 30-minute window is 0.26, as shown in Figure 12. It can be seen that the initial ToD prior for the kettle is significantly higher than the vacuum. This is because the kettle is used more frequently than the vacuum, and therefore has a consistently higher probability of usage through the day. Additionally, due to the minimum standard deviation imposed, the lowest value reached by the vacuum is 0.20 (20% possibility). This ensures the calculation begins open to the possibility of appliance usage at any time of the day.

It is worth noting that the ToD prior lines are very similar to the ratios presented in the usage statistics (see Figure 6). The top plot in the usage statistics is the ratio of the number of usages in a 30-minute window over the maximum number of usages across all 30-minute windows. This is a proportionality metric and it is used in the usage statistics to compare between appliances. The initial ToD priors presented here are a probabilistic calculation, which ensures the resultant prior is the probability the appliance is used.

4.2.3 Iterative Time of Day Prior

The iterative ToD prior is the generation of a household and appliance specific usage profile, from the initial ToD prior as a starting point. The ToD prior is iterated upon daily and is calculated from the previous ToD and the calculated informed matrix likelihoods. The method is as follows:

- Calculate the informed matrix likelihoods,
- Calculate the probability the appliance is on and off (P_{on} and P_{off}),
- Calculate the relative difference between P_{on} and P_{off},
- Update the q_{dist} and calculate the expected value.

4.2.3.1 Informed Matrix Likelihoods

The informed matrix likelihoods are calculated for each appliance and 30-minute window, by multiplying the matrix likelihoods by the ToD prior for the previous day, detailed in Section 4.2.1.4.

For the first day the analysis is run for, this is the initial ToD prior calculated from the usage statistics. Following that, the iterative ToD prior from the previous day is used.

4.2.3.2 Calculating P_{on} and P_{off}

P_{on} is calculated for each appliance, for each 30-minute window, and corresponds to the total probability that the appliance was turned on in that window. It is calculated by adding together the informed matrix likelihoods where the appliance was turned on. For example, take the case where two appliances are considered for a single 30-minute window, the possible matrix combinations are: [0,0], [0,1], [1,0] and [1,1]. Out of these, the kettle is turned on in the combinations [1,0] and [1,1], and therefore P_{on} is the sum of the matrix likelihoods for these two combinations . P_{off} is the sum of the remaining matrix combinations.

Equations below detail how P_{on} and P_{off} are calculated for one 30-minute window, for the m^{th} appliance, where N is the number of matrix combinations and x and y are the normalisation variables:

$$P_{on} = x \times \sum^{N} (informed\ matrix\ likelihoods\ where\ element\ \{0, m\} = 1)$$

$$P_{off} = y \times \sum^{N} (informed\ matrix\ likelihoods\ where\ element\ \{0, m\} = 0)$$

$$P_{on,} + P_{off} = 1$$

4.2.3.3 Relative Difference in P_{on} and P_{off}

P_{on} and P_{off} are compared to calculate the relative probability of the different values of q . This is calculated by representing P_{on} and P_{off} as a linear distribution and adding them together. The linear distributions are defined so the area underneath is one, and are:

$$D[P_{on}] = 2\ P_{on}\ x_i \qquad D[P_{off}] = 2\ P_{off}(1 - x_i)$$

$$D[relative\ difference] = D[P_{on}] + D[P_{off}]$$

Where x_i are all the possible values of q , between zero and one. The addition of these two distributions is the relative difference of P_{on} and P_{off}. Taking an example where $P_{on} = 0.6$ and $P_{off} = 0.4$, the resultant distributions are shown in Figure 14. The resultant difference in this is a line with positive gradient. If instead, the values of P_{on} and P_{off} were both 0.5, then the relative difference distribution will be a flat line as the appliance is equally likely to off on and off, therefore, all values of q are equally likely.

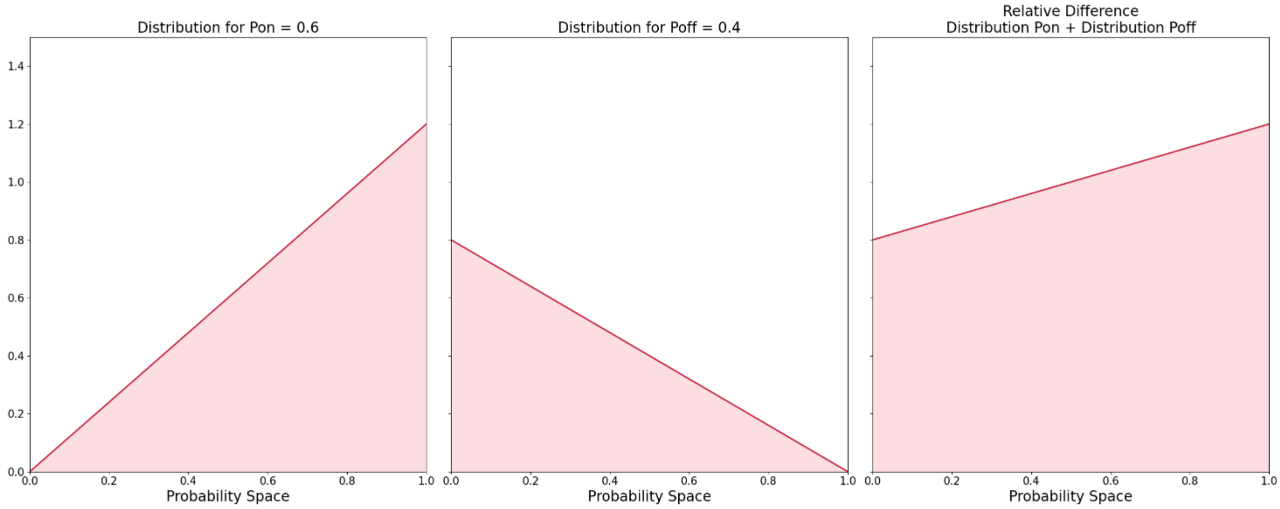

Figure 14: Figures showing the P_{on} and P_{off} distributions where $P_{on} = 0.6$ and $P_{off} = 0.4$. The resultant relative difference distribution is shown in the right figure.

4.2.3.4 Update the q_{dist}

The relative difference distribution is then directly multiplied by the q_{dist} calculated in Section 4.2.2, and renormalised to ensure the area sums to one. This is shown below, where z is the normalisation variable:

$$updated\ q_{dist} = D[relative\ differnce] \times\ q_{dist} \times z$$

Looking at Figure 14, the resultant distribution is a positive gradient. Therefore, when the q_{dist} (a normal distribution) is multiplied by this positively increasing straight line, the peak moves to the right. When the q_{dist} is shifted this way, the expected value, and therefore ToD prior, increases. This new, updated ToD prior, is used when calculating the matrix likelihoods for the next day in the analysis.

Example of the iterative q_{dist}

Continuing with the case presented in Figure 12, the q_{dist} starts as the normal distribution shown by the black line. If we iterate this q_{dist} using a repeated day where the appliance is always turned on in that 30-minute window, then for each day considered $P_{on} > P_{off}$. Therefore, the relative difference distribution will always have a positive incline, the q_{dist} distribution will shift to the right, increasing the ToD prior, and therefore probability the appliance is used in that 30-minute window.

To demonstrate the observed shift, Figure 15 shows the resultant q_{dist} after the relative difference distribution for a day such that $P_{on} > P_{off}$ is repeatedly multiplied by the q_{dist}. In the case presented by Figure 15, the expected value of the q_{dist} , which is the probability that this household uses the kettle in the 35th 30-minute window, increases from 0.26 to 0.41 This change in expected value, and therefore ToD prior for the 35th 30-minute window, means that when the matrix likelihoods are calculated for the next 35th 30-minute window, there will be a higher probability the kettle is used in that window. Therefore, more weight will be given to the matrix combinations where the kettle was used.

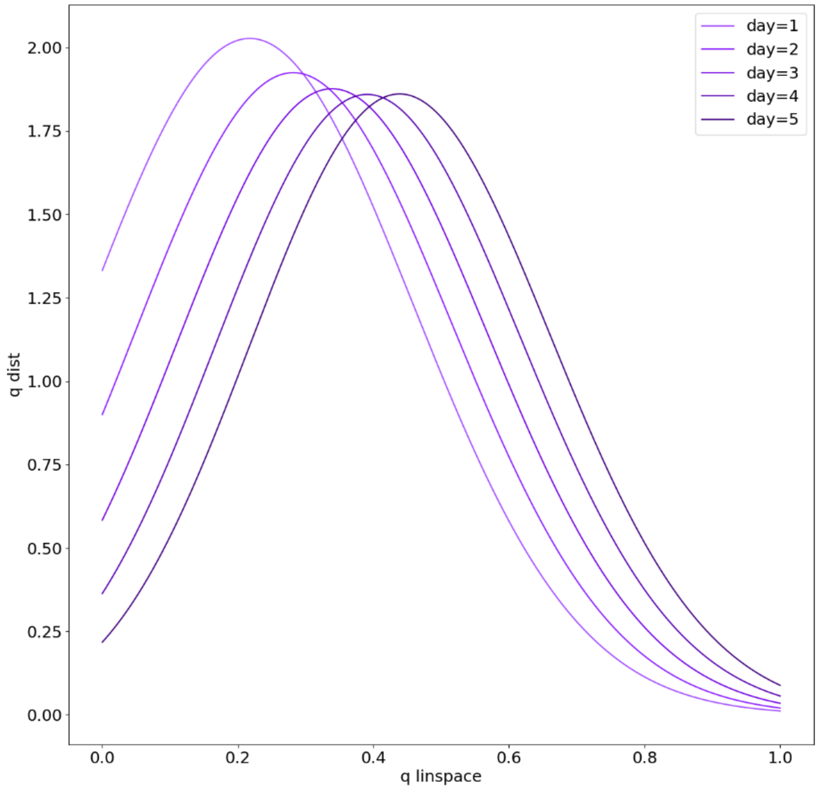

Figure 15: Change in q_{dist} for the kettle in the 35th 30-minute window (17:30 to 18:00), for a case where the kettle is used in that window each day.

With each day the analysis is run for, this effect repeats. The q_{dist} for each 30-minute window will shift depending on P_{on} and P_{off}, and for some days this will increase and for some it will decrease. This updated q_{dist} , and therefore ToD prior (the expected value of the q_{dist}), will impact the informed matrix likelihoods calculation for the following days.

The aim is, after running the analysis for a set number of days, the behavioural patterns of the household will be determined, and the final ToD prior will be representative of the real household usage. For example, if a household uses the kettle every morning before work, the q_{dist} will shift to the right (more likely) for the weekdays, and then shift backwards for the weekend. This affect will continue over multiple months and eventually the q_{dist} will see an overall shift to the right.

This method will not enable us to determine exactly which appliance is being used in each 30-minute window, because humans are erratic! But it will enable us to build a picture of how a household generally uses their appliances. When medical appliances can be included, the initial ToD prior for these will be very low at all times of the day, because a general house is unlikely to have a medical appliance. If the model is run for a household with a medical appliance, the ToD prior should increase, and we could conclude that the household is likely to have a medical device.

4.2.4 Summary of Method

The usage statistics were used to determine an initial ToD prior for each appliance, which is the probability the appliance is used in each 30-minute window for a day. This gives an initial representation of how all the households use each appliance. When the model is run using the smart meter data for a single household, this initial ToD prior is refined for each appliance to become a ToD prior specific for that household. The initial and final ToD priors can be compared to conclude whether these differences could be due to a vulnerability.

The model does this by first generating a distribution for how much power each appliance could draw in a 30-minute window. This distribution accounts for different powers drawn by an appliance when in use, the different usage lengths, and the different start times within the 30-minte window which could result in all, or some, of the usage being observed. These distributions are also generated for the power that could be drawn by the appliance in the current 30-minute window if it were turned on in the previous window and remained on.

Following that, every combination of appliance usage that could have made up the observed power is determined. Using the power distributions, the likelihood of each of these combinations is calculated. For example, take a household where only two appliances are considered and the power they draw is considered only for the current window. In any 30-minte window the only things that could have happened to make up the observed power are: neither appliance was turned on, both appliances were turned on, or just one of them was. If the observed power were relatively high, the likelihood that both appliances were turned on would be the highest.

The model then uses the smart meter data from a household for each 30-mintue window for day at a time. It calculates the likelihood of each appliance combination for each 30-minute window, and this is multiplied by the prior knowledge of the probability each appliance is used in each window. This results in the final likelihood for each appliance usage in each 30-minute window for the day. When the next day is considered, the model uses the previous day as known historic behaviour and updates the ToD prior based on the likely behaviour from the previous day. This is repeated for all the days for a household and a final ToD prior for each appliance is determined.

This model was developed and tested for the appliances: kettle, toaster, microwave, and vacuum. The total power drawn from just these appliances was extracted from one of the houses in the high resolution datasets and modified to look like smart meter data. This meant the only power drawn was from the appliances being considered, but the 30-minute windows when the appliance was used knowns, so this could be used to test the model.

4.3 Testing

4.3.1 Method

This analysis described in Section 4.2 can be tested using the real household data (from UKEDC and REFIT) for the down-selected appliances. The testing approach is outlined below:

1. Choose the appliance categories to run the analysis for (see Table 4) and determine how many 30-minute periods to consider in the matrix combinations (see Section 4.2.1). For appliances with long usage periods (i.e. washing machine), you must consider if the appliance was turned on in the previous 30-minute window, and so on, as it will impact the power drawn in the current window. For appliances with short usages, this consideration may not be necessary.

2. Choose which categories will be included in the baseload (i.e. small bedroom appliances). The power drawn by all these appliances in the real household will be included in the analysis to represent the constant baseload present in a household. This should be a continuous low level, so it does not interfere with the appliance analysis.

3. Randomly select a single day from a single house where all appliance categories were used. This should be a random day chosen from the UKEDC or REFIT datasets, and preferably a day where the appliances were used multiple times.

4. For each appliance category:

    a. From the real household data for the randomly selected day, sum the power drawn by the appliances and baseload for the day and average this power into 30-minute periods so it represents the smart meter data input.

    b. Determine which 30-minute periods the appliances are turned on from the usage statistics generated in Section 3.5. These are the source of truth for the testing.

c. Run the analysis from Section 4.2 for this single day multiple times.

The model has two key outputs: the final ToD prior and the informed matrix likelihoods for a day. By running the model for the same day repeatedly, we can test both outputs:

Final ToD Prior: Review how the ToD prior changes for each 30-minute window, from the initial ToD prior, to the household specific ToD prior from this repeated day. If the model is working as anticipated, the probability that each appliance is used (the ToD prior) will trend towards one when the appliance is known to be turned on, and towards zero when it is not.

Informed matrix likelihoods: Review the impact this changed ToD prior has on the calculated informed matrix likelihoods for each 30-minte window. If the model is working as anticipated, the probability the appliance is turned on (calculated from the informed matrix likelihoods) for each 30-minute window increase when the appliances are known to be turned on and decrease otherwise.

4.3.2 Appliance and Day Selection

The method outlined above was run for the *kettle* and *vacuum* categories, with *continuous* and *small bedroom appliances* forming the baseload (see Table 4). The kettle was chosen for testing as it is the appliance with the most usages across all the households, but it has a very changeable power depending on the specific kettle type. By testing the model with this appliance, we can identify whether this uncertainty is a limiting factor. The vacuum was chosen for the opposite reason: it is an appliance we do not know much about and therefore we can confirm that the distributions are not overfitting to the small amount of historic information we have. The two baseload categories were chosen to add only a low level power so the household usage was never zero, but the baseload is low enough to not interfere with the appliance disaggregation predictions.

The random day chosen for testing was 13th April 2013 from UKEDC House 1, where the kettle and vacuum were used in the 30-minute windows shown in Table 5.

Table 5: The 30-minute window numbers (from 0 to 47) where the kettle and vacuum were used for UKEDC House 1 on the 13th of April 2013.

Appliance	30-minute window the appliance was known to be turned on[8]
Kettle	15, 24, 30, 32, 33
Vacuum	34, 35

Figure 16 shows the power drawn by each appliance category included in the analysis averaged over 30-minute windows, for the day selected for the testing. The total power drawn is the sum of the power drawn by the selected appliance categories (kettle and vacuum) and the baseload appliances. The total power is used as the model input, and the day was iterated ten times for the testing.

[8] The 30-minute window 00:00 to 00:30 is the 0th window and 23:30 to 00:00 is the 47th.

© FNC 2022

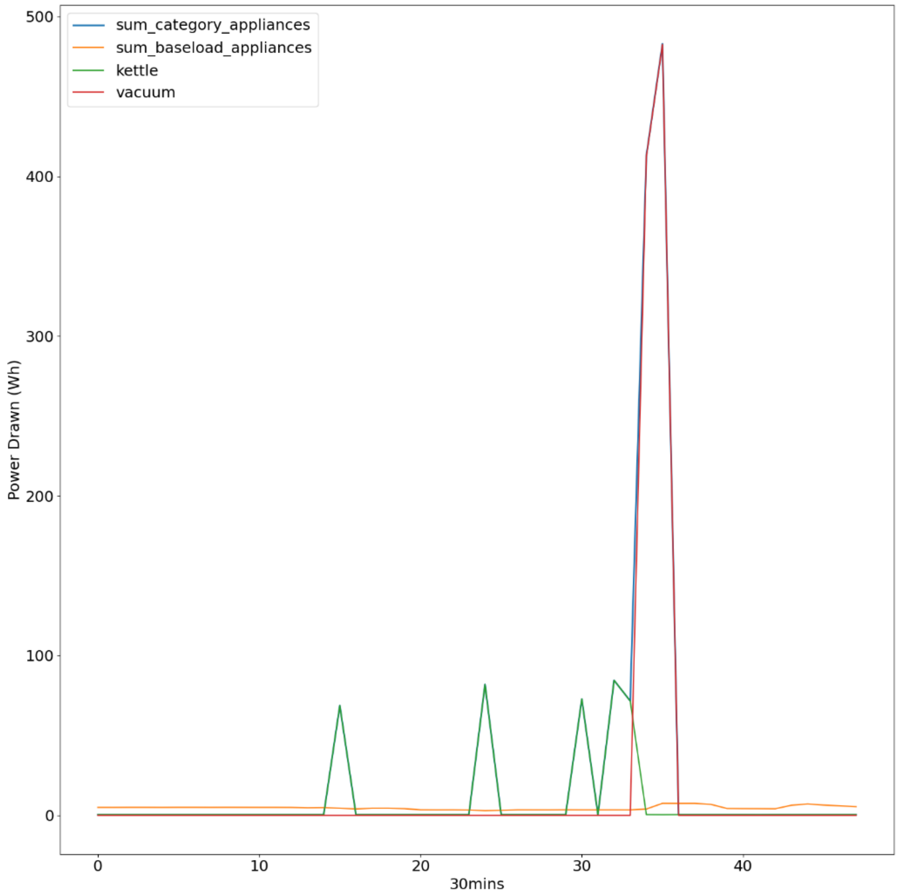

Figure 16: The power drawn by the selected appliance categories (kettle and vacuum) and the baseload categories (continuous and small bedroom appliances) for the UKEDC House 1 on the 13th of April 2013.

4.3.3 Results

4.3.3.1 Final ToD Prior

The aim of this test was to assess if the ToD prior changes as expected when it is provided with a given half an hour of usage data. To assess this efficiently, we supply the ToD prior with the same half an hour of data repeatedly. If the ToD prior appears to converge to the correct value (zero if it was not used in the given half an hour; one if it was), we can be satisfied that the model is able to determine the specific behaviour of a household.

Given the 14th and 15th 30-minute windows for the kettle, we would expect the peak of the q_{dist} to shift to zero and to one, respectively, since it was turned on in window 15 but not 14. Figure 17 and Figure 18 show how the q_{dist} for the kettle changes for the 14th and 15th 30-minute window. In both cases, the expected value shifts as anticipated. It tends towards a value of one in the 15th window when the kettle was known to be used, and to a value of zero in the 14th window when it was not.

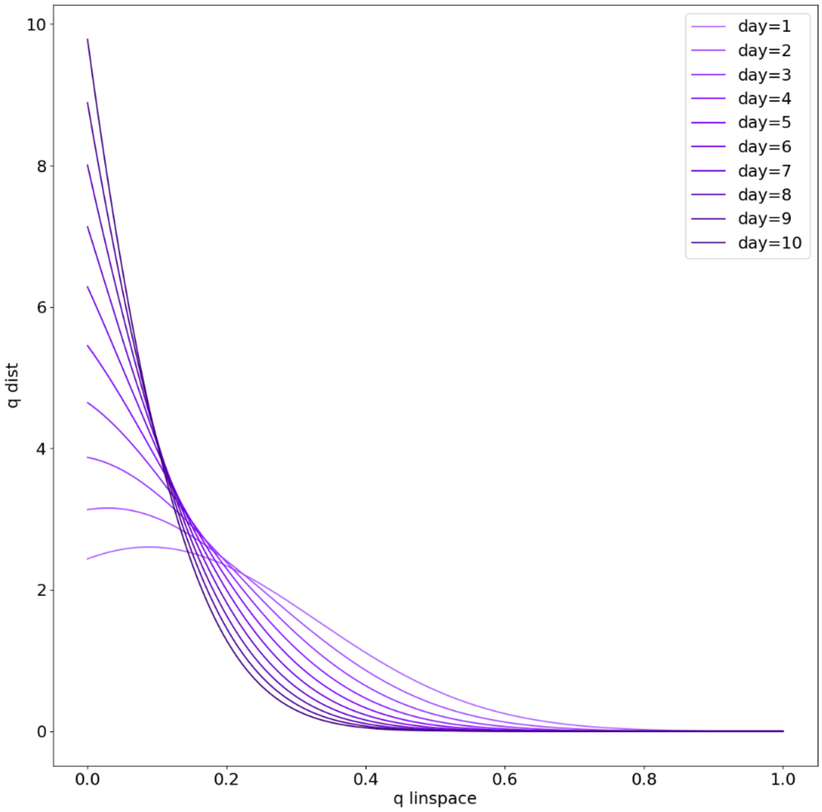

Figure 17: Change in kettle q_{dist} for the 14th 30-minute window when the model is run for the same day repeatedly, where the kettle is not turned on in this window.

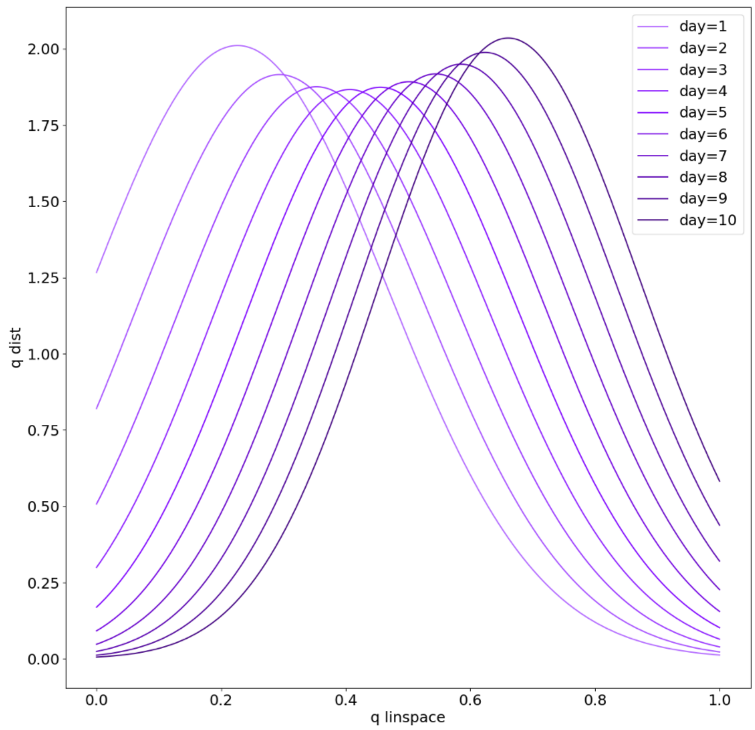

Figure 18: Change in kettle q_{dist} for the 15th 30-minute window when the model is run for the same day repeatedly, where the kettle is turned on in this window.

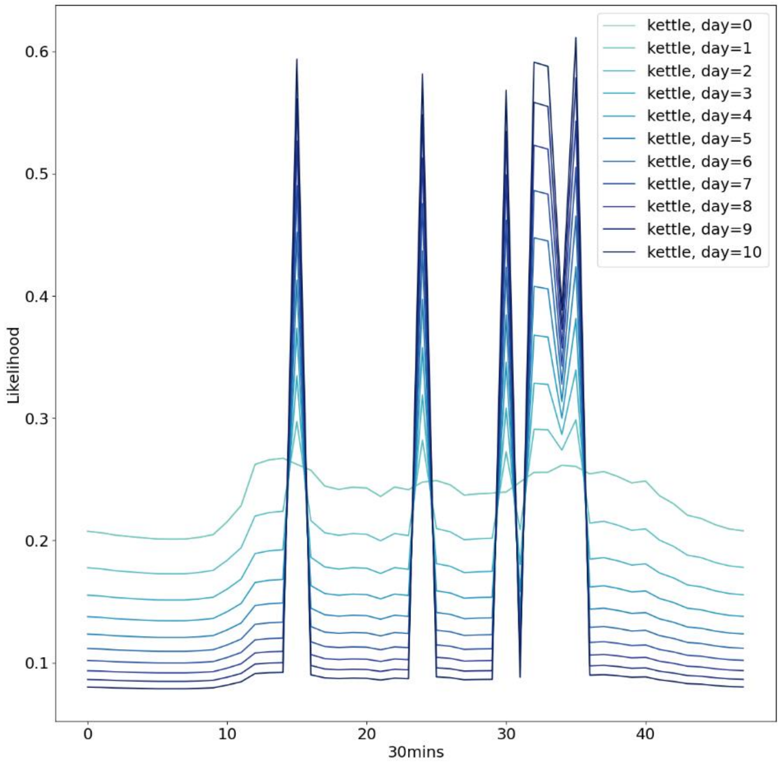

Figure 19: Change in the ToD prior for the kettle when the same day is run through the analysis ten times (day zero is the initial ToD prior and day ten is the ToD prior learnt after ten days of household usage). The kettle was turned on in 30-minute windows 15, 24, 30, 32 and 33.

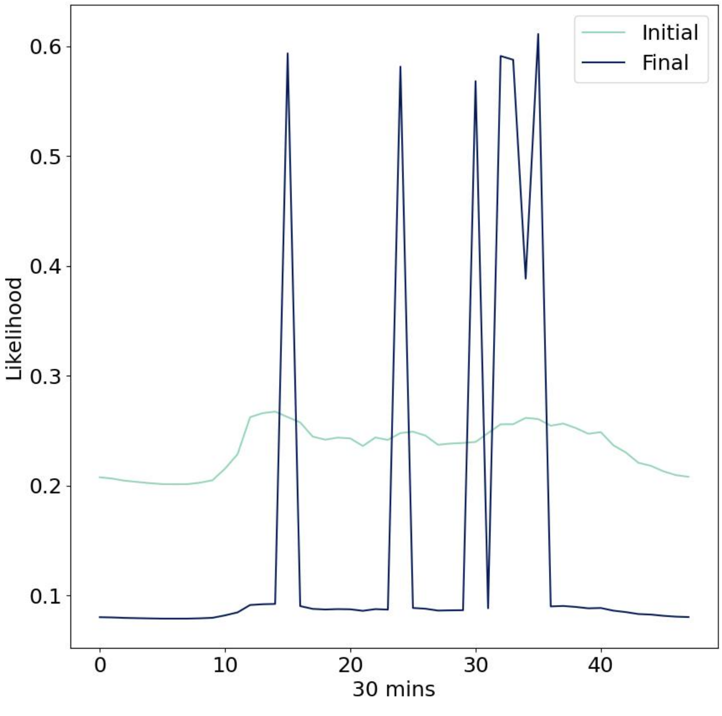

Figure 20: The initial ToD prior for the kettle calculated from the usage statistics alongside the final ToD prior learnt by the model when the same day is repeated. The kettle was turned in on 30-minute windows 15, 24, 30, 32 and 33.

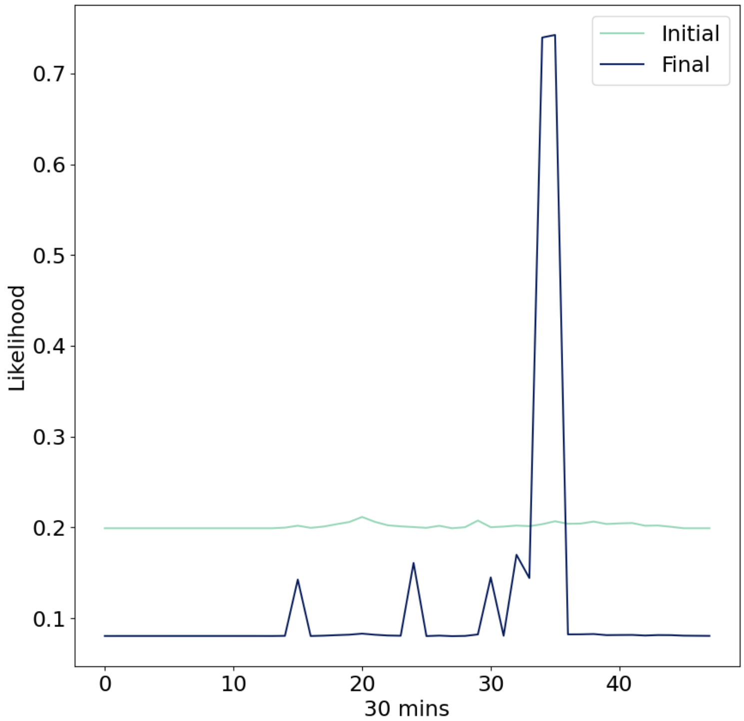

Figure 21: The initial ToD prior for the vacuum calculated from the usage statistics alongside the final ToD prior learnt by the model when the same day is repeated. The vacuum was turned in on 30-minute windows 34 and 35.

Figure 19 shows the resultant change in the ToD prior (the expected value of the q_{dist}) for each time the model was repeatedly given the same day of power data to learn. Day 0 is the initial ToD prior calculated from the usage statistics (see Section 3.5) and each subsequent day is the ToD prior after running the model for the same day. Figure 20 shows the same results but just the comparison of the initial and final ToD priors. Figure 21 shows the same results as in Figure 20 for the vacuum.

For both the kettle and vacuum, it can clearly be seen that the ToD prior shifts towards zero when the appliance is not turned on, and the ToD prior increases towards one when each appliance was turned on. This shows that the ToD priors are changing as anticipated and, therefore, the test has passed.

However, there are two interesting observations:

- There is an increase of ToD prior when the kettle is not used in the 30-minute window 34 and 35 (windows when the vacuum was used).
- The decrease in ToD prior for the vacuum in the windows the kettle is used (windows 15, 24, 30, 32 and 33), is less than the windows the kettle is not used.

Both observations show that the appliances are impacting the results of each other. This can be explained by reviewing how the q_{dist} is calculated: the probability that an appliance is turned on is calculated as the sum of all the matrix likelihoods that correspond to a matrix combination where that appliance was posed as on, which includes the case where both appliances considered were on ([1, 1]). In the case of the vacuum, in windows 15, 24, 30 and 32, there is a proportionally high likelihood that both appliances were turned on in that window, due to the high-power usage by the kettle. This therefore impacts the probability that the vacuum was turned on in those windows, and as a result the vacuum ToD prior does not decrease as much as expected.

Taking a closer look at the usage statistics for the kettle and vacuum for this day, Table 6 shows the duration and power for each usage of the kettle and vacuum for the day. This shows that in both the windows 34 and 35[9], the vacuum was used for approximately nine minutes, which is expected behaviour for a vacuum, but also for a little over 2 minutes. This short usage looks very similar to that of the kettle. The analysis only considers each appliance being turned on once in each window, so once the model has calculated a high probability the vacuum is turned on (accounting for the longer usage), then the only other appliance it can equate the extra power to is the kettle. Therefore, even though the ToD prior has moved in the wrong direction, the model is working exactly as expected.

Table 6: Kettle and vacuum usage details for 13th April 2013 from UKEDC House 1

Appliance	Usage Start Time	Usage Duration	Average Power (W)
Vacuum	17:10	9 mins 20 secs	1,913
	17:20	2 mins 10 secs	1,904
	17:30	2 mins 40 secs	1,913
	17:39	9 mins 50 secs	1,841
Kettle	07:31	2 mins	1,710
	12:18	1 mins 30 secs	2,344
	15:13	1 mins 40 secs	2,366
	16:13	1 mins 50 secs	2,328
	16:48	1 mins 30 secs	2,328

4.3.3.2 Informed Matrix Likelihoods

The aim of this test is to see whether the change from initial to the final ToD prior improves the model's prediction of when each appliance is used.

The probability each appliance is turned on (P_{on}) is calculated using the informed matrix likelihoods (as detailed in Section 4.2.1.4) for the final day the analysis was run for. The probability the kettle is turned on in a 30-minute window is the sum of the informed matrix likelihoods for the matrix combinations [1, 0] and [1, 1].

Figure 22 shows the calculated P_{on} for each 30-minute window throughout the day, where the darker squares show a greater probability. When comparing the 30-minute windows of the darkest squares of each appliance in Figure 22 to the windows the appliances were truly turned on in Table 5, they match-up almost exactly, which is very promising.

[9] Window 34 is 17:00:00 to 17:29:29 and window 35 is 17.30:00 to 18:00:00.

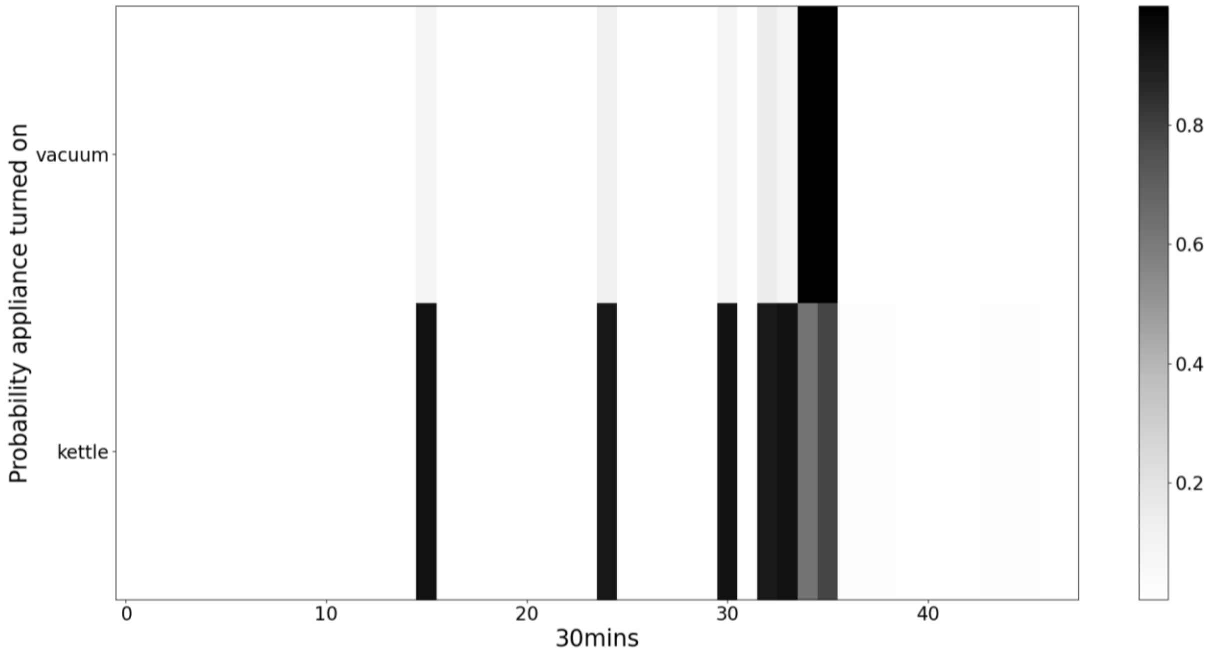

Figure 22: Square plot showing the probability each appliance was turned on for each 30-minute window for the last day the model was run for. The darker the square, the higher the probability.

For the 30-minute windows that the kettle is turned on, Figure 22 also indicates that there is some probability the vacuum was also turned on, and visa-versa for the vacuum in windows. 34 and 35. Table 7 shows the percentage probabilities, calculated from the informed matrix likelihoods[10], for the matrix combinations for the 30-minute windows where the appliances were used. The joint probabilities for the 34th and 35th windows can clearly be seen as the highest likelihood. This is the same behaviour as discussed towards the end of Section 4.3.3.1.

Therefore, from the results presented in Figure 22 and Table 7, we can conclude that the model is working as anticipated for this simplified case, and we can now review the results with real household data.

Table 7: The percent probability for each matrix combination for the windows where appliances we used for the last day. The percent probabilities are calculated from the informed matrix likelihoods. The values in bold show the actual combination.

Matrix Combination	Percent probability per 30-minute window						
	15th	24th	30th	32nd	33rd	34th	35th
0,0	0	0	0	0	0	0	0
0,1	6	7	6	8	6	**40**	**25**
1,0	**91**	**88**	**91**	**87**	**91**	0	0
1,1	3	5	3	5	3	60	75
Total	**100**	**100**	**100**	**100**	**100**	**100**	**100**

[10] Percentage probabilities = informed matrix likelihoods * 100

4.4 Model Results

4.4.1 Method

To obtain the results for a household, the same approach as detailed in Section 4.3.1 is followed, with the exception that the model is run with 30 days of usage data for UKEDC House 1 from the slightly earlier date of 8th April 2013[11], until the 8th May 2013.

The model was run for appliances: kettle, microwave, and toaster. Compared to testing, the vacuum was removed, and the microwave and toaster added. This is because the vacuum is not used frequently in UKEDC House 1, and whilst this feature made it a perfect testing appliance, the actual results will not give much information. By instead using the microwave and toaster, we can see the results for frequently used appliances in the household. The same baseload appliances were used as in Section 4.3 and the power drawn from the appliances was only considered for only one 30-minute window, resulting in eight matrix combinations.

4.4.1.1 Microwave and Toaster Usage Statistics

To understand the usages of the microwave and toaster, Figure 23 and Figure 24 show the usage statistics for each, respectively. The top figure shows the ratio of the number of usages per 30-minute period divided by the maximum number of usages in any 30-minute window. The middle figure shows the average power drawn by each appliance when in use, and the bottom figure shows the how long each appliance is in use when turned on[12]. The usage statistics for the microwave show that across all the households, the power drawn does not vary much- the peaks are in the range 1000W to 1600W whereas the kettle (Figure 6) peaks are between 1600W and 2800W. However, the length of usage varies quite significantly. Conversely, the usage statistics for the toaster show that the power drawn varies significantly across houses, but the length of usage is more consistent. This correlates with inherent knowledge about each appliance: microwaves are consistent but used for different lengths depending on what is being microwaved; whereas a toaster could have two or eight bread slots but is generally used for the same length of time.

[11] The start date has moved so the final day is one where appliances are used so the results can be reviewed.
[12] Note that the REFIT dataset is in 30-second intervals whereas the UKEDC dataset is in 10-second intervals. Therefore, the bottom plots in Figure 23 and Figure 24 look 'peaky'.

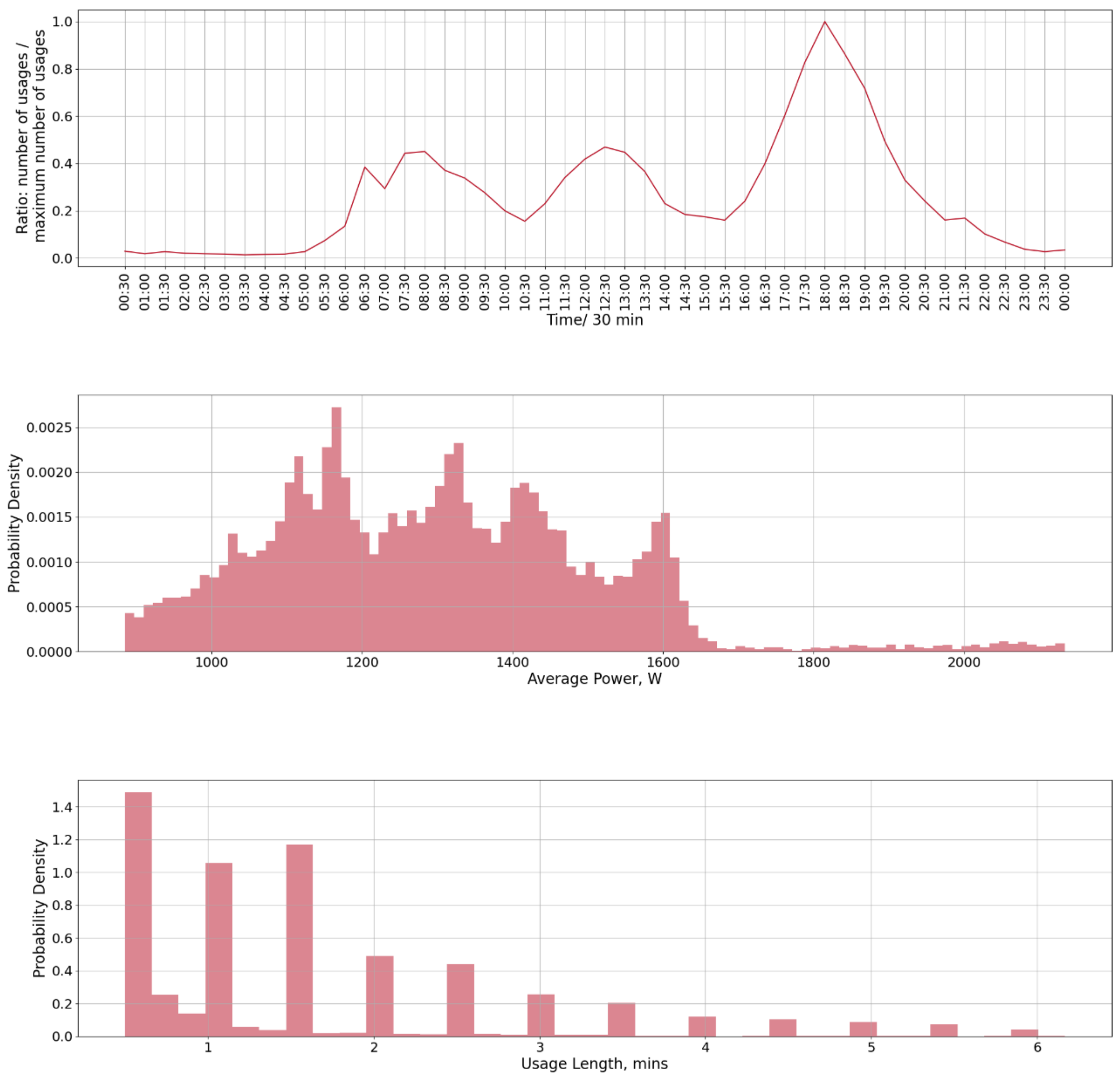

Figure 23: Usage statistics for the microwave across all datasets.

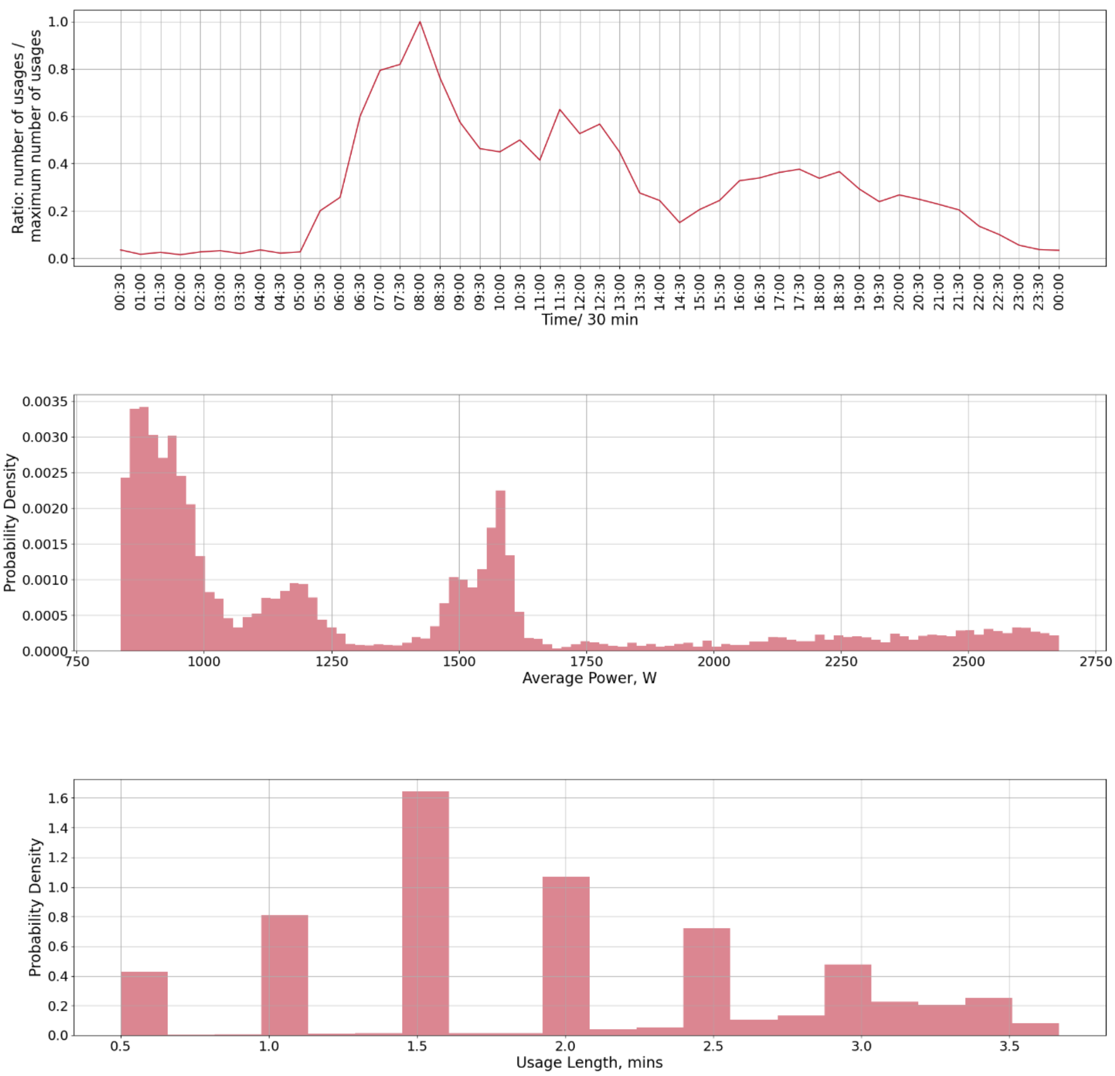

Figure 24: Usage statistics for the toaster across all datasets

4.4.1.2 Matrix Likelihoods

Figure 25 shows the gaussian kernel-density estimations for each matrix combination (see Section 4.2.1.1). There are nine combinations as each of the three appliances are only considered for one 30-minute window. Looking at the lines for [0,1,0] and [0,0,1], which correspond to just the microwave and just the toaster, respectively, the impact of the usage statistics can clearly be seen- they are very close together. This means that if you read off the Figure 25 the matrix likelihood for an observed usage of 50Wh, there is almost an equal probability that the microwave or toaster is used. This is very interesting because the ToD priors will have a very large impact on the relative probability between the microwave and toaster in the results.

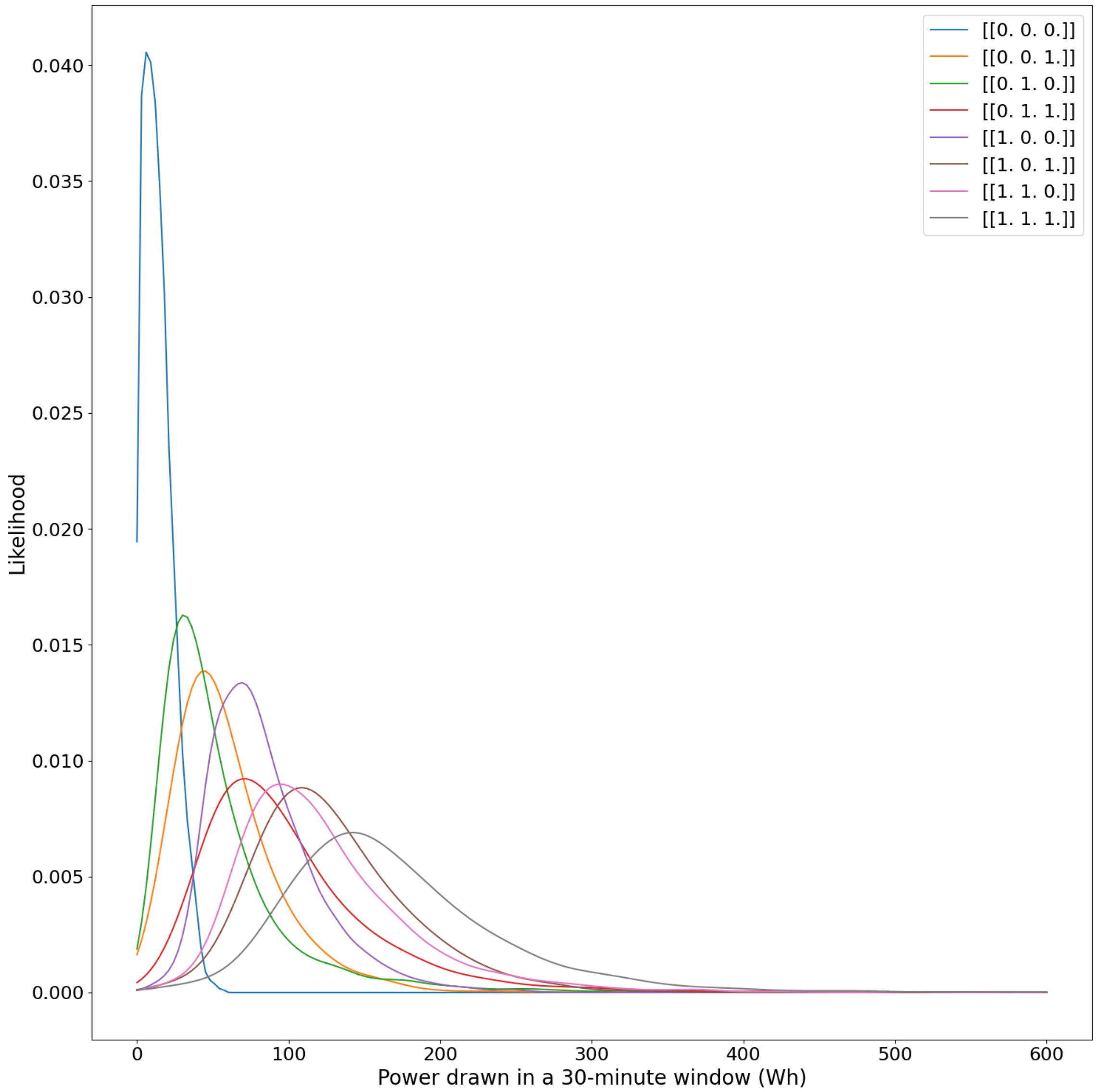

Figure 25: Gaussian kernel-density estimations for the matrix combinations for: kettle, microwave, and toaster, considered for only one 30-minute window.

4.4.2 Results

To analyse the results of the testing, we investigate two outputs:

Final ToD Prior: Review how the ToD prior changes from the initial assumption to the refined version at the end of the month. Whilst, as with the testing, there is no specific source of truth for this output we would expect it to roughly trend towards when the household has used the appliances in the month.

Informed matrix likelihoods: Look at the probability each appliance is on for the final day and compare this to the known reality. This shows whether the changing ToD prior has positively impacted the model's ability to determine which appliances are being used.

4.4.2.1 Final ToD Prior

The model was run for 30 days and the change in ToD prior from the initial prior derived from the usage statistics, to the find ToD prior determined by the model can be compared. We would expect the ToD prior to tend towards the times the appliance was used by the household. To do this, we can use the usage statistics to determine a real ToD prior. The real ToD prior is calculated by counting the number of times the appliance was used by the household in the time range, and normalising by the number of days the appliance could have been used (which in this case is 30 days). This is comparative to the method used to calculate the initial ToD prior, but with no uncertainty in the 30-minute usage (so the q_{dists} are not required).

Figure 26, Figure 28 and Figure 30 show the number of times UKEDC House 1 used the kettle, microwave, and toaster, respectively, for the 30 days from 8th April 2013, in each 30-minute window. These values are normalised to create the real ToD prior. Figure 27, Figure 29 and Figure 31 show the initial ToD prior (calculated from the usage statistics), the real ToD prior calculated from the actual usage counts, and the final ToD prior determined by the model after 30 days of usage data for the kettle, microwave and toaster, respectively.

Figure 26 and Figure 27 show the results for the kettle. The first observation is how well the ToD prior has dropped towards zero when the appliance was not used in the first nine 30-minute windows. The next observation is how well the peaks have remained at a relatively high usage probability and are aligned with the real ToD prior, especially when considering how flat the initial ToD prior is. This shows that the model is very effectively learning the behaviour of the household, almost independent of the initial ToD prior location.

Figure 28 and Figure 29 show the results for the microwave and Figure 30 and Figure 31 show the results for the toaster. The observations for both appliances are very similar to the kettle: the ToD prior decreases when the appliances have not been used at all. The final priors match the distinct peaks in the real usage reasonably well, but not as well as the kettle. This is likely due to the high variability in the power a microwave and toaster could be observed to draw in a 30-minute window. This is interesting because it shows that the model is working better for the kettle than the microwave and toaster. This is likely to be due to the higher power drawn by the kettle, and the similarity in the expected half-hourly usages of the microwave and toaster.

Looking specifically at Figure 28, it can be seen that the microwave was used 29 times in the 35th 30-minute window, which is nearly every day the model was run. We would therefore expect the model to pick this behaviour pattern up as it is a very regular usage pattern, but the final ToD prior is not as high as expected. To investigate this further, the analysis was run with just the microwave and the results showed this peak at the 35th 30-minute window increased significantly. Therefore, we can conclude that the small peak seen in Figure 29 is due to the uncertainty between the toaster and microwave: there is not enough difference in the probabilities to significantly impact the ToD prior.

The final ToD priors for the kettle, microwave and toaster show that the model is working as anticipated but there are some improvements that could be made, as it does not have enough information to correctly distinguish between the microwave and toaster. To improve upon this in the future, the model could be developed further to also learn how long the appliances are generally used for and how much power they draw, for the specific household. With this information, the likelihood gaussian kernel-density estimation curves would become narrower, as the powers specifically seen by that appliance are more certain.

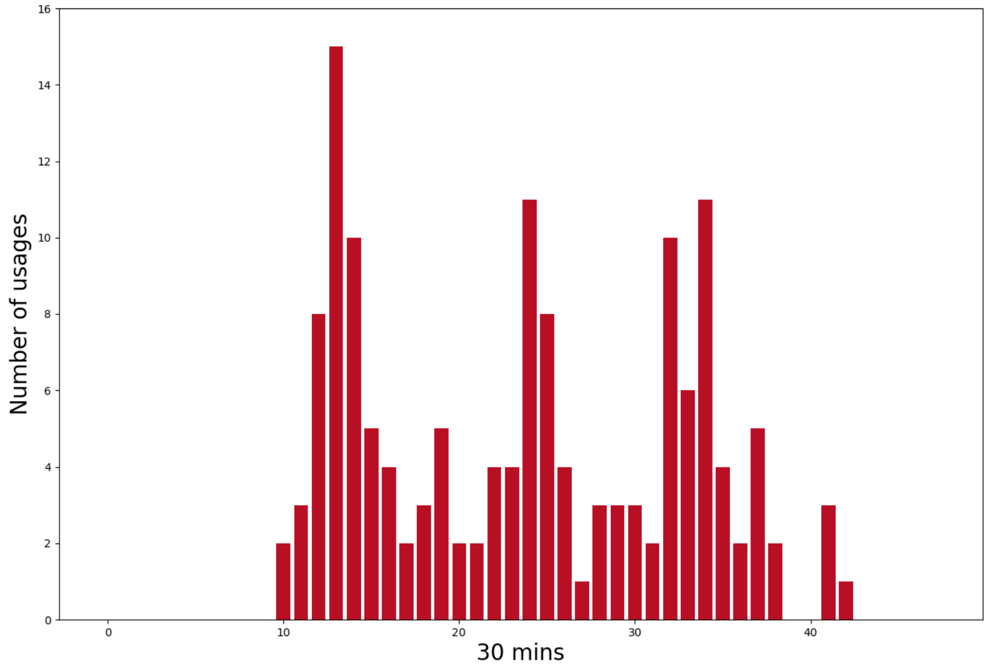

Figure 26: Histogram showing number of times the kettle was turned on in each 30-minute window by UKEDC House 1 for 30 days from the 8th April 2013.

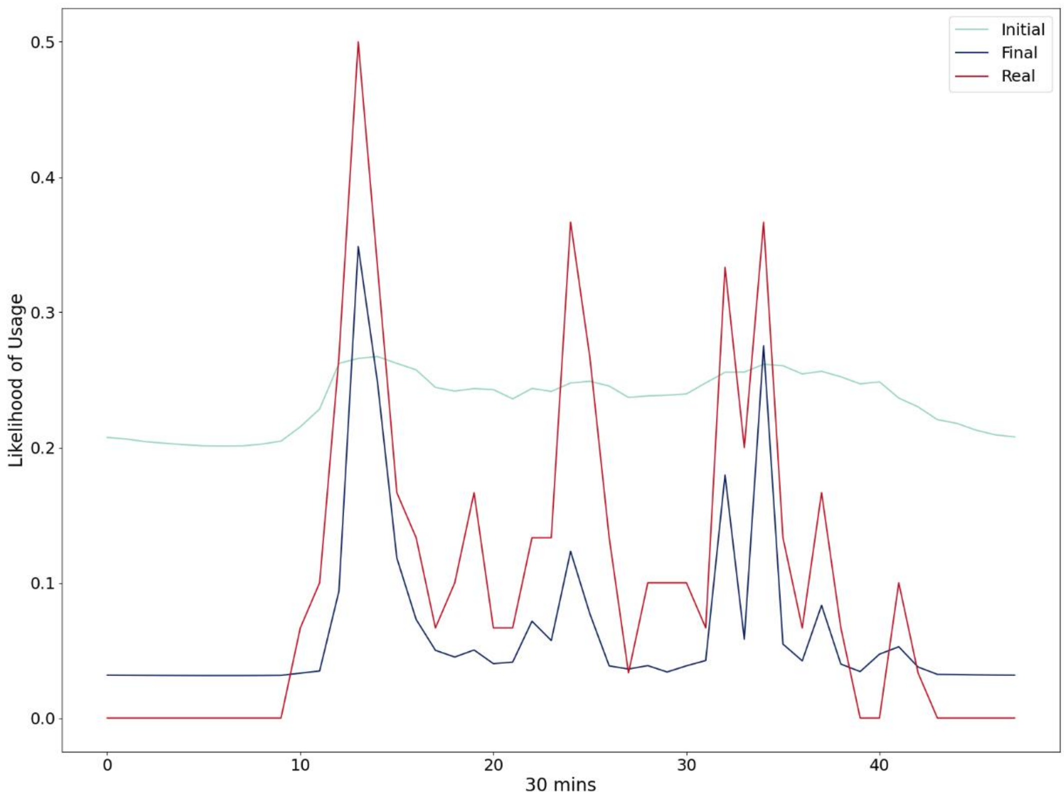

Figure 27: The initial ToD prior for the kettle, calculated from the usage statistics, compared to the final ToD prior for the test household after running the model for 30 days. The real line is the normalised real usage counts.

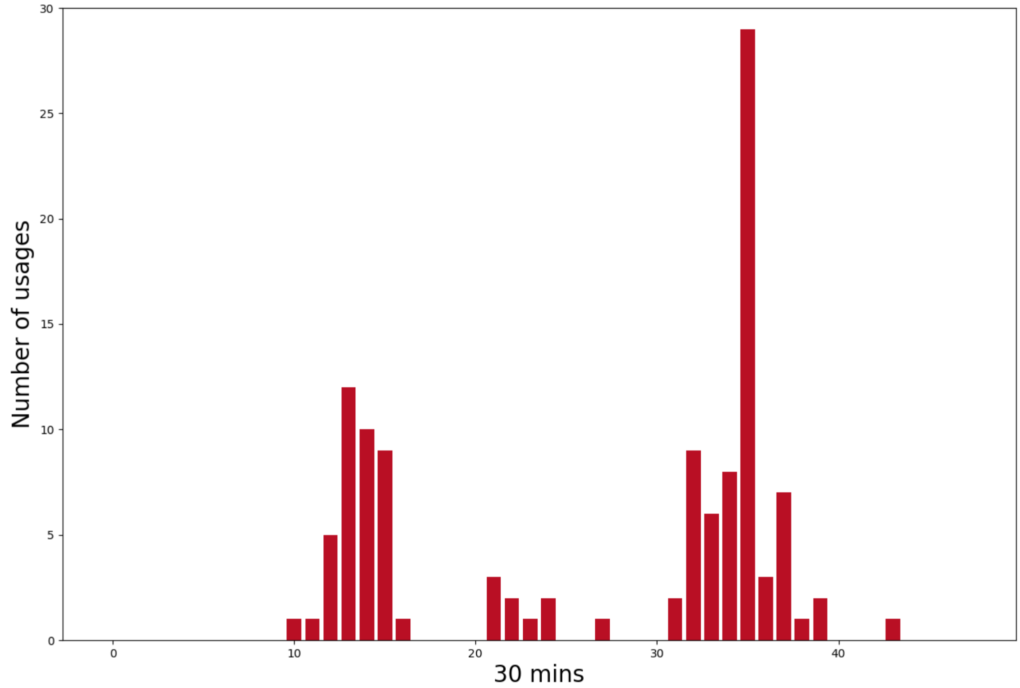

Figure 28: Histogram showing number of times the microwave was turned on in each 30-minute window by UKEDC House 1 for 30 days from the 8th April 2013.

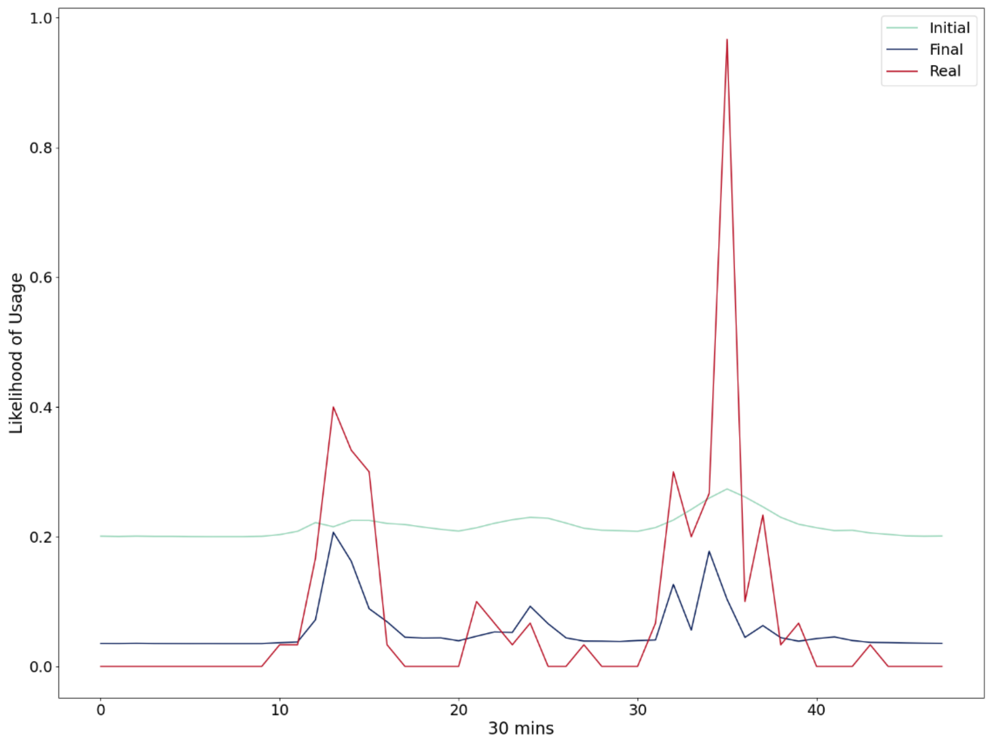

Figure 29: The initial ToD prior for the microwave, calculated from the usage statistics, compared to the final ToD prior for the test household after running the model for 30 days. The real line is the normalised real usage counts.

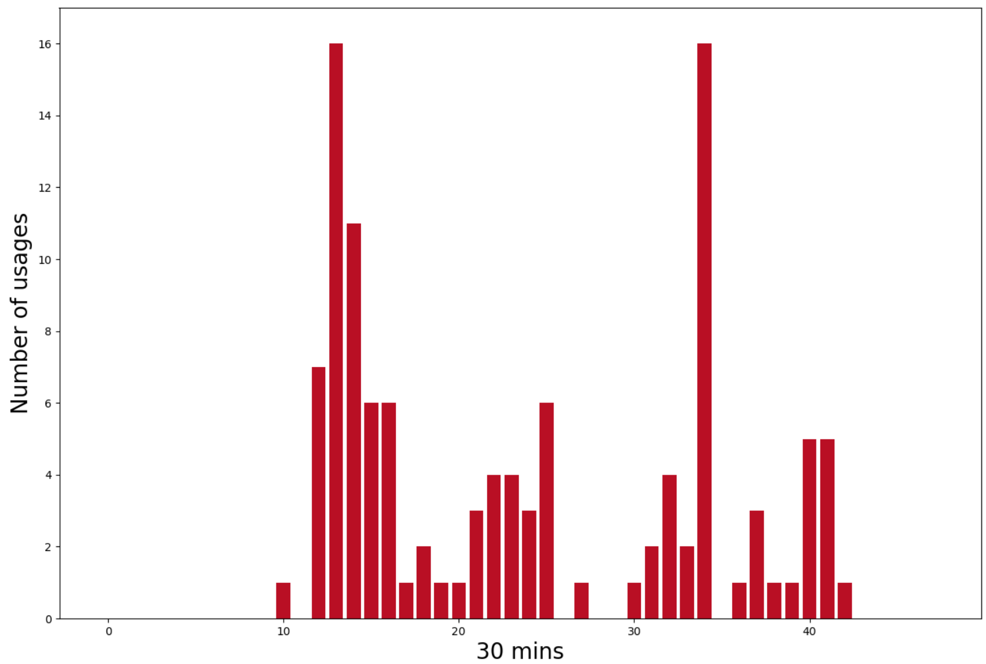

Figure 30: Histogram showing number of times the toaster was turned on in each 30-minute window by UKEDC House 1 for 30 days from the 8th April 2013.

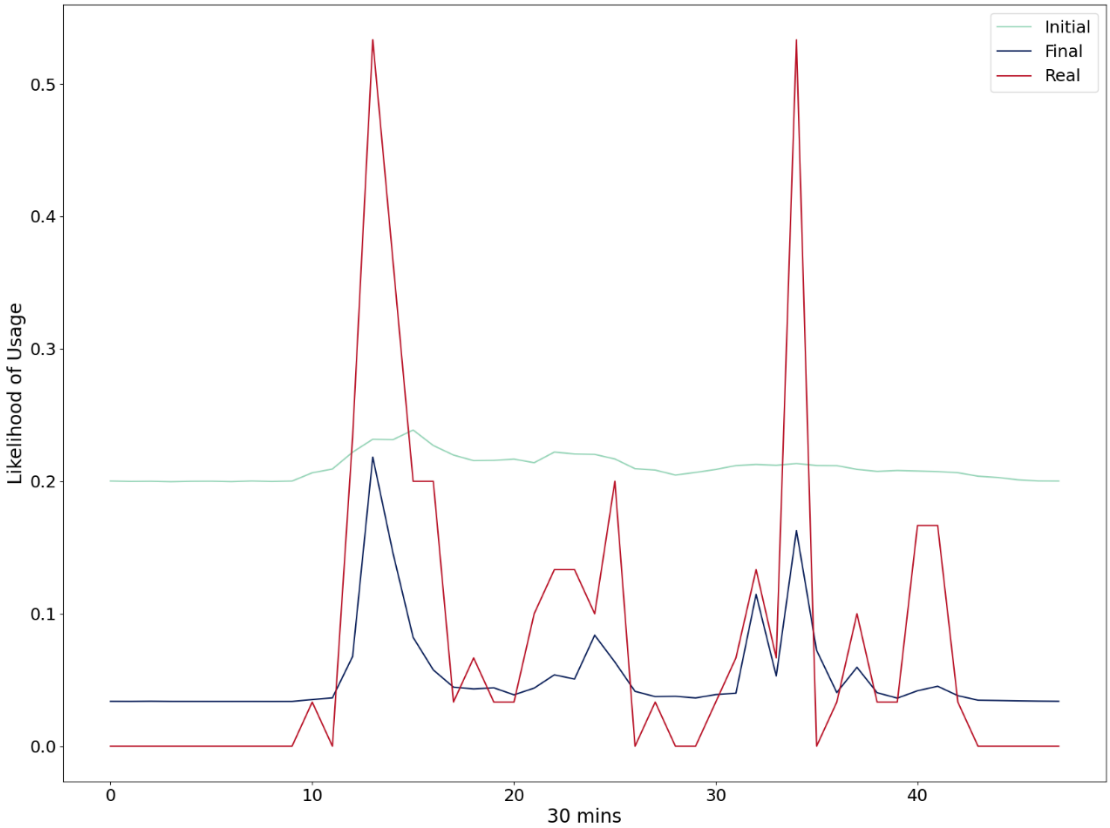

Figure 31: The initial ToD prior for the toaster, calculated from the usage statistics, compared to the final ToD prior for the test household after running the model for 30 days. The real line is the normalised real usage counts.

4.4.2.2 Informed Matrix Likelihoods

To determine if the change in ToD prior for each appliance has an impact on the model's prediction of which appliances are used, we can review the predicted probabilities each appliance is turned on during the 30th day. Table 8 shows the usage details for each of the appliances on that day. Interestingly, all three are used in the 10th 30-minute window, two are used in the 25th and just the kettle is used in the 37th.

Table 8: Kettle, microwave, and toaster usage details for 8th May 2013 from UKEDC House 1.

Appliance	Usage Start Time	30-minute window	Usage Duration	Average Power (W)
Kettle	05:24	10	1 min 50 secs	2,354
	12:32	25	1 min 20 secs	2,351
	18:46	37	1 min 40 secs	2,296
Microwave	05:21	10	1 min 10 secs	1,498
Toaster	05:24	10	4 mins 10 secs	1,499
	12:30	25	3 mins 30 secs	1,593

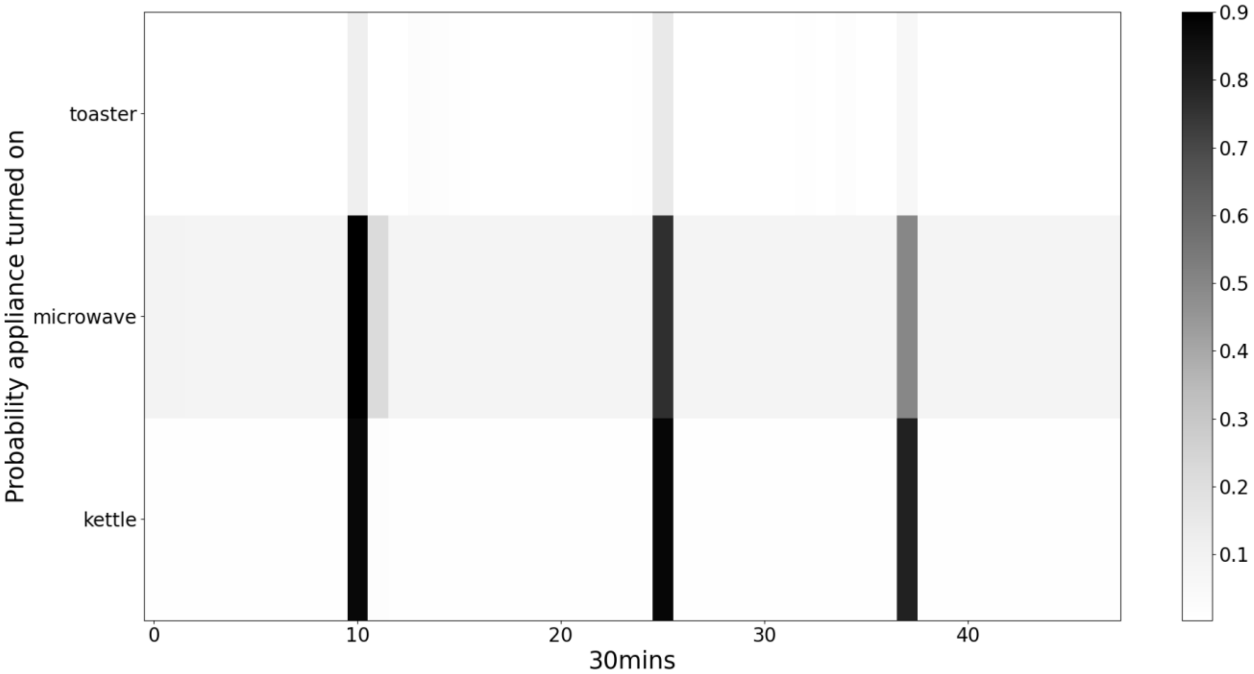

Figure 32: Square plot showing the probability each appliance was turned on for each 30-minute window for the last day the model was run for. P_{on} is calculated using the matrix likelihoods and the final ToD prior. The darker the square, the higher the probability.

Figure 32 shows the calculated probability that each appliance was turned on for each 30-minute window for the 30th day of usage. The darker the box, the higher the probability the appliance was turned on. The figure shows that the model has predicted the appliance usages well for the kettle, and less well for the microwave and toaster. Whilst we would not expect this to be exactly correct as it is only representative on a single day and the model is built to determine overall behaviours, we would hope for a slightly better result with the microwave and toaster.

Table 9: The percent probability for each matrix combination for the windows where appliances we used in the 8th May 2013 by UKEDC House 1. The percent probabilities are calculated from the informed matrix likelihoods. The values in bold show the actual combination.

Matrix Combination	Percent probability per 30-minute window		
	10th window	25th window	37th window
0,0,0	0	0	0
0,0,1	0	1	2
0,1,0	14	9	19
1,0,0	10	18	**45**
0,1,1	1	2	2
1,0,1	2	**5**	1
1,1,0	66	59	30
1,1,1	**6**	7	1
Total	**100**	**100**	**100**

Table 9 shows the percentage probabilities, calculated from the informed matrix likelihoods[13], for the matrix combinations for the 30-minute windows where the appliances were used. Looking at the 10th window, all appliances were used but most likely combination is that the kettle and microwave were used. Looking at the usage details for the 10th window (Table 8), the microwave was used for a relatively short time and the toaster was used for a long time. As the ToD prior for the microwave is higher than the toaster at this time (0.036 for the microwave and 0.034 for the toaster), the model has attributed the usage to the microwave rather than the kettle. This is not an incorrect conclusion for the model to come to, as it does make sense given the information it has. This conclusion highlights the uncertainty inherent in making specific predictions from low-resolution usage data and hence the motivation to consider results probabilistically rather than fixating on a single solution. With an understanding of the relative likelihoods of all potential appliance usages, trends and knowledge can be established with confidence over the long term, despite the challenges with making accurate predictions for specific days.

Therefore, by analysing the informed matrix likelihoods we can conclude that the model is learning behavioural patterns as expected. The kettle is consistently well identified by the model, but the uncertainty in power drawn by the microwave and toaster results in some uncertainty around which one is being used. This uncertainty will be present when the model is run with medical appliances, dependent on the power consumption profile of the appliance. Therefore, some improvements should be made to the model before testing with these. However, the similarity in these results to the actual usage for this day, shows that the model is working as anticipated.

4.5 Conclusions

4.5.1 Aims

The purpose of the appliance disaggregation and prediction model is to determine how a household uses its appliances. This was developed and tested using two datasets with appliance monitoring for common household appliances across 25 households spanning multiple years.

The usage statistics were used to determine an initial Time of Day (ToD) prior for each appliance, which is the probability the appliance is used in each 30-minute window for a day. This gives an initial representation of how all the

[13] Percentage probabilities = informed matrix likelihoods * 100

households use each appliance. When the model is run using the smart meter data for a single household, this initial ToD prior is refined for each appliance to become a ToD prior specific for that household. The initial and final ToD priors can be compared to conclude whether these differences could be due to a vulnerability. For example, for a medical appliance, the initial ToD prior would be very low at all times of the day, because a general house is unlikely to have a medical appliance. If the model is run for a household with a medical appliance, the ToD prior should increase, and we could conclude that the household is likely to have a medical device.

4.5.2 Findings

The testing of the model was completed by considering two appliances for ten days, where they are used at the same time each day. The model predicted the 30-minute windows each appliance was used with complete accuracy and the ToD priors shifted to align with the times the appliances were used very well.

The results were reviewed by running the model with three appliances for one month of household data, modified to only include the three appliances and a baseload. The individual appliance usage time predictions for the final day did not align well with reality. This is because the specific appliance usages in that final day were slightly irregular. This result highlights the importance of producing a model that accounts for all the uncertainty in appliance usage and human behaviour, and the model results should not be considered solely on a single day basis. The final appliance ToD priors aligned very well with reality for some appliances and not others. This is due to the uncertainty in the power drawn by these appliances, and therefore the model should be developed further to reduce this uncertainty if used with more appliances.

4.5.3 Limitations

The limitations of the current model are as follows:

- The model has only been developed for single step appliances.
- The model only considers an appliance being turned on once in a single 30-minute window.
- The model has only been tested with four appliances (kettle, toaster, microwave, and vacuum). In a real household, there would be significantly more appliances that could be used in each 30-mintue window.
- The initial ToD priors, which are used for comparison to determine is a household is using their appliances differently, are solely derived from the 25 houses in the dataset.
- The usage statistics developed for the appliances are solely based on the appliances in the 25 houses in the dataset. These appliances are also from more than seven years ago, and the power consumption profiles may have changed since then.
- No attempt has been made to characterise the difference between initial and final ToD prior which could indicate a vulnerability.
- No medical appliances have been considered. However, we believe the success of the approach in identifying a kettle's usage should be applicable for identifying medical appliances with similar power demands, such as stair lifts.

4.6 Future Development

The analysis presented here has concluded that the model is working as anticipated, but there are improvements that could be made to the current methodology, and new functionality could be added to improve the predictions in future. If these improvements were made, then the model could be thoroughly tested with more representative inputs, and if successful, would be ready for implementation.

The following recommendations are made to improve the appliance disaggregation and prediction model with its current methodology:

- Expand the model to use of double step appliances.
- Expand the model to consider one appliance being turned on multiple times in the current window.
- Improve the determination of usage statistics presented in Section 3.5 so more appliances can be used.
- Generate a methodology and metric for determining how well the final ToD prior matches the known ToD prior for testing purposes.
- Separate the testing household from the appliance usage statistics, so the test is truly blind.

The following recommendations are made to increase the functionality of the model so it can learn more household specific behaviour and improve the final predictions:

- Add functionality so the model learns a household's baseload power from continuous use appliances. This will reduce the uncertainty in appliance combinations that could make up the observed power.
- Include a prior that the household has an appliance, which is updated through time. This will reduce the uncertainty in appliance combinations that could make up the observed power.
- Add functionality to learn the usage statistics for an appliance in a specific household. This would decrease the uncertainty in the power drawn by an appliance when in use. This would likely work with appliances that vary in power between make and model but have a generally consistent usage time.

If these improvements are made to the model, then it could be more thoroughly tested with medical appliances. This would be completed by undertaking the following tasks:

1. Determine the usage statistics for medical appliances that draw a high-level power.
2. Artificially add the power drawn from these medical appliances into the current households. Test the model's ability to predict these appliances and make additional model improvements if required.

4.7 Future Applications

The appliance disaggregation and prediction model has much potential for future applications. Some of these could include:

- Prediction of medical appliances that draw high power in a vulnerable household, to indicate to DNOs when a household is using a medical appliance, so they can tailor their response to the household when there are outages.
- Extracting the usage statistics for high-power drawing appliances, to learn when the appliances are most likely to be used during the day, their average power drawn, and their usage length. Knowledge of how appliances are used could then be used to help predict future changes in demand for usage or predict vulnerability based on appliance usage.
- Disaggregation and prediction of the usage of high-power drawing domestic appliances. Using these predictions and linking them to the behavioural characteristics discussed in Section 2 to inform DNOs of when a household is showing behaviours that could indicate vulnerability. This could be done alongside additional household information, to inform the predictions.

The future applications could be possible after the developments are implemented from the previous section, Section 4.6.

5 Cohort Comparison

The goal of the cohort comparison model was to be able to predict the vulnerability of a consumer based on predictions of their average energy usage given their location, and certain characteristics of their household that change depending on where someone lives. For example, a household that is in a more rural area might have a higher energy usage because their house is older and has a lower EPC rating.

The aim is to be able to predict the average energy usage based on the consumers location and characteristics given, and compare this predicted usage to their actual usage, from smart meter data. If their actual usage is lower than the predicted usage, it might indicate that they are struggling to pay for energy, and if their actual usage is higher than the predicted usage, it might mean that they are struggling to heat their home; both are indicators that someone could potentially be vulnerable and therefore might need extra support. Also, if their energy usage is different to that which is predicted, it may be possible to identify characteristics that indicate vulnerability that are unknown or incorrect when input into the model.

5.1 Approach

To develop this model, we required smart meter data for many houses with known characteristics. However, we could not gain access to this information during the timescales of this project. Therefore, we had to rely on publicly available Lower Layer Super Output Areas (LSOAs) data, which includes average household usage across all homes in the LSOA.

LSOAs are regions in the UK defined by population number, and using the 2011 Census, they detail the population demographics, household types and average household usage, for all residents of each LSOA. We split the number of households in each LSOA into artificial households built up from artificial people, which was used to train the model. Each household contains characteristics presented in the LSOA values, that might influence the average energy usage of a household.

We generated a regression model to calculate the optimised fit parameters and predict the usage of each artificial household. We can review the success of the model by combining the predicted household usage for each artificial house in the LSOA and calculating how different this was from the known average household usage for that LSOA.

5.2 Data Collection

To be able to build a model to predict household usage, we must have a source of truth to train the model against; this source will have to come from real data. The only data available at the time of model development was the average energy usage per household for each LSOA, no other, more granular information is available, as this could pose as a privacy issue to households. Model development had to be tailored to this dataset. LSOAs are areas in the UK that have a population of 1600 people, which equates to approximately 800 households per LSOA. This means that there are over 37,000 LSOAs in England and Wales.

The data collected about the LSOAs came from the 2011 Census; this data is provided by the Office for National Statistics (ONS) (Office for National Statistics, 2011) and their service website Nomis (Nomis, 2011), the official labour market statistics website, which allows access to the data for the public and researchers to use. 17 datasets were gathered and considered for analysis, only three of which came directly from the ONS website (Office for National Statistics, 2011), the others originating from Nomis (Office for National Statistics, 2011).

When collecting the data, it was important to compare the datasets and make sure the number of households and total population was consistent throughout the datasets. Fortunately, because the datasets are all from similar sources, the datasets were mostly the same; some discrepancies arose from some datasets being estimates, rather than concrete data. This was due to flattening of data because of privacy or from some LSOAs having data missing. This was easily mitigated by including “Unknown” fields for each dataset, this is explained further in Section 5.3.

Having a significant difference in energy usage between LSOAs is paramount for our method of modelling to work as intended. To analyse this difference, heat maps of the area covered by the LSOAs were produced to allow an analysis of the discrepancies in average energy usage and further understand the LSOA data.

Figure 33 and Figure 34 show that there was significant variation in the average energy usage per household in each LSOA; yellow represents higher energy usage, and dark purple represents lower energy usage.

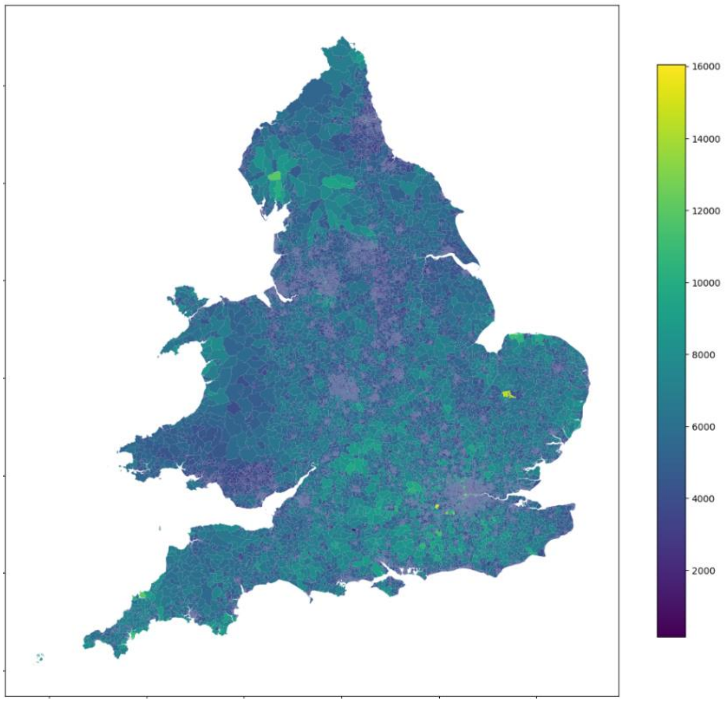

Figure 33: Heatmap of England and Wales showing the average energy usage per household per LSOA.

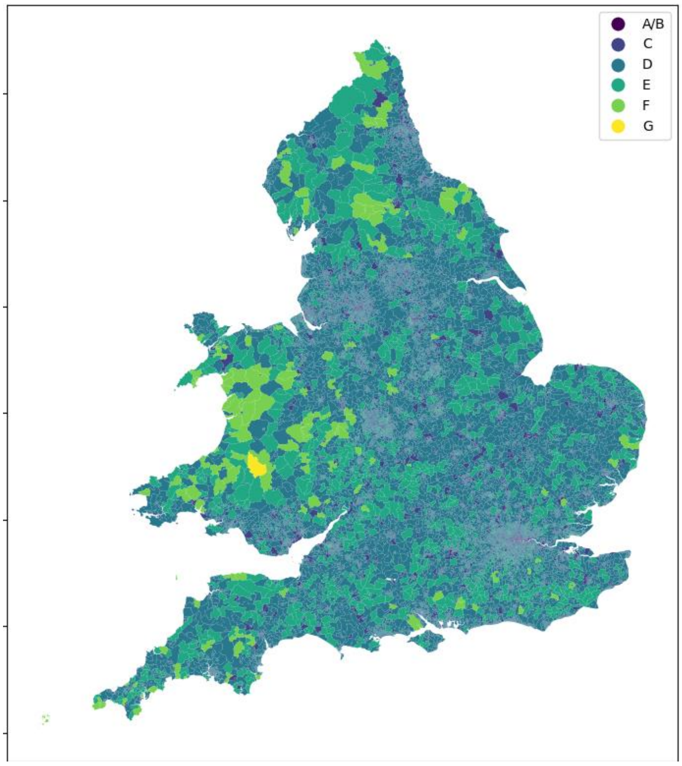

Figure 34: Heatmap of England and Wales showing the most common EPC rating amongst Households on an LSOA level.

Notably there is some correlation between a household's EPC rating and their usage. For example, when comparing the two areas circled in the figures below, it can be observed that a lower EPC rating is correlated to a higher average energy usage.

The correlation between Figure 35 and Figure 36 may not be just due to EPC however, as some of the surrounding areas surrounding usage is much lower but they have similar EPC ratings as areas that have a higher usage. To investigate the potential correlation further, we looked at the correlation of all household characteristics that we could obtain data for to the average energy usage of a household in that LSOA.

Figure 35: Average energy usage per LSOA for Bristol and the surrounding area. The circled area shows an area of higher usage, potentially correlating to EPC usage.

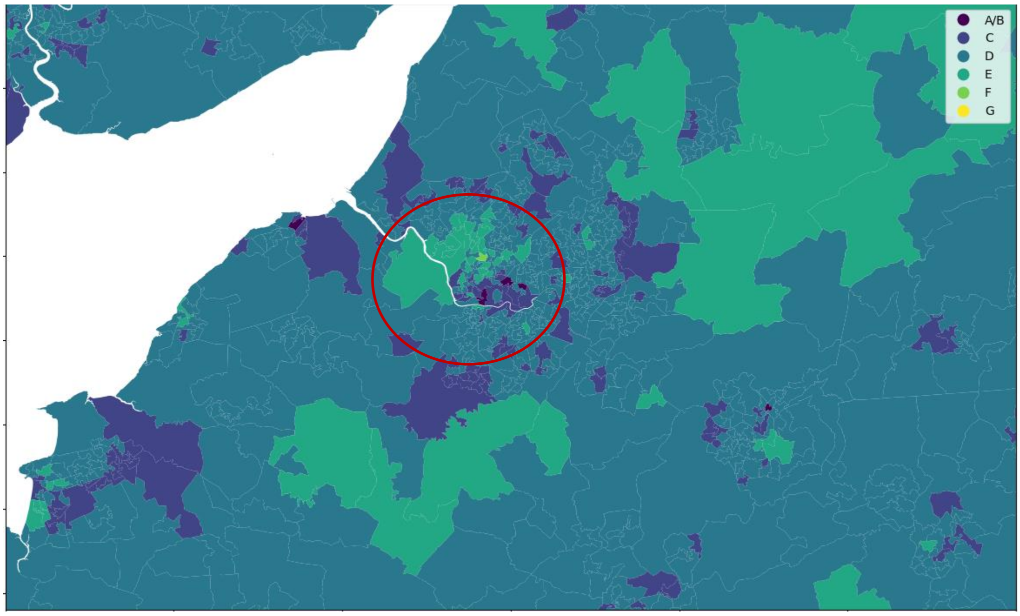

Figure 36: Most common EPC rating amongst households per LSOA. The circled area shows a higher EPC rating, that could be impacting the usage seen in the same area in Figure 35.

5.2.1 Correlation Analysis

To initially review if our method of predicting the usage using the LSOA characteristics was viable, we calculated the correlation of the different characteristics to the average energy usage per household in the LSOA, as discussed in Section 0. The characteristics in the range of datasets were homologated by calculating averages or percentages, depending on the data type, and combined into a single dataset.

The dataset containing the total energy usage per LSOA only provides the number of electricity meters that the measurements were taken from, not the number of households. The number of meters is different from the number of

© FNC 2022

households; it might be that some households share meters or do not have one, since the number of meters was typically lower than the number of households given by other datasets. We concluded that the best way to calculate average energy usage would be to use the number of households rather than number of meters. Note also that households with no usual residents have been excluded.

Figure 37 shows the correlation of all characteristics considered to be relevant to determining vulnerability and potentially influencing energy usage.

The darker red squares show a stronger positive correlation between variables, and the darker blue squares show a stronger negative correlation between variables. The central row of this figure represents the correlation of all the characteristics to the average energy usage.

These correlations show signs of being correct given our inherent knowledge of the underlying data. For example, the correlation between average household size, and percentage of households that contain one person are very strongly, negatively correlated. Following this, some of these characteristics impact energy usage greatly. Average population age of the LSOA, and households that contain people all over 65 seem to have a strong positive correlation to average energy usage, along with households that contain people that work full time, households that contain people that work long hours, and the LSOA's rural urban classification code. Another characteristic that particularly stands out is the strong positive correlation between EPC rating and energy usage. Characteristics that have a strong negative correlation include households that contain people that are unemployed or are of bad health. To summarise, the described correlation trends shown in Figure 37 support the development of a predictive model.

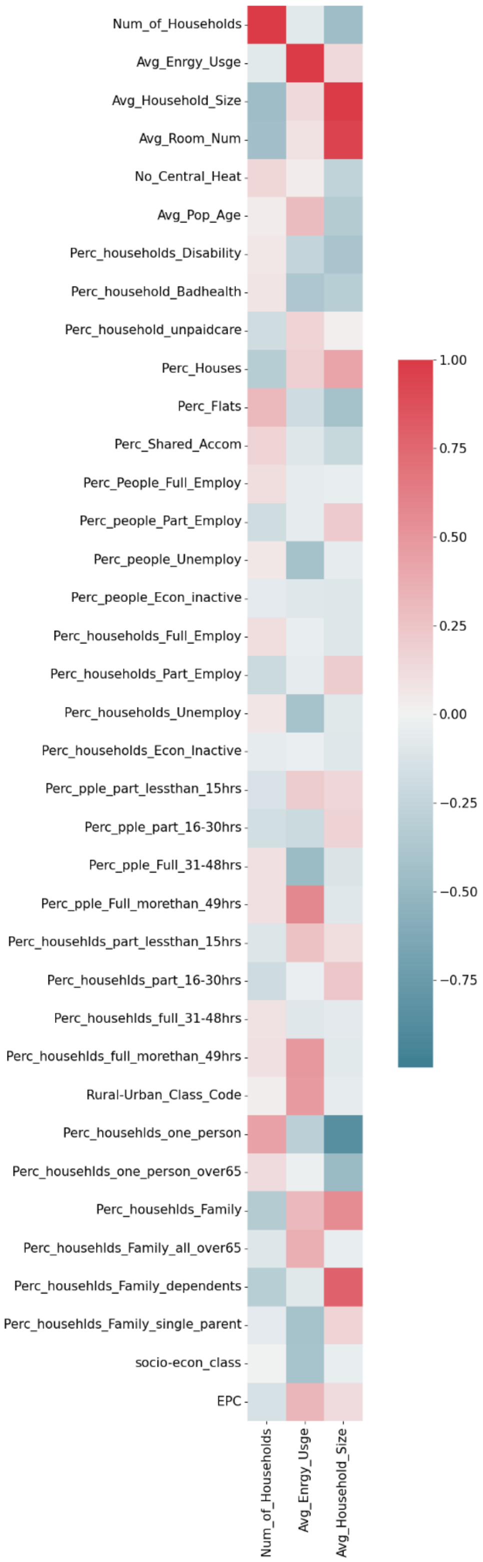

Figure 37: Correlation matrix, showing the correlation between all household characteristics for which data could be retrieved and the average energy usage per household per LSOA.

5.3 LSOA Household Generation

The first step of the model development was to create artificial people based on the real datasets collected from the ONS (Office for National Statistics, 2011) and Nomis (Nomis, 2011). Each person will have certain characteristics. These artificial people make up the artificial households. The data of these artificial households will be used to create the fit parameters for each characteristic that each household and person in the household has. These fit parameters are then be optimized so that when aggregated they produce the average energy usage per LSOA that is closest to the real average energy usage from the datasets discussed in Section 0. These parameters can be used when calculating the predicted average usage of a real household, which can be compared to their real usage to see if they may be considered vulnerable.

The purpose of creating artificial households is to ensure that the “per capita” information captured at an LSOA level is distributed amongst a range of households rather than analysing at the level of the full population of the LSOA.

5.3.1 Artificial People

To create the artificial households in an LSOA, characteristics were assigned to a list of artificial people based on the datasets that were defined by the number of people with that characteristic in an LSOA. For example, if a dataset stated that there are 560 people who are in full-time employment out of a total of 1600 people, then the list of 1600 people contained 560 random people that are in full-time employment. The total number of artificial people created per LSOA matched the known real number of people resident in the LSOA. Table 10 contains examples of randomly generated artificial people.

Table 10: An example of the creation of artificial people.

Person Number	Has a Disability	Social Grade DE	Age 0-4	Age 5-15	Age 16-19	Age 20-59	Age 60+	Full-time	Part-time	Un-employed
1	0	1	1	0	0	0	0	1	0	0
2	1	0	0	1	0	0	0	0	1	0
3	0	0	0	0	1	0	0	1	0	0
4	0	0	0	0	0	1	0	1	0	0
5	0	1	0	0	0	0	1	0	0	1
6	1	1	1	0	0	0	0	0	1	0
7	1	0	0	1	0	0	0	1	0	0

Table 10 shows how the artificial people were assigned characteristics. The ones represent where they do have a characteristic, and the zeros show where they do not. This example does not show all the characteristics applied to an individual person for brevity. The full example can be seen in Section 5.3.3.

5.3.2 Artificial Households

The artificial households were created in a similar way to the artificial people. A list of households equal to the number of households in that LSOA was created and assigned various characteristics that were to do with household, such as number of people in the household, or EPC rating. Then, the individual artificial people were randomly assigned, depending on the number of people that are in that household. Table 11, shows how the artificial people from Table 10 were assigned to the artificial households.

Table 11: Random assignment of artificial people to artificial households

House	Number of tenants	Person Number	Has a Disability	Social Grade DE	Age 0-4	Age 5-15	Age 16-19	Age 20-59	Age 60-90+	Full-time	Part-time	Un-employed
1	4	2	1	0	0	1	0	0	0	0	1	0
		5	0	1	0	0	0	0	1	0	0	1
		1	0	1	1	0	0	0	0	1	0	0
		7	1	0	0	1	0	0	0	1	0	0
2	2	3	0	0	0	0	1	0	0	1	0	0
		6	1	1	1	0	0	0	0	0	1	0
3	1	4	0	0	0	0	0	1	0	1	0	0

As with Table 10, Table 11 does not show all the characteristics discussed previously. A large dataset containing every household for each LSOA was created, showing the following characteristics:

- Household with no heating,
- Either a house or a flat,
- Shared accommodation,
- EPC rating,
- Socio-economic classification of DE,
- Number of tenants,
- Household contains one or more persons that are Age 0-15,
- Household contains one or more persons that are Age 20-58,
- Household contains one or more persons that are 60-90+,
- Household contains one or more persons that has an unknown age,
- Household contains one or more persons that are unemployed,
- Household contains one or more persons that works full-time,
- Household contains one or more persons that works part-time,
- Household contains one or more persons that are self-employed, works part-time and has their own employees,
- Household contains one or more persons that are self-employed, works full-time and has their own employees,
- Household contains one or more persons that are self-employed but has no employees and works part-time,
- Household contains one or more persons that are self-employed, works full-time and has no employees,
- Household contains one or more persons that has an unknown employment type,

- Household contains one or more persons that are disabled,
- Household contains one or more persons that have bad health,
- Household contains one or more persons that provide unpaid care,

Below in Table 12 is an example of the final layout for the three example artificial households. Please note that if one person has the socio-economic classification of DE, then the whole household is considered to have that classification. In the actual output, EPC is represented as 6 columns for each EPC rating, and shows a one or a zero depending on the households rating. Similarly, more columns are in the actual table to represent an unknown socio-economic classification, household type or if they are shared.

Table 12: Final format of artificial households after assigning the artificial people randomly. Here, each column represents a single household.

	Household		
	1	2	3
No Heating	0	0	0
Type	1	1	1
Shared	0	0	0
EPC	2	4	5
DE	0	0	1
Tenants	4	2	1
Age 0-4	1	0	0
Age 5-15	1	1	0
Age 16-19	0	0	0
Age 20-59	0	1	0
Age 60-90+	1	0	1
Unknown Age	0	0	0
Full-Time	1	0	0
Part-Time	1	0	0
Self Employed, employees, part-time	0	0	0
Self Employed, employees, full-time	0	0	0
Self Employed, no employees, part-time	0	0	0
Self Employed, no employees, full-time	0	1	0
Unemployed	1	1	1
Disabled	0	1	1

Bad health	0	0	0
Unpaid care	0	1	0

When looking at whether the household has a social classification of DE or not, 0 represents they have a different classification, 1 represents they are DE classification, and 2 represents that their classification is unknown. EPC rating is rated as zero being Unknown, one being A or B, two being C, three being D and so on.

5.3.3 Final Data for Training the Model

To complete the final analysis and make a direct comparison to LSOA average energy usage, all artificial households were combined, to create values that would represent an "artificial LSOA". Table 13 shows an example of the household data representation used for training.

Table 13: Final table used for training the regression model. Represents the sum of each characteristic of each household for individual LSOAs.

	E01014370	**E01014371**	**E01014372**
No Heating	52	54	69
Type	984	960	1059
Shared	39	24	48
EPC A/B	10	5	25
EPC C	15	55	67
EPC D	12	95	68
…	…	…	…
DE	965	793	680
Tenants	1666	1804	1959
Age 0-4	60	30	29
Age 5-15	16	78	4
Age 16-19	95	79	90
Age 20-59	600	456	247
Age 60-90+	289	337	460
Unknown Age	0	0	0
Full-Time	600	460	350
Part-Time	345	120	68
Self Employed, employees, part-time	21	11	3
Self Employed, employees, full-time	12	4	4
Self Employed, no employees, part-time	9	4	1

Self Employed, no employees, full-time	7	51	360
Unemployed	50	300	34
Disabled	300	150	249
Bad health	150	20	68
Unpaid care	20	65	79

Table 13 represents the sum of each characteristic over an LSOA. Each row of Table 13 is the sum of each row of Table 12 for all households in an LSOA. The method of how we used Table 13 to train the model is discussed in Section 5.4. 800 LSOAs in the WPD region were summed in this way and used to train the model, and then 200 of the LSOAs, also in the WPD region, were used to test the model. 1000 LSOAs were selected from the total 37,500 LSOAs in England and Wales for initial testing. The results from the training and testing datasets are given in Sections 5.5.1 and 5.5.2, respectively.

5.3.4 Limitations

There are several limitations that occurred because of this method of creating the artificial households and people. One of these is that there are artificial households in the LSOAs where the EPC rating, and other characteristics, are unknown, due to discrepancies in the datasets, as discussed in Section 5.2. There were some cases where the number of artificial people would not correlate to the number of people in households because of the way that the data for number of people in households was defined; the number of people in a household scaled from one to eight or more, which gave us more cases where a person's characteristic was unknown. Generally, the total number of households in an LSOA was always greater than the number of households that data was collected for when looking for a certain characteristic, so this seemed like a suitable way to mitigate this issue.

In some cases, for example Person 1 in Table 10, there are unrealistic characteristics for that person, where they are a child but are in full-time employment. This is because we have not accounted for quantifying realistic people. It is thought that this will not provide the main source of uncertainty in the final predictions, so this has not yet been included in model development.

Another limitation with this implementation is that we have not accounted for uncertainty in the way we have disaggregated characteristics. For example, 560 people in an LSOA might be in full-time employment, and 35 people in the LSOA might be disabled, but we do not know the number of disabled people that are in full-time employment. Hence, allocating characteristics in this way does provide some limitations, but still produces suitable results. As well as this, we do not know how many disabled people or employed people might be living in one household.

5.4 LSOA Household Usage Prediction

The aim of the optimisation is to find the optimal fit parameters for the linear regression model used to calculate the average energy usage of that household based on their characteristics that change based on their LSOA. Equation 1 shows the linear regression equation to calculate the predicted average energy usage.

Equation 1: $$Average\ Energy\ Usage = a\ \times heating + b \times type + c \times shared + d \times EPC\ + \cdots +\ constant$$

Where, the constants *a,b,c* and *d* are the fit parameters we want to optimise and heating, type, shared and EPC are equivalent to the values in the rows of Table 13 and the *Average Energy Usage* is the predicted usage of that household. We decided to choose this equation, a linear model, to predict the fit parameters because the correlations discussed in Section 5.2.1 were strong; if the correlations were not as strong this may have been because of a nonlinear correlation. This is not a complex problem, so a linear model is suitable and can produce easily explainable results in this first stage of model development.

Initially two algorithms were used for optimization, these were a particle swarm and the least-squares of the characteristics matrix. The least-squares of the characteristic matrix provided a much lower spread of errors between the predicted and known average energy usage and so was the more successful method of the two. 800 LSOAs were used to train the model and produce the predicted fit parameters.

To test these methods, the predicted average energy usage was calculated for each of the 200 testing LSOAs and subtracted the known average energy usage per LSOA, taken from the datasets discussed in Section 5.2; we wanted to minimise the result of this subtraction, as this represented the error between the predicted and known value, therefore the parameters in Equation 1 will need to represent this as best as possible. This was our best way of testing the model before including real household data, as we had a solid real data source for the Average Usage per LSOA.

5.5 Results

5.5.1 Training Dataset

Figure 38 represents the distribution of error between the known and predicted average energy usage of each LSOA from the 800 LSOAs used to train the predicted parameters. This was a promising distribution to produce from the training set, since the range of errors is quite small, and the mean is close to zero. This meant that we knew that the fit parameters produced would give us suitable predictions for the average energy usage for LSOAs that were not in the training set.

Figure 39, shows the scatter of plots showing the known average energy usage against the predicted energy usage for each LSOA in the training dataset. This scatter is also as predicted, with the trend following a straight line and with most plots being close to that line. The deviation of plots at a higher LSOA usage is representative of the higher errors at around 2000kWh shown in Figure 38.

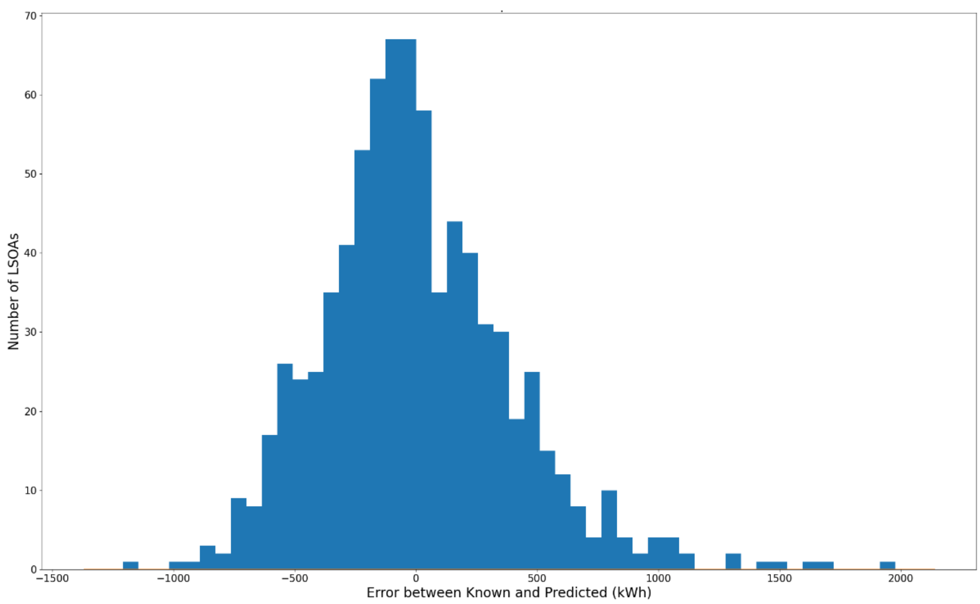

Figure 38: Error distribution between the know average energy usage and the predicted LSOA usage for the training set of LSOAs.

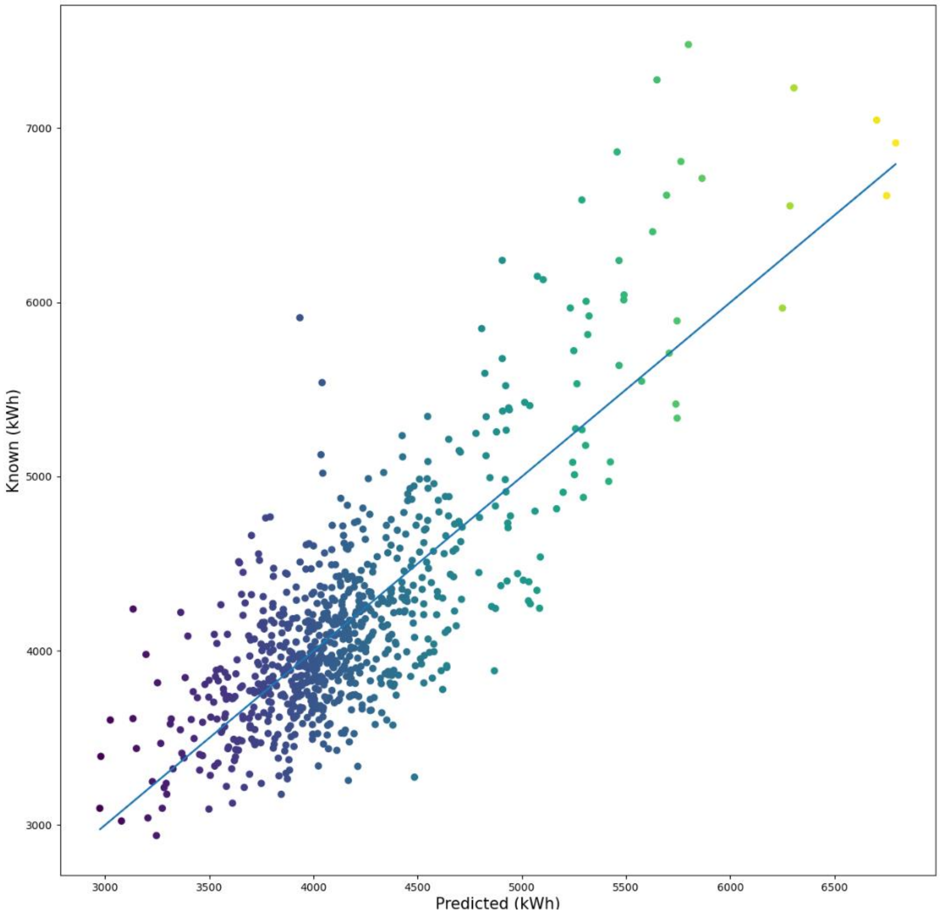

Figure 39: Known average energy usage against the predicted average energy usage for each LSOA in the 800 LSOAs used for the training dataset.

5.5.2 Testing Dataset

Figure 40 shows the error distribution between the predicted average energy usage per LSOA calculated using the optimised fit parameters, using the least-squares of the household characteristics and the known average energy usage for the testing set of LSOAs.

This shows a narrow spread of errors, and the peak is close to zero, which means that only a small error contributes to the predicted average energy usage. The mean and standard deviation for both the testing and the training sets are shown in Table 14.

Table 14: Residual mean and standard deviation between actual LSOA usage and predicted LSOA usage for the training and testing sets of LSOAs

	Mean (kWh)	Standard Deviation (kWh)
Training (800 LSOAs)	$1.2e^{-11}$	393.1
Testing (200 LSOAs)	-140.9	365.9

The standard deviation for the testing set of LSOAs equates to an error of approximately 15% between the predicted average energy usage, and the known average energy usage.

Figure 41 shows the scatter of the known average energy usage per LSOA against the predicted average energy usage. The trend we are expecting is a straight diagonal line through the middle of the plot. The points here follow this trend closely, excluding the points towards the top right and bottom left. These points are equivalent to the tails either side of the distribution in Figure 41, where a deviation of about 2000kWh from the straight line is equivalent to the largest error of about 2000kWh in the histogram. These plots do not follow the trend as well as in Figure 39, but this is to be expected, since the fit parameters used are directly trained using the known average energy usage for the training LSOAs, so the error will be smaller for the training set.

The standardised residual between the known and the predicted average energy usage for the training and the testing LSOAs is shown in Figure 42. This confirms that there were no inherent assumptions in the training and test datasets that affect the confidence that the model, and shows that it has not over-fit to the training set.

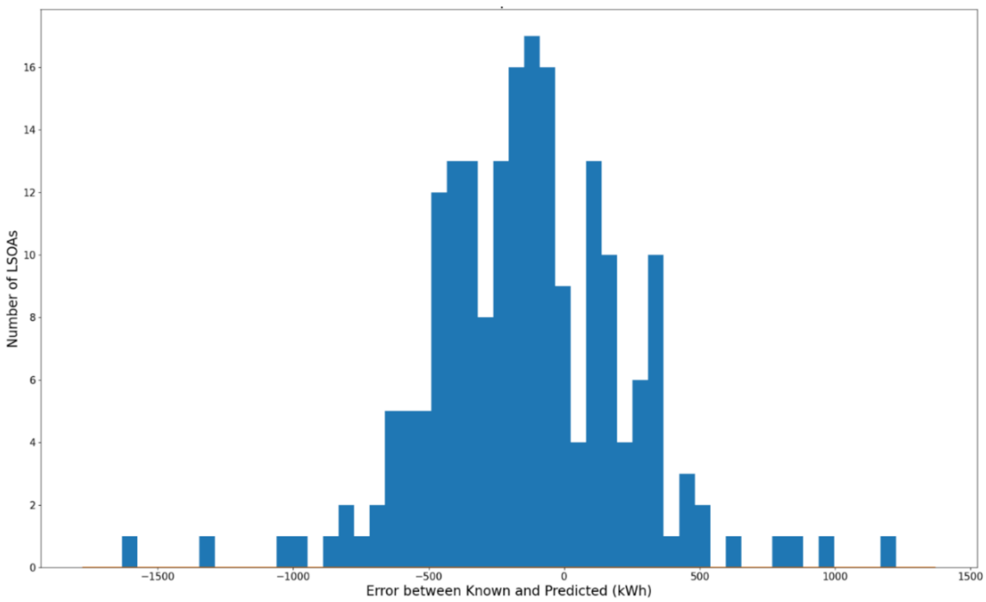

Figure 40: Error distribution between the know average energy usage and the predicted LSOA usage for the testing set of LSOAs.

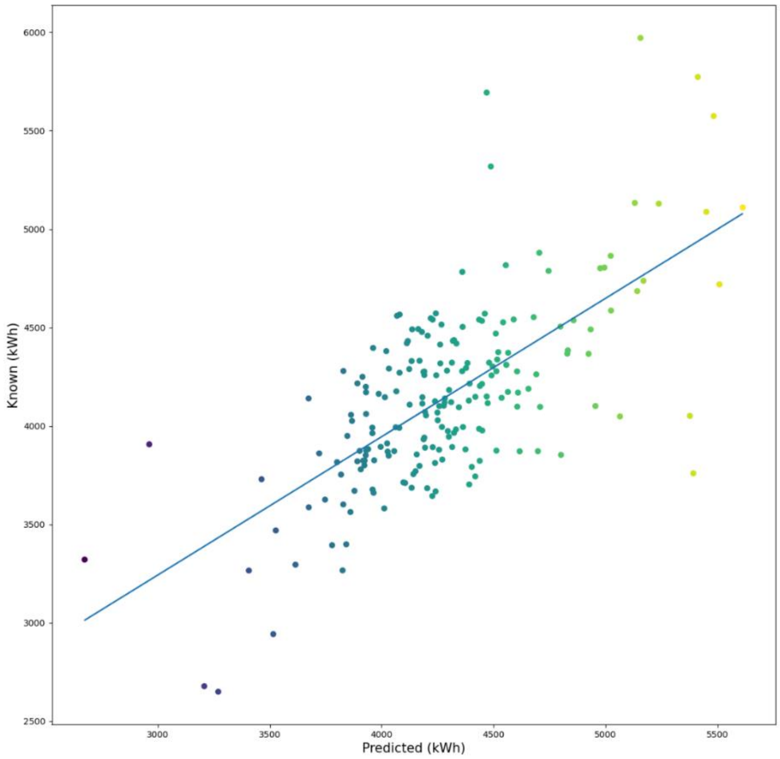

Figure 41: Known average energy usage against the predicted average energy usage for each LSOA in the 200 LSOAs used for the testing dataset.

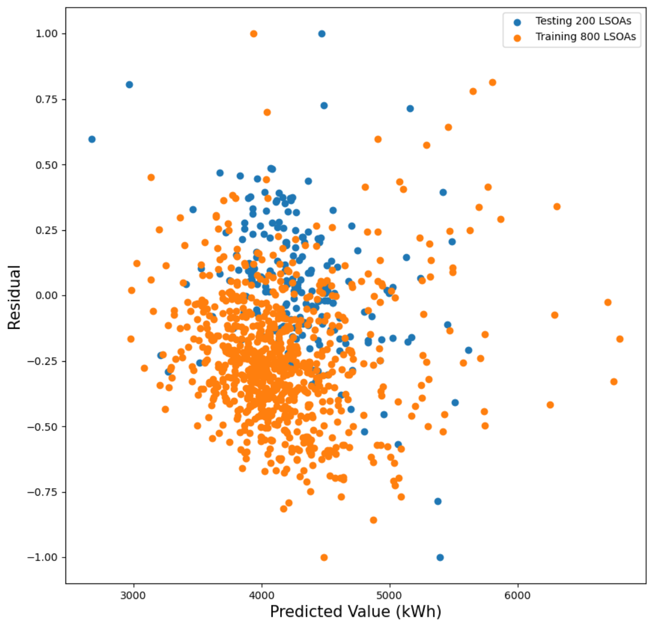

Figure 42: Residual of the known average energy usage per LSOA from the predicted average LSOA against the predicted value for both the training (orange) and the testing (blue) LSOA datasets. The y axis has been normalised using the maximum value of the residual so that the axis varies between 1 and -1.

5.6 Future Work

Collecting real household data containing the household's average energy usage and their household characteristics will allow us to test our model's ability to predict energy usage on a household level rather than per LSOA. Being able to test this will give us finer granularity of detail in how much each characteristic affects usage and will confirm whether we can predict energy usage for individual households.

Obtaining real household data would allow inclusion of constraints on the artificial households and artificial people created to train the model. Adding in these constraints would give more realistic households and characteristics and reduce the error in the predicted energy usage.

As well as improving the artificial households, investigation and review the impact of different characteristics on the predictions is then possible. This would be done by excluding certain characteristics from the prediction calculation and comparing the predicted energy usage without these characteristics to the predicted energy usage that has contribution from all characteristics. Investigating allows confirmation of whether vulnerability characteristics can have a significant enough effect on a household's total energy usage.

Another way to extend the prediction capability would be to include a "LSOA factor", which accounts for changes in usage due to the area that the household is located, rather than just their characteristics. This will give an extra parameter to base our prediction of vulnerability upon and therefore make it more reliable.

5.6.1 Dataset Requirements

To complete these next steps, a large enough dataset with data that is in the correct format for the model will need to be collected. The dataset will have to contain at least 1000 households so that sufficient iterations of testing can be done to minimise error and to maximise accuracy and also to get a wide spread in variation of households that have different household characteristics. We would also need the average energy usage of each household for several months and several years for the same reasons, particularly given the change of energy usage through different seasons.

As well as these, the datasets from each household should have details of a similar range of characteristics as mentioned in Section 5.3. It will be important for us to have households that are non-vulnerable, as well as households that are considered vulnerable, to be able to make comparisons between the predictions and to determine whether there is determinable variation in usage between vulnerable and non-vulnerable households.

5.7 Conclusions

To conclude, our cohort model shows excellent promise in its ability to be able to predict average energy usage of a household based on the household's characteristics that stem from the LSOA in which the household is located.

We have shown that our model can predict the average usage for each LSOA to an accuracy within approximately 15% of the known energy usage across the LSOAs. This error is mainly due to errors arising from constraints on the model development that cannot be resolved until real household data has been obtained.

Because the model has been successful in predicting average energy usage at an LSOA level we can determine that the characteristics we have input into the model, and the fidelity of the model are sufficient to identify energy usage. With further development, and sufficient granularity of data, we can have confidence that a model can be produced to predict energy usage at a household level based on the given characteristics, and meta-analysis of the produced model could be used to identify characteristics of concern from energy usage.

To conclude, this model development has shown that a household's characteristics do impact a households average energy usage.

6 Overall Changes in Usage

6.1 Purpose and Overview

The purpose of this aspect of the modelling work is to develop prototype methods to identify significant changes in energy usage for a given household and identify whether these changes are indicative of a change in vulnerability of that household. In the future, these methods could form a core part of an automated tool to identify and flag when a household potentially becomes vulnerable, based on their energy usage patterns.

Two methods have been developed to identify a statistically significant change in usage patterns for a given household, which in this report have been called the 'short term' method and the 'long term' method. The short term method is quicker to respond to changes, but the long term method in some cases may be more accurate, for reasons outlined in the following sections.

A further method has then been developed to characterise what has changed, which then allows comparison with change profiles associated with certain vulnerabilities. This is discussed in Section 6.3. In this way, the methods outlined demonstrate how a change in usage can be identified and then analysed to see if it corresponds to an increase in vulnerability.

6.2 Change Detection Methods

This Section outlines the data used in the methods development work (Section 6.2.1), and then details the short and long term methods (Section 6.2.2 and 6.2.3 respectively) as well as their relative advantages and disadvantages. The methods are then compared in Section 6.2.4.

6.2.1 Data Used

The data used in this section of the project is taken from the UK Power Network's Low Carbon London (LCL) project (SmartMeter Energy Consumption Data in London Households, 2014), which was run from November 2011 to February 2014. Electricity usage was recorded for each half hour using smart meters for a large number of households across London, across a period of one to two years. However, beyond the date and a MAC address unique to each house, no personal, geographical or other identifying information was recorded. This provides a rich dataset of energy usage patterns against time throughout a relatively long period. An excerpt of the data from LCL is shown in Table 15[14], and Figure 43 and Figure 44 show a comparison of the usage for two houses for two half hours in the morning and evening, plotted as a time series. Figure 45 shows violin plots of the distributions of usage for each half hour for a specific house, indicating the changing distributions of usage with time throughout the day.

[14] We also had access to Customer Lead Network Revolution (CLNR) (Revolution, 2022) dataset, which contains the same level of detail as the LCL datasets. Since we had access to such a detailed dataset from LCL, we decided that we had enough data to complete our work, so we did not use CLNR to prevent confusion between datasets.

Table 15: Excerpt of data used in the change detection work.

House ID	Day	Half hour 0 usage (kWh)	Half hour 1 usage (kWh)	Half hour 2 usage (kWh)	Half hour 3 usage (kWh)	Half hour 4 usage (kWh)
MAC000002	13/10/2012	0.263	0.269	0.275	0.256	0.211
MAC000002	14/10/2012	0.262	0.166	0.226	0.088	0.126
MAC000002	15/10/2012	0.192	0.097	0.141	0.083	0.132
MAC000002	16/10/2012	0.237	0.237	0.193	0.118	0.098
MAC000002	17/10/2012	0.157	0.211	0.155	0.169	0.101

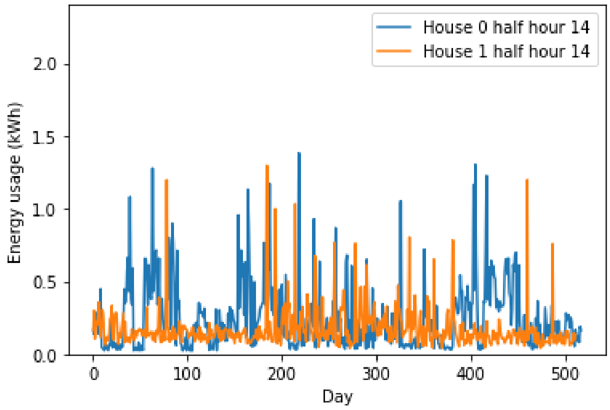

Figure 43: Morning usage comparison.

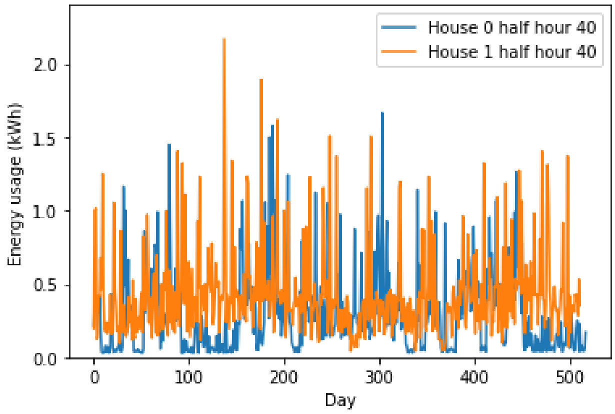

Figure 44: Evening usage comparison.

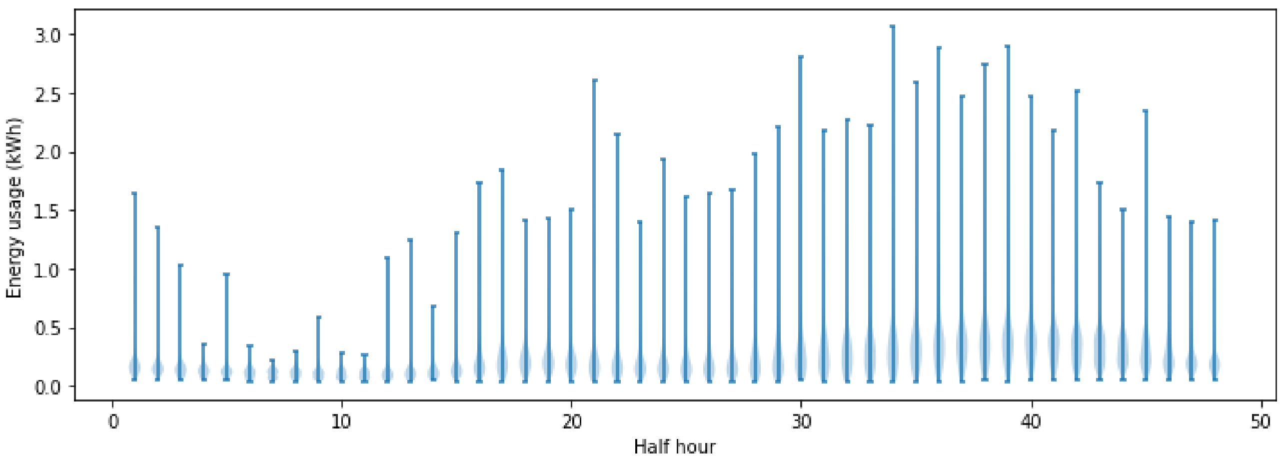

Figure 45: Violin plot showing distributions of half hourly usage for a specific house.

Given no other information accompanies the dataset, it is not possible to know which (if any) houses correspond to vulnerable customers, or whether any changes observed in their usage can be related to a change in circumstance significant from a vulnerability perspective. This limits the utility of the dataset in validating the methods developed –

however, it is still possible to identify changes by comparing one household's usage with that from another, and/or by manually altering usage data to represent known vulnerability usage profiles. This is discussed further in Section 6.3.

6.2.2 Short-term Method

6.2.2.1 Method Description

The short term method works in two phases – a training phase, during which the usage patterns for a particular household is characterised, and then an implementation phase, when one day's usage is analysed to see how well it fits the pattern of previously usage observed. The method should report a change if the day's usage is unlikely to be a continuation of the pattern observed during the training phase. The method is illustrated in Figure 46, for a reduced training period (five consecutive days in this figure, whereas a month or two was found to work well for the examples used in this work), and for a reduced number of half hours, for clarity.

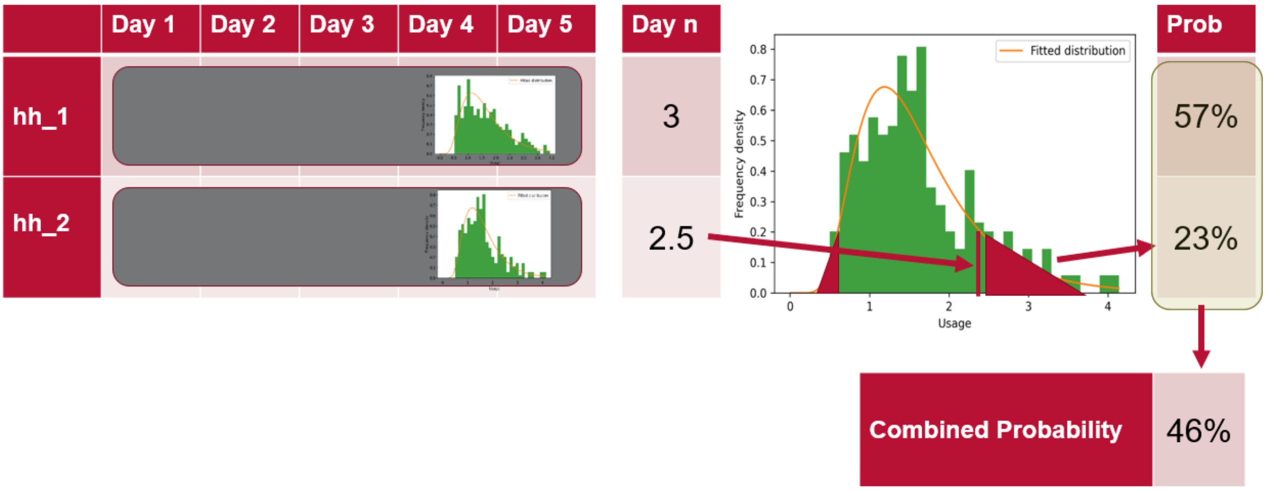

Figure 46: Short term method illustrated for reduced number of days and half hour (hh_n) periods.

During the training phase, for a particular household, a histogram of the energy usage throughout the training phase is produced for each half hour in a 24 hour day, for a total of 48 histograms each covering the entire training period. A statistical distribution is then fitted to each histogram of half-hourly usage. A number of distributions where investigated, but across most houses, a log-normal distribution was found to work best, which is shown in Figure 47. This is due in part to the fact that energy usage cannot be negative, and the usage typically has a longer tail of less likely higher usage values, both of which a log-normal distribution is well able to model.

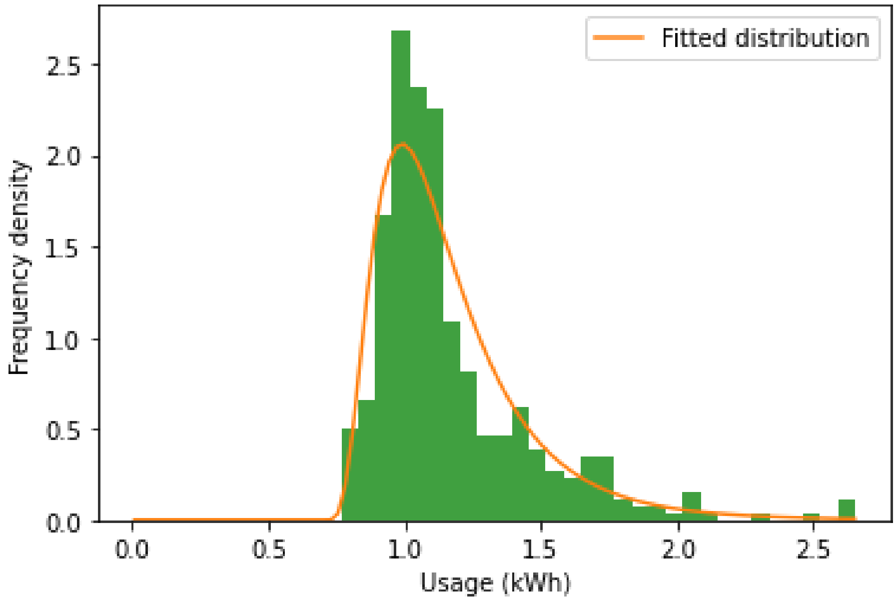

Figure 47: Typical energy usage histogram for a given half hour of usage, compared with the fitted distribution.

With the training phase complete, the implementation phase then assesses each day's half-hourly energy usage profile to see whether the usage pattern for the day fits that observed in the training phase for the same house. This is carried out as illustrated in Figure 46. For a given half hour of usage observed in the implementation phase, a probability metric is calculated that a usage value would be observed that is as likely or less likely than that actually observed, using the half hour's fitted usage distribution. This is done by finding the area under the distribution with a frequency density equal to or lower than that at the observed usage point, accounting for both tails. This means that if a value of usage is observed that lies at the peak of the distribution, a probability metric of one would be returned, whereas if the usage value lies in one of the tails, a much lower probability metric would be returned. This provides a measure of the likelihood of observing the actual half hourly usage, given the histogram fitted during the training phase.

Having calculated a probability metric for each half hour of usage for a given day, these are then combined to find the overall probability metric representing how well the day's usage pattern fits the pattern observed in the training phase. The probability metrics are combined using Fisher's method (scipy.stats.combine_pvalues, n.d.). During testing, the method was found to give noisy results, and we found that taking a rolling average smoothed the results, with seven consecutive days working well as a rolling average period, as shown in Figure 48.

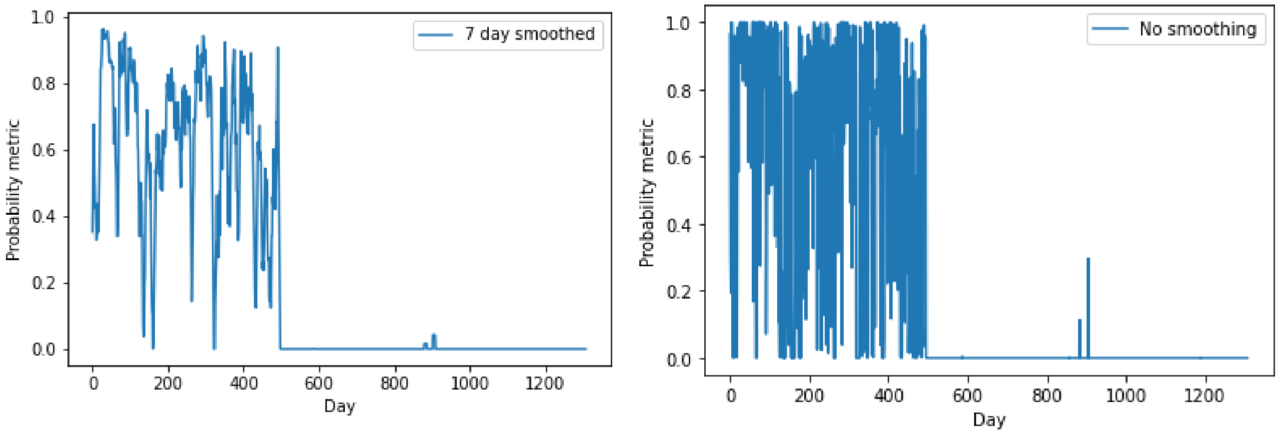

Figure 48: Rolling average smoothing of probability results.

The method was then tested using the dataset discussed in Section 6.2.1. In the absence of data from known vulnerabilities, the method was used to detect change in datasets created by appending another household's data to the data corresponding to the house for which the model had been trained. In this way, the method should return a high probability while it receives data corresponding to the trained house, and a low probability when the house has changed. Figure 49 shows an example output for the short term method, which has been smoothed using a seven day rolling average. The figure shows the method working as expected, with a relatively high probability while the method is received data from the house on which it was trained, and then a probability of almost zero when the new house's data is received.

© FNC 2022

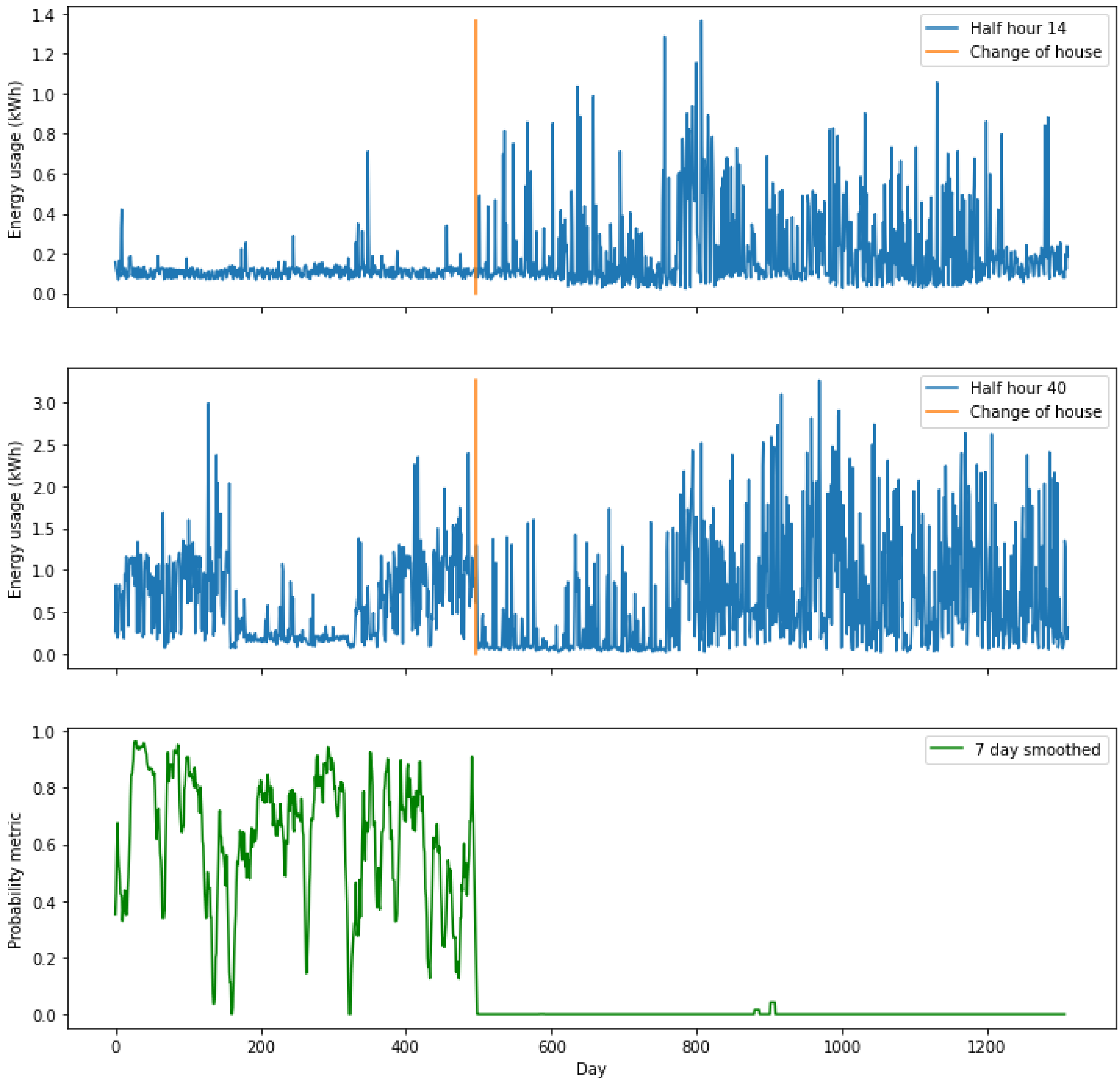

Figure 49: Example short term method output for two houses, with two half hours of underlying data also plotted.

6.2.2.2 Discussion

The method was found to work well by testing with different combinations of trained houses and test houses. The method is responsive, as it calculates a new value for each day, and the only delay to detection of a change is given by the smoothing period. In our case, where we have used a seven day rolling average, this means that changes should be detectable within seven days. There is however a direct trade-off here between responsiveness and number of false positive detections of change, as a longer smoothing period reduces false positives but adds delay to a detection of change. This is something that would need to be tuned and validated before implementing the method in a live system.

The method is also sensitive to the differing degrees of variation between houses and even between individual half hour periods for the same household. This means that the method can detect usage that would be unusual for a specific house, even if it is not unusual in the wider population of houses.

There is one other downside to the method, which is illustrated in Figure 50. If the method is trained on a house which has relatively variable energy usage, then a change to that usage becoming more consistent may not be detectable. For the half hour shown in Figure 50, this is illustrated by the differences in distribution between houses one and two. If the method is trained on the first house, and if it changes to the usage shown by the second house (the red probability distribution), then any value sampled from the second distribution will report a high probability of being from the first distribution. This means that the method will be unable to detect this change, even though the distribution of use has clearly changed. The converse, however, is not true – as there are many relatively probable values in the first house's distribution which are very unlikely to be from the second distribution. This means that if the method were trained on the second house, it would detect the change to the first house.

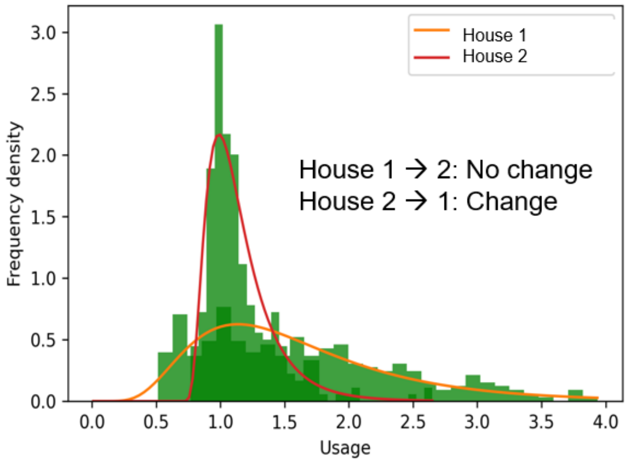

Figure 50: Short term method is not sensitive to changes in variability.

6.2.3 Long-term Method

The long term method seeks to address the problem identified above for the short term method where large variation in usage for a particular house make detection of changes more difficult.

6.2.3.1 Method Description

For the long term method, rather than taking only one day's usage, we take a number of recent days' usage and form another histogram of usage for each half hour. During testing, a period of the most recent one to two months was found to work well[15]. These new half hour histograms can then be compared with the histograms obtained during the training, using a statistical technique called a two-sample Kolmogorov-Smirnov (KS) test (scipy.stats.kstest, n.d.). The KS identifies whether two distributions are drawn from the same underlying distribution, with the reasoning being that if nothing has changed in the household, then this should be true. This allows a change such as that shown in Figure 50 to be identified, as the two distributions are clearly different, although centred about the same location. The method is illustrated in Figure 51, and other than taking the previous month's usage rather than a single day, and other than using the KS test technique, is very similar to the short term method described above, with probabilities calculated for each half hour of usage (usage history in this case) and combined using the same technique as the short term method.

[15] This time period is chosen to be sufficiently long enough to have a representative distribution of the households usage, but to also identify a change in a timely manner.

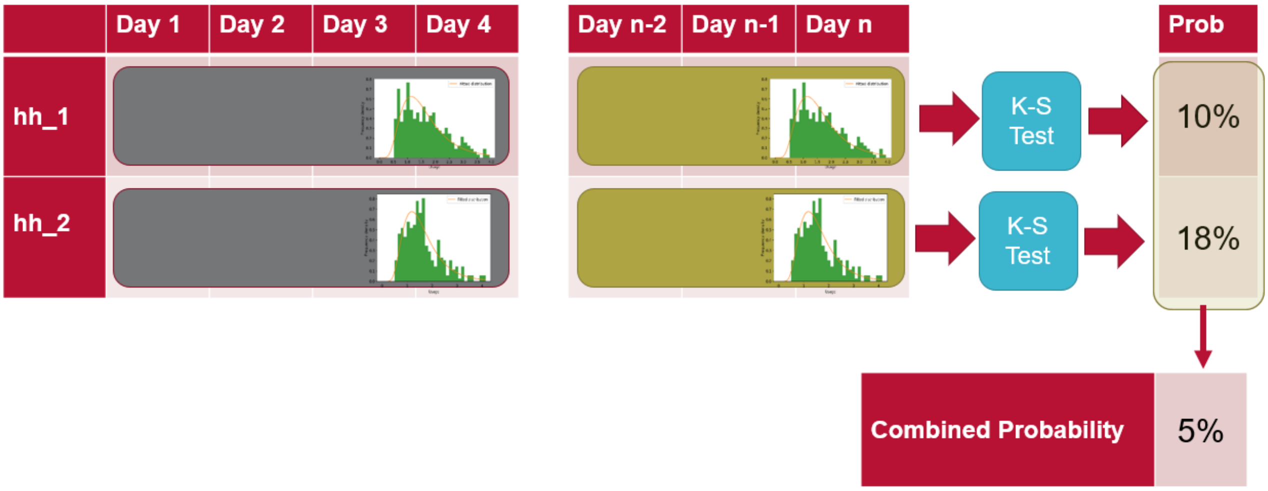

Figure 51: Long term method illustrated for reduced number of days and half hour (hh_n) periods.

6.2.3.2 Discussion

The KS test used by the long term method was found to report significantly lower probability metrics than the short term method, which is not problematic as the two numbers are not directly comparable and we are more interested in the change than in the absolute values. Due to the small numbers involved, the probabilities are best viewed on a log scale, and an example output from the long term method is shown in Figure 52. Since the method works by taking a recent usage period rather than a single day's usage, the output is less noisy and does not require smoothing using a rolling average. However, the method is also slower to respond to a change than the short term method, as it takes on the order of a month for a change to become significant in the histogram of the recent one or two month's usage. This is the reason why the method is called the long term method – it takes longer to detect changes, but is potentially more sensitive to some changes than the short term method.

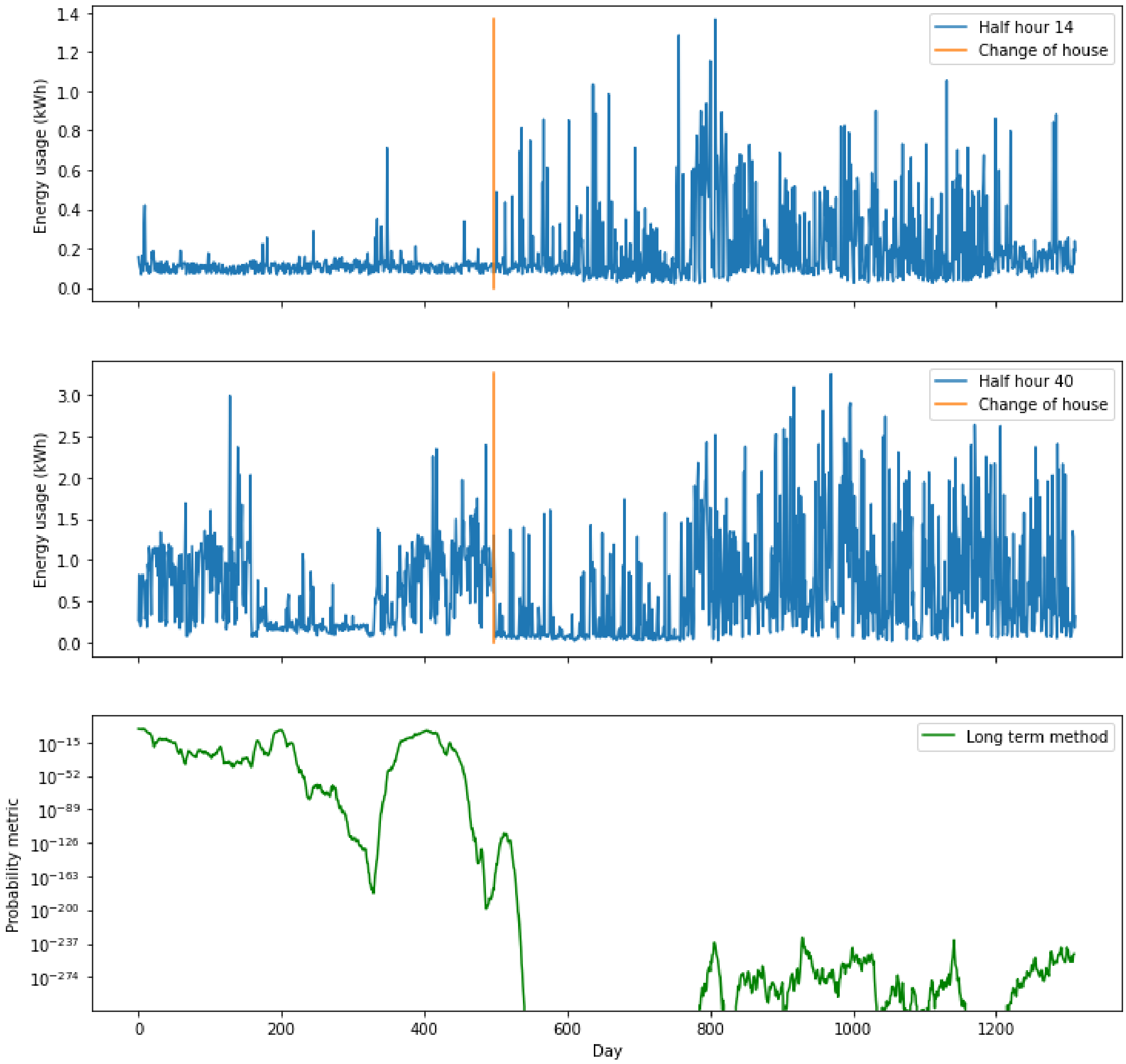

Figure 52: Example long term method output for two houses, with two half hours of underlying data also plotted.

6.2.4 Determination of Effectiveness

The short and long term methods were compared in their effectiveness using a subset of the data described in Section 6.2.1. A subset of twenty households were chosen, and the two methods were trained on each household, one at a time, and then run using the data for all twenty households. If the methods work successfully, they should report a high probability while the data corresponds to the trained house, and a low probability everywhere else. In this way, by training and running on all combinations of the twenty houses, we can compare the methods across a range of households and usage profiles.

By taking the average probability each method returns while the data for each house is being processed, we can build up a matrix of probabilities consisting of one number for each trained house when run against all houses. If the methods work perfectly, the matrices should consist of a diagonal line of probability one (or high probability) where the tested house matches the trained house, and zero (or low) probability everywhere else. These results can then be plotted in three dimensions, as shown in Figure 53 and Figure 54.

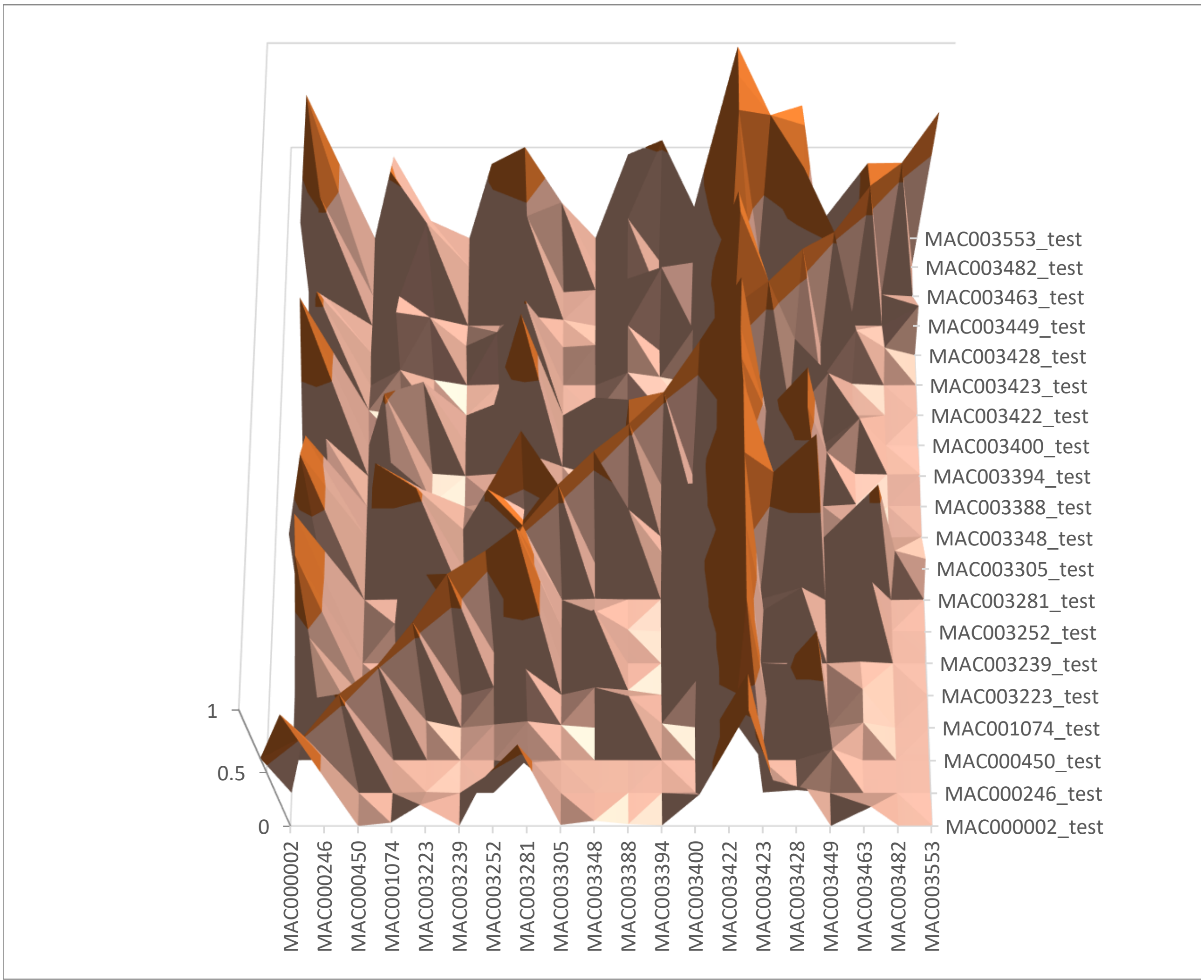

Figure 53: Short term method effectiveness matrix. Trained houses along the bottom axis.

Figure 53 shows the effectiveness matrix results for the short term method. The results show a clear high probability diagonal line where the trained house matches the test house data, which indicates that the method is working well across the different houses. The figure also shows a couple of strong vertical high probability lines, which indicate that for these houses, almost any other house's usage pattern fits what has been observed in the training data. These are good examples of the effect described in Section 6.2.2.2, where almost any house's data, when given to a trained house model with a large degree of variability, will report a high probability. Overall, however, the figure shows that the method works well for most combinations of houses.

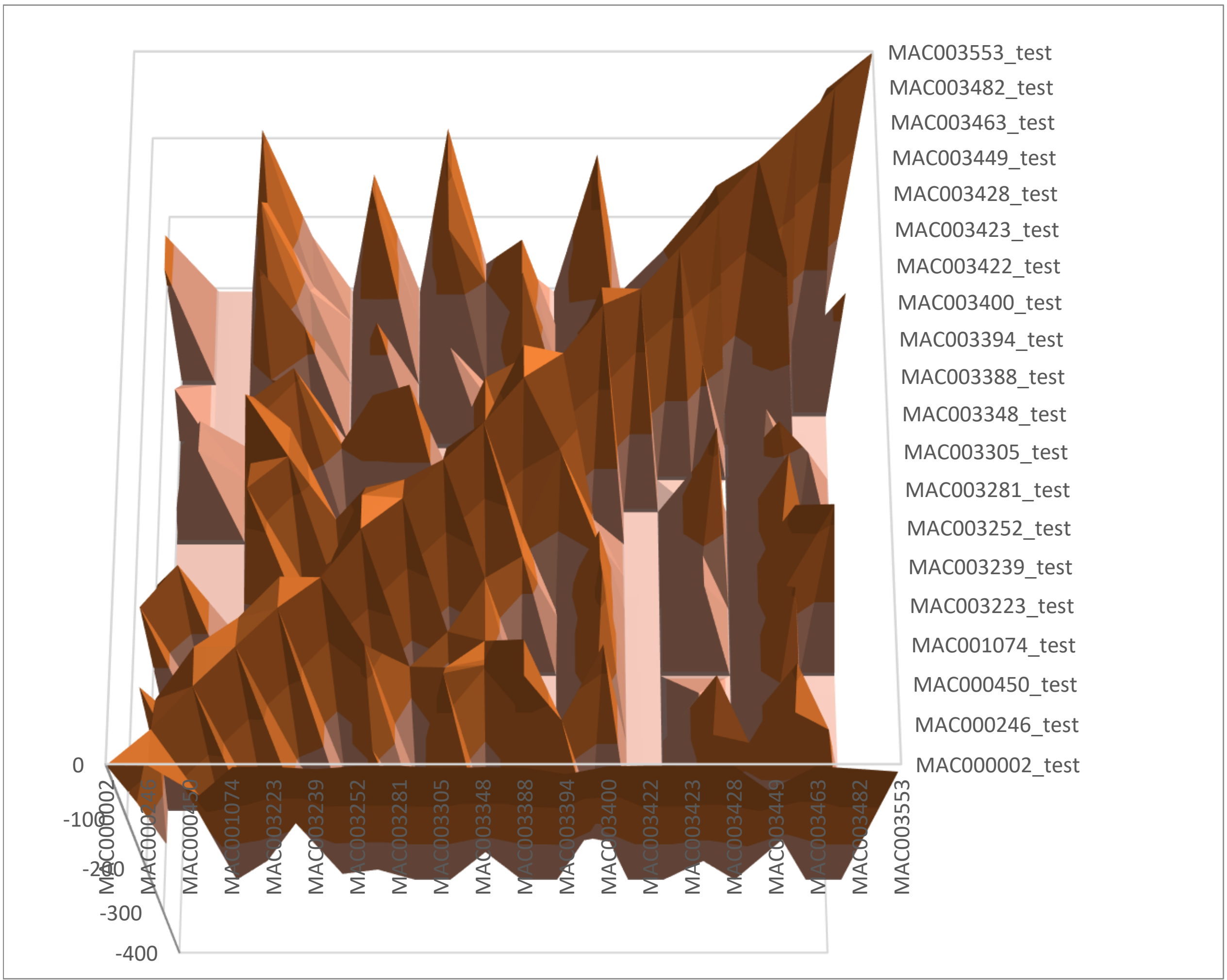

Figure 54: Long term method effectiveness matrix. Trained houses along the bottom axis.

Figure 54 shows the corresponding effectiveness matrix results for the long term method, with the probabilities in this case plotted on a log scale, as discussed in Section 6.2.3. The method again works well, with a very clear high probability diagonal line where the supplied data corresponds to the trained houses. There is one clear horizontal high probability line, the reasons for which have not been identified, but in most other cases, there are smaller isolated peaks where similarities between houses have been identified (as would be expected), but the peaks are not as high as when the correct house has been identified. In addition, the clear vertical lines seen in Figure 53 are now no longer evident – which means that the long term method has successfully accounted for the changes in variation and has managed to identify a change even for a trained house which has very variable usage patterns.

Overall, both methods have been shown to work effectively in detecting a change for the number of trials performed – and although neither is able to identify whether a change relates to a vulnerability, this is discussed further in Section 6.3. There is also a clear trade-off between responsiveness to a change and sensitivity to the change, without introducing false positives. At a high level, this is seen in the advantages and disadvantages of the short and long term methods, with the short term method more responsive, but the long term method slightly more accurate. At a lower level, this also shows up in the choice of rolling average period for the short term method, and the number of previous days taken to build the histograms of recent usage, which are inputs to the KS test in the long term method. These trade-offs would need to be investigated once real validation data relating to vulnerabilities is available, and also considered in the context of the wider project, as it is not clear at this stage how important it is to avoid false positive detections of change or vulnerability.

6.3 Change Identification and Fingerprinting

Using the methods outlined in Section 6.2, we can detect for each house whether a significant change has occurred. However, this does not answer the question of whether this change is related to a change in vulnerability. We therefore require a method to identify what has changed, and whether this change could be related to a change in vulnerability. The method to achieve this is outlined in this section.

6.3.1 Identification Method

Since both the long and short term methods calculate a probability metric per half hour, this presents an opportunity to see which parts of the day contribute most to a detection of change, and whether the usage is above or below average for each half hour. We can do this by plotting a signed log likelihood value for each half hour over a certain window of usage history for which a change has been detected. By taking the negative log of, a large each half hourly probability metric value corresponds to a significant change, and a small value corresponds to a higher probability metric and therefore a less significant change. This allows us to focus on the parts of the day with high values, corresponding to larger changes from the normal behaviour observed during the training period.

Then by giving these log likelihood values a sign such that if the value(s) are greater than normal the sign is positive, and if the value(s) are less than normal then the sign is negative, this allows the pictorial representation in Figure 55 to be produced. In this figure, larger bars correspond to less likely probability metrics (i.e. more significant from a change perspective), and the direction of the bars indicates whether the usage is more or less than normal, with less usage bars going to the left, and more usage to the right. In the example shown Figure 55, most bars are to the right, indicating more power was used than normal during most of the day. In our experience, half hourly magnitudes of log likelihood less than or equal to 0.8 are typical, whereas magnitudes greater than this were less common and potentially more indicative of a significant change.

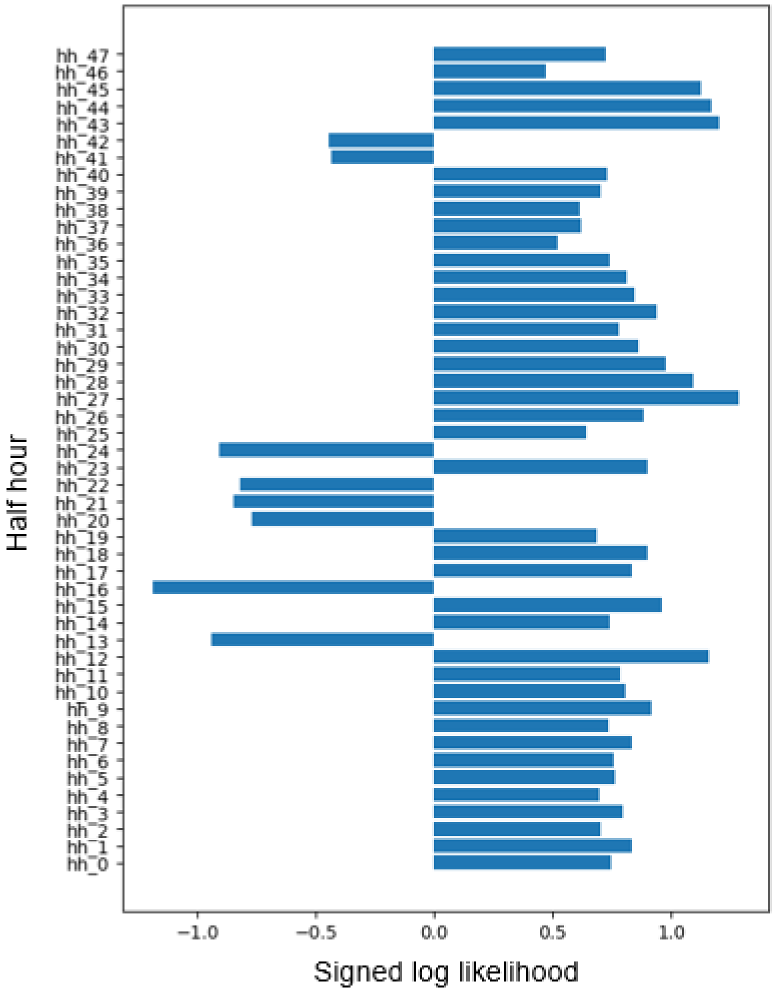

Figure 55: Example change identification plot.

6.3.2 Examples of Change Identification

Following the work reported in the Understanding Vulnerability and Energy Usage report (Lily Darling, Elsie Roberts and Tom Saunders, March 2022), we identified a few vulnerability profiles that may be detectable from changes in overall energy usage. The following subset of cases were chosen to be investigated as part of the model development work:

1. Poor mental health, leading to abnormal and increased energy usage at night time.
2. Job loss, leading to increased daytime energy usage due to not being at work.
3. Decreased usage at the end of the month, where a customer intentionally decreases usage or self-disconnects out of fear of being unable to pay the bills, until they are paid at the end of the month.

Each of these cases were judged to be potentially detectable from the usage patterns recorded by smart meters. However, at the time of writing, recorded data corresponding to these vulnerability profiles is not available, so therefore we were required to generate artificial data profiles which correspond to customers becoming vulnerable in these ways.

© FNC 2022

Artificial data profiles corresponding to these vulnerabilities were constructed by scaling certain half hours of a recorded data block for a household. The scale factors were drawn from a normal distribution centred around the desired scale factor, with a different factor drawn for each half hour. This means that the scaling is not constant, as would be the case in a real scenario. The scaled data blocks were then appended to their equivalent non-scaled ones, to detect a change from normal usage for the household to the household becoming vulnerable in the prescribed way.

6.3.2.1 Increased Night-time Usage

The first example of a customer becoming vulnerable is when due to poor mental health, the customer begins to stay up at night and use more energy than normal. Artificial data was generated for this case as discussed above, and the short term change detection method was trained on the non-scaled data for the customer. Figure 56 shows the change detection plot for this case, which shows a very evident drop in probability when the increased night time usage begins.

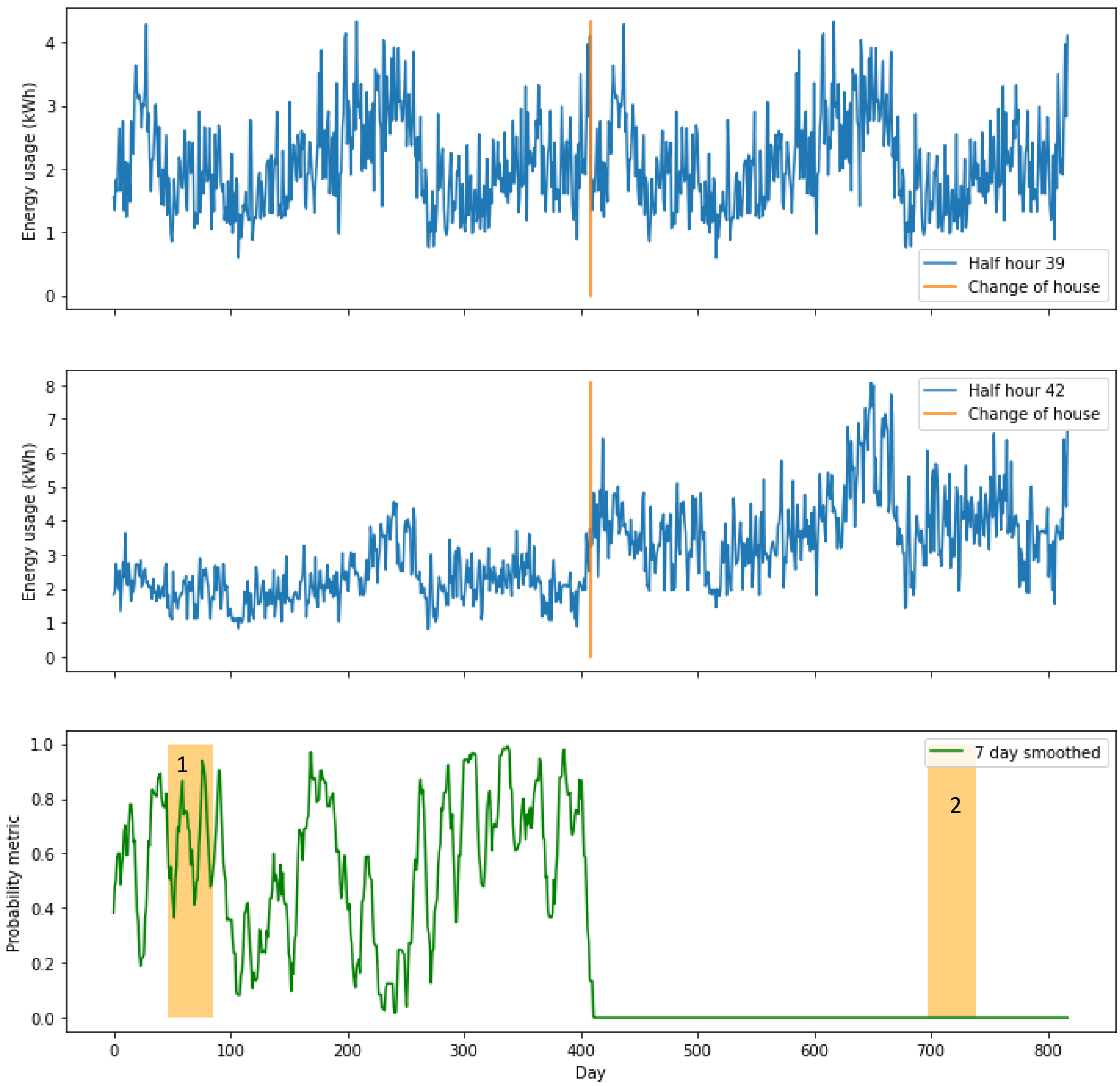

Figure 56: Change detection for increased night time usage (poor mental health).

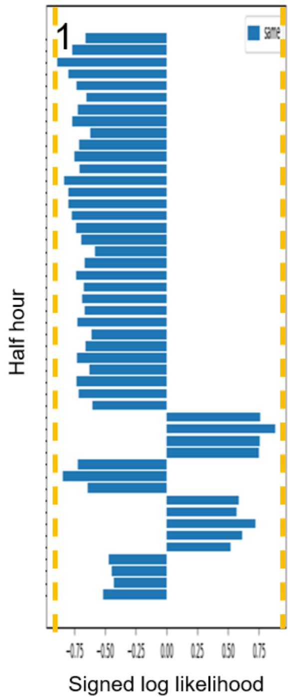

Figure 57: Change fingerprint for region 1 where there is no change.

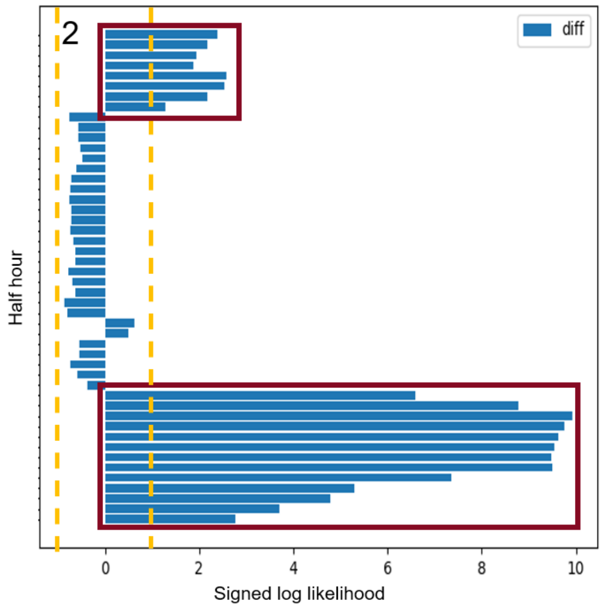

Figure 58: Change fingerprint for region 2 where there is increased night time usage.

Figure 57 shows the change fingerprint plot for the period marked in green with a ‘1’ in Figure 56, which shows small bars with relatively insignificant changes. However, for the green region ‘2’, where the increased night time usage is occurring, Figure 58 shows the equivalent finger print plot. This shows consistently large positive bars during the night time half hours, highlighted in the red boxes. This shows that the method has successfully detected both that the change has occurred, and secondly that the method was able to identify the change as increased night-time usage, which is an indicator of poor mental health.

6.3.2.2 Decreased Usage at the End of the Month

The next notable example of a potentially detectable vulnerability profile is where the customer intentionally reduces their energy consumption while waiting for pay day at the end of the month. Artificial data was again produced for this profile, scaling day time energy usage down for the last week of each month. Figure 59 shows the change identification plot which was calculated for this case. The plot shows relatively normal fluctuations for the first half of the data, but during the second half, when the reduced usage at the end of the month is occurring, there are regular and clear dips at the ends of the months, indicating statistically significant changes in usage.

Figure 60 shows the fingerprint plot for the green region labelled ‘1’ in Figure 59. It shows bars in both directions, most of which are not significantly long. For the region marked ‘2’, which covers one of the end of month dips, Figure 61 shows the usage is consistently lower than normal, with larger bars indicating more unusual behaviour. The combination of regular dips at the end of month, combined with Figure 61 indicating that the dips are due to a consistent and significant reduction in usage, mean that this vulnerability profile can be detected using the methods developed.

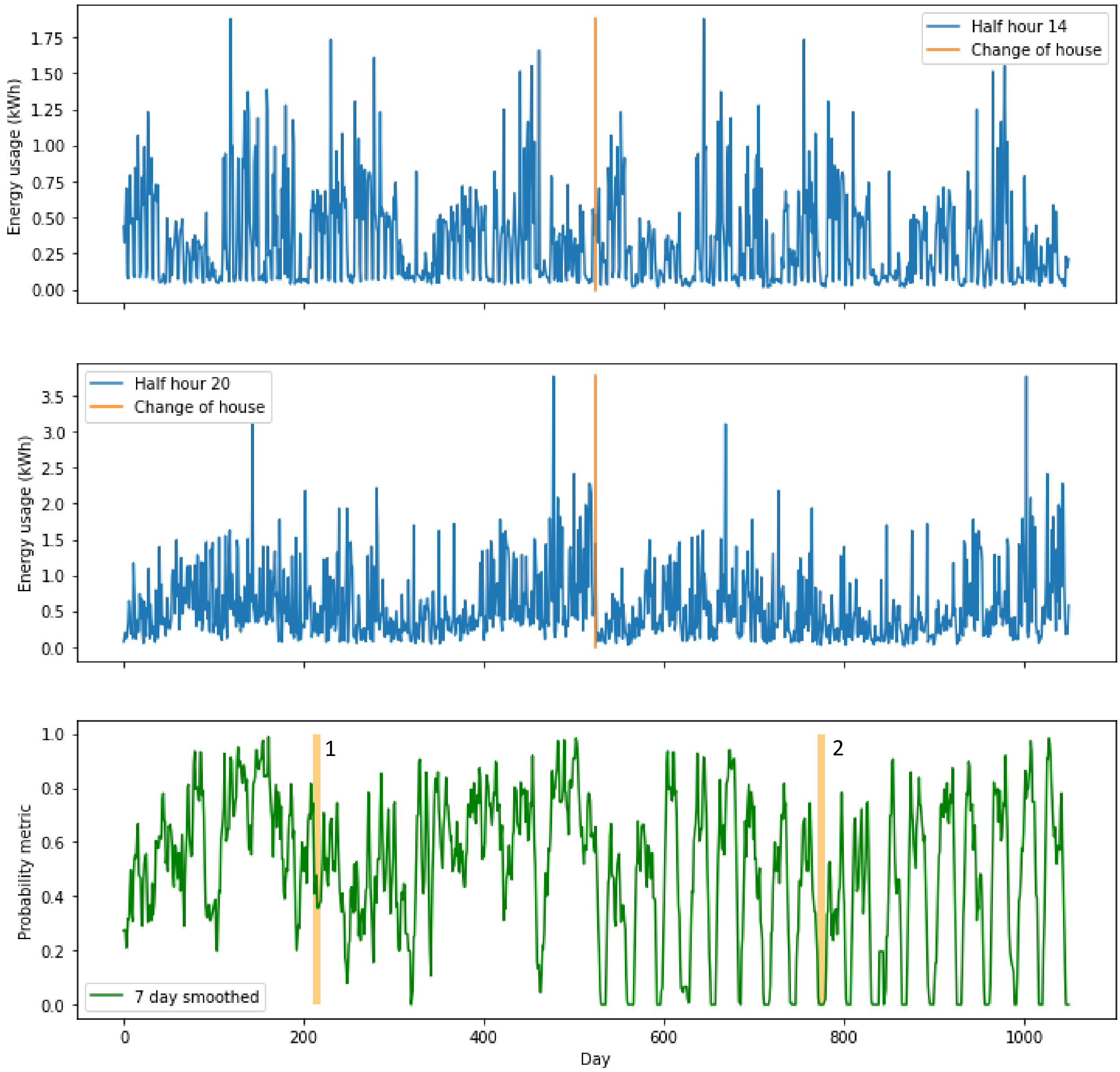

Figure 59: Change detection for increased decreased usage at the end of the month.

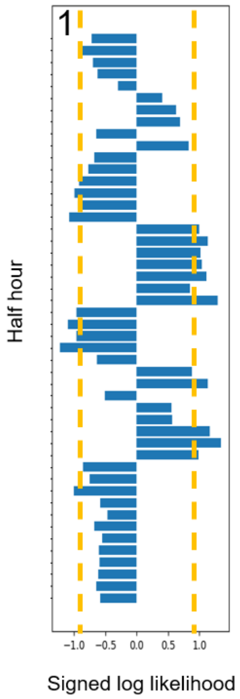

Figure 60: Change fingerprint for region 1 where there is no change.

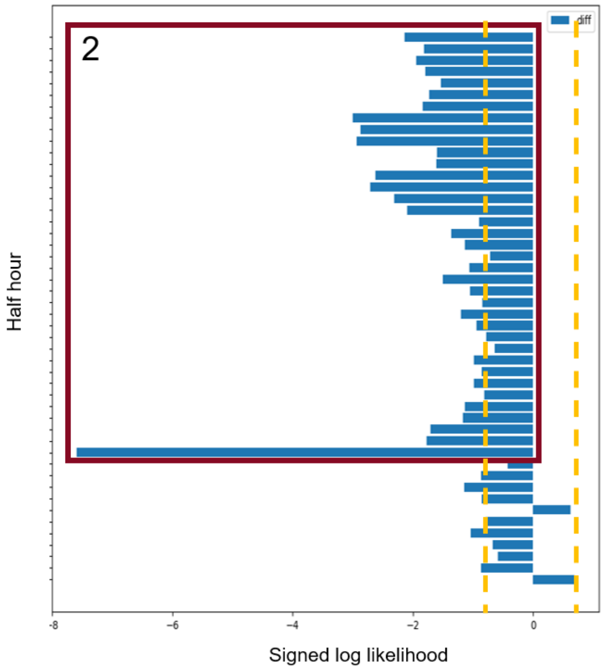

Figure 61: Change fingerprint for region 2 where there is decreased usage at the end of the month.

6.3.2.3 Increased Day-time Usage

The final example investigated is where a customer has lost their job, consequently spending their day time at home, which leads to increased day time energy usage. Figure 62 shows the change detection plot produced by the long term method on this occasion, which again shows a marked drop in probability in the second half of the plot, where the scaling has been applied.

Figure 63 shows the corresponding fingerprint plot for the region marked in orange in Figure 62. It shows consistently large and positive bars during the half hours corresponding to a normal working day, indicating that significantly more energy has been used during this period. This again means that the change has been successfully detected and can be characterised as consistent with this change in vulnerability profile.

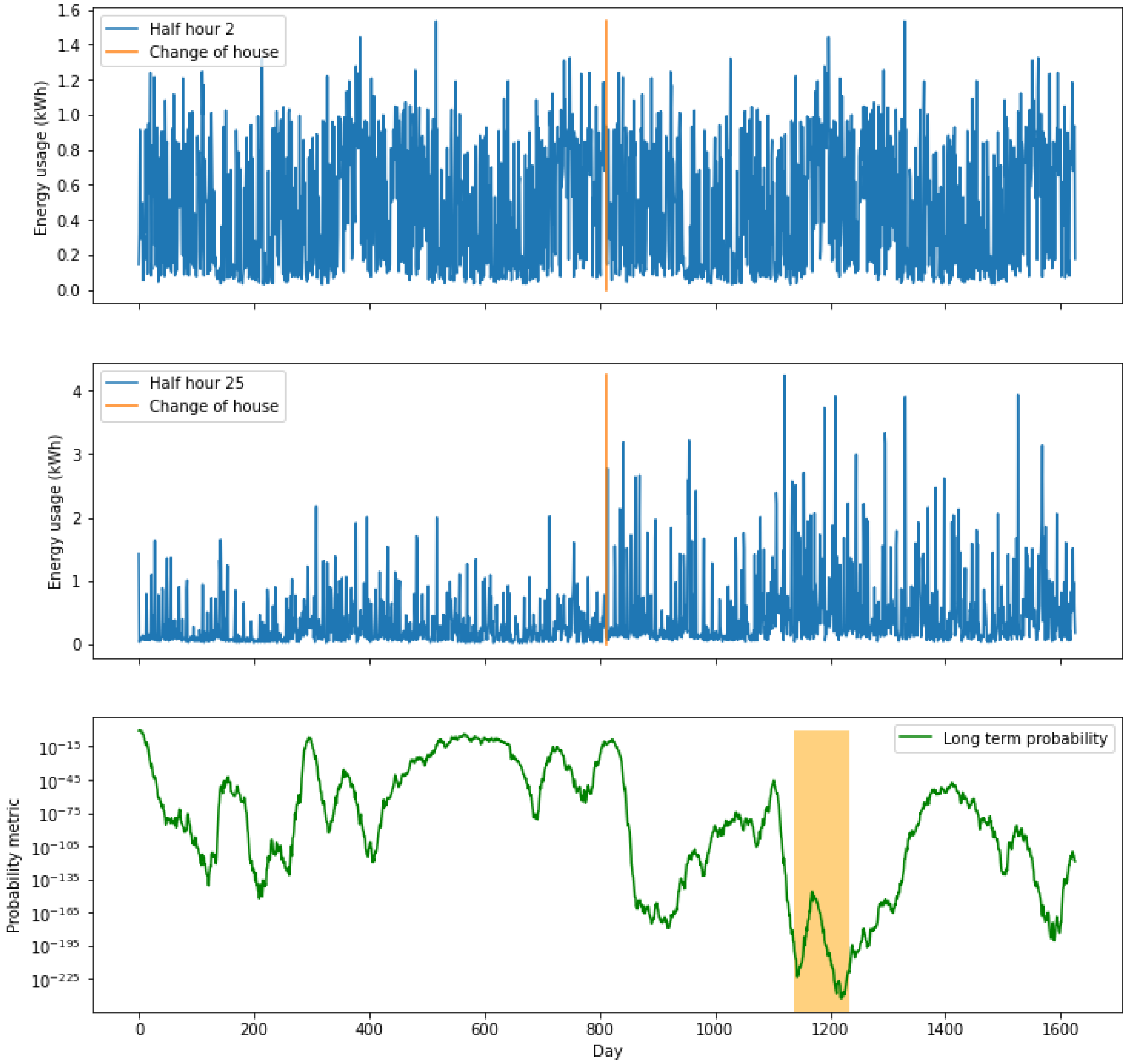

Figure 62: Change detection for increased day time usage (job loss).

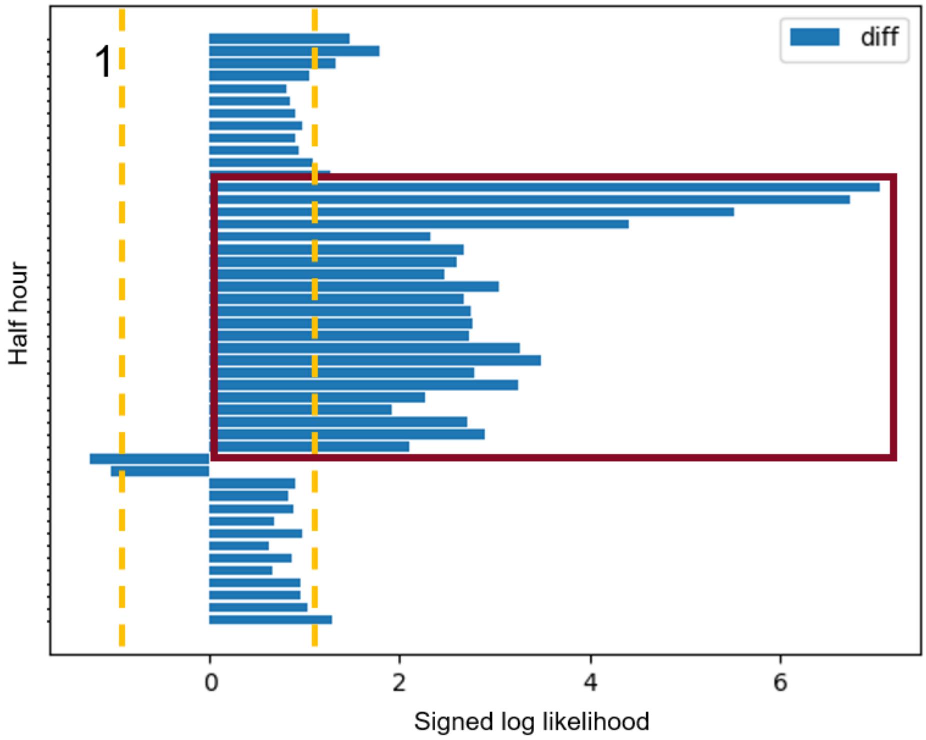

Figure 63: Change fingerprint for increased day time usage (job loss).

6.4 Conclusions

The following conclusions can be made from the change detection and identification model development work:

- Two methods have been developed to detect statistically significant changes in energy usage patterns for a given household. The methods have been tested and compared and found to work well at detecting changes between houses.
- A further method has been developed to characterise what has changed. This method has been found to also work well, and enable identification of changes consistent with changes that may be due to an increase in some vulnerability profiles
- All the conclusions above are subject to the caveat that while the methods have been developed and tested using real smart meter data, we do not currently have access to data from customers with known vulnerabilities. Therefore, while the methods show significant promise, it is vital to validate them with data corresponding to customers with known vulnerabilities before the methods could be developed to a production-ready state.
- Once validation data is available, further tweaks to the methods would be beneficial to identify improvements and tune sensitivities to increase accuracy and reduce the rate of false positive detections of vulnerability.

Overall, noting the absence of vulnerability data, the methods were shown to be effective at detecting changes in usage, and show promise in assisting the identification of vulnerabilities while also giving the possibility of screening out false positive detections.

6.5 Future Work

In this section, we have described the prototype algorithms that we have developed, which have demonstrated using artificial data that the concepts for detecting changes in vulnerability are workable.

The key next steps are as follows:

- Acquire validation data from customers with vulnerabilities. The requirements for this are described in Section 6.5.1.
- Improve the algorithms and validate them using real or artificial data informed by the validation data, as described in Section 6.5.2.
- Integrate the algorithms into a framework which can incorporate other information, as described in Section 6.5.3.
- Develop a framework for how best to support vulnerable customers, as discussed in Section 6.5.4.

6.5.1 Validation Dataset Requirements

The key next step for the model development is to validate the methods using data from customers with vulnerabilities. This would be to check that the usage patterns of vulnerable customers are as expected in this report, and to confirm that the methods are sufficiently sensitive to detect the vulnerabilities. The validation data would need to meet the following requirements:

- Half-hourly electricity usage data over a period of one or more years.
- Gas usage data would also be beneficial but not required.
- Data for customers *becoming* vulnerable – including at least several months' worth of data before and after the onset of vulnerability.
- For vulnerability profiles matching at least one of the cases described in Section 6.3.2.
- Labelled data, such that it is clear which customer usage history corresponds to vulnerability beginning at roughly which time.
- Parallel data for customers which are not vulnerable from similar locations and times (this is desirable but not necessary).

6.5.2 Model Improvement and Validation

Following initial validation, the tweaking the model hyperparameters to identify suitable length scales for training periods, rolling averages and other aspects would further improve the methods, along with exploring the use of kernel density approaches rather than fitted log normal distributions. Improved validation data could then be used to confirm that the tweaks are beneficial. This validation data would either consist of real data as described in Section 6.5.1 if sufficient data could be obtained, or it would be generated from the real data using statistical techniques, to ensure that the data accurately simulates the way that vulnerability changes based on the real data and expert insight.

6.5.3 Integration into Wider Framework

If additional personal data could be known for the customers in the validation data, this would allow the scoping and development of methods to integrate the change detection methods into a wider Bayesian framework, accounting for priors on vulnerability based on the additional personal data. This personal data could include age, gender, occupation, address, income level, payment methods, credit score and more, but noting that one or more of these may not be available or appropriate for privacy reasons, this may not be fully practical.

6.5.4 Development of Support Approaches

Finally, developing the modelling techniques are only worthwhile if a framework can be developed for how to support the customers that have been identified as potentially vulnerable by the algorithms. It would be vital to consider what types of support would be appropriate, how to deal with false positive detections of vulnerability, and what the thresholds for intervention would be.

7 Conclusions and Recommendations

This report presents the analysis undertaken to determine whether vulnerabilities could be identified from a household's smart meter data. The research conducted in (Lily Darling, Elsie Roberts and Tom Saunders, March 2022) determined the types of behavioural characteristics that vulnerable consumers may show in their smart meter data. As a result, three separate models were developed, with each one identifying a different type of behavioural characteristic. These are:

1. Appliance Disaggregation and Prediction: Used to identify vulnerabilities related to appliance usage.
2. Cohort Comparison: Used to identify vulnerabilities related to the overall level of usage a specific household has.
3. Overall Changes in Usage: Used to identify vulnerabilities related to a significant change in its occupants' behaviour.

Each model was developed and tested in isolation using different, open source, datasets. None of the models have been validated using real household data with known vulnerabilities, due to difficulties obtaining this data. Details of the approach, conclusions and recommendations for each model is given below.

7.1 Appliance Disaggregation and Prediction

7.1.1 Overview

Following a review into the challenges faced in current research into appliance disaggregation, we decided to develop a probabilistic model. This allowed us to directly account for all the uncertainties in the model development, rather than building a model that works perfectly with the limited, very specific, datasets we had.

Two open source datasets were used to develop and test the model. The datasets contained appliance load monitoring for a variation of appliances over multiple years within 25 different houses in the London and Loughborough regions. After the datasets were pre-processed, we developed an algorithm to identify appliance usages and extract distributions for: how long the appliance is used for, when during the day it is used, and the average power drawn by the appliance when in use. These formed a set of usage statistics for appliances within the datasets.

The appliance disaggregation model uses the usage statistics to predict the likelihood of every possible appliance combination, given the power observed in a 30-minute window. These probabilities are used to predict which combination of appliances were used. After many days of data for a single household, the model calculates the probability each appliance is used in each 30-minute window, and this is the final output. These probabilities can be compared between households and conclusions drawn as to whether the different behaviours are signs of vulnerability or not. For example, a medical appliance would have a continually low probability for each 30-minute window for a normal household, whereas after the model was run for a household with a medical appliance, the probability would increase significantly.

7.1.2 Conclusions

The model was developed for appliances which draw one level of power when turned on (single step), and it was tested with the following appliances: kettle, microwave, toaster, and vacuum. The testing was completed by aggregating the open source data on power for just the appliances being considered and a constant baseload.

The model was tested with the same day repeated multiple times, and then the results were reviewed when it was given a month of data. The testing on a repeated day showed the model performs as anticipated, namely, it correctly uses the historic behaviour of the household to inform the probability of appliance usage for the next day. The results from when the model was run with one month of data showed that model was still working as anticipated but the uncertainty in the power drawn by the microwave and toaster were limiting prediction capability.

7.1.3 Recommendations

These conclusions show that the modelling techniques are working, but the uncertainty in power drawn by appliances is the largest limiting factor. To improve this, some improvements can be made to the model. Most notably, the functionality can be increased to include additional factors for the model to learn from a household. For example, it could learn the power drawn by the toaster based off historic usage and then this would reduce the uncertainty in prediction through time. If these, and the additional developments listed in Section 4.6, were implemented, then the model could be tested further with medical appliances and additional, more complex, household appliances.

7.2 Cohort Comparison

This model aims to determine a households anticipated electricity usage based on its characteristics and compare this to the known average usage. This accounts for all comparative usage level vulnerabilities, for example, low winter usage my imply they are struggling with payments.

7.2.1 Overview

This model was developed to predict the average usage of a household, given its known characteristics. 17 characteristics were included in total, and each was thought to be readily available household information and to have an impact on energy usage. For example, EPC rating, number of inhabitations and average income.

The model was trained and tested using open source LSOA data, which contained average household usage across all households in each LSOA (approximately 800). Each LSOA was disaggregated into artificial sets of people and households, and each person was randomly assigned to a household. This produced a set of households which had the same average and total percentage population and characteristics as the LSOA was known to have. The model predicted the usage for each artificial household and then averaged these to compare to the known average household usage for the LSOA. This was completed for 1000 LSOAs, with 800 in the training test and 200 in the testing.

7.2.2 Conclusions

The model concluded that some of the variation in average household usage could be attributed to known household characteristics. The results were never anticipated to be perfect due to the uncertainty in LSOA household disaggregation and the unpredictability of human behaviour. However, this conclusion showed that some of the variation in household usage could be attributed to its known circumstances.

7.2.3 Recommendations

The model was not developed further due to the inherent uncertainties in the proxy dataset. It is essential that the average energy usage for individual households with known characteristics is obtained before developing this model further.

If obtained, this would enable us to develop and test the model's prediction capability, and the inherent, currently unaccounted for, uncertainties in the method could be quantified. If this were completed, the model would provide a distribution of average energy for the household with the width of this distribution due to the unpredictability human behaviour. Households which were outliers to this distribution could then be investigated as potentially showing characteristics of vulnerability.

7.3 Overall Changes in Usage

This model aims to identify and characterise significant changes in a household's usage patterns. This accounts for time dependent and transient vulnerabilities, for example, job loss.

7.3.1 Overview

The model detects statistically significant changes in energy usage patterns for a given household, and then characterises what has changed.

The model works by creating a distribution of energy usage over a baseline period for each house for each half hour. Then, for a given day's energy usage, each half hour's usage can be compared with the baseline distributions to form a metric which quantifies the extent to which the usage fits the pattern observed in the baseline period. These half hourly metrics can be combined to form an overall metric where a high value indicates no significant change to usage and a low value indicates that a change is more likely. In addition, they can be plotted individually to show when identified changes are coming from, whether they indicate higher or lower usage than normal, and whether they fit a pattern that would be characteristic of vulnerability.

All development and testing was completed using open source smart meter data from UK Power Network's Low Carbon London project (SmartMeter Energy Consumption Data in London Households, 2014).

7.3.2 Conclusions

The model's ability to recognise a step change was tested by combining two randomly selected houses and predicting the time when the houses changed. This showed that for most household combinations, the model predicted correctly when the households switched. The houses where the model did not predict the change well were those that had very sporadic and varied behaviour.

The model's ability to quantify the reason for the change was tested by modifying the household smart meter data to include a known characteristic, for example, increased daytime usage. The original and modified datasets were combined, and the model determined whether there was a statistically significant increase or decrease in usage for each 30-minute window. When the results of this test are manually compared to the known added characteristic, the model correctly identified which characteristics had changed.

All of the conclusions above are subject to the caveat that while the methods have been developed and tested using real smart meter data, we do not currently have access to data from customers with known vulnerabilities. Therefore, as detailed above, the real data was adjusted in ways known to correspond to usage patterns of vulnerable data. These changes were successfully detected but it is important to note that, while the methods show significant promise, it is vital to validate them with data corresponding to customers with known vulnerabilities. Most notably, this is because the magnitude and consistency of the expected changes in vulnerable customers is currently unknown.

Overall, noting the absence of vulnerability data, the methods were shown to be effective at detecting changes in usage, and show promise in assisting the identification of vulnerabilities while also giving the possibility of screening out false positive detections.

7.3.3 Recommendations

If this model was to be developed further, suitable validation data much be obtained. This should be smart meter data where it is known if the household is vulnerable, and when a change in circumstance occurred. Once validation data is available, further developments to the methods would be beneficial to identify improvements and tune sensitivities to increase accuracy and reduce the rate of false positive detections of vulnerability.

Finally, if additional personal data could be known for the customers in the validation data, this would allow the scoping and development of methods to integrate the change detection methods into a wider Bayesian framework, accounting for priors on vulnerability based on the additional personal data, and potentially making the model significantly more effective.

7.4 Conclusion

Three models were developed to identify different behavioural characteristics that may be an indicator of vulnerability from a household's smart meter data. All three models showed that the behavioural characteristic was detectable using the data sources available. The next step would be to test each model using real world data, however, this has not been possible to date due to lack of smart meter data for normal and vulnerable households. If this information were obtained, and the model's capability assessed, they could be combined and deployed as a novel and efficient means to detect household vulnerability, to aid the DNOs discharging their responsibility.

8 References

BEIS. (March 2022). *Smart Meter Statistics in Great Britian: Quarterly Report to end December 2021.* BEIS.

David Murray et al. (2015). A Data Management Platform for Personalised Real-time Energy Feedback. In *Proceedings of the 8th International Conference on Energy Efficiency in Domestic Appliances and Lighting.*

Hana Altrabalsi et al. (December 2014). *Low-complexity Energy Disaggregation using Appliance Load Modelling.* University of Strathclyde.

J. Paxman, M. Jones, Dr E. Constanza and J. Mannin. (November 2020). *Smart Future of Healthcare.* 2020Health.

Lily Darling, Elsie Roberts and Tom Saunders. (March 2022). *VENICE: Understanding Vulnerability and Energy Usage.* 018037-103/52287R Issue 2: Frazer-Nash Consultancy.

Nomis. (2011, July). *Nomis*. (ONS) Retrieved 02 2022, from https://www.nomisweb.co.uk/

Office for National Statistics. (2011, July). *ONS*. (UK Government) Retrieved 02 2022, from https://www.ons.gov.uk/

Ofgem. (October 2019). *Consumer Vulnerability Strategy.* OFG1161.

Revolution, C.-L. N. (2022). *Customer-Led Network Revolution Project Data*. (Customer-Led Network Revolution) Retrieved June 22, 2022, from http://www.networkrevolution.co.uk/resources/project-data/

scipy.stats.combine_pvalues. (n.d.). Retrieved March 16, 2022, from https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.combine_pvalues.html

scipy.stats.kstest. (n.d.). Retrieved March 17, 2022, from https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.kstest.html

SmartMeter Energy Consumption Data in London Households. (2014). Retrieved from https://data.london.gov.uk/dataset/smartmeter-energy-use-data-in-london-households

UK Energy Research Centre Energy Data Centre. (2017). *UK Domestic Appliance Level Electricity (UK-DALE) - High speed (16 kHz) aggregated whole house current/voltage*. (Jack Kelly, Imperial College) Retrieved September 2021, from https://data.ukedc.rl.ac.uk/browse/edc/efficiency/residential/EnergyConsumption/Domestic/UK-DALE-2017/UK-DALE-FULL-disaggregated

Western Power Distribution. (July 2021). *Smart Meter Strategy.* WPD.

A.1 Appliance Power Probability Distributions

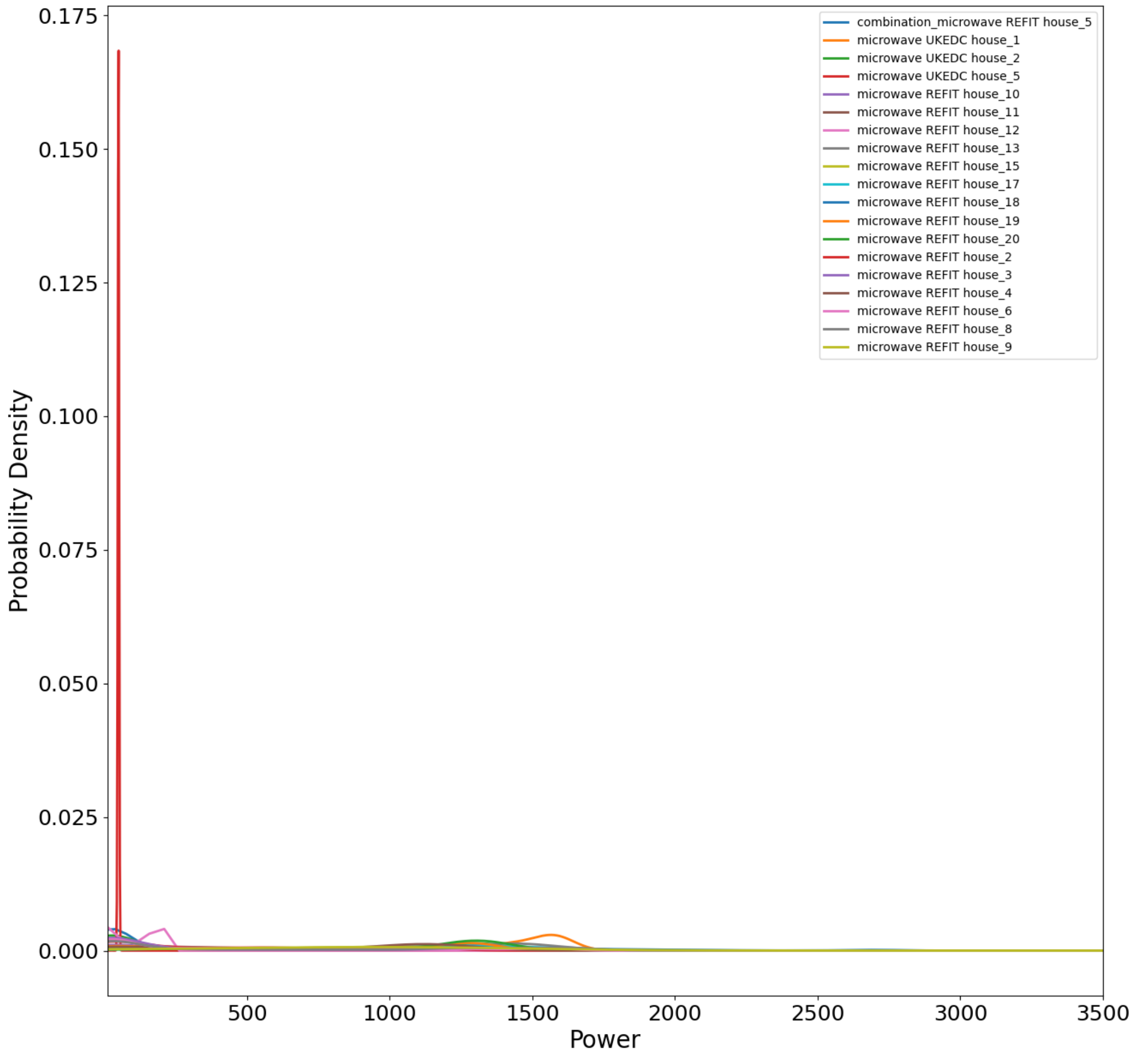

Figure 64: Gaussian kernel-density estimates of the power drawn by each appliance within the microwave category.

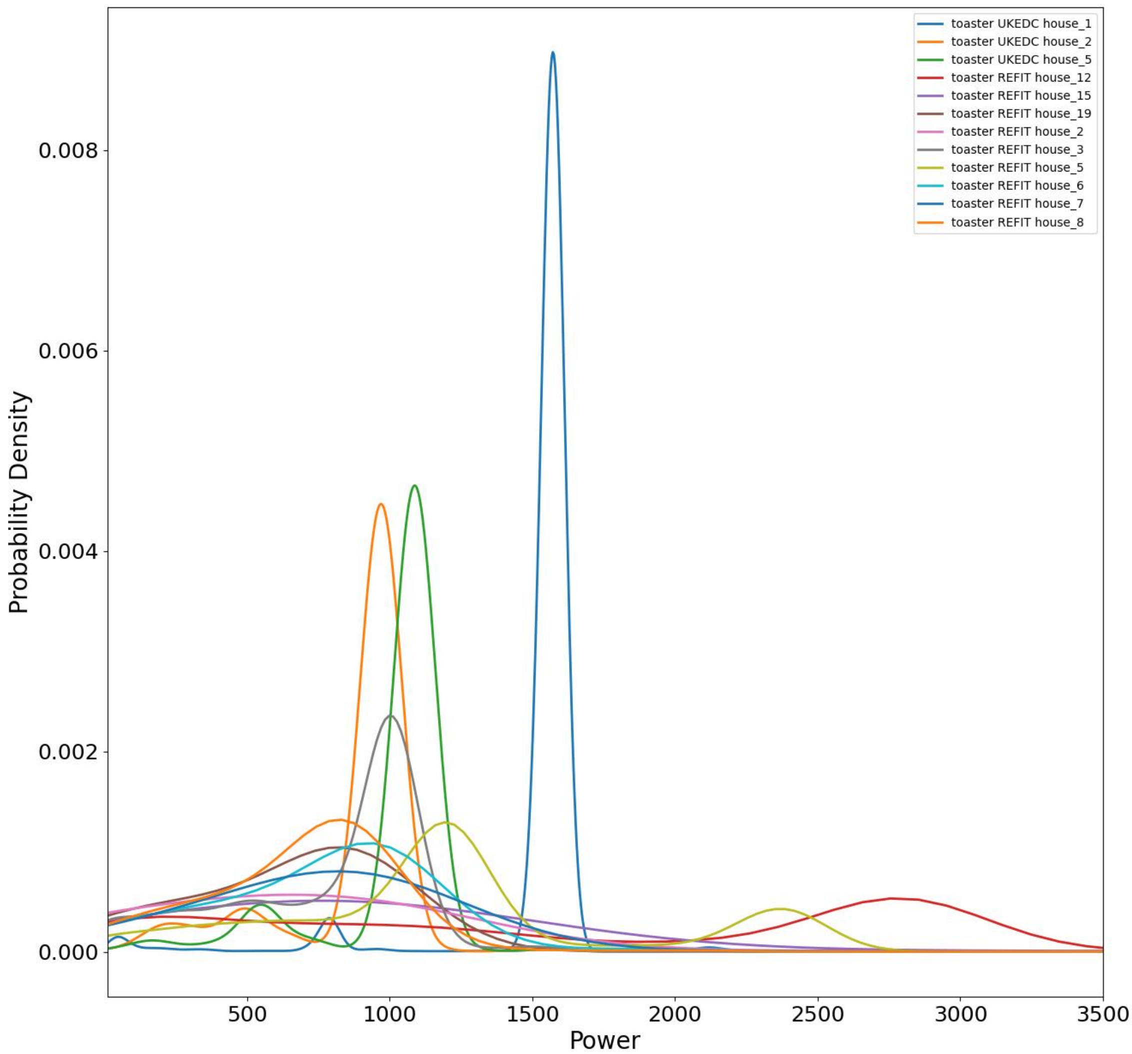

Figure 65: Gaussian kernel-density estimates of the power drawn by each appliance within the toaster category.

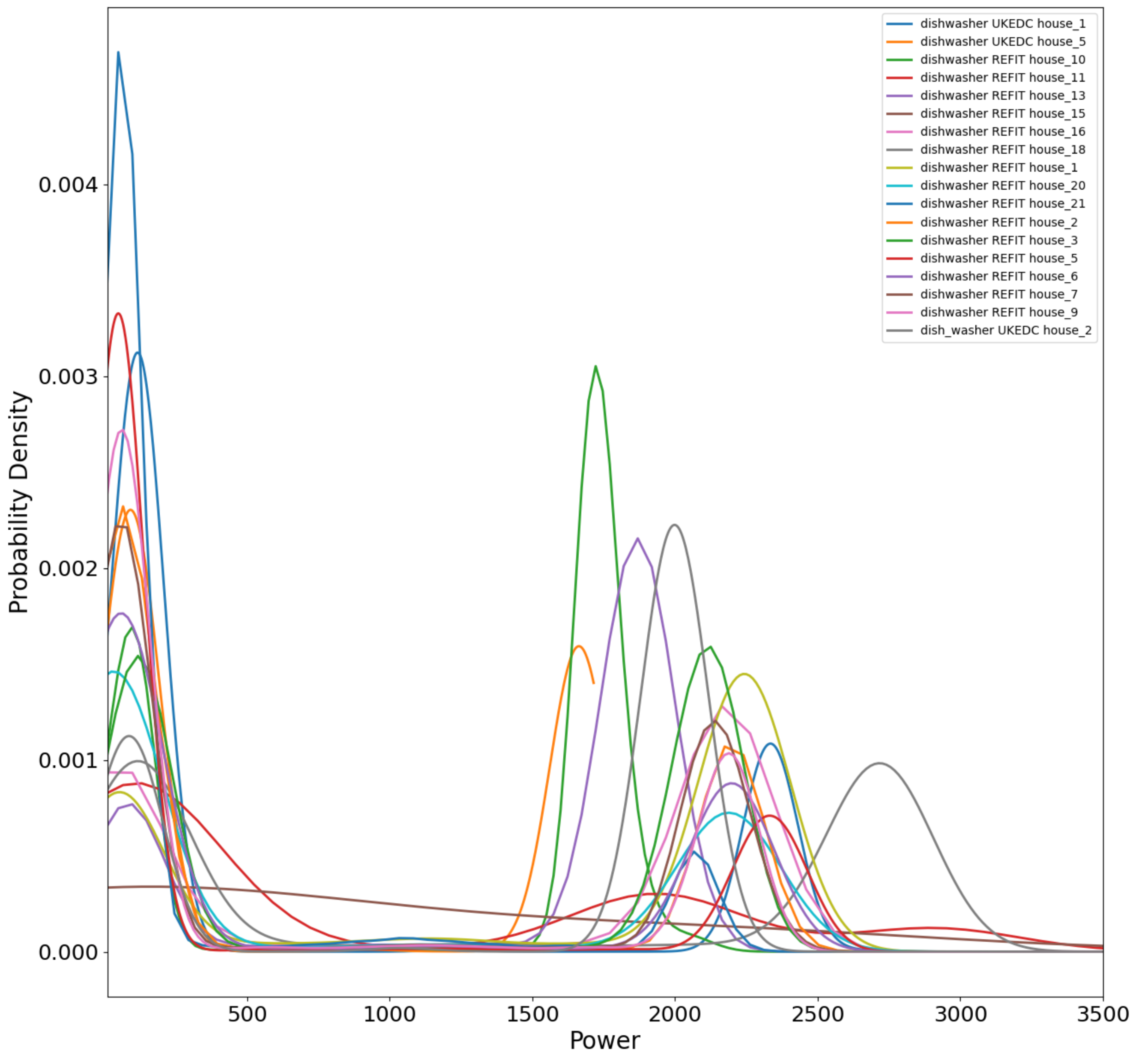

Figure 66: Gaussian kernel-density estimates of the power drawn by each appliance within the dishwasher category.

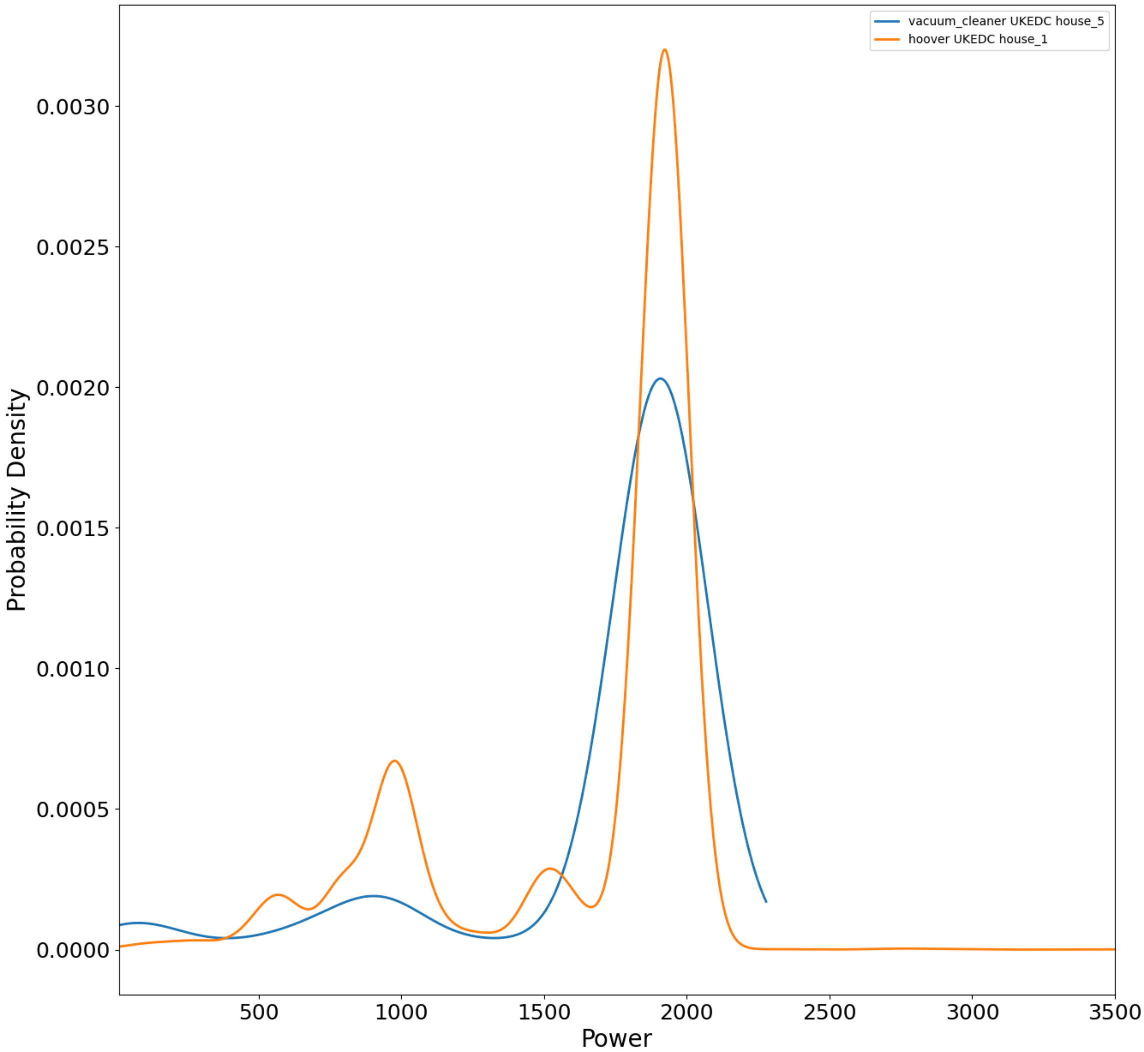

Figure 67: Gaussian kernel-density estimates of the power drawn by each appliance within the vacuum category.

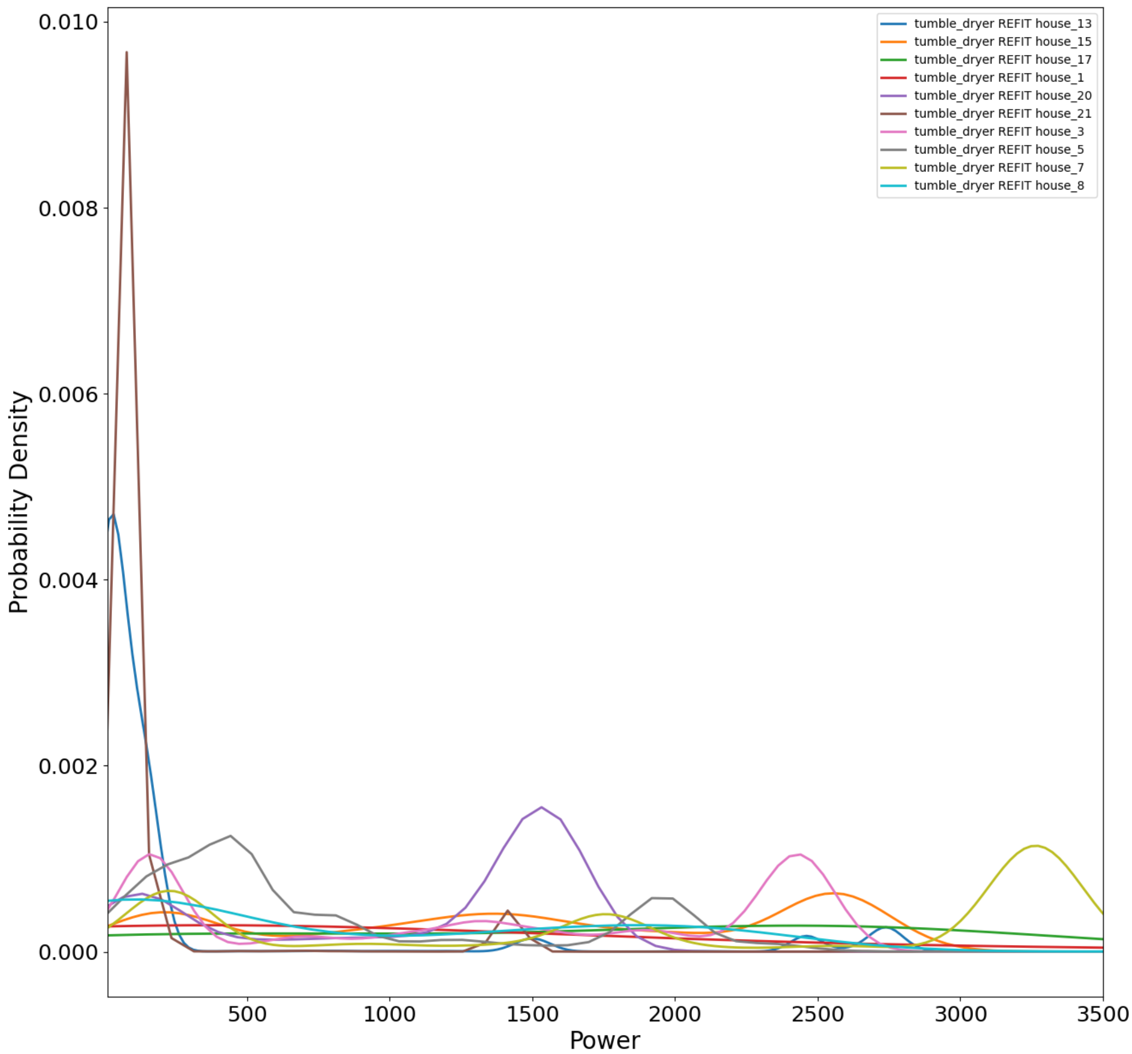

Figure 68: Gaussian kernel-density estimates of the power drawn by each appliance within the tumble dryer category.

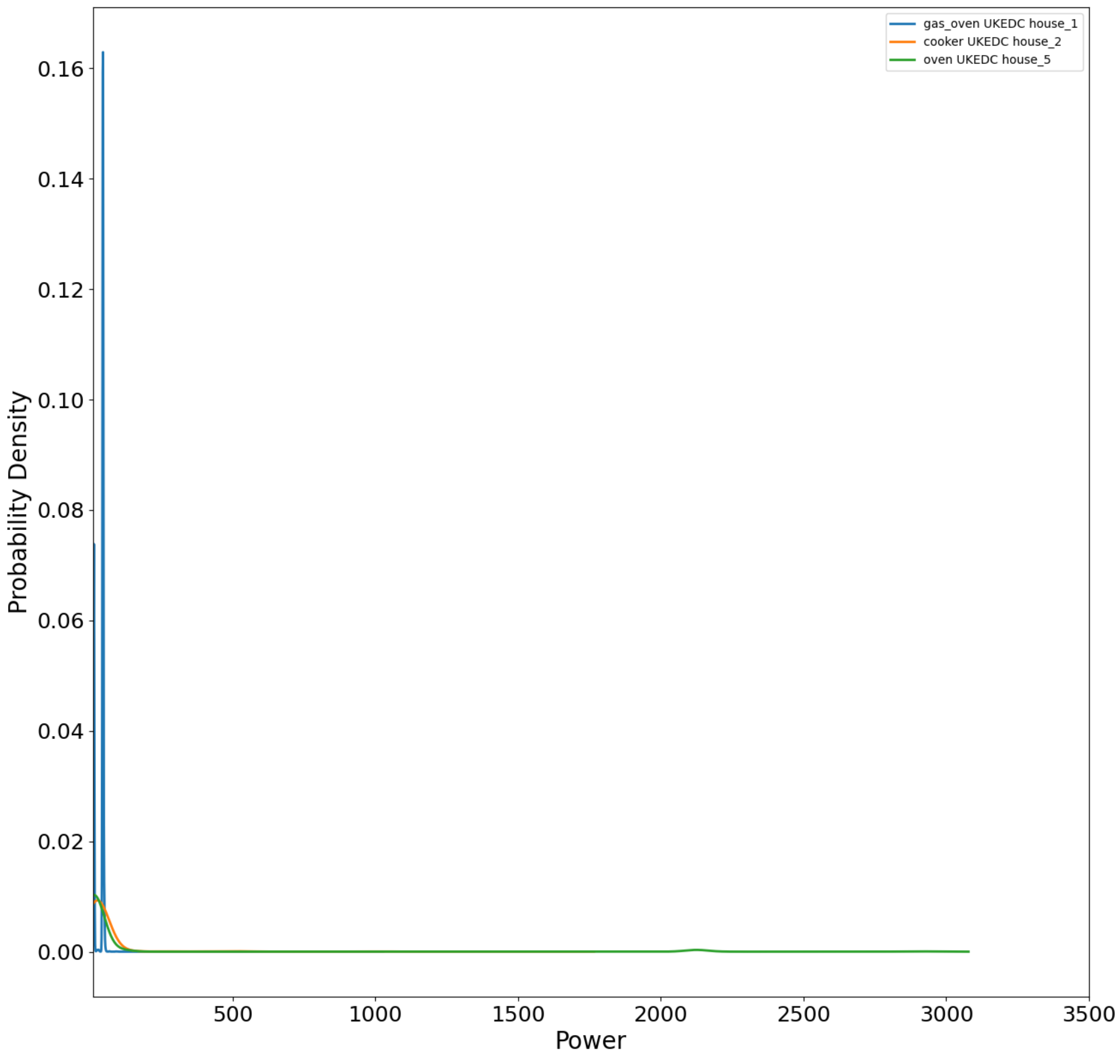

Figure 69: Gaussian kernel-density estimates of the power drawn by each appliance within the oven category.

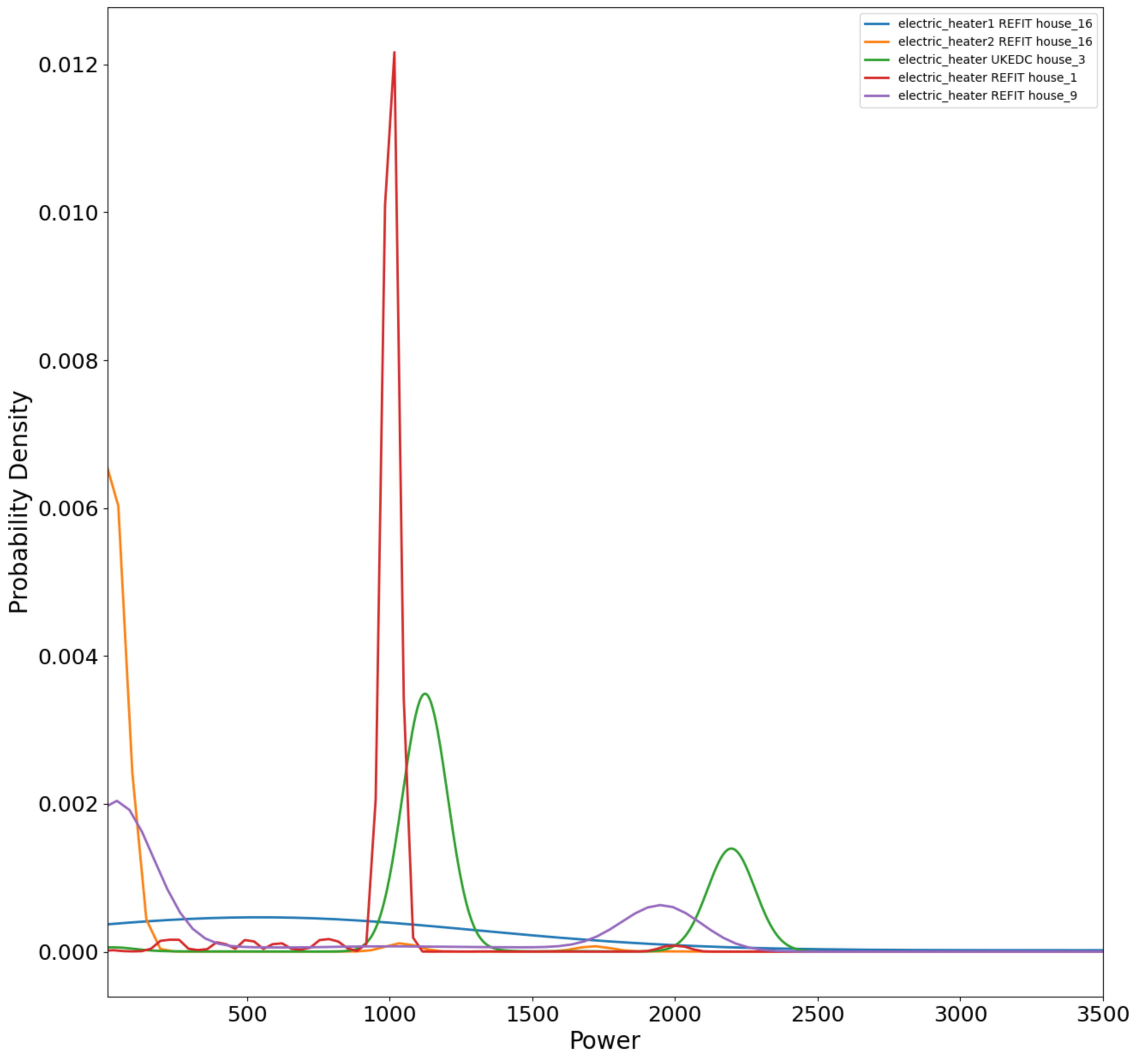

Figure 70: Gaussian kernel-density estimates of the power drawn by each appliance within the electric heater category.

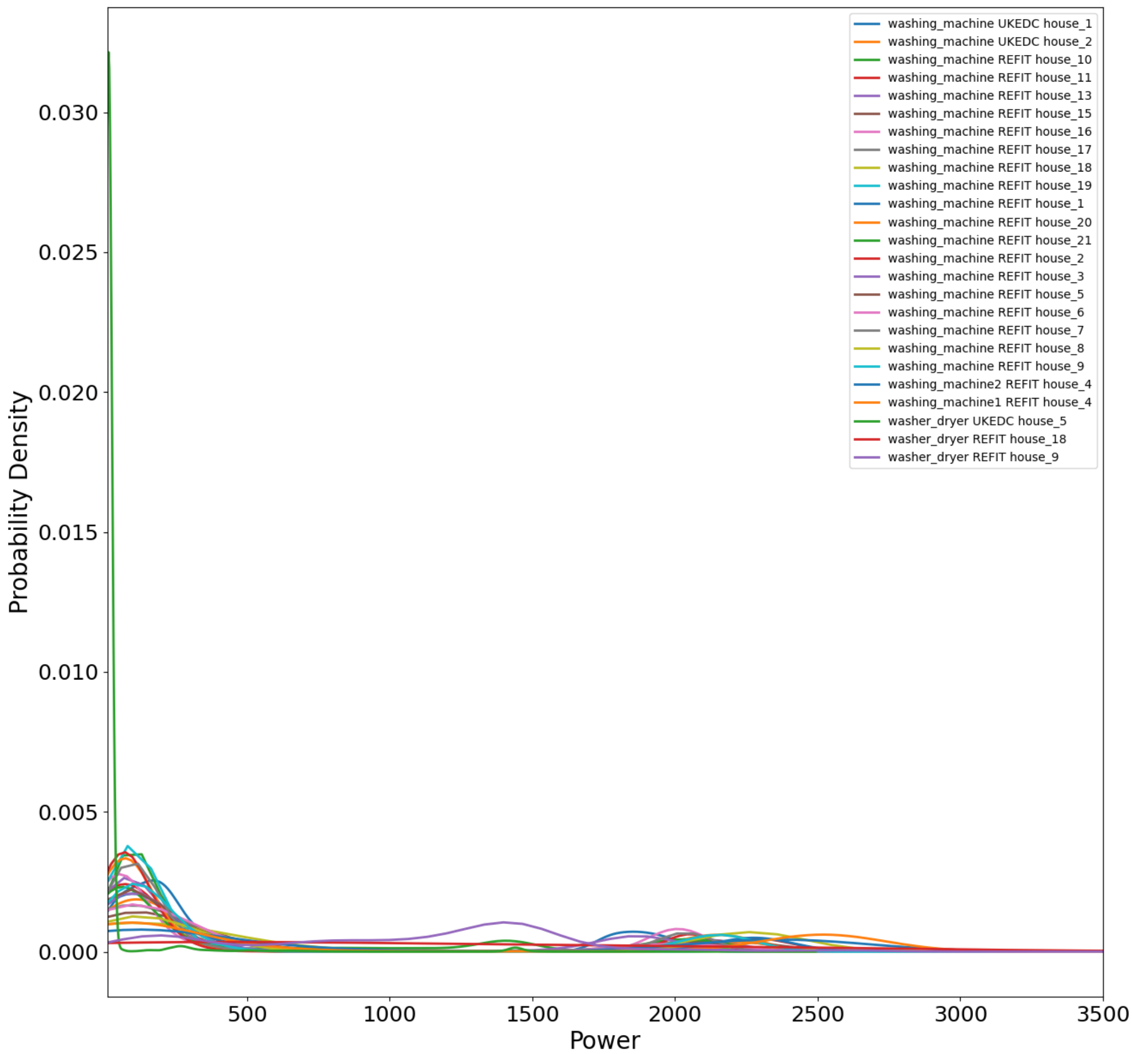

Figure 71: Gaussian kernel-density estimates of the power drawn by each appliance within the washing machine category.

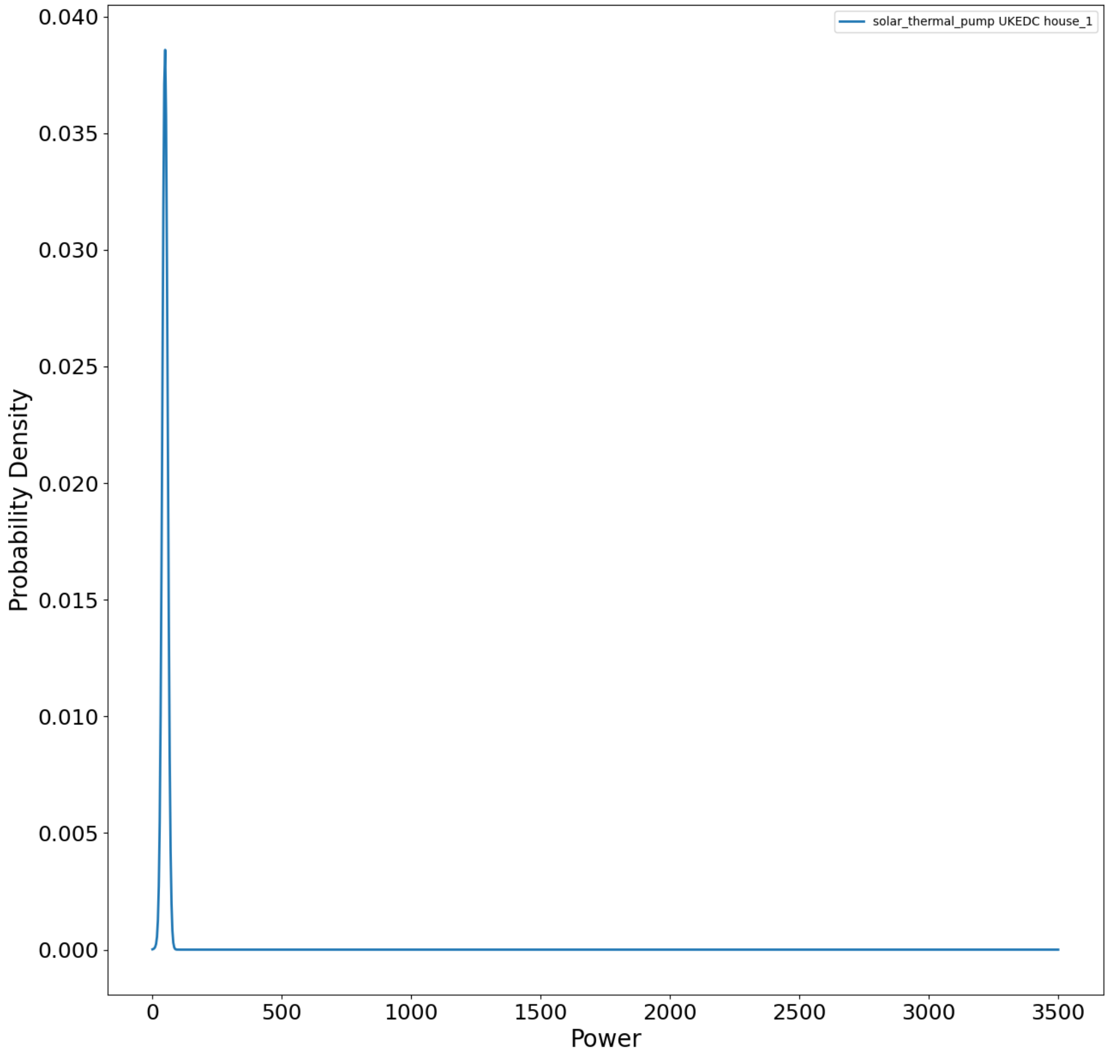

Figure 72: Gaussian kernel-density estimates of the power drawn by each appliance within the heat pump category.

A.2 Appliance Usage Statistics

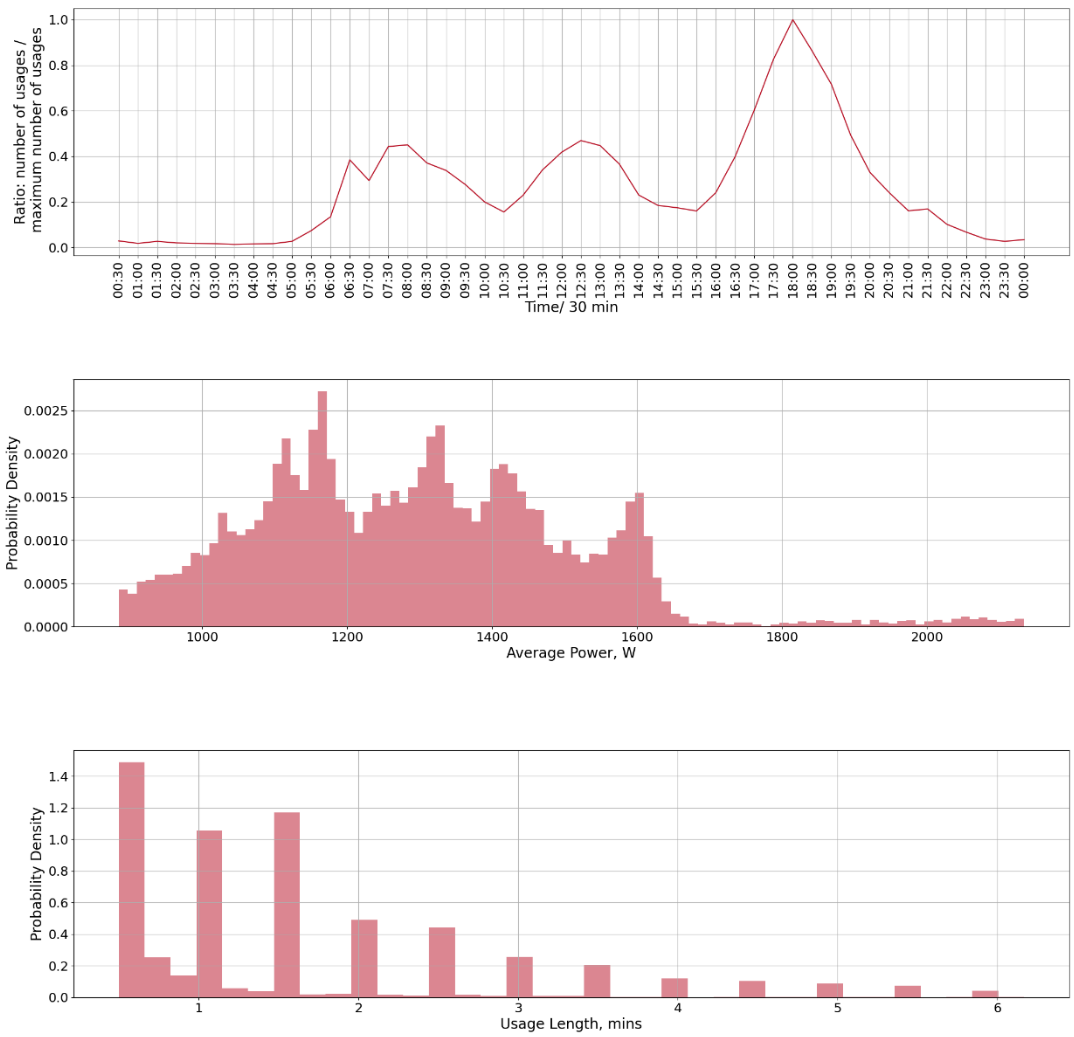

Figure 73: Usage statistics extracted for the microwave across all datasets.

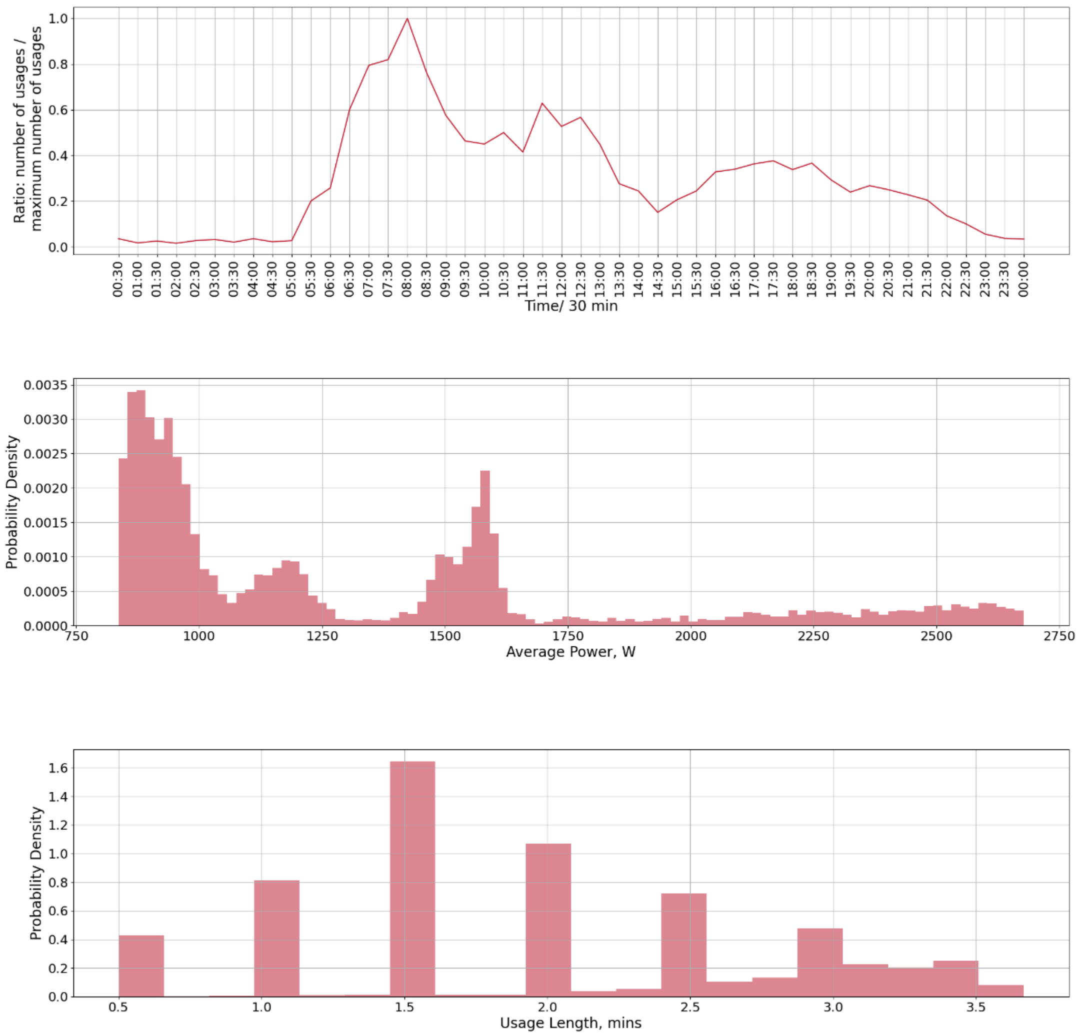

Figure 74: Usage statistics extracted for the toaster across all datasets.

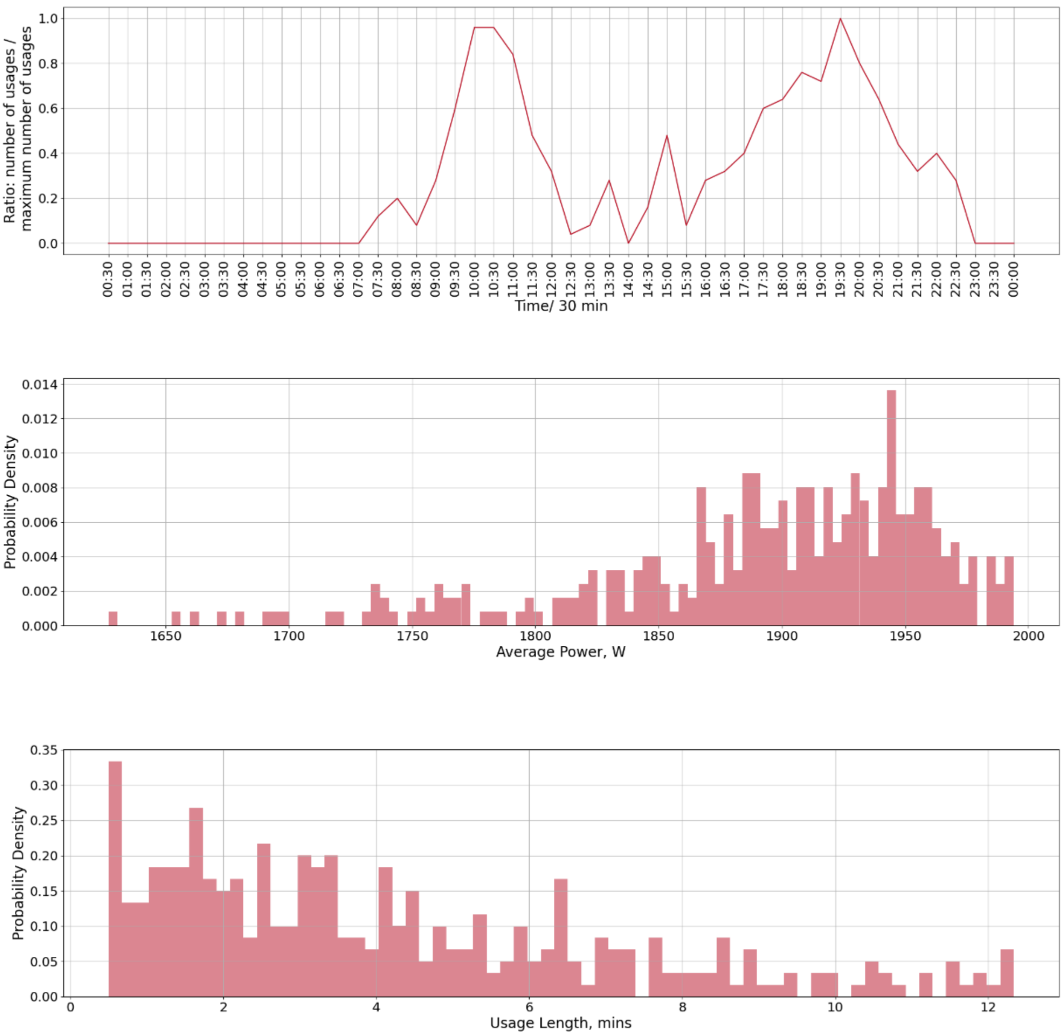

Figure 75: Usage statistics extracted for the vacuum across all datasets.

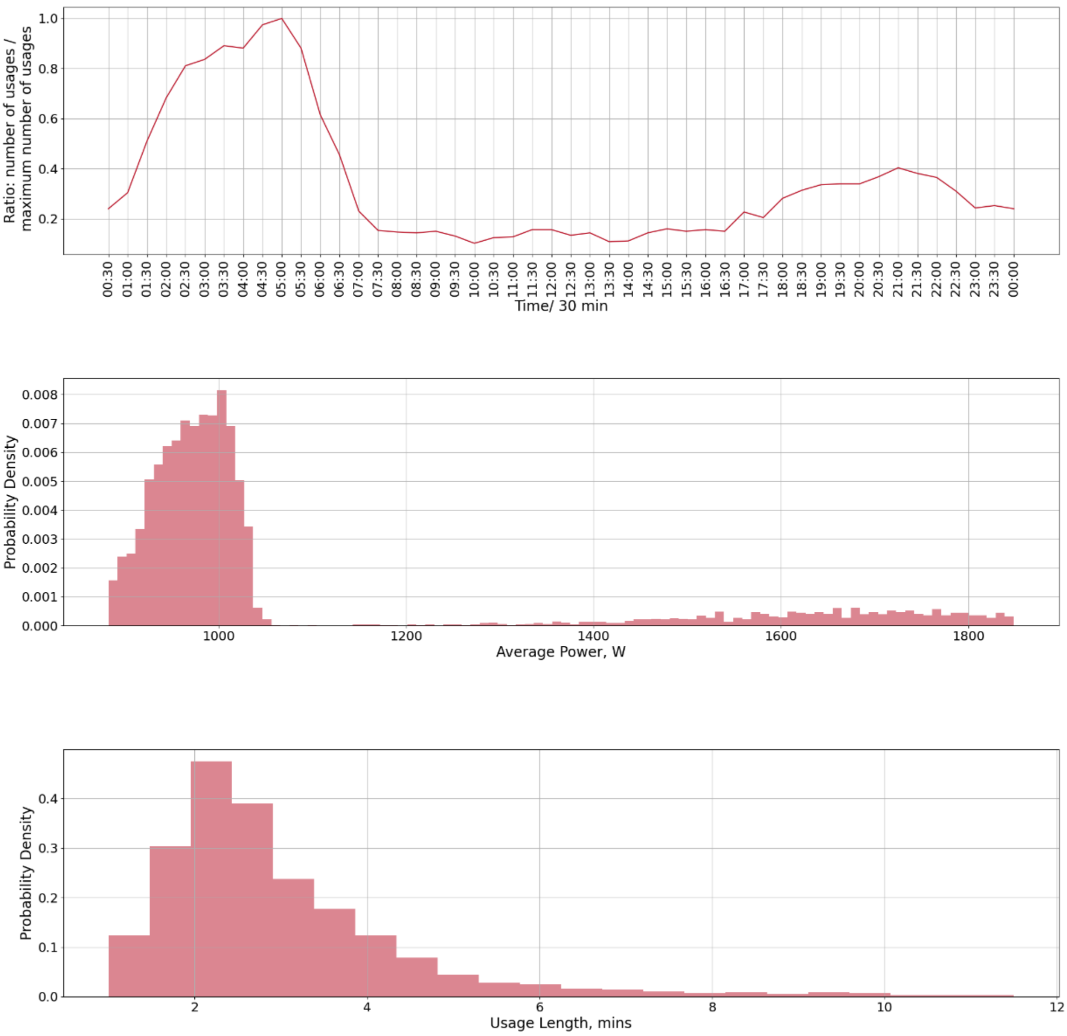

Figure 76: Usage statistics extracted for the electric heater across all datasets.

FRAZER-NASH CONSULTANCY
A KBR COMPANY

Renaissance House
Basing View
Basingstoke
Hampshire
RG21 4EQ

Tel: +44 (0)1306 885050

fnc.co.uk

Offices at:
Basingstoke, Bristol, Burton-on-Trent, Dorchester, Dorking, Glasgow, Gloucester, Middlesbrough, Plymouth and Warrington